



Comprendre le monde,
construire l'avenir®



UNIVERSITÉ PARIS-SUD

ECOLE DOCTORALE STITS

CEA, LIST, LABORATOIRE ANALYSE DE DONNÉES
ET INTELLIGENCE DES SYSTÈMES

DISCIPLINE : GÉNIE ÉLECTRIQUE, INFORMATIQUE,
AUTOMATIQUE ET TRAITEMENT DU SIGNAL

THÈSE DE DOCTORAT

soutenue le 19 Décembre 2014 par

Jérémy Rapin

**Décompositions parcimonieuses
pour l'analyse avancée de données
en spectrométrie pour la Santé**

Directeur de thèse : M. Jean-Luc Starck CEA, IRFU, CosmoStat

Encadrants : Anthony Larue CEA, LIST, LADIS

Jérôme Bobin (invité) CEA, IRFU, SEDI

Composition du jury :

Rapporteurs : M. Christian Jutten Université Joseph Fournier

M. Gabriel Peyré Université Paris-Dauphine

Examineurs : David Brie Université de Lorraine

Pascal Larzabal Université Paris XI

Pierre Vandergheynst École Polytechnique Fédérale
de Lausanne

Remerciements

Je remercie tout d'abord les membres de mon jury de thèse, et notamment mes rapporteurs, pour le temps qu'ils ont consacré à mon manuscrit et pour leurs pertinentes remarques.

Je tiens à remercier tout particulièrement Jérôme, mon encadrant, qui a été présent tout au long de ma thèse et grâce auquel j'ai énormément appris. Je remercie également Anthony qui a suivi avec attention tous mes travaux et était toujours disponible en cas de besoin, ainsi que Jean-Luc, mon directeur de thèse, qui m'a accueilli au sein de son équipe. Je souhaite enfin remercier Antoine pour le travail effectué sur les données LC/MS, ainsi que Christophe Junot et Minale Ouethrani qui nous ont permis d'utiliser ces données.

Plus généralement, je remercie les membres du laboratoire CosmoStat et du LADIS au CEA, qui étaient toujours disponibles pour donner un coup de main ou juste pour discuter. Une pensée particulière pour tous les thésards avec qui j'ai partagé de nombreuses pauses cafés, et notamment ceux avec qui j'ai même partagé l'intimité d'un bureau. Et surtout, surtout, un grand merci à tous ceux qui ont apporté des petits-déj' ou fait des pots, ce qui m'a permis d'appliquer à la lettre la devise "un esprit sain dans un corps sain" !

Enfin, un grand merci à mes amis et ma famille, qui étaient toujours là pour m'épauler tout au long de la thèse. J'ai bien sûr une pensée spéciale pour Guillaume, parti faire du volontariat au Tchad, Nolan, mon tout petit neveu, et Simon, mon encore plus petit filleul. Peut-être qu'un jour l'un de vous fera une thèse à son tour !

Pour finir, puisqu'il est souvent de bon ton de finir sur une citation, voilà qui est chose faite :

*"Wise words are like arrows flung at your forehead.
What do you do? Why, you duck of course."*

— Steven Erikson, *House of Chains*

Décompositions parcimonieuses pour l'analyse avancée de données en spectrométrie pour la Santé

Résumé : La séparation de sources en aveugle (SSA) vise à rechercher des signaux sources inconnus et mélangés de manière inconnue au sein de plusieurs observations. Cette approche très générique et non-supervisée ne fournit cependant pas nécessairement des résultats exploitables. Il est alors nécessaire d'ajouter des contraintes, notamment physiques, afin de privilégier la recherche de sources ayant une structure particulière. La factorisation en matrices positives (non-negative matrix factorization, NMF) qui fait plus précisément l'objet de cette thèse recherche ainsi des sources positives observées au travers de mélanges linéaires positifs.

L'ajout de davantage d'information reste cependant souvent nécessaire afin de pouvoir séparer les sources. Nous nous intéressons ainsi au concept de parcimonie qui permet d'améliorer le contraste entre celles-ci tout en produisant des approches très robustes, en particulier au bruit. Nous montrons qu'afin d'obtenir des solutions stables, les contraintes de positivité et la régularisation parcimonieuse doivent être appliqués de manière adéquate. Aussi, l'utilisation de la parcimonie dans un espace transformé potentiellement redondant, permettant de capturer la structure de la plupart des signaux naturels, se révèle difficile à appliquer au côté de la contrainte de positivité dans l'espace direct. Nous proposons ainsi un nouvel algorithme de NMF parcimonieuse, appelé nGMCA (non-negative Generalized Morphological Component Analysis), qui surmonte ces difficultés via l'utilisation de techniques de calcul proximal. Des expérimentations sur des données simulées montrent que cet algorithme est robuste à une contamination par du bruit additif Gaussien, à l'aide d'une gestion automatique du paramètre de parcimonie. Des comparaisons avec des algorithmes de l'état-de-l'art en NMF sur des données réalistes montrent l'efficacité ainsi que la robustesse de l'approche proposée.

Finalement, nous appliquerons nGMCA sur des données de chromatographie en phase liquide - spectrométrie de masse (liquid chromatography - mass spectrometry, LC/MS). L'observation de ces données montre qu'elles sont contaminées par du bruit multiplicatif, lequel détériore grandement les résultats des algorithmes de NMF. Une extension de nGMCA conçue pour prendre en compte ce type de bruit à l'aide d'un a priori non-stationnaire permet alors d'obtenir d'excellents résultats sur des données réelles annotées.

Mots clés : NMF, SSA, LC/MS, parcimonie, ondelettes, positivité

Sparse decompositions for advanced data analysis of hyperspectral data in biological applications

Abstract : Blind source separation aims at extracting unknown source signals from observations where these sources are mixed together by an unknown process. However, this very generic and non-supervised approach does not always provide exploitable results. Therefore, it is often necessary to add more constraints, generally arising from physical considerations, in order to favor the recovery of sources with a particular sought-after structure. Non-negative matrix factorization (NMF), which is the main focus of this thesis, aims at searching for non-negative sources which are observed through non-negative linear mixtures.

In some cases, further information still remains necessary in order to correctly separate the sources. Here, we focus on the sparsity concept, which helps improving the contrast between the sources, while providing very robust approaches, even when the data are contaminated by noise. We show that in order to obtain stable solutions, the non-negativity and sparse constraints must be applied adequately. In addition, using sparsity in a potentially redundant transformed domain could allow to capture the structure of most of natural image, but this kind of regularization proves difficult to apply together with the non-negativity constraint in the direct domain. We therefore propose a sparse NMF algorithm, named nGMCA (non-negative Generalized Morphological Component Analysis), which overcomes these difficulties by making use of proximal calculus techniques. Experiments on simulated data show that this algorithm is robust to additive Gaussian noise contamination, with an automatic control of the sparsity parameter. This novel algorithm also proves to be more efficient and robust than other state-of-the-art NMF algorithms on realistic data.

Finally, we apply nGMCA on liquid chromatography - mass spectrometry data. Observation of these data show that they are contaminated by multiplicative noise, which greatly deteriorates the results of the NMF algorithms. An extension of nGMCA was designed to take into account this type of noise, thanks to the use of a non-stationary prior. This extension is then able to obtain excellent results on annotated real data.

Keywords : NMF, BSS, LC/MS, sparsity, wavelets, non-negativity

Table des matières

Liste des Abréviations	viii
Notations	ix
I Introduction	1
II Introduction à la parcimonie	3
A Représenter un signal de manière parcimonieuse	3
A.1 Définition de la parcimonie	3
A.2 Approximation par des signaux parcimonieux	4
A.3 Représentation parcimonieuse dans un espace transformé	6
A.4 Dictionnaires et mesure ℓ_1 de la parcimonie	9
B Parcimonie pour la résolution de problèmes inverses	12
B.1 Formulations des problèmes inverses parcimonieux	12
B.2 Débruitage : une application directe	14
B.3 Formulations de la parcimonie dans les espaces redondants	15
B.4 Séparation de sources	16
B.5 Apprentissage de dictionnaire	17
C Algorithmes de résolution des problèmes parcimonieux	18
C.1 Rappels d'analyse convexe	18
C.2 Présentation des opérateurs proximaux	19
C.3 Quelques algorithmes proximaux	20
D Récapitulatif	24
III État-de-l'art en NMF	25
A Séparation de sources en aveugle	26
A.1 Modélisation des données	26
A.2 Identifiabilité des sources en SSA	30
B Factorisation en matrices non-négatives	32
B.1 Choix du terme d'attache aux données	33
B.2 Algorithmes de NMF - Approches de descente par blocs	34
C Régularisations en NMF	39
C.1 A priori lisse	39
C.2 A priori parcimonieux	40
D Récapitulatif des algorithmes	43

E	Critères de séparation	45
E.1	Calcul des critères de séparation	45
E.2	Exemple de décomposition	46
IV	nGMCA dans l'espace direct	51
A	Une première adaptation de GMCA	52
A.1	Présentation de GMCA	52
A.2	GMCA et positivité (nGMCA ^{naive})	53
B	Conception de nGMCA	54
B.1	Formulation et résolution	55
B.2	Gestion du seuil	56
B.3	Initialisation de l'algorithme	57
B.4	Renormalisations des sources et des mélanges	57
B.5	Convergence de l'algorithme	58
B.6	Version "seuillage dur"	59
C	Expérimentations sur des données simulées	59
C.1	Simulation des données	59
C.2	Comparaison avec les algorithmes standards dans le cas non-bruité	60
C.3	Expérimentations avec du bruit	65
C.4	Application à des données RMN synthétiques	71
D	Justification a posteriori de l'approche choisie	72
D.1	Choix de la formulation régularisée	74
D.2	Approche algorithmique	76
E	Conclusion	78
V	nGMCA dans un espace transformé	79
A	Parcimonie dans un espace transformé	80
A.1	Formulation en synthèse	80
A.2	Formulation en analyse	82
A.3	Complexité numérique	83
B	Expérimentations	83
B.1	Conditions expérimentales	84
B.2	Comparaison des approches	85
B.3	Conclusions des expérimentations	89
C	Mélanges d'images et repondération- ℓ_1	90
C.1	Influence du biais sur la séparation	90
C.2	La repondération comme compromis	92
C.3	Expérimentations	96
D	Conclusion	100
VI	Application aux données LC/MS	101
A	Présentation des données	101
A.1	Spectres de masse et profils d'élution	102

A.2	À propos du bruit et du modèle de mélange	103
A.3	Chaîne de traitement des données	105
A.4	Création de données synthétiques	106
B	Adaptations de nGMCA aux données LC/MS	108
B.1	Formulation avec du bruit non-stationnaire	108
B.2	Validation de l'approche non-stationnaire	110
B.3	Mises-à-jour stochastiques	112
C	Comparaison des algorithmes	113
C.1	Premiers résultats sur les mélanges synthétiques	113
C.2	Étude qualitative sur les données réelles - Cas de nGMCA	115
C.3	Étude quantitative sur les données réelles	119
D	Perspectives d'amélioration du modèle	121
D.1	Modèle de bruit	122
D.2	Modèle de mélange	122
D.3	Modélisation des spectres et des profils d'élution	123
VII	Conclusion	125
A	Opérateurs proximaux	129
A	Récapitulatif des opérateurs utilisés	129
B	Positivé et contrainte de norme	129
C	Positivité et trames ajustées	130
D	Parcimonie dans un espace transformé	131
E	Contrainte de sous-espace et de positivité	132
B	Détails des algorithmes	133
A	Mise-à-jour positive régularisée en synthèse	133
B	Mise-à-jour positive régularisée en analyse	134
C	Analyse formulée en synthèse contrainte	134
D	Reconditionnement élément par élément de l'algorithme FB avec une régularization ℓ_1	135
C	Séparation de sources partiellement corrélées	137
	Bibliographie	143

Liste des Abréviations

ACI	Analyse en Composantes Indépendantes
ALS	Alternating Least Squares (algorithme des moindres carrés alternés)
CEA	Commissariat à l'Énergie Atomique
FB	Forward-Backward (algorithme)
GFB	Generalized Forward-Backward (algorithme)
GMCA	Generalized Morphological Component Analysis (analyse généralisée en composantes morphologiques)
HALS	Hierarchical Alternating Least Squares (algorithme hiérarchique des moindres carrés alternés)
IRFU	Institut de Recherche sur les lois Fondamentales de l'Univers
IS	Itakura-Saito (divergence de)
KL	Kullback-Leibler (divergence de)
LADIS	Laboratoire Analyse de Données et Intelligence des Systèmes
LC-MS	Liquid Chromatography - Mass-Spectrometry (Chromatographie en phase liquide - Spectrométrie de masse)
LIST	Laboratoire d'Intégration de Systèmes et des Technologies
MAD	Median Absolute Deviation (écart médian absolu)
MCA	Morphological Component Analysis (analyse en composantes morphologiques)
MM	Majoration-Minimisation
nGMCA	non-negative Generalized Morphological Component Analysis (analyse généralisée en composantes morphologiques non-négatives)
NMF	Non-negative Matrix Factorization (factorisation en matrices positives)
RdP	Ratio de Parcimonie
RMN	Résonance Magnétique Nucléaire
SAR	Signal to Artifacts Ratio (rapport signal sur artefacts)
SDR	Signal to Distortion Ratio (rapport signal sur distorsion)
SIR	Signal to Interferences Ratio (rapport signal sur interférences)
SNR	Signal to Noise Ratio (rapport signal sur bruit)
SSA	Séparation de Sources en Aveugle
SVD	Singular Value Decomposition (décomposition en valeurs singulières)

Conventions

Conventions d'écriture

- x : un vecteur colonne ou un scalaire.
- x_i : $i^{\text{ième}}$ coefficient de x .
- $x_{(i)}$ le $i^{\text{ième}}$ coefficient de $x \in \mathbb{R}^n$ par ordre décroissant des valeurs absolues, de telle sorte que $|x_{(1)}| \geq |x_{(2)}| \geq \dots \geq |x_{(n)}|$.
- \mathbf{X} : une matrice.
- $\mathbf{X}_{i,j}$: la valeur de l'élément (i, j) de \mathbf{X} .
- $\mathbf{X}_{i,\cdot}$: la $i^{\text{ième}}$ ligne de \mathbf{X} .
- $\mathbf{X}_{\cdot,j}$: la $j^{\text{ième}}$ colonne de \mathbf{X} .

Opérateurs

- $\text{supp}(x)$: support du vecteur x (ensemble des coefficients non-nuls de x).
- $\text{card}(\mathcal{E})$: cardinal de l'ensemble \mathcal{E} (nombre d'éléments de l'ensemble \mathcal{E}).
- $\|x\|_p$: avec $p > 0$, norme ℓ_p de x , définie par $\|x\|_p = \sqrt[p]{\sum_i |x_i|^p}$. On définit de plus $\|x\|_0 = \text{card}(\text{supp}(x))$ (pseudo-norme ℓ_0) ainsi que $\|x\|_\infty = \max(x)$ (norme infinie).
- $\|\mathbf{X}\|_p$: avec $p > 0$, norme matricielle définie par $\|\mathbf{X}\|_p = \sqrt[p]{\sum_{i,j} |\mathbf{X}_{i,j}|^p}$ (norme de Frobenius pour $p = 2$). On définit de également la pseudo-norme ℓ_0 et la norme infinie.
- $\|\mathbf{X}\|_{s,2}$: norme subordonnée à la norme ℓ_2 de \mathbf{X} , définie par $\|\mathbf{X}\|_{s,2} = \max_{v \neq 0} \frac{\|\mathbf{X}v\|_2}{\|v\|_2}$.
- \odot : multiplication de matrices élément-à-élément (produit d'Hadamard).
- \oslash : division de matrices élément-à-élément (division d'Hadamard).
- $*$: opérateur de convolution circulaire.
- $\mathbf{X} \geq \mathbf{Y}$: la comparaison $\mathbf{X}_{i,j} \geq \mathbf{Y}_{i,j}, \forall (i, j)$.
- \mathbf{X}^T : la transposée de \mathbf{X} . Toutes les données considérées dans ce manuscrit étant réelles, cela équivaut également à son adjoint.
- $\langle \cdot, \cdot \rangle$: produit scalaire.
- $[\cdot]_+$: partie positive.
- Hard_λ : l'opérateur de seuillage dur (hard-thresholding) défini comme suit :

$$\text{Hard}_\lambda : x \mapsto \begin{cases} 0 & \text{si } |x| < \lambda, \\ x & \text{sinon.} \end{cases} \quad (1)$$

- Soft_λ : l'opérateur de seuillage doux (soft-thresholding) défini par :

$$\text{Soft}_\lambda : x \mapsto \text{sign}(x) [|x| - \lambda]_+. \quad (2)$$

où "sign" est la fonction valant 1 lorsque x est positif ou nul, et -1 dans le cas contraire.

- prox_f : l'opérateur proximale d'une fonction f , défini par :

$$\text{prox}_f(x) = \underset{y}{\text{argmin}} \frac{1}{2} \|y - x\|_2^2 + f(y). \quad (3)$$

- $i_{\mathcal{C}}$: la fonction caractéristique d'un ensemble \mathcal{C} . Elle est définie comme suit :

$$i_{\mathcal{C}} : x \mapsto \begin{cases} 0 & \text{si } x \in \mathcal{C}, \\ +\infty & \text{sinon.} \end{cases} \quad (4)$$

Conventions de nommage

- \mathbb{R} : ensemble des réels.
- m : un nombre d'observations.
- n : le nombre d'échantillons d'un signal.
- r : un nombre de sources.
- $\mathbf{Y} \in \mathbb{R}^{m \times n}$: une matrice de données, dans laquelle chaque ligne est une observation.
- $\mathbf{S} \in \mathbb{R}^{r \times n}$: la matrice des sources, dans laquelle chaque ligne est une source/un spectre.
- $\mathbf{A} \in \mathbb{R}^{m \times r}$: la matrice de mélanges, qui définit la contribution de chaque sources dans chacune des observations.
- $\mathbf{Z} \in \mathbb{R}^{m \times n}$: la matrice de bruit, qui représente le bruit expérimental et les imperfections de modèle.
- $\mathbf{1}_{rn} \in \mathbb{R}^{r \times n}$: matrice dont tous les coefficients valent 1.
- $\mathbf{I}_n \in \mathbb{R}^{n \times n}$: la matrice identité de taille $n \times n$.
- $\mathbf{W} \in \mathbb{R}^{p \times n}$: une matrice de transformation de \mathbb{R}^n dans \mathbb{R}^p .
- p : la dimension de l'espace transformé.
- x_w : la transformée du vecteur x par \mathbf{W} , $x_w = \mathbf{W}x$. Quand plusieurs indices sont nécessaires, ils sont séparés par le symbole |. Ainsi, le $j^{\text{ième}}$ élément de x_w est noté $x_{w|j}$.
- \mathbf{S}_w : la transformée des signaux de \mathbf{S} par \mathbf{W} . Les signaux étant des vecteurs lignes, $\mathbf{S}_w = \mathbf{W}^T \mathbf{S}$.

Introduction

Le travail présenté dans ce manuscrit a été effectué au sein du laboratoire d'analyse des données et d'intelligence des systèmes (LADIS) du CEA-LIST, ainsi qu'au sein du laboratoire CosmoStat du Service d'Astrophysique du CEA-IRFU, tous les deux basés à Saclay. Le but de cette thèse, intitulée "Décomposition parcimonieuse pour l'analyse avancée de données en spectrométrie pour la Santé", est de transposer et approfondir des développements effectués au sein du laboratoire CosmoStat sur le traitement de données multispectrales et de les mettre en pratique sur des données biologiques traitées au sein du LADIS.

Cette thèse porte ainsi sur le traitement du signal et plus spécifiquement la séparation de sources en aveugle (SSA). Ce domaine des mathématiques vise à identifier un nombre réduit de signaux "sources" à partir d'observations où ceux-ci sont mélangés entre eux d'une manière qui n'est pas connue à l'avance. Chacune des sources correspond généralement à un signal caractéristique d'une entité particulière : le spectre de masse ou électromagnétique pour une molécule, ou le spectre sonore d'un instrument ou d'une voix par exemple. Retrouver ces spectres à partir des mélanges via la décomposition des données fournie par la SSA permet alors d'identifier ces entités. Les problèmes de SSA sont cependant très mal définis, pouvant admettre de nombreuses solutions qui ne correspondent pas nécessairement aux sources recherchées. Il est donc nécessaire de rajouter des informations sur les sources recherchées afin de pouvoir les identifier. Plusieurs types d'a priori ont été envisagés menant à des approches différentes. Nous nous intéresserons plus particulièrement dans ce manuscrit à l'intersection entre deux de ces approches.

La première approche revient à chercher des sources parcimonieuses. Un signal est dit parcimonieux quand il peut être convenablement approximé à l'aide d'un faible nombre de coefficients dans une base donnée. C'est le cas pour un grand nombre de signaux de la vie courante, comme les signaux périodiques, qui sont parcimonieux dans l'espace de Fourier, ou bien les images naturelles, qui sont typiquement parcimonieuses dans l'espace des ondelettes. La recherche de signaux parcimonieux revient alors à chercher des signaux ayant une structure particulière, que l'on sait bien modéliser du point de vue mathématique.

La deuxième approche est la factorisation en matrices positives (Non-negative Matrix Factorization, NMF). Cette approche revient à faire l'hypothèse que les sources sont des signaux positifs, et qu'elles apparaissent dans les observations via des combinaisons linéaires positives. Le modèle est ainsi purement additif. On retrouve cette situation dans de nombreux types de données. En imagerie hyperspectrale par exemple, les sources sont les spectres d'intensité électromagnétique de

composées chimiques, tels que l'eau ou le CO_2 . En tant qu'intensités, ces spectres sont ainsi nécessairement positifs. Les différents mélanges observés correspondent quant à eux à des concentrations différentes de ces composés. Ceux-ci sont donc également positifs.

Le premier chapitre de ce manuscrit est une introduction à la parcimonie, permettant de mieux appréhender ce domaine. Il fournira ainsi les outils nécessaires au développement des algorithmes proposés dans ce manuscrit. Le chapitre suivant fera l'état-de-l'art de la factorisation en matrices positives, en détaillant en particulier l'usage qui a été fait de la parcimonie sous différentes formes dans ce domaine.

Nous présenterons ensuite un nouvel algorithme de NMF, non-negative Generalized Morphological Component Analysis (nGMCA) recherchant des signaux parcimonieux dans l'espace direct et utilisant des approches d'optimisation dites "proximales". Ce premier algorithme permettra en particulier d'étudier comment cet a priori parcimonieux peut être utilisé afin de gérer une contamination par du bruit additif, et comment il interagit avec la recherche des sources.

Le cadre de conception de nGMCA sera ensuite étendu afin de pouvoir également rechercher des signaux parcimonieux dans un espace transformé. Peu d'algorithmes de NMF permettent d'utiliser un tel a priori, celui-ci engendrant une complexité accrue du processus d'optimisation. Un tel a priori peut cependant s'avérer extrêmement utile afin de pouvoir modéliser des signaux plus complexes, comme des images. Nous verrons ainsi que nGMCA permet d'utiliser plusieurs formulations de cette parcimonie, et que celle-ci permet d'améliorer très sensiblement les résultats pour la recherche de sources qui ne sont pas parcimonieuses dans l'espace direct.

La dernière partie de cette thèse met en application les outils développés afin de traiter des données de chromatographie en phase liquide-spectrométrie de masse. Cette technique permet d'analyser un fluide en le faisant passer dans une colonne de rétention dans laquelle les composés chimiques du fluide circulent à des vitesses différentes et en sortent ainsi à des temps différents. À chaque instant, un spectromètre analyse les composés sortant de la colonne de rétention. On obtient alors une première double séparation approximative des composés chimiques, en temps et en masse. La NMF est parfaitement appropriée à l'étude de ces données, car les spectres de masse sont positifs et les mélanges dépendent des concentrations des composés du fluide en sortie de colonnes, et sont donc également positifs. Ces données comportent cependant des particularités qui les rendent difficiles à traiter. Ainsi, elles sont par exemple contaminées par du bruit multiplicatif, et nous développerons donc une version de nGMCA spécifique à ce type de données, permettant d'améliorer très nettement la recherche des composés du fluide analysé.

Introduction à la parcimonie

Sommaire

A	Représenter un signal de manière parcimonieuse	3
A.1	Définition de la parcimonie	3
A.2	Approximation par des signaux parcimonieux	4
A.3	Représentation parcimonieuse dans un espace transformé	6
A.4	Dictionnaires et mesure ℓ_1 de la parcimonie	9
B	Parcimonie pour la résolution de problèmes inverses	12
B.1	Formulations des problèmes inverses parcimonieux	12
B.2	Débruitage : une application directe	14
B.3	Formulations de la parcimonie dans les espaces redondants	15
B.4	Séparation de sources	16
B.5	Apprentissage de dictionnaire	17
C	Algorithmes de résolution des problèmes parcimonieux	18
C.1	Rappels d'analyse convexe	18
C.2	Présentation des opérateurs proximaux	19
C.3	Quelques algorithmes proximaux	20
D	Récapitulatif	24

En mathématique, le concept de parcimonie revient à considérer qu'un signal peut être exprimé à l'aide d'un faible nombre d'éléments. L'usage de ce concept a permis de grandes avancées dans la modélisation et la manipulation de la plupart des signaux naturels. Il est ainsi utilisé tout au long de ce manuscrit afin d'apporter une information structurelle sur les signaux recherchés. Ce chapitre vise à présenter cette notion de parcimonie et son utilisation au sein d'algorithmes d'optimisation. Nous commencerons par définir plus précisément cette notion et voir son intérêt pour le cas particulier de la représentation des signaux, puis nous nous intéresserons à son intégration dans des problèmes inverses. Ces problèmes inverses nous amèneront à décrire des algorithmes d'optimisation spécifiques permettant de les résoudre.

A Représenter un signal de manière parcimonieuse

A.1 Définition de la parcimonie

Les signaux peuvent être modélisés de façon continue ou de façon discrète. Nous nous placerons ici dans le cas de signaux discrets et de longueur finie, qui en pratique

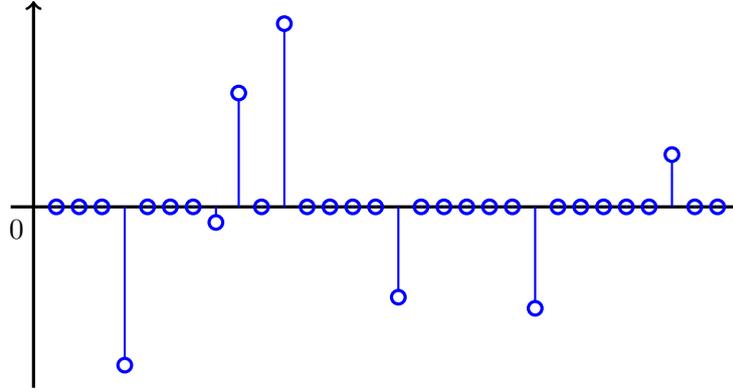


Figure II.1 – Exemple de signal parcimonieux.

sont des éléments/vecteurs de \mathbb{R}^n , où n est leur nombre d'échantillons. Il est courant de vouloir approximer de tels signaux via une classe de signaux plus facilement manipulables. C'est le cas des signaux dits parcimonieux. Au sens strict, un signal est dit parcimonieux quand il ne comporte qu'un faible nombre de coefficients non-nuls, comme le signal de la figure II.1. Un tel signal est ainsi plus simple à manipuler : la connaissance des valeurs et positions de ces quelques coefficients non-nuls suffit à parfaitement le définir. Pour cette raison, ces signaux sont également dits "compressibles" et la parcimonie a des applications très directes en compression de signaux et d'images par exemple [Donoho *et al.* 1998].

Plus quantitativement, on définit le support "supp" d'un vecteur comme l'ensemble des coefficients non-nuls de ce vecteur, et la pseudo norme ℓ_0 comme le cardinal "card" du support. En d'autres termes, la pseudo-norme ℓ_0 est le nombre de coefficients non-nuls de ce signal. Pour un vecteur x , on a donc :

$$\|x\|_0 = \text{card}(\text{supp}(x)) \quad (\text{II.1})$$

Un signal est ainsi d'autant plus parcimonieux que sa pseudo-norme ℓ_0 est faible.

A.2 Approximation par des signaux parcimonieux

En pratique, peu de signaux sont exactement parcimonieux tel que décrit précédemment. On peut cependant se demander si, à un signal donné y , on peut trouver un signal parcimonieux x qui soit proche de y . Pour cela, on se munit d'une mesure de distorsion $(x, y) \mapsto d(x, y) \in \mathbb{R}^+$. Elle sert à mesurer l'erreur que l'on engendre en approximant y par x . Elle augmente lorsqu'on s'éloigne du signal d'origine et vaut 0 uniquement dans le cas où $x = y$. Une mesure couramment utilisée est la mesure de distorsion ℓ_2 (erreur quadratique) :

$$d(x, y) = \|y - x\|_2^2. \quad (\text{II.2})$$

Ainsi, retrouver le signal le plus proche de y au sens de la distorsion précédente

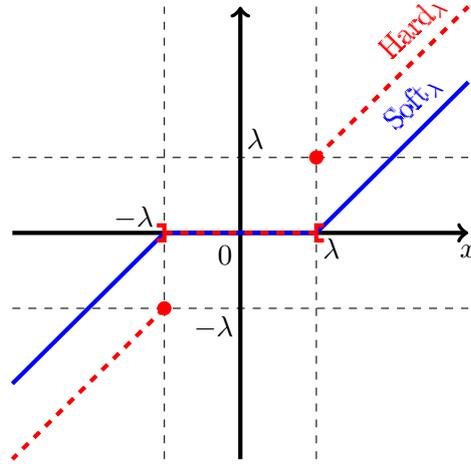


Figure II.2 – Seuillage dur (Hard) et seuillage doux (Soft).

et ne comportant que q coefficients non nuls revient à résoudre le problème suivant :

$$\operatorname{argmin}_x \|y - x\|_2^2, \|x\|_0 \leq q. \quad (\text{II.3})$$

Ce problème d'approximation est dit non-linéaire (voir [DeVore 1998] pour une vue d'ensemble sur le sujet), l'espace d'approximation, qui correspond aux signaux ayant au plus q coefficients non nuls, n'étant pas un espace linéaire. Il a une solution simple qui consiste à garder les q plus grands coefficients de y . En notant $y_{(k)}$ le $k^{\text{ième}}$ coefficient de y par ordre décroissant des valeurs absolues, de telle sorte que $|y_{(1)}| \geq |y_{(2)}| \geq \dots \geq |y_{(n)}|$, on peut exprimer la solution $S_q(y)$ de (II.3) sous la forme :

$$S_q(y) = \operatorname{Hard}_{|y_{(q)}|}(y), \quad (\text{II.4})$$

où $\operatorname{Hard}_\lambda$ est l'opérateur de seuillage dur (hard-thresholding), qui met à 0 les coefficients dont la valeur absolue est plus faible que λ et laisse inchangés les autres. Cet opérateur est représenté figure II.2 et est défini comme suit :

$$\operatorname{Hard}_\lambda : x \mapsto \begin{cases} 0 & \text{si } |x| < \lambda, \\ x & \text{sinon.} \end{cases} \quad (\text{II.5})$$

La question qui se pose alors est de savoir à quel point cette approximation $S_q(y)$ de y est proche de y , et ainsi quelles conditions permettent de rendre cette approximation valide. La distorsion de l'approximation est donnée par :

$$d(S_q(y), y) = \sum_{i=q+1}^n y_{(i)}^2. \quad (\text{II.6})$$

Cette quantité correspond à la somme des carrés des éléments coupés par le seuillage dur, c'est-à-dire des $n - q$ plus faibles coefficients de y . Ainsi, plus la décroissance

des coefficients de y est rapide, plus le signal peut être approximé correctement avec un faible nombre de coefficients. En plus des signaux ayant un faible nombre de coefficients non nuls, on élargit donc généralement la classe des signaux parcimonieux aux signaux pouvant être convenablement approximés à l'aide d'un faible nombre de coefficients.

De façon plus formelle, on peut définir la classe de signaux :

$$w\ell_\alpha(R) = \{x, x_{(k)} \leq Rk^{-\alpha}\}, \quad (\text{II.7})$$

Cet ensemble est appelé boule ℓ_α -faible de rayon R et comprend des signaux dont la vitesse de décroissance des coefficients est une loi puissance en $k^{-\alpha}$. Pour de tels signaux, si $\alpha > \frac{1}{2}$, on peut majorer l'erreur d'approximation :

$$d(S_q(y), y) \leq \frac{R^2}{2\alpha - 1} q^{-2\alpha+1}. \quad (\text{II.8})$$

Ainsi, plus α est élevé, plus l'erreur d'approximation est faible. Le coefficient α conditionnant la vitesse de décroissance des coefficients, plus cette vitesse est grande, meilleure est l'approximation. Les ensembles $w\ell_\alpha(R)$ permettent ainsi une caractérisation moins stricte des signaux parcimonieux.

L'approche de la parcimonie ci-dessus s'intéressait à la décroissance des valeurs du signal. Considérer l'erreur relative de l'approximation apporte un autre point de vue :

$$\text{Err}_q(y) = \frac{\sum_{i=q+1}^n y_{(i)}^2}{\|y\|_2^2}, \quad (\text{II.9})$$

$$= 1 - \frac{\sum_{i=1}^q y_{(i)}^2}{\|y\|_2^2}, \quad (\text{II.10})$$

$$= 1 - \frac{\|S_q(y)\|_2^2}{\|y\|_2^2}. \quad (\text{II.11})$$

La norme quadratique $\|\cdot\|_2^2$ d'un vecteur est appelée énergie de ce vecteur. La formulation de l'erreur relative précédente permet alors de dire que l'approximation de y s'améliore quand $S_q(y)$ capte davantage d'énergie de y . Ainsi, plus l'énergie du signal est concentrée dans les q plus grands coefficients, meilleure est l'approximation parcimonieuse.

A.3 Représentation parcimonieuse dans un espace transformé

Même après élargissement de la définition, la classe des signaux parcimonieux ne couvre qu'une faible partie des données réelles. Prenons le cas de l'image de Saturne figure II.3a. L'approximation de cette image obtenue en ne gardant que ses 0,5% plus grands coefficients est visible figure II.3b. Cette image ne ressemble pas du tout à l'image d'origine, l'erreur d'approximation est extrêmement grande (environ 95%).

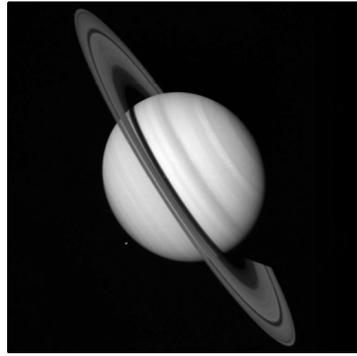
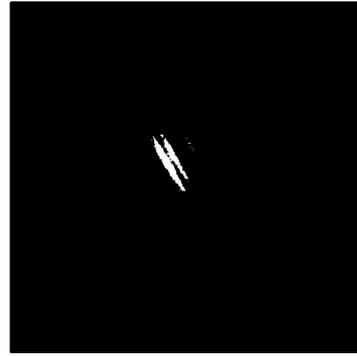
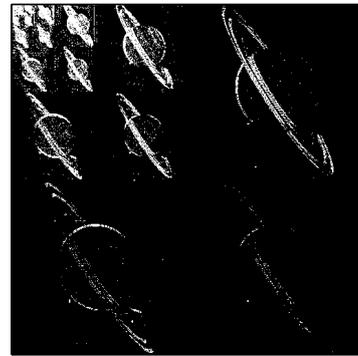
(a) image initiale : y .(b) approximation par les 0.5% plus grands coefficients de l'image : $S_{q_{0.5\%}}(y)$ ($\text{Err}_{q_{0.5\%}} \approx 95\%$).(c) approximation par les 0.5% plus grands coefficients d'ondelette : $\mathbf{R}^T S_{q_{0.5\%}}(\mathbf{R}y)$ ($\text{Err}_{q_{0.5\%}, \mathbf{R}} \approx 2, 2\%$).(d) Emplacement des 5% plus grands coefficients de $\mathbf{R}y$.

Figure II.3 – Exemple d'approximation d'une image grâce aux ondelettes (512×512). Crédit image : NASA/JPL (PIA02225).

Considérons maintenant une transformée orthonormale \mathbf{R} . Que se passe-t-il alors si l'on veut approximer y par un signal parcimonieux dans l'espace transformé ? Cette approximation est donnée par :

$$S_{q, \mathbf{R}}(y) = \mathbf{R}^T S_q(\mathbf{R}y), \quad (\text{II.12})$$

où \mathbf{R}^T est la transposée de \mathbf{R} , et donc également son inverse étant donné que \mathbf{R} est une matrice orthonormée. Une caractéristique intéressante de l'énergie d'un signal est son invariance par transformations orthonormales, elle ne dépend pas de la base orthonormée dans lequel le signal est représenté et on a donc $\|\mathbf{R}x\|_2^2 = \|\mathbf{R}^T x\|_2^2 = \|x\|_2^2$. On peut alors réécrire l'erreur relative d'approximation dans un

espace transformé comme suit :

$$\text{Err}_{q,\mathbf{R}}(y) = 1 - \frac{\|\mathbf{R}^T S_q(\mathbf{R}y)\|_2^2}{\|\mathbf{R}y\|_2^2}, \quad (\text{II.13})$$

$$= 1 - \frac{\|S_q(\mathbf{R}y)\|_2^2}{\|y\|_2^2}, \quad (\text{II.14})$$

$$= 1 - \frac{1}{\|y\|_2^2} \sum_{i=1}^q (\mathbf{R}y)_{(i)}. \quad (\text{II.15})$$

L'erreur d'approximation ne dépend donc que de la décroissance des coefficients dans le domaine transformé. Plus la transformée concentre l'énergie du signal dans les q plus grands coefficients, meilleure sera l'approximation. Ainsi, si l'on connaît une transformée capturant bien la morphologie du signal à approximer, c'est-à-dire concentrant son énergie dans un faible nombre de coefficients, on est capable de l'approximer avec une grande précision par un signal parcimonieux dans cette espace transformé.

Une question se pose alors : connaît-on de telles transformées ? La réponse est oui pour de nombreux types de signaux. Ainsi, une sinusoïde n'est par exemple pas du tout parcimonieuse dans l'espace direct car pratiquement tous ses coefficients sont non nuls. Elle ne peut donc pas être approximée de manière parcimonieuse dans cet espace. Cependant, l'espace de Fourier capture très bien les structures périodiques et permet donc de trouver de très bonnes approximations parcimonieuses de ces signaux. L'utilisation de cette transformée est classique en traitement du signal, afin de séparer ou filtrer des signaux concentrés sur une bande de fréquence donnée.

L'image de Saturne précédente peut quant à elle être très correctement représentée de manière parcimonieuse à l'aide d'une transformée en ondelettes orthonormales [Mallat 2008]. En n'utilisant que les 0,5% plus grands coefficients en valeur absolue dans l'espace transformé, on obtient l'approximation en figure II.3c. Celle-ci n'est pas parfaite, mais est très correcte (erreur relative d'approximation de 2,2% environ), notamment en regard de l'approximation parcimonieuse dans l'espace direct de la figure II.3b. La figure II.3d montre l'emplacement des 5% plus grands coefficients de la transformée en ondelettes de l'image de Saturne (c'est-à-dire 10 fois plus de coefficients que ceux utilisés pour l'estimation de la figure II.3c afin de mieux observer la structure sous-jacente). Cette image permet d'avoir un aperçu l'intérêt de cette transformée :

(i) Saturne apparaît plusieurs fois, à plusieurs échelles dans l'image transformée. Les transformées en ondelettes sont en effet multi-échelles et peuvent donc capturer l'énergie des images contenues à différentes échelles.

(ii) les 5% plus grands coefficients sont disposés essentiellement le long des contours de Saturne. Ainsi, une partie importante de l'énergie est contenue dans les discontinuités locales des images, et les ondelettes sont particulièrement appropriées pour capturer celles-ci.

La théorie des transformées en ondelettes dépassent de loin le cadre de cette thèse. Plusieurs types d'ondelettes existent en pratique, avec des transformées pouvant être orthonormales ou non, et les ondelettes peuvent également être définies pour des signaux continus. Les lecteurs intéressés pourront se référer à [Mallat 2008]. Il est cependant important de se souvenir que les transformées en ondelettes rendent parcimonieux les signaux polynomiaux par morceaux, qui fournissent une bonne approximation de la plupart des signaux et images naturels. Pour cette raison, nous utiliserons ce type de transformées de nombreuses fois au cours de cette thèse.

Il peut également être instructif d'observer l'énergie relative contenue dans les q premiers coefficients de l'image de Saturne codée dans différentes bases, c'est-à-dire :

$$\text{Ene}_{q,\mathbf{R}}(y) = \frac{1}{\|y\|_2^2} \sum_{i=1}^q (\mathbf{R}y)_{(i)}^2. \quad (\text{II.16})$$

C'est ce que montre la figure II.4a pour l'espace direct (i.e. sans transformée, en bleu), l'espace de Fourier (en vert), et l'espace des ondelettes (en rouge). L'énergie augmentant bien plus vite dans les deux derniers cas, la figure II.4b montre le même graphique avec q allant jusqu'à 256. Dans tous les cas cependant, l'énergie relative totale est égal à 1 une fois tous les coefficients utilisés. Ce ratio est directement lié à l'erreur quadratique relative, via l'équation $\text{Err}_{q,\mathbf{R}}(y) = 1 - \text{Ene}_{q,\mathbf{R}}(y)$. Ainsi, plus l'énergie est élevée, meilleure est la reconstruction. Dans l'espace direct, le plus grand coefficient représente $4,2 \times 10^{-3}\%$ de l'énergie de l'image. Pour les transformées de Fourier et en ondelettes, il représente respectivement 21,6% et 64,9% de l'énergie. Si l'on étend ce calcul aux 0,1% plus grands coefficients (262 coefficients), on trouve qu'ils représentent respectivement 1,1%, 97,3% et 98,6% de l'énergie. On remarque donc encore une fois que pour cette image, la transformée de Fourier et les transformées en ondelettes permettent de concentrer une grande partie de l'énergie dans quelques grands coefficients.

A.4 Dictionnaires et mesure ℓ_1 de la parcimonie

Nous avons vu que choisir une base orthonormée appropriée au signal pouvait permettre d'en obtenir une représentation parcimonieuse. Il n'est cependant pas nécessaire d'utiliser une base pour représenter un signal de manière parcimonieuse. Supposons que l'on ait une famille génératrice $(d^{(i)})_{i=1,\dots,p}$ de vecteurs de \mathbb{R}^n . On peut alors décomposer tout vecteur y de cet espace comme la somme de ces éléments de \mathbb{R}^n que l'on nomme généralement "atomes" :

$$y = \sum_{i=1}^p x_i d^{(i)}, \quad (\text{II.17})$$

où les x_i sont les coefficients de la décomposition. L'utilisation d'une transformée orthonormale correspond au cas particulier où les vecteurs $d^{(i)}$ sont normés et orthogonaux entre eux. En pratique, rien n'interdit d'utiliser davantage d'atomes qu'il n'y a de degrés de liberté dans \mathbb{R}^n , soit $p > n$. Dans ce cas là, la famille étant

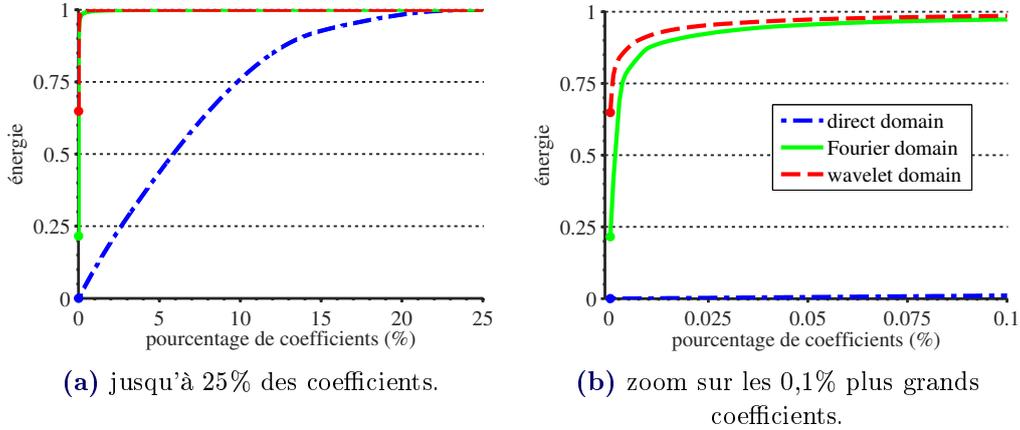


Figure II.4 – Énergie relative en fonction du pourcentage de coefficients conservés, pour l'image de Saturne.

liée, plusieurs combinaisons d'atomes sont possibles pour représenter y . Il est alors possible de rajouter une contrainte sur le choix des atomes utilisés. On peut ainsi s'intéresser à trouver quelle est la décomposition la plus parcimonieuse, c'est-à-dire utilisant le moins possible d'atomes de la famille $(d^{(i)})_{i=1,\dots,p}$. Soit \mathbf{D} la matrice dont les colonnes sont les $d^{(i)}$, généralement appelée "dictionnaire". Le problème peut se mettre sous la forme suivante :

$$\operatorname{argmin}_x \|x\|_p, \quad y = \mathbf{D}x, \quad (\text{II.18})$$

avec $p = 0$. En d'autres termes, cela revient à trouver le vecteur x comprenant le plus de coefficients nuls et tel que l'on obtienne la décomposition (II.17). Ce problème est représenté figure II.5 avec $y \in \mathbb{R}$, $\mathbf{D} = [d_1, d_2] \in \mathbb{R}^{1 \times 2}$ et $x \in \mathbb{R}^2$. La contrainte $y = \mathbf{D}x$ est représentée par la droite rouge. On peut trouver graphiquement ses solutions en recherchant la boule ℓ_p de rayon q :

$$\mathcal{B}_p(q) = \{x, \|x\|_p \leq q\}, \quad (\text{II.19})$$

intersectant la droite de contrainte et ayant le plus petit rayon q . Dans le cas $p = 0$, on peut énumérer les possibilités :

(i) $q = 0$: il n'y a pas d'intersection entre la droite et cet espace, qui ne comprend que le point $(0, 0)$.

(ii) $q = 1$: les points ayant un unique élément non nul sont ceux situés sur les axes et il y a deux intersections entre cet ensemble et la contrainte $y = \mathbf{D}x$.

Il y a ainsi deux solutions au problème (II.18) dans le cas présenté sur le graphique II.5, qui sont les deux intersections entre la droite de contrainte et les axes (le point a , et l'intersection entre l'axe x_1 et la contrainte, qui n'est pas visible sur le graphique).

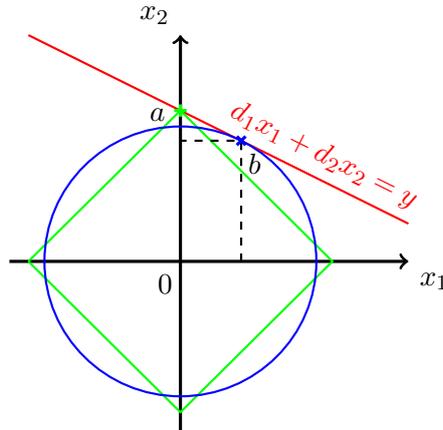


Figure II.5 – Boules ℓ_1 (vert) et ℓ_2 (bleu).

En plus grande dimension, énumérer tous les cas possibles afin de rechercher des solutions n'est pas envisageable car combinatoire. Cela vient du fait qu'avec $p = 0$, le problème n'est pas convexe, cette propriété permettant de concevoir des algorithmes efficaces de minimisation. Nous reviendrons sur cette notion en section C. Il est alors courant d'utiliser la norme ℓ_1 à la place de la pseudo-norme ℓ_0 , car celle-ci en est la norme convexe la plus proche [Chen *et al.* 2001]. Le problème (II.18) avec $p = 1$ est généralement appelé "basis pursuit". De la même manière que précédemment, cela revient à rechercher graphiquement la boule ℓ_1 de plus faible rayon intersectant la droite de contrainte. Cette boule est représentée en vert sur la figure II.5 et intersecte la contrainte au point a . La solution pour le problème avec $p = 1$ coïncide ici avec une des solutions du problème avec $p = 0$. Plus généralement, il a été prouvé que les solutions pour $p = 1$ sont également solutions du problème avec $p = 0$ lorsqu'elles sont très parcimonieuses [Donoho & Huo 2006]. La norme ℓ_1 a ainsi tendance à privilégier les solutions parcimonieuses. Le graphique II.5 montre également en bleu la boule ℓ_2 de rayon minimal intersectant la contrainte. La solution pour $p = 2$ est ainsi le point b , qui lui n'est pas parcimonieux ayant ses deux composantes non nulles. De façon plus générale, la norme ℓ_2 ne privilégie aucune direction et ne permet donc pas d'obtenir des solutions parcimonieuses.

En pratique, de nombreux choix de dictionnaires sont envisageables. Les ondelettes orthonormales ne présentent par exemple pas la propriété d'invariance par translation, qui signifie que lorsqu'un signal est translaté, sa transformée est également translatée. Des ondelettes redondantes $\mathbf{W} \in \mathbb{R}^{n \times p}$, avec $p > n$, peuvent cependant permettre cette invariance et celle-ci se montre bénéfique afin d'obtenir une représentation plus parcimonieuse [Coifman & Donoho 1995]. On peut alors choisir comme dictionnaire $\mathbf{D} = \mathbf{W}^T$. Il est également possible d'apprendre de tels dictionnaires de façon automatique (cf. section B.5). Une fois ce dictionnaire sélectionné, nous avons vu que remplacer la pseudo-norme ℓ_0 par la norme ℓ_1 permettait d'obtenir un problème convexe. Nous verrons section C des algorithmes permettant de résoudre ces problèmes.

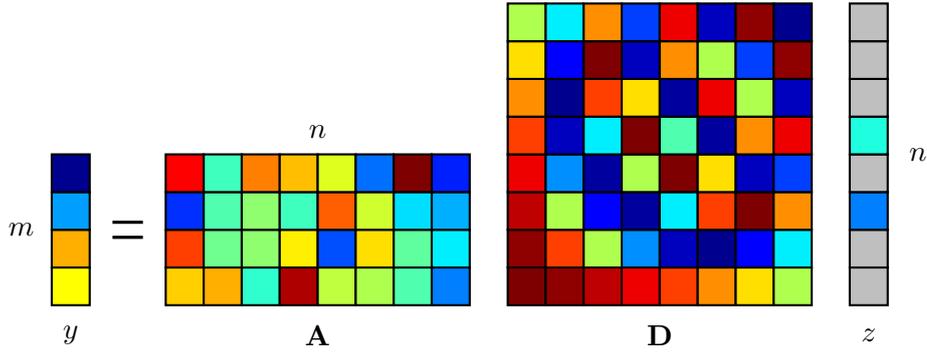


Figure II.6 – Problème inverse avec une représentation z parcimonieuse.

B Parcimonie pour la résolution de problèmes inverses

La partie précédente a présenté ce qu'était la parcimonie et a permis de montrer que de nombreux types de signaux pouvaient être convenablement représentés de façon parcimonieuse. Cette information morphologique sur les signaux peut également aider à résoudre de nombreux types de problèmes inverses [Starck *et al.* 2010]. Nous présentons dans cette section plusieurs problèmes courants où la parcimonie intervient, comme le débruitage et la séparation de sources, ainsi que les formulations utilisées.

B.1 Formulations des problèmes inverses parcimonieux

Un signal $x \in \mathbb{R}^n$ est souvent observé au travers d'un nombre m d'observations qui sont modélisées par des mesures linéaires $y_i = \mathbf{A}_{i,\cdot} x$ où $\mathbf{A}_{i,\cdot}$ est la $i^{\text{ème}}$ ligne de la matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$. Les observations y ne sont pas nécessairement suffisantes pour retrouver x avec la précision souhaitée. En effet, les mesures peuvent être bruitées, le nombre de mesures m peut être insuffisant, ou le modèle d'acquisition des données trop imprécis. Le fait que x puisse être représenté de manière parcimonieuse dans un dictionnaire \mathbf{D} est une information sur sa morphologie qu'il peut alors être utile de prendre en compte. Ainsi, on peut rechercher les coefficients parcimonieux $z \in \mathbb{R}^m$ de la représentation de telle sorte que $x = \mathbf{D}z$. Le problème revient donc à retrouver le vecteur z le plus parcimonieux permettant d'obtenir les mesures y :

$$\underset{z}{\operatorname{argmin}} \|z\|_p, \quad y = \mathbf{A}z, \quad (\text{II.20})$$

D'une solution \hat{z} de ce problème on obtient une estimation du signal recherché $\hat{x} = \mathbf{D}\hat{z}$. On pourra se reporter à la figure II.6 pour une interprétation graphique de ce problème, où les coefficients z de la représentation sont parcimonieux.

En posant $\Phi = \mathbf{A}\mathbf{D}$, ce problème est similaire au problème (II.18) avec Φ au lieu de \mathbf{D} . Comme énoncé précédemment, il est difficile à résoudre car non-convexe pour $p = 0$. Afin de rendre ce problème convexe tout en privilégiant des solutions parcimonieuses, nous avons vu qu'il était courant d'utiliser $p = 1$.

La contrainte dure peut être relâchée afin de pouvoir permettre un écart par rapport à l'égalité stricte, quand les mesures sont bruitées par exemple. Cela s'écrit alors :

$$\operatorname{argmin}_z \|z\|_1, \|\Phi z - y\|_2^2 \leq \varepsilon^2, \quad (\text{II.21})$$

quand $\varepsilon > 0$, ce problème est équivalent à la résolution du Lagrangien, donnant la version régularisée suivante :

$$\operatorname{argmin}_z \frac{1}{2} \|\Phi z - y\|_2^2 + \lambda \|z\|_1, \quad (\text{II.22})$$

où λ dépend de ε de façon non triviale. Ce type de problème a été et est toujours abondamment étudié, autant pour ses applications comme la déconvolution par exemple [Starck *et al.* 2010], que du point de vue théorique. Ainsi, de récentes études ont porté sur le choix des matrices de mesures \mathbf{A} permettant de retrouver un signal x parcimonieux avec une grande probabilité au coût d'un minimum de mesures (acquisitions compressées, compressed sensing [Donoho 2006, Candès & Tao 2006]). On la retrouve également dans de nombreux domaines des mathématiques comme les statistiques [Tibshirani 1994]. Nous retrouverons ce type de problème tout au long du manuscrit, et la section C fournira les outils permettant de le résoudre.

De manière plus générale, l'ensemble des problèmes précédents peuvent être mis sous la forme suivante :

$$\operatorname{argmin}_x f(y, x) + g(x). \quad (\text{II.23})$$

La fonction $x \mapsto f(y, x) + g(x)$ est appelée fonction de coût, et ses deux termes ont des rôles bien distincts :

- le terme d'attache aux données $f(y, x)$: celui-ci exprime l'écart entre les observations y et les observations qui auraient été effectuées avec un signal x quelconque, c'est-à-dire Φx . Comme à la section A.2, il est courant d'utiliser la distance ℓ_2 (cf. équation (II.2)) par exemple.
- la régularisation (ou pénalisation) $g(x)$: ce terme a pour but d'apporter une information a priori sur le signal recherché, indépendamment des observations. C'est ce terme qui est utilisé pour privilégier la parcimonie du signal recherché.

Pour le problème (II.23), le terme d'attache aux données est directement $f(y, z) = \frac{1}{2} \|y - \Phi z\|_2^2$ et la régularisation est $g(z) = \lambda \|z\|_1$. Le problème (II.21) peut également être mis sous la même forme en utilisant une fonction caractéristique. La fonction caractéristique d'un ensemble \mathcal{C} , que l'on notera $i_{\mathcal{C}}$, est définie comme suit :

$$i_{\mathcal{C}} : x \mapsto \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{otherwise.} \end{cases} \quad (\text{II.24})$$

Ainsi, le problème (II.21) est complètement équivalent au problème suivant :

$$\operatorname{argmin}_x \|z\|_1 + i_{\|\Phi \cdot - y\|_2^2 \leq \varepsilon^2}(z), \quad (\text{II.25})$$

où l'on note $i_{\|\Phi_{\cdot}-y\|_2^2 \leq \varepsilon^2}$ au lieu de $i_{\{z, \|\Phi z - y\|_2^2 \leq \varepsilon^2\}}$ afin de simplifier l'écriture de la fonction. On peut donc identifier $f(x, y) = i_{\|\Phi_{\cdot}-y\|_2^2 \leq \varepsilon^2}(z)$ qui est ainsi un terme d'attache aux données sous forme de contrainte, imposant que x ne s'éloigne pas d'une distance supérieure à ε de y , et $g(z) = \lambda \|z\|_1$ comme précédemment.

B.2 Débruitage : une application directe

Un des intérêts de la parcimonie est les possibilités qu'elle ouvre en matière de débruitage [Donoho & Johnstone 1994, Coifman & Donoho 1995, Aharon *et al.* 2006]. Dans la suite de ce manuscrit, nous considérerons la plupart du temps que les signaux observés sont contaminés par du bruit blanc Gaussien centré et d'écart-type σ . La distribution d'une telle variable aléatoire $b \in \mathbb{R}^n$ est donnée par :

$$f(b) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{1}{2\sigma^2} \|b\|_2^2} \quad (\text{II.26})$$

Le signal visible figure II.7 correspond au signal de la figure II.1 contaminé par la réalisation d'un tel bruit. Une des particularités du bruit blanc Gaussien est que son énergie moyenne se répartit uniformément dans toutes les directions de l'espace, n'en privilégiant donc aucune. À l'inverse quand le signal recherché est parcimonieux, son énergie est concentrée sur quelques grands coefficients non nuls. On peut donc espérer séparer le signal du bruit en recherchant ces quelques grands coefficients. Cela revient à résoudre le problème suivant, qui recherche un signal x aussi parcimonieux que possible tout en restant à une distance de y inférieure à ε :

$$\underset{x}{\operatorname{argmin}} \|x\|_0, \quad \|y - x\|_2^2 \leq \varepsilon^2. \quad (\text{II.27})$$

Le choix de ε dépend du niveau de bruit.

Ce problème peut être résolu de façon similaire au problème (II.3), en coupant les plus faibles coefficients de y en valeur absolue. Il existe un seuil τ dépendant de ε tel que la solution de (II.27) soit donnée par le seuillage dur à τ de y , noté $\operatorname{Hard}_\tau(y)$. Cette solution est la partie du signal qui est en bleu sur la figure II.7. On remarque en effet que les coefficients dont la valeur absolue est plus faible que τ ne sont majoritairement pas présents dans le signal recherché (figure II.1), et qu'à l'inverse il y a peu de coefficients dépassant le seuil qui ne proviennent pas du signal.

Dans la pratique, utiliser $\tau = \tau_\sigma \sigma$ où σ est l'écart-type du bruit et τ_σ est choisi entre 1 et 3 permet un débruitage efficace puisque dans le cas d'un bruit blanc Gaussien, la probabilité qu'un coefficients uniquement dû au bruit soit supérieur à σ (resp. 3σ) en valeur absolue est de 31,8% (resp. 0,4%).

La technique présentée ci-dessus peut s'appliquer pour n'importe quel signal dont on connaît une transformation le rendant parcimonieux. En effet, puisque le bruit blanc Gaussien ne privilégie aucune direction, il reste du bruit blanc Gaussien après un changement de base orthonormale, et ce changement de base permet donc de se ramener au problème (II.27). En pratique, cette technique est très liée aux méthodes classiques de filtrage dans l'espace de Fourier visant à ne conserver qu'une bande

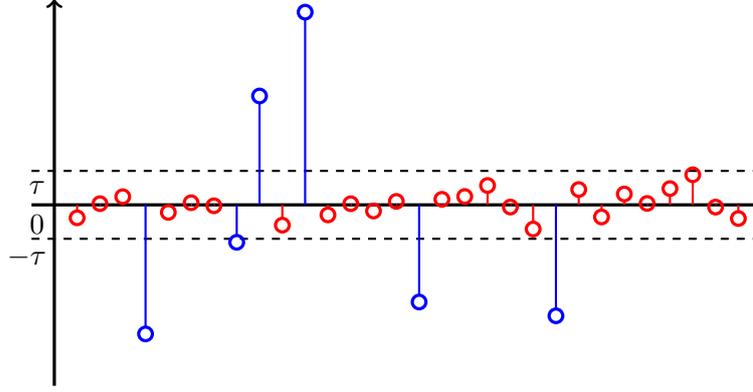


Figure II.7 – Exemple de signal parcimonieux et bruité (coefficients d'amplitude inférieure à τ en rouge, et supérieure en bleu).

de fréquence où se concentre l'énergie du signal à débruiter. L'usage d'une méthode non-linéaire comme la parcimonie peut cependant améliorer les performances du débruitage.

B.3 Formulations de la parcimonie dans les espaces redondants

Lorsque l'on utilise une transformation redondante $\mathbf{W} \in \mathbb{R}^{p \times n}$, la parcimonie peut être formulée de deux façons différentes, à savoir la synthèse et l'analyse.

Considérons un terme d'attache aux données $f : \mathbb{R}^n \rightarrow \mathbb{R}$, nous avons vu que l'on pouvait estimer le signal recherché via sa représentation parcimonieuse dans un dictionnaire $\mathbf{D} = \mathbf{W}^T$. Cela s'écrit :

$$\operatorname{argmin}_{z \in \mathbb{R}^p} f(\mathbf{W}^T z) + \lambda \|z\|_1. \quad (\text{II.28})$$

Cette formulation, que nous avons utilisée dans la section précédente, est une formulation en "synthèse". Avec cette formulation, l'inconnue du problème de minimisation n'est pas directement le signal mais une représentation parcimonieuse dans le domaine transformé. À partir de la solution \hat{z} , on retrouve l'estimation du signal par $\hat{x} = \mathbf{W}^T \hat{z}$. C'est un modèle génératif, d'où le nom "synthèse" : on reconstruit le signal recherché en utilisant des briques de l'espace du signal, les atomes.

Il est cependant essentiel de se rendre compte que le fait que l'estimation \hat{x}_w du signal (II.28) puisse être représentée de façon parcimonieuse ne veut pas dire que $\mathbf{W}\hat{x} = \mathbf{W}\mathbf{W}^T \hat{x}_w$ est nécessairement parcimonieux, puisque généralement $\mathbf{W}\mathbf{W}^T \neq \mathbf{I}$. En pratique, cela veut dire que la formulation en synthèse trouve une estimée qui a une représentation parcimonieuse dans le domaine transformé, mais pas que la transformée de cette estimée est parcimonieuse. À l'inverse, avec la formulation en analyse à l'inverse, on effectue la minimisation directement dans le domaine du signal/le domaine direct, afin de trouver une solution \hat{x} dont la transformée $\mathbf{W}\hat{x}$ est

parcimonieuse. Cela s'exprime de la manière suivante :

$$\operatorname{argmin}_{x \in \mathbb{R}^n} f(x) + \lambda \|\mathbf{W}x\|_1. \quad (\text{II.29})$$

Le terme $\lambda \|\mathbf{W}x\|_1$ pénalise les corrélations entre x et les atomes de $\mathbf{D} = \mathbf{W}^T$. En d'autres termes, alors qu'avec la formulation en synthèse le signal était exprimé comme la somme d'un nombre limité d'atomes, le but avec la formulation en analyse est d'obtenir un signal qui est très corrélé avec un faible nombre d'atomes de \mathbf{D} et faiblement corrélé avec les autres.

Quand la transformée \mathbf{W} est orthonormale, comme dans le cas de la transformée de Fourier et de la transformée en ondelettes utilisée section A.3, les formulations en synthèse et en analyse sont strictement équivalentes. En effet, elles sont alors liées par un simple changement de variable. Utiliser des transformées présentant de la redondance peut cependant être très bénéfique, comme indiqué dans [Coifman & Donoho 1995] dans le cas d'ondelettes redondantes. Cet avantage vient du plus grand nombre d'atomes disponibles pour représenter les signaux de façon parcimonieuse. Dans le cas des ondelettes redondantes, cela permet d'obtenir l'invariance par translation de la transformée. Avec de telles transformées redondantes, les formulations en analyse et en synthèse ne sont cependant plus équivalentes et ont des comportements différents [Elad *et al.* 2007, Selesnick & Figueiredo 2009, Nam *et al.* 2013]. En effet, puisque $p \neq n$, les espaces dans lesquels s'effectuent les minimisations en synthèse et en analyse sont nécessairement différents. Les études citées ci-dessus ont montré que la formulation en analyse avait tendance à être plus flexible et mieux adaptée à des signaux naturels, qui ne peuvent généralement pas être synthétisés avec un faible nombre d'atomes.

B.4 Séparation de sources

Les sections précédentes présentaient des problèmes où l'on ne recherchait qu'un unique signal. Certaines observations peuvent cependant résulter de plusieurs signaux "sources" mélangés entre eux. La séparation de sources vise à les démêler. Un critère est cependant nécessaire afin de distinguer ces sources les unes des autres. La parcimonie peut apporter ce critère [Takigawa *et al.* 2004, Golbabaee *et al.* 2013] et fournit ainsi une information structurelle/morphologique. En effet, les signaux sont plus parcimonieux dans les espaces qui capturent bien leur morphologie. Ainsi, les signaux sinusoïdaux sont par exemple très parcimonieux dans l'espace de Fourier mais le sont très peu dans l'espace direct. Les impulsions sont elles bien codées dans l'espace direct, et les ondelettes capturent bien les discontinuités locales. L'analyse en composante morphologique (MCA, [Starck *et al.* 2004]) se propose pour ces raisons de décomposer les données y en une somme de différentes composantes, chacune parcimonieuse dans un espace différent. Cela peut prendre la forme suivante :

$$\operatorname{argmin}_{x_d, x_f, x_w} \|y - x_d - \mathbf{W}^T x_w - \mathbf{F}^T x_f\|_2^2 + \lambda_d \|x_d\|_1 + \lambda_w \|x_w\|_1 + \lambda_f \|x_f\|_1 \quad (\text{II.30})$$

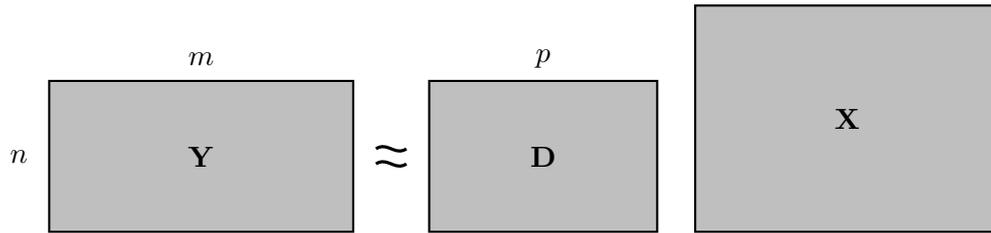


Figure II.8 – Apprentissage de dictionnaire

On retrouvera ainsi une décomposition des données sous la forme d'une somme de composantes parcimonieuses dans l'espace direct, x_d , dans l'espace de \mathbf{F} (Fourier par exemple), x_f , et une dans l'espace de \mathbf{W} (des ondelettes par exemple), x_w .

B.5 Apprentissage de dictionnaire

Le choix d'une transformée ou d'un dictionnaire dans lequel les signaux qui nous intéressent sont parcimonieux est primordial. L'apprentissage de dictionnaire vise à apprendre un tel dictionnaire directement à partir des données au lieu d'en utiliser un préconçu [Olshausen & Field 1996]. Ce dictionnaire peut ainsi être bien mieux adapté aux données. En pratique, cela revient à trouver p atomes $\mathbf{D}_{\cdot,k} \in \mathbb{R}^n$ permettant de coder n signaux $\mathbf{Y}_{\cdot,i} \in \mathbb{R}^n$ de façon parcimonieuse. Ainsi on peut par exemple fixer une limite s d'atomes utilisables pour coder chaque signal. Aussi, il faut imposer une limite de norme pour les atomes afin d'éviter un problème d'invariance par échelle (comme en séparation de sources en aveugle, nous reviendrons dessus en section A.1 du chapitre III, page 26). Sous forme matricielle, cela revient à résoudre le problème suivant :

$$\underset{\mathbf{D}, \mathbf{X}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2, \text{ avec } \begin{cases} \forall i = 1, \dots, p, \|\mathbf{D}_{\cdot,k}\| \leq 1, \\ \forall j = 1, \dots, m, \|\mathbf{X}_{\cdot,j}\|_0 \leq s. \end{cases} \quad (\text{II.31})$$

On apprend ainsi à la fois le dictionnaire \mathbf{D} et les coefficients \mathbf{X} . En pratique, le nombre d'observation m doit être plus grand que le nombre d'atomes p pour que le problème ne soit pas trivial, et p est généralement plus grand que n afin d'obtenir un dictionnaire redondant. La figure II.8 représente ainsi les conditions du problème. Cette application est par exemple utilisée pour le débruitage d'images [Aharon *et al.* 2006], pour le traitement de données d'électro-encéphalographie [Mailhe *et al.* 2009] et pour l'apprentissage de signaux de mouvement 2D [Barthélemy *et al.* 2012]. Il existe également des versions où l'apprentissage est formulé en analyse [Peyré & Fadili 2011, Rubinstein *et al.* 2013].

C Algorithmes de résolution des problèmes parcimonieux

La plupart des problèmes énoncés ci-dessus, tels que le problème inverse (II.21) ou le problème de séparation de sources (II.30), ne peuvent pas être résolus analytiquement et nécessitent donc l'utilisation d'un algorithme de minimisation. Deux grandes familles de méthodes apparaissent alors. La première correspond aux méthodes dites gloutonnes qui sont conçues afin de résoudre les problèmes minimisant une pseudo-norme ℓ_0 . Le "matching pursuit" en est l'exemple type [Mallat & Zhang 1993]. Le principe est de faire rentrer un à un de nouveaux éléments dans la variable courante, jusqu'à obtenir la parcimonie souhaitée, ou un résidu suffisamment faible. Ce type d'approche est très efficace quand les signaux recherchés sont extrêmement parcimonieux, mais fonctionne beaucoup moins bien pour des signaux plus complexes ou quand les atomes du dictionnaire utilisé sont très corrélés.

Dans cette section, nous nous intéressons donc essentiellement aux méthodes dites de "basis pursuit" qui utilisent des outils d'optimisation convexe pour résoudre les problèmes régularisés en norme ℓ_1 . Nous ferons ainsi un bref rappel d'analyse convexe avant d'introduire un type d'opérateur spécifique, l'opérateur proximal [Moreau 1962], qui sera abondamment utilisé dans les algorithmes d'optimisation présentés dans la dernière partie de cette section.

C.1 Rappels d'analyse convexe

Soit f une fonction sur \mathbb{R}^n à valeur dans $\mathbb{R} \cup \{-\infty, +\infty\}$, on peut définir les termes suivants :

- domaine effectif : le domaine effectif d'une fonction est le domaine sur lequel celle-ci ne vaut pas $+\infty$, c'est-à-dire $\text{dom}(f) = \{x \in \mathbb{R}^n, f(x) < +\infty\}$.
- convexité d'un ensemble : un ensemble $\mathcal{E} \subset \mathbb{R}^n$ est dit convexe si $\forall (x, y) \in \mathcal{E}^2$ et $\forall \lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in \mathcal{E}$.
- convexité d'une fonction : on dit que f est convexe si son domaine effectif est convexe et $\forall (x, y) \in \text{dom}(f)^2$ et $\forall \lambda \in [0, 1]$, on a l'inégalité suivante :

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (\text{II.32})$$

- fonction propre : une fonction est dite propre si elle ne prend pas la valeur $-\infty$ et si son domaine effectif n'est pas vide.
- semi-continuité inférieurement : une fonction est dite semi-continue inférieurement si les ensembles $\{x \in \mathbb{R}^n, f(x) \leq a\}$ sont fermés pour tout $a \in \mathbb{R}$.
- sous-gradient : un vecteur g est appelé sous gradient de f au point $x_0 \in \text{dom}(f)$ si $\forall x \in \text{dom}(f)$, on a :

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle. \quad (\text{II.33})$$

Le sous-gradient est ainsi une extension du gradient au cas des fonctions non-différentiables.

- $\partial f(x)$: ensemble des sous-gradients de f au point x . En effet, le sous-gradient en un point x n'est pas nécessairement unique.
- point critique/stationnaire : point $x \in \mathbb{R}^n$ tel que $0 \in \partial f(x)$.

En terme d'optimisation, la convexité apporte des propriétés extrêmement importantes. En particulier :

$$f(x^*) = \min_{x \in \text{dom}(f)} f(x) \Leftrightarrow 0 \in \partial f(x^*). \quad (\text{II.34})$$

Cette propriété permet de qualifier les minima de f et permet également de dire que tout point critique de f est un minimum global. Ces propriétés sont fondamentales pour la conception des algorithmes proximaux de minimisation présentés ci-dessous. Elle ne seront cependant pas détaillées ici et le lecteur intéressé pourra se référer à [Nesterov 2004].

C.2 Présentation des opérateurs proximaux

L'opérateur proximal est défini pour une fonction g convexe propre et semi-continue inférieurement. Pour cette raison, on dira par la suite qu'une fonction est proximable quand elle présente ces propriétés. L'opérateur proximal est alors donné par la relation suivante :

$$\text{prox}_g(x) = \underset{y}{\text{argmin}} \frac{1}{2} \|y - x\|_2^2 + g(y). \quad (\text{II.35})$$

Cet opérateur peut ainsi être vu comme une minimisation locale de la fonction g . En effet, le terme à minimiser est la somme de la fonction g , et de la fonction $x \mapsto \|y - x\|_2^2$ qui pénalise l'éloignement entre y et le point courant x .

Afin de mieux comprendre ce à quoi peut correspondre cet objet, il est intéressant de le calculer dans des cas particuliers. Supposons d'abord que g est dérivable. On trouve alors par simple dérivation que :

$$\text{prox}_g(x) = x - \nabla g(\text{prox}_g(x)) \quad (\text{II.36})$$

L'application de l'opérateur proximal sur x ressemble donc à l'utilisation d'une étape de descente de gradient. Le gradient est cependant pris en $\text{prox}_g(x)$, et est donc un gradient implicite. Encore une fois, une étape de descente de gradient, qu'elle soit implicite ou explicite, revient à minimiser localement la fonction g .

Le deuxième cas particulier est le cas d'une contrainte convexe. Soit \mathcal{C} un ensemble convexe. Nous avons déjà mentionné que la résolution du problème :

$$\underset{x \in \mathcal{C}}{\text{argmin}} f(x), \quad (\text{II.37})$$

était parfaitement équivalente à la minimisation du problème non-contraint suivant :

$$\underset{x \in \mathcal{C}}{\text{argmin}} f(x) + i_{\mathcal{C}}(x). \quad (\text{II.38})$$

De manière intéressante, quand \mathcal{C} est convexe, la fonction caractéristique $i_{\mathcal{C}}$ est proximable et son opérateur proximal est la projection orthonormale sur l'ensemble \mathcal{C} . Ainsi, par exemple, l'opérateur proximal de la fonction caractéristique de la contrainte de positivité est l'opérateur partie positive $[\cdot]_+$.

Le seuillage dur peut être aussi vu comme l'opérateur proximal de la pseudo-norme ℓ_0 , même s'il n'en a pas les propriétés puisque la pseudo-norme ℓ_0 n'est pas convexe. La norme ℓ_1 est quant à elle proximable. L'opérateur proximal de $g(x) = \tau\|x\|_1$ est appelé seuillage doux et définit de la manière suivante :

$$\text{Soft}_{\tau} : x \mapsto \text{sign}(x) \odot [|x| - \tau]_+, \quad (\text{II.39})$$

où "sign" est la fonction valant 1 lorsque x est positif ou nul, et -1 dans le cas contraire. La fonction "Soft $_{\tau}$ " est représentée figure [II.2](#). Le seuillage doux se comporte comme le seuillage dur, il coupe les plus faibles coefficients de x . Une différence subsiste cependant : tous les coefficients plus grands que le seuil τ sont contractés d'une valeur τ vers 0. Cette différence majeure fait que cet opérateur est continu alors que le seuillage dur ne l'est pas, et peut avoir des effets aussi bien bénéfiques que néfastes comme nous le verrons pour les méthodes développées au cours des chapitres suivants.

Plus généralement, de nombreuses fonctions admettent un opérateur proximal explicite ou pouvant être facilement calculés. Les opérateurs proximaux des principales fonctions utilisées dans cette thèse sont résumés dans le tableau [A.1](#) de l'annexe A (page 130). Certains calculs d'opérateurs y sont également explicités. Une liste plus complètes d'opérateurs proximaux utiles peut être trouvée dans [\[Combettes & Pesquet 2009\]](#).

Dans la section suivante, nous détaillons quatre algorithmes dits proximaux car utilisant de tels opérateurs. Ces algorithmes sont utilisés en tant que briques de base au sein des algorithmes de séparation de sources en aveugle développés dans cette thèse.

C.3 Quelques algorithmes proximaux

C.3.1 Forward-Backward

L'algorithme Forward-Backward (FB) [\[Combettes & Wajs 2005\]](#) vise à résoudre un problème qui prend la forme suivante :

$$\underset{x}{\text{argmin}} f(x) + g(x), \quad (\text{II.40})$$

où f est une fonction convexe propre différentiable, et g est une fonction proximable. Souvent f est le terme d'attache aux données, et g une régularisation. C'est le cas du problème [\(II.22\)](#) par exemple, qui peut s'écrire de cette façon en prenant $f(x) = \|y - \Phi x\|_2^2$ qui est convexe et différentiable, et $g(x) = \lambda\|x\|_1$ qui est convexe, non-différentiable, mais proximable. L'algorithme s'écrit comme suit :

Algorithme II.1 : Forward-backward

-
- 1: **initialiser** $x^{(0)}$, $k = 0$, $\gamma > 0$.
 - 2: **tant que** (non-convergé) **faire**
 - 3: $x^{(k+1)} = \text{prox}_{\gamma g}(x^{(k)} - \gamma \nabla f(x^{(k)}))$
 - 4: $k = k + 1$
 - 5: **renvoyer** $x^{(k)}$
-

Cet algorithme converge vers un minimum global de $f+g$ à condition que $\gamma < \frac{2}{L}$, où L est le coefficient de Lipschitz de ∇f . Le nom "Forward-Backward" vient du fait que cet algorithme alterne entre une étape forward/explicite, la descente de gradient, et une étape backward/implicite, l'application de l'opérateur proximal.

La vitesse de convergence de l'algorithme FB est en $\frac{1}{k}$, mais peut-être accélérée à $\frac{1}{k^2}$ en remplaçant $x^{(k)}$ à la ligne 3 de l'**Algorithme II.1** par une combinaison linéaire bien choisie des itérés précédents [Beck & Teboulle 2009, Nesterov 2013].

Nous avons vu que si g est la fonction caractéristique d'un ensemble convexe, son opérateur proximal est la projection sur cet ensemble. Ainsi, dans ce cas, l'algorithme FB est en tout point similaire à l'algorithme de gradient projeté. L'algorithme FB peut donc être vu comme une généralisation du gradient projeté.

C.3.2 Forward-backward généralisé

Lorsque l'on souhaite appliquer plusieurs régularisations, il arrive souvent que g n'admette pas d'opérateur proximal explicite. On retrouve cette configuration pour le problème suivant par exemple :

$$\underset{x}{\operatorname{argmin}} \|\mathbf{A}x - y\|_2^2 + \lambda \|\mathbf{W}x\|_1 + i_{\geq 0}(x). \quad (\text{II.41})$$

En effet, on prendrait alors $f(x) = \|\mathbf{A}x - y\|_2^2$ et $g(x) = \lambda \|\mathbf{W}x\|_1 + i_{\geq 0}(x)$ mais cette dernière fonction n'admet pas d'opérateur proximal simple à exprimer. g peut cependant s'écrire comme la somme de deux fonctions $g_1(x) = \lambda \|\mathbf{W}x\|_1$ et $g_2(x) = i_{\geq 0}(x)$ qui sont toutes les deux proximables et dont les opérateurs proximaux admettent une forme analytique si \mathbf{W} est orthonormale. Dans ce cas, l'opérateur proximal de g_2 est alors :

$$\operatorname{prox}_{g_2} = \mathbf{W}^T \operatorname{Soft}_\lambda(\mathbf{W}x). \quad (\text{II.42})$$

Cet opérateur applique le seuillage doux aux coefficients dans l'espace transformé avant de revenir dans l'espace direct. On remarquera encore une fois le parallèle avec l'approximation parcimonieuse dans l'espace transformé (II.12) qui effectuait le même processus avec du seuillage dur.

Lorsque g peut être décomposée comme une somme de n fonctions proximables et que le problème s'écrit donc :

$$\underset{x}{\operatorname{argmin}} f(x) + \sum_{i=1}^n g_i(x), \quad (\text{II.43})$$

il est possible d'employer l'algorithme de Forward-Backward généralisé [Raguet *et al.* 2013] qui nécessite la connaissance du gradient de f et des opérateurs proximaux de chacune des fonctions g_i , et peut être défini de manière simplifiée comme suit :

Algorithme II.2 : Forward-backward Généralisé

- 1: **initialiser** $x^{(0)}$, $\gamma > 0$, $\mu > 0$ et $\omega_i \in]0, 1[$ tels que $\sum_{i=1}^n \omega_i = 1$, $k = 0$.
 - 2: **tant que** (non-convergé) **faire**
 - 3: **Pour** i allant de 1 à n **faire**
 - 4: $z^{(k+1,i)} = z^{(k,i)} + \mu \operatorname{prox}_{\frac{\gamma}{\omega_i} g_i} (2x^{(k)} - z^{(k,i)} - \gamma \nabla f(x^{(k)})) - \mu x^{(k)}$
 - 5: $x^{(k+1)} = \sum_{i=1}^n \omega_i z^{(k+1,i)}$
 - 6: $k = k + 1$
 - 7: **renvoyer** $x^{(k)}$
-

Cet algorithme alterne ainsi entre une descente de gradient et une étape proximale sur des variables intermédiaires. La suite $(x^{(k)})$ converge vers le minimum du problème (II.43) lorsque $\gamma < \frac{2}{L}$ et $\mu \in]0, \min(\frac{3}{2}, \frac{1+2/(L\gamma)}{2})[$, où L est la constante de Lipschitz de ∇f .

C.3.3 Douglas-Rachford

L'algorithme de Douglas-Rachford [Douglas & Rachford 1956] vise également à résoudre des problèmes prenant la forme (II.40). À la différence de l'algorithme FB, il ne nécessite cependant plus que f soit différentiable, mais seulement que cette fonction soit proximable, ce qui est une condition plus souple. La minimisation du problème s'effectue alors de la manière suivante ne nécessitant que la connaissance des opérateurs proximaux de f et g :

Algorithme II.3 : Douglas-Rachford

- 1: **initialiser** $x^{(0)}$, $\gamma > 0$ et $\lambda \in]0, 2[$, $k = 0$.
 - 2: **tant que** (non-convergé) **faire**
 - 3: $x^{(k)} = \operatorname{prox}_{\gamma g}(y^{(k)})$
 - 4: $y^{(k+1)} = y^{(k)} + \lambda (\operatorname{prox}_{\gamma f}(2x^{(k)} - y^{(k)}) - x^{(k)})$
 - 5: $k = k + 1$
 - 6: **renvoyer** $x^{(k)}$
-

Plus de détails sont fournis dans [Combettes & Pesquet 2009]. Cet algorithme permet ainsi la minimisation du problème (II.21). En effet, celui-ci peut s'écrire sous la forme $f + g$ avec $f(x) = \|x\|_1$ et $g = i_{\|\Phi \cdot - y\|_2^2 \leq \varepsilon^2}$ et ces deux fonctions sont non-différentiables mais sont toutes les deux proximables. L'opérateur proximal de g n'est cependant pas toujours, explicite suivant le choix de Φ , et son calcul peut nécessiter l'usage de sous-itérations.

C.3.4 Chambolle-Pock

Supposons maintenant que l'on veuille utiliser la parcimonie dans un espace transformé redondant avec une formulation en analyse :

$$\operatorname{argmin}_x \|y - \mathbf{A}x\|_2^2 + \lambda \|\mathbf{W}x\|_1. \quad (\text{II.44})$$

Dans ce cas là, l'opérateur de $g(x) = \lambda \|\mathbf{W}x\|_1$ n'est plus analytique. Son calcul nécessiterait l'usage d'une sous-routine (cf. appendix D) mais cette approche peut se révéler très lourde vu qu'elle nécessite des sous-itérations.

Dans [Chambolle & Pock 2010], Chambolle et Pock ont proposé un algorithme permettant de résoudre le problème précédant sans l'usage de sous itérations. Celui-ci vise à résoudre des problèmes de type :

$$\operatorname{argmin}_x f(x) + g(\mathbf{W}x), \quad (\text{II.45})$$

où f et g sont deux fonctions proximables. Pour ce faire, les auteurs réécrivent ce problème sous la forme primal-dual suivante :

$$\operatorname{argmin}_x \max_y \langle \mathbf{W}x, y \rangle + f(x) + g^*(y), \quad (\text{II.46})$$

où g^* est le conjugué convexe de g , Le conjugué d'une fonction g étant défini par la relation :

$$g^* : x \mapsto \sup_y \langle x, y \rangle - g(y) \quad (\text{II.47})$$

La résolution du problème (II.46) s'effectue comme indiqué dans l'**Algorithme II.4** ci-dessous.

Algorithme II.4 : Chambolle-Pock

- 1: **initialiser** $x^{(0)}, y^{(0)}, \bar{x}^{(0)} = x^{(0)}, \tau > 0, \sigma > 0, k = 0$.
 - 2: **tant que** (non-convergé) **faire**
 - 3: $y^{(k+1)} = \operatorname{prox}_{g^*}(y^{(k)} + \sigma \mathbf{W}\bar{x}^{(k)})$
 - 4: $x^{(k+1)} = \operatorname{prox}_f(x^{(k)} - \tau \mathbf{W}^T y^{(k+1)})$
 - 5: $\bar{x}^{(k+1)} = 2x^{(k+1)} - x^{(k)}$
 - 6: $k = k + 1$
 - 7: **renvoyer** $x^{(k)}$
-

Cette algorithme converge pour $\tau\sigma L^2 < 1$, avec $L = \|\mathbf{W}\|_{s,2}$ la norme opérateur de \mathbf{W} . Ceci s'applique facilement au problème (II.44) qui peut être mis sous la forme :

$$\operatorname{argmin}_x \max_y \langle \mathbf{W}x, y \rangle + \|y - Ax\|_2^2 + i_{\|\cdot\|_\infty \leq \lambda}(y) \quad (\text{II.48})$$

L'algorithme nécessite les opérateurs proximaux de f et de g^* . On pourra noter que $\operatorname{prox}_{g^*}(x) = x - \operatorname{prox}_g(x)$ [Moreau 1962]. Cette propriété permet ainsi de connaître l'opérateur proximal du conjugué d'une fonction dont on connaît l'opérateur proximal. En particulier, l'opérateur proximal de la fonction $g^* = i_{\|\cdot\|_\infty \leq \lambda}$ est ainsi $\operatorname{prox}_{g^*} = x - \operatorname{Soft}_\lambda(x)$.

D Récapitulatif

Ce chapitre a introduit la notion de parcimonie, permettant une représentation efficace de la plupart des signaux naturels, dont les images naturelles par exemple. L'information morphologique que la parcimonie d'un signal apporte permet d'aider à résoudre de nombreux types de problèmes inverses. Son utilisation s'effectue souvent via la minimisation d'un problème régularisé avec un terme induisant la parcimonie du signal recherché. La dernière section du chapitre a présenté des algorithmes proximaux de minimisation de ces problèmes, qui seront abondamment utilisés par la suite. Le tableau **II.1** récapitule les conditions sur les fonctions f et g pour lesquels chacun des algorithmes converge vers la solution du problème suivant :

$$\operatorname{argmin}_x f(x) + g(x). \quad (\text{II.49})$$

Algorithme	f	g
Forward-Backward	convexe et différentiable	proximable
Douglas-Rachford	proximable	proximable
Generalized Forward-Backward	convexe et différentiable	somme de fonctions proximables
Chambolle-Pock	proximable	composée d'une application linéaire et d'une application proximable

Tableau II.1 – Récapitulatif des conditions d'utilisation des algorithmes présentés.

État-de-l'art en NMF

Sommaire

A	Séparation de sources en aveugle	26
A.1	Modélisation des données	26
A.2	Identifiabilité des sources en SSA	30
B	Factorisation en matrices non-négatives	32
B.1	Choix du terme d'attache aux données	33
B.2	Algorithmes de NMF - Approches de descente par blocs . .	34
C	Régularisations en NMF	39
C.1	A priori lisse	39
C.2	A priori parcimonieux	40
D	Récapitulatif des algorithmes	43
E	Critères de séparation	45
E.1	Calcul des critères de séparation	45
E.2	Exemple de décomposition	46

Le concept de parcimonie que nous avons décrit au chapitre précédent a grandement bénéficié au traitement de données spectrométriques, en particulier dans le domaine de la séparation de sources en aveugle (SSA) qui fait l'objet de ce chapitre. Le développement de techniques de spectrométrie comme la spectrométrie infrarouge, de masse et la résonance magnétique nucléaire a permis de caractériser les entités physiques que sont les atomes et les molécules. Cette connaissance permet d'analyser des matériaux ou fluides ou ces entités apparaissent mélangées. Cependant, les signatures des composés du mélange ne sont pas nécessairement connues à l'avance. L'ajout d'une dimension en plus de la dimension spectrale peut alors permettre de les retrouver.

Le type de données obtenues, dites données multispectrales, apparaît dans de nombreux domaines. En astrophysique par exemple, l'instrument OMEGA du satellite Mars Express¹ a cartographié Mars sous plusieurs longueurs d'onde. Le cube de données visible sur la figure [III.1a](#) a ainsi été collecté par ce satellite et correspond à un empilement des images d'un même lieu de Mars, prises sous des longueurs d'onde proches infrarouges, ajoutant une dimension spatiale au spectre infrarouge. Certaines des images obtenues sont montrées sur la figure [III.1b](#). On retrouve le

1. <http://sci.esa.int/mars-express/>, <http://smc.cnes.fr/MEX/Fr/>

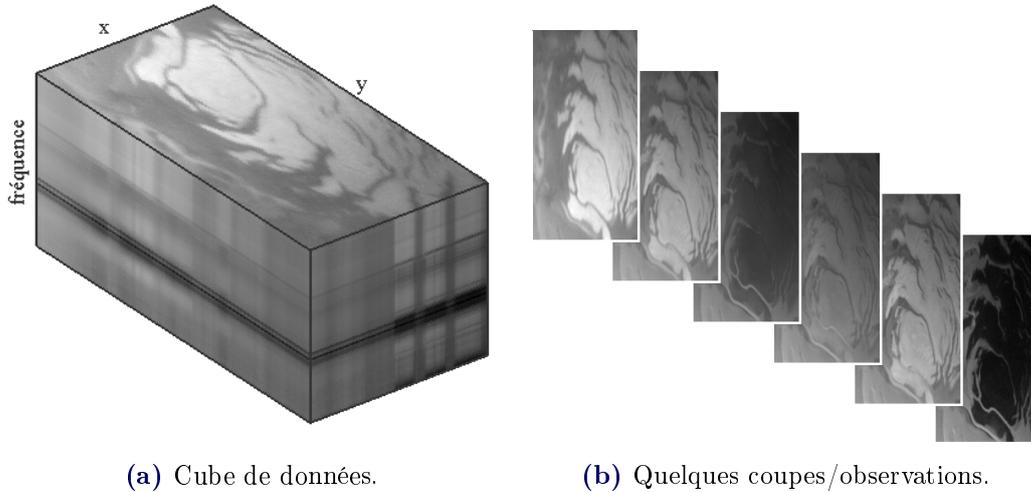


Figure III.1 – Données Mars Express (spectromètre OMEGA).

même type de situation avec des données de chromatographie en phase liquide-spectrométrie de masse telles que celles de la figure [III.2](#). Ces données sont constituées de spectres de masses mesurés à des temps différents en sortie d'une colonne de rétention dans laquelle a été placée un liquide à analyser, ajoutant ainsi une dimension temporelle.

Ce chapitre présente les techniques de séparation de sources en aveugle permettant d'extraire l'information utile de ce type de données, et en particulier les signatures des composés présents, que l'on appelle sources. Il détaille ensuite une technique en particulier, la factorisation en matrices positives (Non-negative Matrix Factorization, NMF) qui vise à prendre en compte que les composantes recherchées sont positives. Nous verrons ainsi que cette technique peut être formulée de différentes manières, et que l'on peut inclure de l'information à propos des sources recherchées via l'ajout d'un terme de régularisation approprié. La dernière section de ce chapitre présente pour finir les outils nécessaires à l'évaluation des résultats fournis par les algorithmes de SSA.

A Séparation de sources en aveugle

A.1 Modélisation des données

La séparation de sources en aveugle a pour but d'identifier des signaux à partir d'observations où ceux-ci sont mélangés. Chacune des observations de Mars Express montrées figure [III.1b](#) dépend en effet de la valeur à la longueur d'onde observée des spectres des espèces chimiques qui sont présents à la surface de la planète, comme l'eau et le CO_2 . Leurs spectres sont montrés respectivement figures [III.3a](#) et [III.3b](#). Ces deux espèces ayant des concentrations spatiales différentes, comme le montrent les figures [III.3c](#) et [III.3d](#), chaque pixel de l'image correspond à un mélange diffé-

rent des composés. Ainsi, chaque observation est fonction des concentrations et des spectres des composés présents. Retrouver les spectres et les concentrations associées à partir des données peut alors être une source d'information importante sur la composition du sol observé.

Les données LC/MS de la figure [III.2](#) contiennent quant à elles de nombreuses molécules appelées métabolites ayant chacune un spectre de masse spécifique. Via la séparation de sources en aveugle, on peut ainsi espérer retrouver ces spectres qui sont initialement mélangés, et donc identifier les constituants du fluide analysé. La figure [III.4](#) montre deux des spectres de masse présents et les profils temporels correspondants. Le traitement des données LC/MS fait l'objet du chapitre [VI](#).

Chacune des m observations (figure [III.1b](#)), mises sous forme de vecteurs lignes $\mathbf{Y}_{i,\cdot} \in \mathbb{R}^{1 \times n}$ quelle que soit leurs dimensionalités initiales, est ainsi fonction des r sources $\mathbf{S}_{j,\cdot} \in \mathbb{R}^{1 \times n}$ (comme celles des figures [III.3a](#) et [III.3b](#) par exemple) sous forme vectorielle également. Les propriétés physiques du système étudié permettent souvent de simplifier ces fonctions en les approximant par des applications linéaires. L'hypothèse des mélanges linéaires instantanés considère ainsi que chaque observation est une combinaison linéaire des signaux recherchés. Étant donné que les observations sont généralement bruitées et que le modèle est souvent imparfait, cela peut être formulé ainsi :

$$\mathbf{Y}_{i,\cdot} = \sum_{j=1}^r \mathbf{A}_{ij} \mathbf{S}_{j,\cdot} + \mathbf{Z}_{i,\cdot}, \quad \forall i \in \{1, \dots, m\}, \quad (\text{III.1})$$

où les \mathbf{A}_{ij} sont les coefficients de mélange, et les $\mathbf{Z}_{i,\cdot}$ permettent de prendre en compte le bruit et les imperfections du modèle. Ce modèle de mélange peut être réécrit de façon plus compacte sous la forme matricielle suivante :

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{Z}, \quad (\text{III.2})$$

où les $\mathbf{Y}_{i,\cdot}$, $\mathbf{S}_{j,\cdot}$ et $\mathbf{Z}_{i,\cdot}$ sont respectivement les lignes de \mathbf{Y} , \mathbf{S} et \mathbf{Z} ; et la matrice de mélange \mathbf{A} est constituée des profils de mélanges $\mathbf{A}_{\cdot,j} \in \mathbb{R}^m$ tels que ceux des

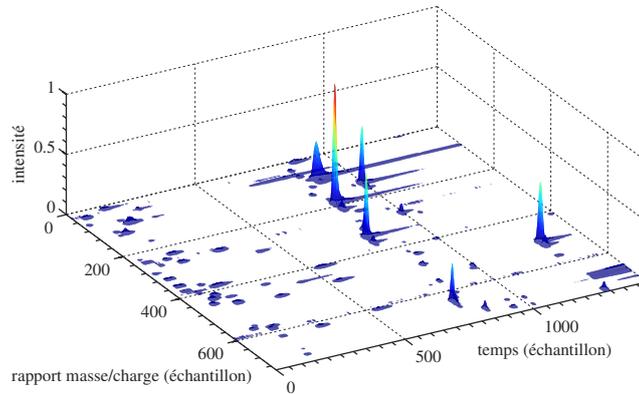


Figure III.2 – Données de chromatographie en phase liquide-spectrométrie de masse (filtrées à des fins de visualisation)

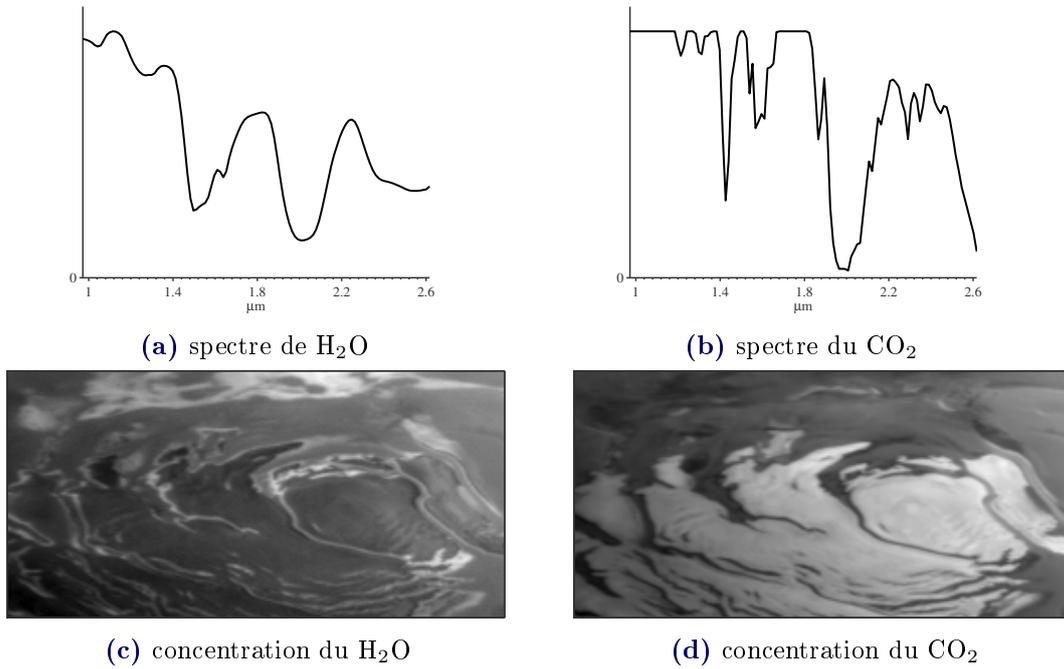


Figure III.3 – Sources dans les données Mars Express (spectromètre OMEGA).

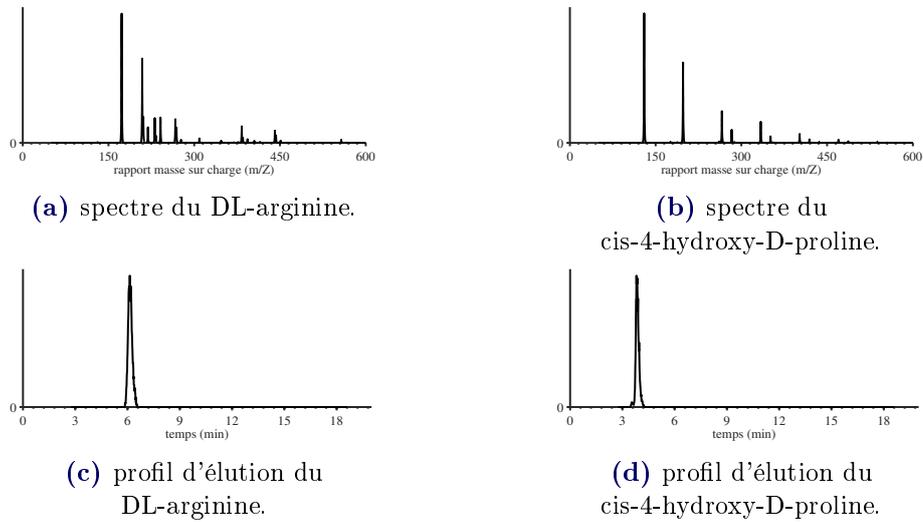


Figure III.4 – Composés du mélange LC/MS (signaux synthétiques).

figures III.3c et III.3d, mis sous forme de vecteurs colonnes. Dans le cas de la séparation de sources en aveugle, à la fois \mathbf{A} , \mathbf{S} et \mathbf{Z} sont inconnus. Si l'on considère que \mathbf{Z} est une matrice aléatoire Gaussienne dont les coefficients sont indépendants et identiquement distribués, l'estimation du maximum de vraisemblance est donnée par :

$$\operatorname{argmin}_{\mathbf{A}, \mathbf{S}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AS}\|_2^2. \quad (\text{III.3})$$

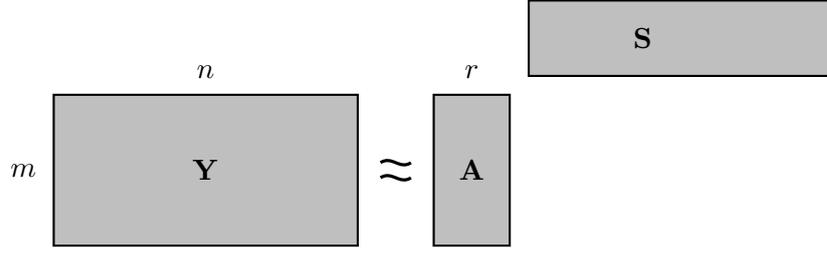


Figure III.5 – Séparation de sources en aveugle.

Le choix du nombre de sources r n'est pas trivial, et doit être estimé ou fixé au préalable en fonction des données. Par la suite, nous considérons le nombre de sources comme connu. Les caractéristiques du modèle sont illustrés à la figure III.5. On remarquera que ce problème ressemble fortement au problème d'apprentissage de dictionnaire représenté figure II.8 en page 17. Ces deux problèmes sont en effet fortement liés [Cardoso & Donoho 1999] mais leurs finalités sont différentes. En apprentissage de dictionnaire, la dimension intérieure de la factorisation p est choisie plus grande que la longueur n des signaux, afin de décomposer les signaux sur un dictionnaire sur-complet. À l'inverse, celle r de la séparation de sources est plus faible que n afin de décomposer les observations en un faible nombre de sources. Souvent, le nombre d'observations m est également grand devant le nombre r de sources recherchées. Dans ce cas là, la SSA est une technique de réduction de dimension.

Le problème (III.3) présente une infinité de solutions. Celles-ci peuvent être explicitées à l'aide d'une décomposition en valeurs singulières (Singular Value Decomposition, SVD), qui permet d'écrire \mathbf{Y} sous la forme :

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad (\text{III.4})$$

où $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_m$, $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$ (matrices orthonormales), et \mathbf{D} est diagonale à coefficients positifs ou nuls ordonnés de manière décroissante. À un nombre de sources recherchées r donné et lorsque que \mathbf{Y} n'a pas de valeurs propres multiples, le théorème d'Eckart-Young stipule que la meilleure approximation de rang r de \mathbf{Y} au sens de la norme ℓ_2 est unique. Elle est alors donnée par la matrice $\mathbf{U}_{:,1:r}\mathbf{D}_{1:r,1:r}\mathbf{V}_{:,1:r}^T$, où la notation $\mathbf{X}_{:,1:r}$ correspond à la matrice formée des r premières colonnes de \mathbf{X} . On peut ainsi prendre comme solution de (III.3) la factorisation :

$$(\mathbf{A}, \mathbf{S}) = (\mathbf{U}_{:,1:r}\mathbf{D}_{1:r,1:r}, \mathbf{V}_{:,1:r}^T), \quad (\text{III.5})$$

Cependant, cette factorisation n'est quant à elle pas unique, et le problème (III.3) présente une infinité de solutions équivalentes en considérant les matrices inversibles \mathbf{P} [Ho 2008]. En effet, si (\mathbf{A}, \mathbf{S}) est une solution de (III.3), alors $(\mathbf{A}\mathbf{P}^{-1}, \mathbf{P}\mathbf{S})$ l'est aussi puisque l'on a toujours :

$$\mathbf{A}\mathbf{S} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{S}. \quad (\text{III.6})$$

Deux types de matrices inversibles ont une importance particulière. En effet, à moins d'avoir des informations spécifiques à chaque source, le problème présentera

toujours au moins les invariances suivantes du fait que l'on n'observe que le produit \mathbf{AS} :

(i) *les permutations* : si π est une permutation, $\mathbf{AS} = \sum_i \mathbf{A}_{:,i} \mathbf{S}_i = \sum_i \mathbf{A}_{:,\pi(i)} \mathbf{S}_{\pi(i)}$, donc l'ordre des lignes de \mathbf{S} et des colonnes de \mathbf{A} peut être modifié de façon jointe pour obtenir une nouvelle solution.

(ii) *les changements d'échelle* : si $\mathbf{\Delta} \in \mathbb{R}^{r \times r}$ est une matrice diagonale à coefficients non nuls, $\mathbf{AS} = (\mathbf{A}\mathbf{\Delta})(\mathbf{\Delta}^{-1}\mathbf{S})$, ainsi, si (\mathbf{A}, \mathbf{S}) est une solution de (III.8), $(\mathbf{A}\mathbf{\Delta}, \mathbf{\Delta}^{-1}\mathbf{S})$ l'est aussi.

A.2 Identifiabilité des sources en SSA

Même en faisant abstraction des deux indéterminations mentionnées ci-dessus, le problème précédent présente toujours une infinité de solutions qui ne sont pas nécessairement intéressantes pour une application donnée. Il est alors courant d'utiliser des critères permettant d'identifier les sources recherchées parmi les solutions possibles. On dira ainsi que des sources sont identifiables si on peut les retrouver à une permutation et un facteur près, qui sont les deux invariances présentées au paragraphe précédent. Plusieurs branches de la SSA se sont développées en utilisant des informations différentes afin d'identifier les sources, comme l'indépendance statistique, la parcimonie et la positivité des sources et des mélanges.

A.2.1 Analyse en composantes indépendantes

Souvent, les sources correspondent à des signaux ayant des origines différentes et peuvent donc être considérées comme statistiquement indépendantes. La SVD fournit des solutions qui sont décorréelées. L'analyse en composantes indépendantes (ACI, [Jutten & Herault 1991], [Comon 1994]) va plus loin et vise donc à rechercher des sources statistiquement indépendantes. À partir de la solution blanchie de la SVD $\mathbf{S} = \mathbf{V}_{:,1:r}^T$, la plupart des algorithmes d'ACI cherchent à trouver une transformation \mathbf{P} de l'espace des rotations $\{\mathbf{P} \in \mathbb{R}^{r \times r}, \mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}_r\}$ qui rende les lignes de $\mathbf{P}\mathbf{S}$ aussi indépendantes que possible selon un critère donné (voir [Lee et al. 2000] pour plus de détails sur ces critères et leurs liens). Parmi les exemples les plus connus d'implémentation, on trouve Jade [Cardoso 1999] dont le critère consiste à diagonaliser la matrice des cumulants d'ordre 4, et FastICA [Hyvärinen 1999, Hyvärinen et al. 2001] qui tend à minimiser l'information mutuelle des sources obtenues. Si les profils de mélange forment une famille libre et qu'au plus une des sources est Gaussienne, l'ACI permet l'identifiabilité des sources recherchées [Eriksson & Koivunen 2004]. La présence de bruit dans les données peut cependant rapidement compliquer cette identification [Davies 2004]. Parmi les nombreuses applications possibles, on pourra citer entre autres l'étude des données Mars Express de la figure III.1a [Moussaoui et al. 2008].

A.2.2 SSA de sources parcimonieuses

Les sources à rechercher ne sont cependant pas nécessairement décorrélées. La parcimonie peut servir de mesure de contraste entre les sources et permettre de les séparer [Zibulevsky & Pearlmutter 2001], en apportant une information structurelle au lieu de l'hypothèse d'indépendance. Elle peut aussi s'appliquer à une très grande variété de signaux comme nous l'avons vu au chapitre précédent. Ce type d'approche a par exemple été testée sur des signaux provenant d'électro-encéphalogrammes [Li *et al.* 2003], et est bien adaptée à la séparation de sources sous déterminées (cas $m < r$) [Georgiev *et al.* 2005, Georgiev *et al.* 2007].

Dans [Bobin *et al.* 2007] et [Bobin *et al.* 2008], Bobin et al. introduisent l'hypothèse de diversité morphologique. Celle-ci stipule qu'il existe un seuil au dessus duquel les différentes sources ont des supports disjoints dans l'espace qui les rend parcimonieuses. L'exploitation des supports disjoints a souvent été utilisée en séparation de sources, notamment dans le cas sous-déterminé [Rivet *et al.* 2007]. L'hypothèse de diversité morphologique est néanmoins plus souple, ne nécessitant pas la connaissance préalable des supports et autorisant de faibles corrélations. Les sources peuvent alors être identifiées en recherchant leurs quelques grands coefficients. Dans la pratique, cette hypothèse semble très souvent respectée pour des signaux parcimonieux, et cette approche obtient d'excellents résultats et se montre robuste au bruit. Elle a également été utilisée sur les données Mars Express [Moudden & Bobin 2011].

A.2.3 Factorisation en matrices positives

Dans de nombreuses applications, nous savons de plus que les sources et les mélanges sont à valeurs positives. La factorisation en matrices positives ("positive matrix factorization" [Paatero & Tapper 1994] ou plus couramment "non-negative matrix factorization", NMF [Lee & Seung 1999]) vise ainsi à rechercher des matrices de sources \mathbf{S} et de mélange \mathbf{A} dont tous les coefficients sont positifs. Lee & Seung mirent l'accent sur l'importance de la positivité qui permet d'obtenir une représentation "par morceaux" des données. Dans [Donoho & Stodden 2003], Donoho & Stodden expliquent que le problème de NMF peut être vu géométriquement comme la recherche d'un cône simplicial qui contiendrait l'ensemble des données et serait lui-même contenu dans l'orthant positif (l'ensemble des points à coefficients positifs ou nuls). Le cône simplicial d'une famille (s_i) de vecteur est donné par $C = \{x = \sum_i a_i s_i, a_i \in \mathbb{R}^+\}$. Ils proposent en particulier l'hypothèse de séparabilité, ou de dominance locale, qui stipule que l'on peut trouver les sources parmi les observations (c'est-à-dire que chacune des sources soit observée au moins une fois de façon "pure") et/ou que l'on retrouve les directions de mélange parmi les colonnes de \mathbf{Y} . Cette hypothèse fait parmi des conditions permettant de montrer l'unicité du cône simplicial recherché, et par là même de la décomposition.

Dans la suite de cette thèse nous nous intéressons particulièrement à la factorisation en matrices positives, qui est par exemple bien adaptée au traitement des données d'imagerie hyperspectrale [Ma *et al.* 2014] dont encore une fois celles de Mars Express [Moussaoui *et al.* 2008], de résonance magnétique nucléaire (RMN), [Snyder

et al. 2008, Toumi *et al.* 2013]) ou de LC/MS [Dubroca *et al.* 2012, Ouedraogo 2012]. En effet, dans ce dernier cas, les spectres de masse des composés physiques sont nécessairement positifs, et les mélanges observés en LC/MS sont dus à des concentrations, qui sont positives également. Les hypothèses d'identifiabilité proposées dans [Donoho & Stodden 2003] ne sont cependant pas toujours vérifiées en pratique. Nous verrons dans la suite de ce chapitre qu'il est possible de combiner cette approche de séparation de sources en aveugle avec celles utilisant l'indépendance ou la parcimonie que nous avons exposées précédemment, permettant ainsi d'améliorer les chances de retrouver une factorisation pertinente. La section qui suit détaille les principales démarches utilisées pour effectuer la NMF d'une matrice.

B Factorisation en matrices non-négatives

L'hypothèse de positivité des sources et des mélanges apparaît naturellement dans beaucoup d'autres applications que la LC/MS, comme la fouille de texte [Berry & Browne 2005], le clustering [Kim & Park 2008], le traitement de signaux audio [Févotte *et al.* 2009], les données de résonance magnétique nucléaire (RMN) [Toumi *et al.* 2013] ou l'imagerie hyperspectrale [Jia & Qian 2009]. Les données Mars Express précédente (figure III.1) en sont ainsi un nouvel exemple. La résolution du problème de NMF est cependant très complexe puisqu'elle est NP-difficile [Vavasis 2009]. Plusieurs démarches ont été envisagées pour tenter de le résoudre.

Ainsi, la première démarche consiste à se servir de la géométrie du problème [Babaie-Zadeh *et al.* 2004]. Pour la plupart, les algorithmes qui en découlent trouvent leur justification dans le travail de Donoho & Stodden [Donoho & Stodden 2003] et ainsi visent à rechercher des observations ou mélanges "purs", c'est-à-dire à trouver les rayons extrêmes du cône simplicial contenant les données [Chan *et al.* 2008, Arora *et al.* 2012, Bittorf *et al.* 2012, Recht *et al.* 2012, Liu & Aeron 2013, Gillis & Vavasis 2014]. Ces approches ont l'avantage de pouvoir être adaptée à des données de grandes dimensions (cf. [Liu & Aeron 2013, Gillis & Vavasis 2014] par exemple). Ils sont cependant tributaires du fait que chaque source doit apparaître de manière pure dans les données et prennent difficilement en compte l'influence du bruit. D'autres approches géométrique relâche la condition de séparabilité en cherchant par exemple le cône de volume minimum [Ouedraogo *et al.* 2012], mais ces approches sont également sensibles au bruit.

La démarche de l'ACI peut aussi être adaptée au cas de sources et de mélanges positifs [Plumbley & Oja 2004, Raj *et al.* 2010]. Celle-ci vise ainsi à rechercher des sources positives qui soient également indépendantes. Ces approches partagent cependant les mêmes faiblesses que l'ACI, c'est-à-dire la difficulté de la gestion du bruit, et un modèle qui n'est pas nécessairement vérifié en pratique. Les méthodes de type MCMC peuvent également être utilisées afin de rechercher les sources et les directions de mélange à l'aide de modèles bayésiens sur \mathbf{A} et \mathbf{S} [Moussaoui *et al.* 2006, Schmidt *et al.* 2009]

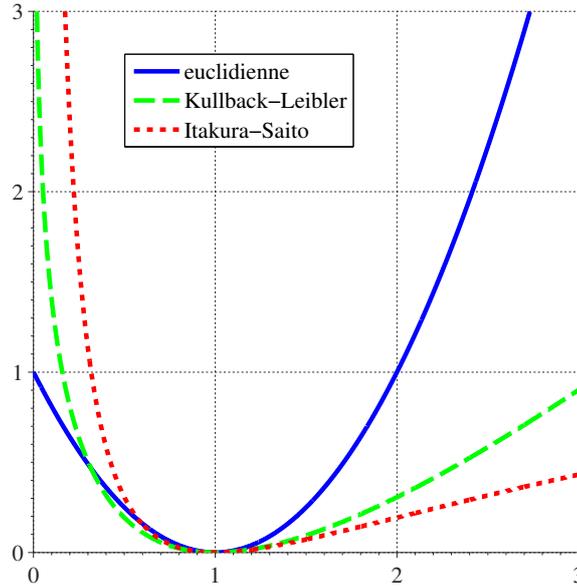


Figure III.6 – Exemples de divergences, avec $y = 1$.

Une dernier type de démarche vise à formuler la NMF comme un problème inverse en \mathbf{A} et \mathbf{S} et à le résoudre via des algorithmes d'optimisation. Le problème de factorisation en matrices positives prend en effet la forme suivante :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \mathcal{D}(\mathbf{Y} || \mathbf{AS}). \quad (\text{III.7})$$

où $\mathcal{D}(\cdot || \cdot)$ est une divergence qui mesure la différence entre les données \mathbf{Y} et l'approximation \mathbf{AS} que l'on en fait. Celle-ci a pour minimum 0 et l'atteint uniquement pour $\mathbf{Y} = \mathbf{AS}$. L'avantage de cette approche est sa flexibilité. En effet, le terme d'attache aux données $\mathcal{D}(\mathbf{Y} || \mathbf{AS})$ peut tout d'abord être adapté au problème envisagé. Il est également possible de rajouter des régularisations afin de favoriser les sources présentant des caractéristiques recherchées. Cette démarche est donc la plus adaptée à l'utilisation d'a priori parcimonieux sous des formes variées et c'est donc celle que nous choisirons d'utiliser dans les prochains chapitres. Les sections suivantes portent ainsi sur les différents études envisageant la NMF sous cet aspect et présentent les différents termes d'attache aux données, algorithmes, et régularisations couramment utilisés.

B.1 Choix du terme d'attache aux données

Le choix de la divergence est déterminant autant pour son impact sur la modélisation des observations que sur les algorithmes d'optimisation applicables pour la minimisation de (III.7). Les divergences utilisées sont séparables, de telle sorte que pour deux matrices \mathbf{Y} et \mathbf{X} , $\mathcal{D}(\mathbf{Y} || \mathbf{X}) = \sum_{ij} d(\mathbf{X}_{ij} || \mathbf{Y}_{ij})$. Plusieurs possibilités pour d ont été étudiées en NMF et sont illustrées figure III.6 :

Distance euclidienne (ℓ_2) : $d(y||x) = (y - x)^2$.

La distance euclidienne, initialement proposée dans [Paatero & Tapper 1994] et [Lee & Seung 2001], est le choix qui est le plus souvent considéré. Comme nous l'avons vu dans le cas de la BSS, celui-ci apparaît naturellement via le maximum de vraisemblance quand les sources sont contaminées par du bruit Gaussien. En raison de cette caractéristique ainsi que de la simplicité de la fonction quadratique, c'est la divergence qui est la plus souvent utilisée [Hoyer 2002, Merritt & Zhang 2005, Cichocki *et al.* 2007].

Kullback-Leibler (KL) : $d(y||x) = y \log(\frac{y}{x}) - y + x$.

La divergence de Kullback-Leibler est, à l'origine, utilisée pour quantifier la similarité entre des densités de probabilité. Elle a été proposée en NMF par Lee & Seung dans leur premier article [Lee & Seung 1999] et réutilisée de nombreuses fois [Virtanen 2007, Zdunek & Cichocki 2007c]. Cette divergence peut s'interpréter comme un maximum de vraisemblance dans le cas d'une contamination par du bruit de Poisson.

Itakura-Saito (IS) : $d(y||x) = \frac{y}{x} - \log(\frac{y}{x}) - 1$.

La divergence d'Itakura-Saito est également très utilisée dans la communauté de traitement de la parole car elle a la propriété d'être invariable par changement d'échelle [Févotte *et al.* 2009]. Elle peut également être vue comme le maximum de vraisemblance dans le cas d'une contamination par du bruit multiplicatif gamma.

Généralisation : Dans [Dhillon & Sra 2005], Dhillon et Sra introduisent des algorithmes multiplicatifs s'appliquant à la NMF formulée avec les divergences de Bregman. La distance euclidienne, la divergence de Kullback-Leibler et la divergence d'Itakura-Saito en sont des cas particuliers. Plus généralement, Cichocki *et al.* expliquent dans [Cichocki *et al.* 2006] qu'il est également possible d'utiliser les divergences ϕ de Csiszár et les divergences α d'Amari, et montrent comment les traiter à l'aide d'algorithmes multiplicatifs.

La suite de notre étude se concentre principalement sur une contamination par du bruit Gaussien additif, et l'accent sera donc mis sur la formulation avec la distance euclidienne :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \|\mathbf{Y} - \mathbf{AS}\|_2^2. \quad (\text{III.8})$$

La distance euclidienne a également l'avantage d'être plus facile à manipuler, étant simplement quadratique, et plus régulière que les autres divergences, ayant par exemple l'avantage d'avoir un gradient Lipschitzien.

B.2 Algorithmes de NMF - Approches de descente par blocs

Une fois le problème formulé vient le choix des algorithmes de minimisation. Ceux-ci utilisent les méthodes de descente par blocs [Sardy *et al.* 2000], qui alternent

généralement entre des minimisations en \mathbf{A} à \mathbf{S} fixé, et des minimisations en \mathbf{S} à \mathbf{A} fixé. En effet, alors que le problème (III.7) est extrêmement compliqué à manipuler en raison du produit \mathbf{AS} qui le rend non-convexe, le sous-problème en \mathbf{A} , qui s'écrit :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}} \|\mathbf{Y} - \mathbf{AS}\|_2^2, \quad (\text{III.9})$$

et le sous-problème équivalent en \mathbf{S} :

$$\operatorname{argmin}_{\mathbf{S} \geq \mathbf{0}} \|\mathbf{Y} - \mathbf{AS}\|_2^2, \quad (\text{III.10})$$

sont tous les deux plus aisés à traiter car ils sont convexes. À chaque étape, on peut cependant choisir de rechercher un minimum global de chacun de ces sous-problèmes ou de ne chercher qu'à faire décroître la fonction de coût. On peut également choisir de tenir compte des contraintes pendant les minimisations ou de ne les appliquer que dans un deuxième temps à l'aide d'une projection. On distingue ainsi quatre catégories d'algorithmes qui découlent de ces choix :

(i) globale-projetée : les sous-problèmes sont résolus de manière globale à chaque itération sans considérer les contraintes, puis la solution est projetée sur les contraintes.

(ii) partielle-contrainte : les sous-problèmes sous contraintes sont résolus de manière partielle à chaque itération.

(iii) partielle-projetée : à chaque itération, les sous-problèmes sont résolus de manière partielle sans considérer les contraintes, et le résultat est ensuite projeté sur les contraintes.

(iv) globale-contrainte : les sous-problèmes sous contraintes sont résolus de manière globale à chaque itération.

Chacune de ces catégories est détaillée ci-dessous avec des exemples d'algorithmes.

B.2.1 Minimisation globale puis projection

Le premier algorithme proposé pour résoudre le problème de NMF est l'algorithme des moindres carrés alternés (Alternating Least Square, ALS [Paatero & Tapper 1994]). L'**Algorithme III.1** résume les étapes de cette minimisation, avec la notation $[x]_+ = \max(x, 0)$ pour la projection sur l'orthant positif.

Aux étapes 3 et 4, cet algorithme résout de manière exacte chaque sous-problème formulé avec la distance euclidienne sans tenir compte de la contrainte de positivité, via un moindre carré, puis projette la solution obtenue afin de la rendre à nouveau positive via l'opérateur $[\cdot]_+$. ALS est très utilisé en raison de sa facilité d'implémentation et de son efficacité à faire décroître la fonction de coût. Cet algorithme ne converge cependant pas nécessairement.

Algorithme III.1 : ALS**Paramètre(s) d'entrée** : \mathbf{Y}

- 1: initialiser $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}, k = 0$.
- 2: tant que (non-convergé) faire
- 3: $\mathbf{A}^{(k+1)} = \left[\mathbf{Y}\mathbf{S}^{(k)T} (\mathbf{S}^{(k)}\mathbf{S}^{(k)T})^{-1} \right]_+$
- 4: $\mathbf{S}^{(k+1)} = \left[(\mathbf{A}^{(k+1)T}\mathbf{A}^{(k+1)})^{-1} \mathbf{A}^{(k+1)T}\mathbf{Y} \right]_+$
- 5: $k = k + 1$
- 6: renvoyer $\mathbf{A}^{(k)}, \mathbf{S}^{(k)}$

B.2.2 Minimisation partielle sous contraintes

Lee & Seung furent les premiers à concevoir un algorithme convergeant pour la résolution de la NMF, en proposant de mettre à jour \mathbf{A} et \mathbf{S} via une descente de gradient pondérée [Lee & Seung 1999, Lee & Seung 2001]. Les poids choisis permettent de s'assurer que les étapes de descente de gradient ne font pas augmenter la fonction de coût (III.7) tout en faisant en sorte de garder les coefficients de \mathbf{A} et \mathbf{S} positifs, puisque la mise-à-jour peut s'exprimer comme le produit terme-à-terme de matrices positives. En utilisant les notations \odot pour la multiplication terme-à-terme de matrices et \oslash leur division terme-à-terme, l'algorithme de mise-à-jour multiplicative est donné par l'Algorithme III.2 dans le cas d'une divergence euclidienne.

Algorithme III.2 : Mise-à-jour multiplicative (ℓ_2)**Paramètre(s) d'entrée** : $\mathbf{Y} \geq \mathbf{0}$

- 1: initialiser $\mathbf{A}^{(0)} \geq \mathbf{0}, \mathbf{S}^{(0)} \geq \mathbf{0}, k = 0$.
- 2: tant que (non-convergé) faire
- 3: $\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} \odot (\mathbf{Y}\mathbf{S}^{(k)T}) \oslash (\mathbf{A}^{(k)}\mathbf{S}^{(k)}\mathbf{S}^{(k)T})$
- 4: $\mathbf{S}^{(k+1)} = \mathbf{S}^{(k)} \odot (\mathbf{A}^{(k+1)T}\mathbf{Y}) \oslash (\mathbf{A}^{(k+1)T}\mathbf{A}^{(k+1)}\mathbf{S}^{(k)})$
- 5: $k = k + 1$
- 6: renvoyer $\mathbf{A}^{(k)}, \mathbf{S}^{(k)}$

L'algorithme de mise-à-jour multiplicative est généralement considéré comme un standard en NMF autant pour son aspect historique que pour sa facilité d'utilisation, étant donné qu'il converge sans avoir besoin de régler le moindre paramètre. Cet algorithme est cependant considéré comme lent même s'il peut être accéléré [Gonzalez & Zhang 2005]. Enfin, un problème soulevé dans de nombreux articles est que le point limite de cet algorithme n'est pas nécessairement un point stationnaire [Gonzalez & Zhang 2005, Berry *et al.* 2007, Lin 2007a]. Cette stationnarité est pourtant recherchée car elle est une condition nécessaire à l'obtention d'un minimum local, et donc d'autant plus pour l'obtention d'un minimum global.

On retrouve également dans cette catégorie beaucoup d'algorithmes utilisant une divergence comme celles d'Itakura-Saito ou de Kullback-Leibler. Ces divergences sont plus difficiles à manipuler que la distance euclidienne. Les algorithmes de mise-

à-jour multiplicative, qui ne nécessitent pas de paramètres, sont alors particulièrement bien adaptés [Dhillon & Sra 2005, Cichocki *et al.* 2006]. Ainsi, l’**Algorithme III.3** montre le cas de la mise-à-jour multiplicative pour la divergence de Kullback-Leibler tel que proposé dans [Lee & Seung 2001]. Dans ce dernier algorithme, le trait de division est utilisé de manière équivalente à la division terme à terme \odot afin de simplifier l’écriture, et la matrice $\mathbf{1}_{mn}$ est la matrice de taille $m \times n$ dont tous les coefficients ont pour valeur 1. Des approches de majoration-minimisation sont également utilisées pour gérer ce type de divergences [Févotte 2011, Essid & Févotte 2013]. Celle-ci vise à majorer localement la fonction de coût par une fonction plus simple, que l’on peut ensuite minimiser sous contraintes.

Algorithme III.3 : Mise-à-jour multiplicative (KL)

Paramètre(s) d’entrée : $\mathbf{Y} \geq \mathbf{0}$

- 1: **initialiser** $\mathbf{A}^{(0)} \geq \mathbf{0}$, $\mathbf{S}^{(0)} \geq \mathbf{0}$, $k = 0$.
 - 2: **tant que** (non-convergé) **faire**
 - 3: $\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} \odot \left(\frac{\mathbf{Y}}{\mathbf{A}^{(k)}\mathbf{S}^{(k)}} \mathbf{S}^{(k)T} \right) \odot (\mathbf{1}_{mn} \mathbf{S}^{(k)T})$
 - 4: $\mathbf{S}^{(k+1)} = \mathbf{S}^{(k)} \odot \left(\mathbf{A}^{(k+1)T} \frac{\mathbf{Y}}{\mathbf{A}^{(k+1)}\mathbf{S}^{(k)}} \right) \odot (\mathbf{A}^{(k+1)T} \mathbf{1}_{mn})$
 - 5: $k = k + 1$
 - 6: **renvoyer** $\mathbf{A}^{(k)}$, $\mathbf{S}^{(k)}$
-

On peut finalement trouver dans cette catégorie les algorithmes de point intérieur comme celui de [Merritt & Zhang 2005]. Chaque étape de cet algorithme est une étape de descente de gradient dont le pas est choisi afin de conserver la positivité du point courant.

B.2.3 Minimisation partielle puis projection

Dans la catégorie précédente, chaque étape de descente restait dans le domaine positif. Dans cette nouvelle catégorie, on retrouve les algorithmes utilisant une étape de minimisation partielle en \mathbf{A} ou \mathbf{S} sans considérer la contrainte de positivité, suivie d’une projection sur l’espace des matrices à coefficients positifs. Dans [Chu *et al.* 2004] et [Xu 2011], la minimisation partielle s’effectue via une étape de descente de gradient. Chaque mise-à-jour correspond ainsi à une étape de gradient projeté. Comme pour les autres catégories d’algorithmes de NMF, ces mises-à-jour sont alternativement effectuées en \mathbf{A} et \mathbf{S} , comme indiqué dans l’**Algorithme III.4**.

Dans [Zdunek & Cichocki 2007c], les auteurs proposent une méthode du second ordre, utilisant la matrice Hessienne du terme d’attache aux données, afin d’obtenir une minimisation plus rapide. Ils projettent ensuite également sur la contrainte de positivité dans un deuxième temps.

Algorithme III.4 : NMF par gradient projeté

Paramètre(s) d'entrée : \mathbf{Y}

- 1: initialiser $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}, k = 0$.
 - 2: tant que (non-convergé) faire
 - 3: choisir le pas $\gamma_A^{(k)}$
 - 4: $\mathbf{A}^{(k+1)} = \left[\mathbf{A}^{(k)} - \gamma_A^{(k)} (\mathbf{A}\mathbf{S} - \mathbf{Y}) \mathbf{S}^{(k)T} \right]_+$
 - 5: choisir le pas $\gamma_S^{(k)}$
 - 6: $\mathbf{S}^{(k+1)} = \left[\mathbf{S}^{(k)} - \gamma_S^{(k)} \mathbf{A}^{(k)T} (\mathbf{A}\mathbf{S} - \mathbf{Y}) \right]_+$
 - 7: $k = k + 1$
 - 8: renvoyer $\mathbf{A}^{(k)}, \mathbf{S}^{(k)}$
-

B.2.4 Minimisation globale sous contraintes

Pour finir, il est possible de résoudre exactement chaque sous-problème contraint, sur le modèle de l'**Algorithme III.5**.

Algorithme III.5 : Minimisation exacte sous contraintes

Paramètre(s) d'entrée : \mathbf{Y}

- 1: initialiser $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}, k = 0$.
 - 2: tant que (non-convergé) faire
 - 3: $\mathbf{A}^{(k+1)} = \underset{\mathbf{A} \geq \mathbf{0}}{\operatorname{argmin}} \mathcal{D}(\mathbf{Y} || \mathbf{A}\mathbf{S}^{(k)})$
 - 4: $\mathbf{S}^{(k+1)} = \underset{\mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \mathcal{D}(\mathbf{Y} || \mathbf{A}^{(k+1)}\mathbf{S})$
 - 5: $k = k + 1$
 - 6: renvoyer $\mathbf{A}^{(k)}, \mathbf{S}^{(k)}$
-

Dans [Lin 2007b] par exemple, Lin utilise l'algorithme du gradient projeté pour résoudre chaque sous-problème. Cela est équivalent à l'utilisation de l'algorithme forward-backward (voir section II-C.3.1, page 20), comme montré dans [Tandon & Sra 2010]. En pratique, ces algorithmes consistent à répéter l'étape 4 de l'**Algorithme III.4** jusqu'à convergence vers le minimum du sous-problème en \mathbf{A} , puis de même avec l'étape 6 pour la mise-à-jour de \mathbf{S} , et ainsi de suite. Guan et al. proposèrent ensuite d'utiliser des méthodes du premier-ordre plus rapide [Guan et al. 2012] en s'inspirant de [Nesterov 2013] et équivalente à l'utilisation de FISTA [Beck & Teboulle 2009] pour la résolution des sous-problèmes.

Les algorithmes ci-dessus nécessitent l'utilisation de sous-routines pour résoudre les sous-problèmes. Dans l'algorithme d'ALS hiérarchique (Hierarchical ALS, HALS [Cichocki et al. 2007, Cichocki et al. 2009]) les auteurs ont proposé d'utiliser des blocs différents pour la descente par bloc. Ainsi, au lieu de mettre à jour \mathbf{A} entièrement et \mathbf{S} entièrement, les colonnes de \mathbf{A} et les lignes de \mathbf{S} sont mises à jour une à une. Chacun des sous-problèmes sous contraintes de positivité obtenu a alors une solution explicite très simple à calculer via un produit matrice-vecteur comme indiqué

dans l’**Algorithme III.6**. Cette mise-à-jour est ainsi significativement plus rapide comparée aux multiples produits matrice-matrice nécessaires avec des approches de gradient itéré par exemple.

Algorithme III.6 : HALS

Paramètre(s) d’entrée : \mathbf{Y}

- 1: **initialiser** $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}, k = 0$.
 - 2: **tant que** (non-convergé) **faire**
 - 3: $\mathbf{R} = \mathbf{Y} - \mathbf{A}^{(k)}\mathbf{S}^{(k)}$ %calcul du résidu
 - 4: **Pour** s allant de 1 à r **faire**
 - 5: $\mathbf{R} = \mathbf{R} + \mathbf{A}_{\cdot,s}^{(k)}\mathbf{S}_{s,\cdot}^{(k)}$ %résidu sans la source s
 - 6: $\mathbf{A}_{\cdot,s}^{(k+1)} = \frac{1}{\mathbf{S}_{s,\cdot}^{(k)}\mathbf{S}_{s,\cdot}^{(k)T}} \left[\mathbf{R}\mathbf{S}_{s,\cdot}^{(k)T} \right]_+$
 - 7: $\mathbf{S}_{s,\cdot}^{(k+1)} = \frac{1}{\mathbf{A}_{\cdot,s}^{(k+1)T}\mathbf{A}_{\cdot,s}^{(k+1)}} \left[\mathbf{A}_{\cdot,s}^{(k+1)T}\mathbf{R} \right]_+$
 - 8: $\mathbf{R} = \mathbf{R} - \mathbf{A}_{\cdot,s}^{(k+1)}\mathbf{S}_{s,\cdot}^{(k+1)}$ %mise-à-jour du résidu
 - 9: $k = k + 1$
 - 10: **renvoyer** $\mathbf{A}^{(k)}, \mathbf{S}^{(k)}$
-

C Régularisations en NMF

Les solutions obtenues par les algorithmes de minimisation présentés ci-dessus ne correspondent cependant pas nécessairement aux solutions attendues pour une application donnée. En effet, le problème de problème (III.8) étant non-convexe, il peut présenter un grand nombre de points stationnaires qui peuvent avoir des intérêts variés. Étant de plus NP-difficile [Vavasis 2009], trouver une solution optimale se révèle souvent hors de portée. Donoho et Stodden [Donoho & Stodden 2003] n’ont ainsi prouvé l’identifiabilité des sources que dans le cas où celles-ci apparaissent de manière pure, ce qui n’est pas toujours le cas. L’ajout de contraintes additionnelles peut alors encore une fois être bénéfique afin de retrouver les sources recherchées. Dans cette section, nous présentons les différentes régularisations généralement utilisées en plus de la contrainte de positivité dans des algorithmes de NMF, notamment les régularisations lisses et parcimonieuses.

C.1 A priori lisse

Certains algorithmes de NMF cherchent à favoriser l’estimation de sources lisses via une régularisation appropriée. De telles régularisations peuvent être bénéfiques pour la recherche de sources comme le spectre infrarouge de l’eau à la figure III.3a. À l’inverse, elles ne sont pas appropriées pour les signaux comportant des singularités tels comme les pics des sources de LC/MS de la figure III.4 ou les discontinuités du spectre infrarouge du CO₂ figure III.3b.

Dans [Zdunek & Cichocki 2007b] et [Zdunek & Cichocki 2007a], Zdunek & Cichocki ajoutent ainsi au problème (III.8) un terme de lissage correspondant à une

fonction potentiel :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \|\mathbf{Y} - \mathbf{AS}\|_2^2 + \alpha U_\delta(\mathbf{S}), \quad (\text{III.11})$$

où U_δ est par exemple définie comme la fonction de potentiel de Green :

$$U_\delta(\mathbf{S}) = \delta \sum_{i=1}^r \sum_{j=2}^n \log \left(\cosh \left(\frac{\mathbf{S}_{i,j} - \mathbf{S}_{i,j-1}}{\delta} \right) \right). \quad (\text{III.12})$$

En pénalisant les grandes différences entre la valeur d'un coefficient et celle de ses voisins, cette régularisation privilégie ainsi bien la recherche de sources lisses. Dans cet algorithme, la minimisation est effectuée en utilisant des mises-à-jour multiplicatives. Ajuster les paramètres δ et α peut cependant se révéler fastidieux. Le code implémentant cet algorithme est disponible en ligne².

Les différences quadratiques peuvent aussi être utilisées comme terme de lissage :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \mathcal{D}(\mathbf{Y} \|\mathbf{AS}) + \alpha \sum_{j=2}^n \|\mathbf{S}_{:,j} - \mathbf{S}_{:,j-1}\|_2^2. \quad (\text{III.13})$$

Dans [Virtanen 2007] et [Essid & Févotte 2013], les auteurs étudient ce type de régularisation avec une divergence de Kullback-Leibler.

Des visions plus orientées statistiques sont également proposées. Févotte modélise par exemple les signaux avec des chaînes de Markov et en dérive la régularisation correspondante [Févotte 2011] ; et Wilson et al. utilisent une phase d'apprentissage pour calculer les paramètres d'un modèle Gaussien et en déduisent aussi une régularisation [Wilson *et al.* 2008].

C.2 A priori parcimonieux

Dans beaucoup d'applications, la parcimonie apparaît naturellement au côté de la positivité, comme par exemple en résonance magnétique nucléaire ou en spectrométrie de masse, où les sources sont très piquées comme le montre la figure III.4. De récents travaux ont souligné le fait que cette connaissance pouvait se révéler utile dans le but d'obtenir des factorisations plus pertinentes [Kim & Park 2008, Eggert & Körner 2004, Hoyer 2002]. Cette observation est confortée par le fait que la parcimonie facilite déjà l'identifiabilité des sources en SSA. Celle-ci a été utilisée en NMF via l'ajout de plusieurs types de régularisations.

Régularisation ℓ_1 : La formulation la plus directe consiste à rajouter une régularisation ℓ_1 , comme nous l'avons vu au chapitre précédent :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \|\mathbf{Y} - \mathbf{AS}\|_2^2 + \lambda \|\mathbf{S}\|_1 \quad (\text{III.14})$$

Afin de traiter les sous-problèmes comme la minimisation d'une fonction différentiable sous contraintes de positivité, la régularisation ℓ_1 peut ici être remplacée

2. <http://www.bsp.brain.riken.jp/ICALAB/nmflab.html>

de façon équivalente par $\lambda \sum_{i,j} \mathbf{S}_{ij}$ car tous les coefficients de \mathbf{S} sont positifs. Ce type d'a priori parcimonieux est parfaitement adapté pour les sources du type des spectres de LC/MS (figure III.4), mais ne serait cependant pas adéquates pour des sources comme celles des spectres infrarouges de la figure III.3 qui ne sont pas du tout parcimonieux dans l'espace direct.

Ce type de régularisation a par exemple été utilisé par Hoyer dans [Hoyer 2002] avec un algorithme de gradient projeté, et dans [Cichocki & Zdunek 2007] avec un algorithme de type ALS. Il est aussi envisagé dans [Guan *et al.* 2012].

Une version parcimonieuse d'HALS existe également [Cichocki *et al.* 2007, Cichocki *et al.* 2009], visant à résoudre le problème avec une régularisation ℓ_1 . Les mises-à-jour reste également explicites dans ce cas. Dans une implémentation récente d'HALS [Gillis & Glineur 2012, Gillis 2012]³, le paramètre λ est géré automatiquement afin d'obtenir un niveau de parcimonie donné. Ce niveau est défini comme la proportion du nombre de coefficients dont la valeur est plus faible que 10^{-6} fois la valeur du plus grand coefficient. Il n'est cependant généralement pas connu à l'avance.

Ratio de parcimonie : Dans [Hoyer 2004], Hoyer utilise une approche différente en imposant un certain ratio de parcimonie (RdP) aux sources. Pour un vecteur $x \in \mathbb{R}^n$, il définit ce ratio de la manière suivante :

$$\text{RdP}(x) = \frac{\sqrt{n} - \frac{\|x\|_1}{\|x\|_2}}{\sqrt{n} - 1}. \quad (\text{III.15})$$

Cette fonction, qui utilise le fait que les normes ℓ_1 privilégie la parcimonie contrairement à la norme ℓ_2 , prend la valeur 1 quand x est parfaitement parcimonieux (un seul coefficient non nul), et 0 quand il ne l'est pas du tout (tous les coefficients identiques et non nuls). L'idée de cet algorithme est de contraindre le RdP de chaque source et/ou de chaque mélange à un niveau donné :

$$\begin{cases} \underset{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}}{\text{argmin}} \|\mathbf{Y} - \mathbf{AS}\|_2^2, \\ \text{RdP}(\mathbf{A}_{\cdot,t}) = \lambda_A, \forall t \in \{1, \dots, r\} \text{ (optionnel)}, \\ \text{RdP}(\mathbf{S}_{t,\cdot}) = \lambda_S, \forall t \in \{1, \dots, r\} \text{ (optionnel)}. \end{cases} \quad (\text{III.16})$$

Étant donné que le ratio peut être adapté aux sources, il est capable de gérer différentes morphologies, mais la contrainte n'est vraiment contraignante que quand les données sont parcimonieuses, et n'apportera pas beaucoup d'informations le cas échéant.

La recherche d'un minimum de ce problème est effectuée en utilisant un gradient projeté suivi d'une projection sur la contrainte de parcimonie si celle-ci est active, ou avec une mise-à-jour multiplicative si la contrainte n'est pas active. Ainsi, il se place dans le cas des algorithmes effectuant une minimisation locale suivie d'une projection

3. <https://sites.google.com/site/nicolasgillis/code>

(paragraphe B.2.2). Comme pour le niveau de parcimonie présenté précédemment, le ratio de parcimonie que l'on attend des sources n'est cependant généralement pas connu à l'avance. De plus, imposer une contrainte dure du RdP peut rendre la solution très dépendante du choix des paramètres de parcimonie λ_A et λ_S . Une implémentation de cet algorithme est disponible en ligne ⁴.

Régularisation $\ell_{2,1}$: Dans [Kim & Park 2008], Kim & Park ont proposé de formuler le problème de NMF comme suit :

$$\underset{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{AS}\|_2^2 + \eta \|\mathbf{A}\|_2^2 + \beta \sum_{t=1}^n \|\mathbf{S}_{:,t}\|_1^2. \quad (\text{III.17})$$

Dans cette équation, la régularisation $\sum_{t=1}^n \|\mathbf{S}_{:,t}\|_1^2$ impose un certain type de parcimonie qui privilégie les solutions où une source domine les autres à chaque échantillon. À la différence des régularisations précédentes, elle n'impose donc pas une parcimonie intrinsèque à chaque source, mais plutôt une parcimonie inter-sources. Ainsi, cet a priori peut aider à rechercher un ensemble de sources telles que celles de la figure III.4, qui ne présentent pas ou très peu de grands coefficients en commun. Comme les autres a priori parcimonieux, il ne serait cependant pas adéquat pour les spectres infrarouges de la figure III.3 puisque ceux-ci partagent de nombreux coefficients non nuls.

La régularisation ℓ_2 sur \mathbf{A} dans III.17 permet d'équilibrer le problème. En effet, la régularisation de \mathbf{S} tend à faire décroître sa norme, ce qui peut mener à la solution dégénérée $\mathbf{S} = \mathbf{0}$ et $\mathbf{A} = \infty$ en raison de l'invariance par échelle. Nous reviendrons sur cette considération en section IV-B.4 et IV-B.5. Kim & Park ont utilisé une méthode de type "active-set" afin de résoudre le problème. Cette technique peut minimiser exactement chacun des sous-problèmes contraints et rentre donc dans la catégorie des algorithmes de minimisation globale-contrainte basés sur le modèle de l'Algorithme III.5. Il n'y a cependant pas de manière simple de régler les paramètres η et β . Les auteurs fournissent une implémentation sur leur site web ⁵.

Dans [Zdunek & Cichocki 2007c], Zdunek & Cichocki utilisent une régularisation similaire pour le terme \mathbf{S} mais avec un paramètre β qui décroît au fur et à mesure des itérations afin d'être plus robuste aux minima locaux et points stationnaires. Il ne résout cependant pas exactement les sous-problèmes et se trouve donc dans la catégorie partielle-projetée des algorithmes (section B.2.3).

D'autres normes composites peuvent également être utilisées. Dans [Tandon & Sra 2010], les auteurs en recensent plusieurs, dont la norme composite $\ell_{1,2}$ par exemple.

Domaine transformé : Une régularisation parcimonieuse dans un espace transformé permet de gérer aussi bien la source très lisse de la figure III.3a que le spectre

4. <http://www.cs.helsinki.fi/u/phoyer/software.html>

5. <http://www.cc.gatech.edu/~hpark/nmfsoftware.php>

de la figure III.3b qui présente de fortes discontinuités, à l'aide de transformées en ondelettes par exemple. Les transformées en ondelettes étant multi-résolutions, elles peuvent également être utile pour des sources parcimonieuses dans l'espace direct comme celles de la figure III.4 mais un a priori parcimonieux dans l'espace direct reste cependant plus adapté dans ce cas-là.

Il est cependant difficile de gérer en même temps la positivité dans l'espace direct et la parcimonie dans un autre espace. À notre connaissance, seul Jiang et al. [Jiang & Yin 2012] proposent d'utiliser une telle régularisation en NMF, avec une transformée orthonormale \mathbf{W} . Pour ce faire, les auteurs résolvent le sous-problème formulé en "basis pursuit" dans l'espace transformé \mathbf{S}_w :

$$\operatorname{argmin}_{\mathbf{S}_w} \|\mathbf{S}_w\|_1, \|\mathbf{Y} - \mathbf{A}\mathbf{S}_w\mathbf{W}\|_2 \leq \varepsilon \quad (\text{III.18})$$

Celui-ci est équivalent, dans le cas d'une transformé orthonormale, à la résolution du Lagrangien :

$$\operatorname{argmin}_{\mathbf{S}} \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2 + \lambda\|\mathbf{S}\mathbf{W}^T\|_1 \quad (\text{III.19})$$

Dans l'algorithme de Jiang et al., la contrainte de positivité n'est appliquée par projection que dans un deuxième temps, ce qui fait que cet algorithme se rapproche davantage de l'approche globale-projetée B.2.1.

D Récapitulatif des algorithmes

Le tableau III.7 de la page 44 récapitule les algorithmes de descente par blocs présentés ci-dessus, en identifiant :

(i) l'approche utilisée parmi les quatre recensées en B.2, en mentionnant si la minimisation est globale (le cas échéant elle est partielle), et si elle est effectuée sous contraintes (le cas échéant elle est appliquée par projection).

(ii) la méthode utilisée parmi celles décrites précédemment, comme les mise-à-jour multiplicatives ou la descente de gradient par exemple.

(iii) la divergence utilisée pour le terme d'attache aux données, parmi celle présentée en B.1.

(iv) les régularisations appliquées aux données, comme indiquées dans la section C.

Cette classification peut cependant être parfois arbitraire. En effet, plusieurs versions sont souvent proposées, et des méthodes différentes peuvent être utilisées pour les mises-à-jour de \mathbf{A} et pour la mise-à-jour de \mathbf{S} . Nous noterons dans ce tableau les versions mises particulièrement en avant dans les articles.

Les algorithmes proposés dans cette thèse sont également indiqués dans la dernière partie du tableau. Ces algorithmes visent à résoudre le problème de NMF formulé avec un terme d'attache aux données quadratique. Ils incluent une régularisation parcimonieuse dans l'espace direct [Rapin *et al.* 2013a] ou dans un espace

référence	approche		méthode d'optimisation	divergence	régularisation
	glob.	courr.			
[Paterno & Tapper 1994] (ALS)	✓		moindres carrés	ℓ_2	
[Lee & Seung 2001]		✓	multiplicative	ℓ_2 , KL	
[Hoyer 2002]		✓	multiplicative	ℓ_2	$\ \mathbf{S}\ _1$
[Hoyer 2004]			gradient projeté	ℓ_2	RdP
[Chu <i>et al.</i> 2004]			gradient projeté	ℓ_2	
[Dhillon & Sra 2005]		✓	multiplicative	Bregman	
[Merritt & Zhang 2005]		✓	point intérieur	ℓ_2	
[Cichocki <i>et al.</i> 2006]			multiplicative	divergence généralisée	possibles
[Cichocki <i>et al.</i> 2007] (HALS)	✓	✓	"hiérarchique"	ℓ_2	$\ \mathbf{S}\ _1$
[Cichocki & Zdunek 2007]			moindres carrés	ℓ_2	$\ \mathbf{S}\ _1, \ \mathbf{L}\mathbf{S}\ _2^2$
[Lin 2007b]	✓	✓	gradient projeté	ℓ_2	
[Virtanen 2007]		✓	multiplicative	KL	$\sum_{j=2}^n \ \mathbf{S}_{:,j} - \mathbf{S}_{:,j-1}\ _2^2, \ \mathbf{S}\ _1$
[Zdunek & Cichocki 2007a]			multiplicative	ℓ_2	U_δ
[Zdunek & Cichocki 2007c]			2 nd ordre projeté	KL ℓ_2	$\sum_j \ \mathbf{S}_{:,j}\ _1^2$
[Kim & Park 2008]	✓	✓	active-set	ℓ_2	$\sum_j \ \mathbf{S}_{:,j}\ _1^2, \ \mathbf{S}\ _2^2$
[Fevotte <i>et al.</i> 2009]		✓	multiplicative	IS	envisagées
[Tandon & Sra 2010]	✓	✓	proximale	ℓ_2	normes composites
[Xu 2011]			gradient projeté	ℓ_2	
[Guan <i>et al.</i> 2012]	✓	✓	gradient projeté accéléré	ℓ_2	$\ \mathbf{S}\ _1$ et $\ \mathbf{S}\ _2^2$ envisagées
[Jiang & Yin 2012]	✓		basis pursuit	ℓ_2	$\ \mathbf{S}\mathbf{W}^T\ _1$
[Essid & Fevotte 2013]		✓	Majoration-Minimisation	KL	$\sum_{j=2}^n \ \mathbf{S}_{:,j} - \mathbf{S}_{:,j-1}\ _2^2$
[Rapin <i>et al.</i> 2013a] (nGMCA)	✓	✓	proximale	ℓ_2	$\ \mathbf{S}\ _1$
[Rapin <i>et al.</i> 2014] (nGMCA)	✓	✓	proximale	ℓ_2	$\ \mathbf{S}\mathbf{W}^T\ _1, \ \mathbf{S}_w\ _1$

Figure III.7 – Synthèse des principaux algorithmes de descente par blocs en NMF.

transformé potentiellement redondant, formulée aussi bien en analyse qu'en synthèse [Rapin *et al.* 2014]. Ces algorithmes rentrent dans le cadre de la classe des algorithmes de minimisation globale des sous-problèmes sous contraintes (globale-contrainte, section B.2.4), chaque étape étant minimisée jusqu'au minimum global sous contraintes, via l'utilisation d'algorithmes proximaux.

E Critères de séparation

La suite de cette thèse comprend de nombreuses expérimentations visant à comparer plusieurs de ces algorithmes de NMF entre eux. Des critères de qualité sont alors nécessaires afin d'évaluer les estimations qu'ils fournissent. Ainsi, dans [Zibulevsky & Pearlmutter 2001] par exemple, les auteurs ont pris pour critère $\frac{\|s-\hat{s}\|_2}{\|s\|}$ où s est la source estimée et \hat{s} la source recherchée. D'autres propositions sont abordées dans [Schobben *et al.* 1999].

Nous choisissons ici d'utiliser les critères proposés par Vincent *et al.* [Vincent *et al.* 2006] qui ont l'avantage d'être invariants par échelle et de fournir une description précise des types d'erreurs affectant les estimations.

E.1 Calcul des critères de séparation

Afin de calculer les critères proposés dans [Vincent *et al.* 2006], chacune des sources estimées $s^{\text{estimation}}$ est d'abord décomposée en une somme de plusieurs composantes :

$$s^{\text{estimation}} = s_{\text{cible}} + s_{\text{interférences}} + s_{\text{bruit}} + s_{\text{artefacts}}.$$

Ces composantes sont obtenues via une suite de projections, et peuvent être interprétées comme suit :

(i) s_{cible} est la projection de $s^{\text{estimation}}$ sur la source recherchée (vérité-terrain). Autrement dit, c'est la seule part de la décomposition qui correspond à ce que l'on souhaite retrouver. Les autres termes sont des résidus.

(ii) $s_{\text{interférences}}$ correspond aux interférences dues aux autres sources.

(iii) s_{bruit} est la partie de la reconstruction qui est due au bruit.

(iv) $s_{\text{artefacts}}$ correspond aux artefacts restants, qui ne peuvent être imputés ni aux interférences ni au bruit.

À partir de cette décomposition, les auteurs ont créé des ratios d'énergie invariants par échelle, afin d'évaluer l'influence préjudiciable sur les résultats des interférences (via un critère appelé SIR), du bruit (SNR), et des artefacts (SAR). Ils ont aussi conçu un critère plus global appelé rapport signal sur distorsion (source distortion ratio, SDR). Tous ces critères sont définis comme suit :

$$\text{SIR}(s^{\text{estimation}}) = 10 \log_{10} \left(\frac{\|s_{\text{cible}}\|_2^2}{\|s_{\text{interférences}}\|_2^2} \right). \quad (\text{III.20})$$

$$\text{SNR}(s^{\text{estimation}}) = 10 \log_{10} \left(\frac{\|s_{\text{cible}} + s_{\text{interférences}}\|_2^2}{\|s_{\text{bruit}}\|_2^2} \right). \quad (\text{III.21})$$

$$\text{SAR}(s^{\text{estimation}}) = 10 \log_{10} \left(\frac{\|s_{\text{cible}} + s_{\text{interférences}} + s_{\text{bruit}}\|_2^2}{\|s_{\text{artefacts}}\|_2^2} \right). \quad (\text{III.22})$$

$$\text{SDR}(s^{\text{estimation}}) = 10 \log_{10} \left(\frac{\|s_{\text{cible}}\|_2^2}{\|s_{\text{interférences}} + s_{\text{bruit}} + s_{\text{artefacts}}\|_2^2} \right). \quad (\text{III.23})$$

Le SIR augmente lorsque les interférences diminuent, le SNR quand le bruit diminue, et le SAR quand les artefacts diminuent. Le SDR augmente plus généralement quand la qualité des reconstructions s'améliore, c'est-à-dire quand ces trois types de contaminations diminuent. Ce dernier critère est utilisé de manière intensive dans les chapitres qui suivent afin d'évaluer les algorithmes, puisque il prend en compte tous les aspects d'une bonne reconstruction (faibles interférences, faible bruit et peu d'artefacts).

En pratique, en raison de l'invariance par permutation, la source s_{cible} n'est cependant pas connue en avance pour chaque source estimée. Afin de lever cette indétermination, le SDR est calculé entre chacune des sources estimée et chacune des sources de référence, et un algorithme du Hongrois permet d'affecter les sources estimées et les sources de référence une-à-une.

E.2 Exemple de décomposition

Afin de mieux appréhender l'intérêt de cette décomposition, il est intéressant d'en voir le résultat sur des images. Dans la suite de cette thèse, en section C, nous nous attellerons à identifier les sources de la figure III.8 à partir d'un mélange bruité de celles-ci.

L'image III.9a est une des sources obtenue en résolvant le problème d'inversion sous contraintes de positivité avec une régularisation parcimonieuse en ondelette. Celle-ci représente l'image "cible" #1 de la figure III.8. Elle présente cependant de fortes interférences provenant des autres images. Celles-ci sont extraites dans le terme d'interférence $s_{\text{interférences}}$ de la figure III.9b. Du bruit est également visible, et extrait dans la figure III.9c qui correspond à s_{bruit} . Enfin la partie de l'estimation qui n'est due ni à la source cible, ni aux interférences et ni au bruit est rassemblée dans le terme $s_{\text{artefacts}}$ qui est visible figure III.9d. Ces artefacts, particulièrement présents le long des contours, sont ici dus au biais induit par la régularisation parcimonieuse en ondelettes. Les SDR, SIR, SNR et SAR sont fournis dans les légendes des figures. Le facteur limitant la qualité de la reconstruction sont ainsi les artefacts, étant donné qu'ils représentent la source de contamination la plus importante (critère le plus bas en dehors du SDR, qui est global). Les détails de cette expérimentation sont fournis section C.



Figure III.8 – 4 sources en dimension 2.

La figure **III.10** représente le même type de visualisation mais pour une source provenant d'une inversion sous contraintes de positivité sans parcimonie. Le SDR de la source obtenue (figure **III.10a**) est identique à celui du résultat précédent. La source est cependant très différente étant donnée que la principale source de contamination est ici le bruit, alors que les termes d'interférences et d'artefacts sont ici très faibles (SIR et SAR élevés). On voit ainsi qu'une même qualité de reconstruction peut correspondre à des situations variées, que ces critères permettent d'appréhender.

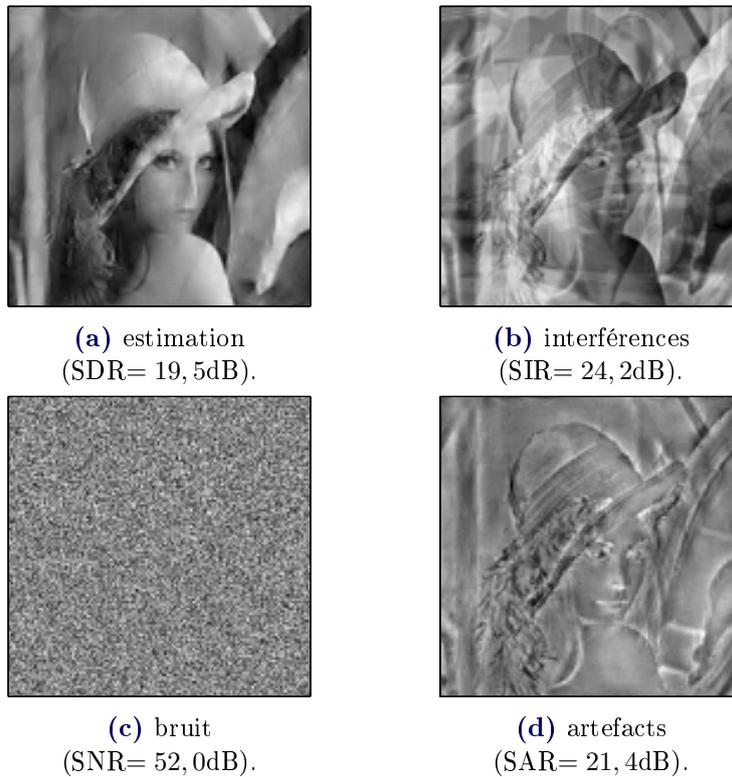


Figure III.9 – Composantes de l'estimation de la source #1 (inversion sous contraintes de positivité avec parcimonie en ondelettes, images renormalisées entre 0 et 1 pour la visualisation).

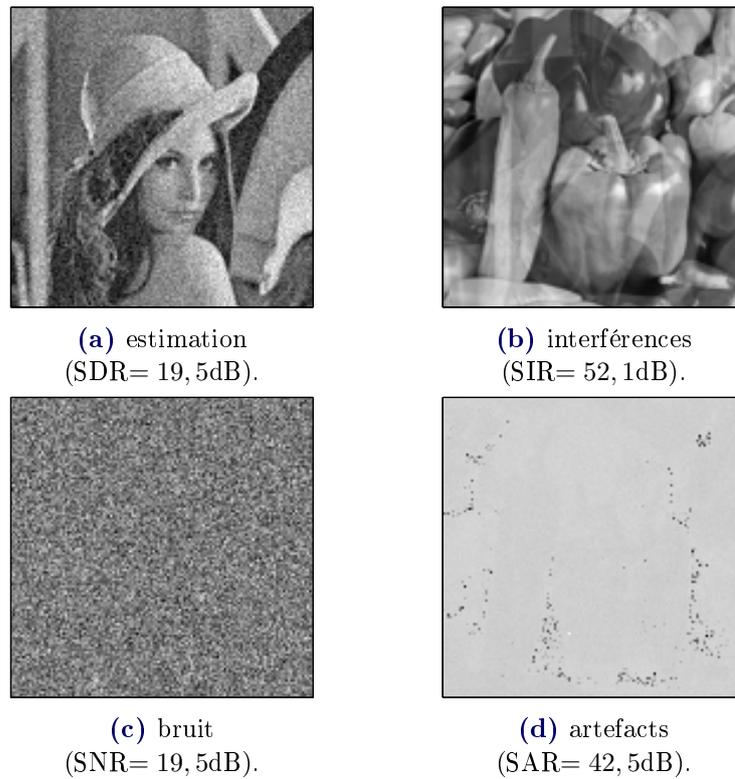


Figure III.10 – Composantes de l'estimation de la source #1 (inversion sous contraintes de positivité, images renormalisées entre 0 et 1 pour la visualisation).

nGMCA dans l'espace direct

Sommaire

A	Une première adaptation de GMCA	52
A.1	Présentation de GMCA	52
A.2	GMCA et positivité (nGMCA ^{naive})	53
B	Conception de nGMCA	54
B.1	Formulation et résolution	55
B.2	Gestion du seuil	56
B.3	Initialisation de l'algorithme	57
B.4	Renormalisations des sources et des mélanges	57
B.5	Convergence de l'algorithme	58
B.6	Version "seuillage dur"	59
C	Expérimentations sur des données simulées	59
C.1	Simulation des données	59
C.2	Comparaison avec les algorithmes standards dans le cas non-bruité	60
C.3	Expérimentations avec du bruit	65
C.4	Application à des données RMN synthétiques	71
D	Justification a posteriori de l'approche choisie	72
D.1	Choix de la formulation régularisée	74
D.2	Approche algorithmique	76
E	Conclusion	78

Ce chapitre présente un nouvel algorithme de NMF appelé non-negative Generalized Morphological Component Analysis (nGMCA). Cet algorithme s'inspire des travaux [Bobin *et al.* 2007, Bobin *et al.* 2008] qui présentent une technique de SSA parcimonieuse appelée Generalized Morphological Component Analysis (GMCA) que nous avons vue au chapitre précédent (section III-A.2.2, page 31) et qui sera détaillé dans la section A ci-dessous. Cette méthode ne permet cependant pas de prendre en compte de manière rigoureuse l'a priori de positivité de la NMF, d'où la nécessité de la modifier et l'étendre à ce type de situation (section B). Pour ce faire, nGMCA fera usage de méthodes proximales permettant la résolution exacte des sous-problèmes sous contraintes (approche de la section B.2.4). Aussi, l'accent sera mis sur la gestion du bruit additif à l'aide d'une gestion automatique de l'a priori de parcimonie, ce qui n'a pas été proposé dans la littérature en NMF. Un

ensemble d'expérimentations en section C permettent d'évaluer la robustesse des algorithmes au bruit ainsi que l'influence du nombre de sources et du conditionnement du problème.

A Une première adaptation de GMCA

A.1 Présentation de GMCA

GMCA s'inspire du travail pionnier de Zibulevsky et Pearlmutter [Zibulevsky & Pearlmutter 2001] utilisant la parcimonie pour la séparation de sources en aveugle. L'idée de GMCA est de séparer les sources à partir de leurs structures géométriques : des sources séparables ayant des morphologies différentes ne partagent pas les mêmes coefficients significatifs dans une représentation parcimonieuse bien choisie. Cela signifie que les plus grands coefficients de chaque source ne sont pas les mêmes dans cette représentation. Même si les supports des sources ne sont pas entièrement disjoints, les supports des plus grands coefficients le sont.

Le problème que GMCA vise à résoudre est le suivant :

$$\operatorname{argmin}_{\mathbf{A}, \mathbf{S}_w} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}_w\mathbf{W}\|_2^2 + \|\mathbf{\Lambda}^2 \odot \mathbf{S}_w\|_0. \quad (\text{IV.1})$$

Ce problème est donc formulé en synthèse et se décompose en deux termes : le terme d'attache aux données $f(\mathbf{A}, \mathbf{S}_w) = \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}_w\mathbf{W}\|_2^2$ et la régularisation des sources $g(\mathbf{S}_w) = \|\mathbf{\Lambda}^2 \odot \mathbf{S}_w\|_0$. On considère ainsi ici un paramètre de parcimonie $\mathbf{\Lambda}_{ij}$ potentiellement différent pour chaque élément \mathbf{S}_{ij} , afin de pouvoir l'adapter, en général source par source. L'approche utilisée consiste à mettre à jour successivement \mathbf{A} et \mathbf{S} en inversant le terme d'attache aux données, ce qui peut être fait de manière explicite, puis en appliquant la régularisation dans un deuxième temps pour \mathbf{S}_w . En ce sens, cet algorithme suit le même principe qu'ALS (catégorie globale-projetée, section III-B.2.1, page 35). La régularisation est appliquée via du seuillage dur $\text{Hard}_{\mathbf{\Lambda}}$, qui est l'équivalent d'un opérateur proximal pour la pénalisation ℓ_0 . Le pseudo-code de GMCA est donné en tant qu'**Algorithme IV.1**.

Un point essentiel de cet algorithme est la gestion du seuil pour le seuillage dur. Dans GMCA, au lieu de voir g comme une régularisation dont il est très difficile d'avoir une compréhension fine d'un point de vue Bayésien, les auteurs ont choisi de concentrer leurs efforts sur le niveau de seuillage qu'il implique. On ne s'intéresse donc pas vraiment à résoudre le problème (IV.1) avec un λ fixé d'avance, mais à trouver une solution ayant certaines caractéristiques. Dans les détails, ce seuil décroît itération après itération. Le recuit simulé a déjà inspiré les décroissances de régularisations ℓ_2 et $\ell_{1,2}$ en NMF [Zdunek & Cichocki 2007c, Cichocki & Zdunek 2007] mais les motivations derrière un seuil décroissant sont :

1. démélanger : en estimant dans un premier temps les directions de mélange uniquement à l'aide des plus grands coefficients, qui sont les plus susceptibles de n'appartenir qu'à une unique source.

Algorithme IV.1 : GMCA**Paramètre(s) d'entrée** : \mathbf{Y}, K

- 1: **initialiser** $\mathbf{A}^{(0)}$
- 2: **Pour** $k \leftarrow 1, K$ **faire**
- 3: **normaliser** les colonnes de $\mathbf{A}^{(k-1)}$
- 4: $\mathbf{S}_{w,\text{all}} = (\mathbf{A}^{(k-1)T} \mathbf{A}^{(k-1)})^{-1} \mathbf{A}^{(k-1)T} \mathbf{Y} \mathbf{W}^T$
- 5: **choisir** les seuils $\Lambda^{(k)}$ à partir de $\mathbf{S}_{w,\text{all}}$
- 6: $\mathbf{S}_w^{(k)} \leftarrow \text{Hard}_\Lambda((\mathbf{A}^{(k)T} \mathbf{A}^{(k)})^{-1} \mathbf{A}^{(k)T} \mathbf{Y} \mathbf{W}^T)$
- 7: $\mathbf{S}^{(k)} \leftarrow \mathbf{S}_w^{(k-1)} \mathbf{W}$
- 8: $\mathbf{A}^{(k)} \leftarrow \mathbf{Y} \mathbf{S}^T (\mathbf{S} \mathbf{S}^T)^{-1}$
- 9: **renvoyer** $\mathbf{A}^{(K)}, \mathbf{S}^{(K)}$

2. débruiter : en ne gardant pas les plus petits coefficients, qui sont les plus susceptibles de contenir du bruit.
3. être plus robuste aux minima locaux : ceux-ci sont déstabilisés lorsque le seuil est modifié.

Comme dans le recuit simulé cependant, le seuil décroît petit à petit afin de faire rentrer de nouveaux coefficients et d'affiner la solution en gardant une continuité des itérés [Hale *et al.* 2007]. Le seuil final est choisi indépendamment pour chaque sources d'index i à $\lambda_i = \tau_\sigma \sigma_i^{\text{source}}$, où σ_i^{source} est un estimateur empirique de l'écart-type du bruit sur cette source. En pratique, l'estimateur utilisé est l'écart médian absolu (MAD) défini comme :

$$\text{MAD}(x) = \text{median}(|x - \text{median}(x)|), \quad (\text{IV.2})$$

c'est-à-dire la médiane de l'écart absolu à la médiane. En effet, il permet une estimation robuste aux données aberrantes (outliers) de l'écart-type d'un bruit blanc Gaussien via l'approximation $\sigma(x) \approx 1,48 \text{MAD}(x)$.

τ_σ est généralement pris dans l'intervalle $[1, 3]$ comme compromis entre un débruitage efficace et une séparation correcte. En effet, garder un seuil trop élevé ne permet pas de retrouver les sources avec une grande précision. D'un autre côté, pour un signal contaminé par du bruit Gaussien indépendant et identiquement distribué ayant pour écart-type σ , seuiller à 3σ permet de rejeter les échantillons bruités avec une probabilité de 0,99 comme nous l'avons déjà vu en section II-B.2 (page 14).

A.2 GMCA et positivité (nGMCA^{naive})

Dans cette section, nous adaptions GMCA à la NMF en rajoutant une contrainte de positivité. Dans ce chapitre, seule la parcimonie dans l'espace direct est considérée afin d'observer dans un premier temps l'influence de la gestion de la positivité. Cette adaptation directe et naïve de GMCA sera ainsi nommée nGMCA^{naive}. Le problème à résoudre est maintenant le suivant :

Algorithme IV.2 : nGMCA^{naive}**Paramètre(s) d'entrée** : \mathbf{Y}, K

- 1: initialiser $\mathbf{A}^{(0)}$
- 2: **Pour** $k \leftarrow 1, K$ faire
- 3: **normaliser** les colonnes de $\mathbf{A}^{(k-1)}$
- 4: $\mathbf{S}_{\text{all}} = (\mathbf{A}^{(k-1)T} \mathbf{A}^{(k-1)})^{-1} \mathbf{A}^{(k-1)T} \mathbf{Y}$
- 5: **choisir** les seuils $\mathbf{\Lambda}^{(k)}$ à partir de \mathbf{S}_{all}
- 6: $\mathbf{S}^{(k)} \leftarrow [\text{Hard}_{\mathbf{\Lambda}^{(k)}}(\mathbf{S}_{\text{all}})]_+$
- 7: $\mathbf{A}^{(k)} \leftarrow [\mathbf{Y} \mathbf{S}^{(k)T} (\mathbf{S}^{(k)} \mathbf{S}^{(k)T})^{-1}]_+$
- 8: renvoyer $\mathbf{A}^{(K)}, \mathbf{S}^{(K)}$

$$\underset{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{A} \mathbf{S} - \mathbf{Y}\|_2^2 + \|\mathbf{\Lambda}^2 \odot \mathbf{S}\|_0. \quad (\text{IV.3})$$

Les seules différences avec l'équation (IV.1) sont ainsi l'ajout des contraintes de positivité, et la suppression de la transformée \mathbf{W} . Ce problème peut s'écrire de façon équivalente en intégrant les contraintes dans la fonction de coût :

$$\underset{\mathbf{A}, \mathbf{S}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{A} \mathbf{S} - \mathbf{Y}\|_2^2 + i_{\geq \mathbf{0}}(\mathbf{A}) + \|\mathbf{\Lambda}^2 \odot \mathbf{S}\|_0 + i_{\geq \mathbf{0}}(\mathbf{S}). \quad (\text{IV.4})$$

En suivant le raisonnement précédent, nGMCA^{naive} alterne entre une mise-à-jour de \mathbf{A} et une mise-à-jour de \mathbf{S} . Chaque mise-à-jour s'effectue en trouvant le minimum de la fonction de coût via une inversion puis en projetant dans un deuxième temps sur les contraintes grâce à un opérateur proximal. L'opérateur proximal de $i_{\geq \mathbf{0}}$ étant l'opérateur partie positive $[\cdot]_+$, cela donne l'**Algorithme IV.2**, qui diffère uniquement de l'**Algorithme IV.1** par la suppression de la transformée \mathbf{W} et l'ajout de la projection sur l'orthant positif $[\cdot]_+$ pour \mathbf{A} et \mathbf{S} .

nGMCA^{naive} souffre de deux faiblesses inhérentes à GMCA. La première est le fait que les sous-problèmes à résoudre ne sont pas convexes. En effet, le terme $\|\mathbf{\Lambda}^2 \odot \mathbf{S}\|_0$ n'est pas convexe, comme expliqué au chapitre II. Cela peut amplifier les risques de tomber dans des minima locaux dans le problème global qui n'était déjà pas convexe. La seconde est que la fonction de coût et les régularisations sont prises en compte de manières indépendantes, ce qui ne garantit pas dans ce cas la convergence de l'algorithme. La section suivante propose des modifications afin d'outrepasser ces faiblesses.

B Conception de nGMCA

Dans cette partie, nous proposons de résoudre exactement les sous-problèmes convexifiés en utilisant des algorithmes proximaux, et s'inscrivant ainsi dans le cadre des algorithmes de la catégorie globale-contraainte présentée en section III-B.2.4 (page 38).

Algorithme IV.3 : nGMCA**Paramètre(s) d'entrée** : \mathbf{Y}, K

- 1: **initialiser** $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}$ et $\mathbf{\Lambda}^{(1)}$
- 2: **Pour** $k \leftarrow 1, K$ **faire**
- 3: **normaliser** les colonnes de $\mathbf{A}^{(k-1)}$
- 4: $\mathbf{S}^{(k)} \leftarrow \underset{\mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}^{(k-1)} \mathbf{S}\|_2^2 + \|\mathbf{\Lambda}^{(k)} \odot \mathbf{S}\|_1$
- 5: **normaliser** les lignes de $\mathbf{S}^{(k)}$
- 6: $\mathbf{A}^{(k)} \leftarrow \underset{\mathbf{A} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A} \mathbf{S}^{(k)}\|_2^2$
- 7: **choisir** $\mathbf{\Lambda}^{(k+1)} \leq \mathbf{\Lambda}^{(k)}$
- 8: **renvoyer** $\mathbf{A}^{(K)}, \mathbf{S}^{(K)}$

B.1 Formulation et résolution

Afin d'aboutir à des sous-problèmes convexes, il est nécessaire de convertir la régularisation ℓ_0 en régularisation ℓ_1 qui est son pendant convexe (cf. chapitre II). Le problème global est donc maintenant le suivant :

$$\underset{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{A} \mathbf{S} - \mathbf{Y}\|_2^2 + \|\mathbf{\Lambda} \odot \mathbf{S}\|_1. \quad (\text{IV.5})$$

Le sous-problème en \mathbf{S} à \mathbf{A} fixé s'écrit ainsi directement :

$$\underset{\mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{A} \mathbf{S} - \mathbf{Y}\|_2^2 + \|\mathbf{\Lambda} \odot \mathbf{S}\|_1. \quad (\text{IV.6})$$

Ce problème est convexe et peut être résolu efficacement grâce à une approche proximale. En effet, il peut se mettre sous la forme :

$$\underset{\mathbf{S}}{\operatorname{argmin}} f(\mathbf{S}) + g(\mathbf{S}), \quad (\text{IV.7})$$

où $f(\mathbf{S}) = \frac{1}{2} \|\mathbf{A} \mathbf{S} - \mathbf{Y}\|_2^2$ est le terme d'attache aux données, et $g(\mathbf{S}) = \|\mathbf{\Lambda} \odot \mathbf{S}\|_1 + i_{\geq \mathbf{0}}(\mathbf{S})$ impose la parcimonie et la positivité de \mathbf{S} . On observe alors que :

- f est convexe et différentiable, avec pour gradient $\nabla_{\mathbf{S}} f(\mathbf{S}) = \mathbf{A}^T (\mathbf{A} \mathbf{S} - \mathbf{Y})$ qui est L -Lipschitz avec $L = \|\mathbf{A}^T \mathbf{A}\|_{s,2}$ où $\|\cdot\|_{s,2}$ est la norme opérateur subordonnée à la norme 2.
- g est convexe, propre et semi-continue inférieurement. Elle admet donc un opérateur proximal qui est ici explicite : $\operatorname{prox}_g = [\operatorname{Soft}_\lambda(\cdot)]_+$.

Le problème de mise-à-jour de \mathbf{S} peut donc être directement résolu à l'aide de l'algorithme FB présenté en section II-C.3.1 (page 20). La mise-à-jour de \mathbf{A} peut être effectuée de la même manière. Les grandes lignes de l'algorithme global, nommé nGMCA, sont données dans l'**Algorithme IV.3**. Les détails de l'implémentation sont décrits dans les sections qui suivent.

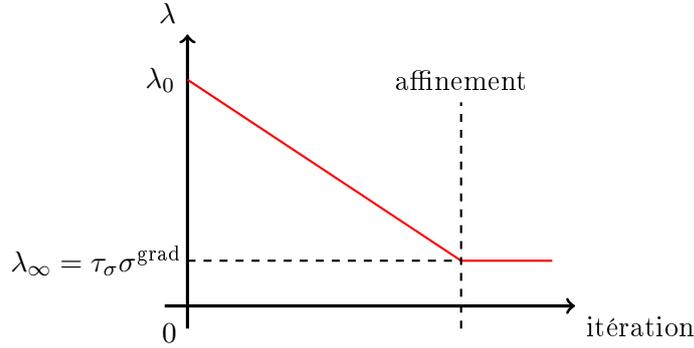


Figure IV.1 – Gestion du seuil au cours des itérations de nGMCA.

B.2 Gestion du seuil

De la même manière que dans la version originale de GMCA, le paramètre de parcimonie $\mathbf{\Lambda}$ décroît au fur et à mesure de l'avancement de l'algorithme. La stratégie de seuillage pour cette version doit cependant différer de celle présentée pour l'approche naïve. Dans cette dernière, le seuillage s'applique directement sur les sources telles que données en résolvant le problème des moindres carrés. Un coefficient initialement nul ne s'active alors que si le résultat des moindres carrés est plus élevé que le seuil. La mise-à-jour dans nGMCA est très différente. Une itération de l'algorithme FB servant à résoudre le sous-problème en \mathbf{S} peut en effet s'écrire sous la forme suivante :

$$\mathbf{S}^{(k+1)} \leftarrow \left[\mathbf{S}^{(k)} - \frac{1}{L} \left(\nabla_{\mathbf{S}} f(\mathbf{S}^{(k)}) - \lambda \mathbf{1}_{rn} \right) \right]_+, \quad (\text{IV.8})$$

où $\mathbf{1}_{rn}$ est la matrice de $\mathbb{R}^{r \times n}$ dont tous les coefficients valent 1. Ainsi, un élément initialement à 0 dans \mathbf{S} ne s'activera que si son gradient, et non plus la source elle-même, est plus élevé que le seuil. Aussi, contrairement au seuillage dur, qui n'affecte que les coefficients plus faibles que le seuil, le seuillage doux les modifie dans leur ensemble.

En prenant en compte ces deux considérations, la stratégie de seuillage est définie comme illustrée à la figure IV.1 :

(i) commencer à un niveau $\lambda_0 = \|\nabla_{\mathbf{S}} f_0(\mathbf{S}^{(0)})\|_{\infty} = \|\mathbf{A}^{(0)T}(\mathbf{A}^{(0)}\mathbf{S}^{(0)} - \mathbf{Y})\|_{\infty}$. Ce choix permet de s'assurer que la valeur des coefficients de \mathbf{S} n'augmente pas à la toute première itération de la première mise-à-jour, et en pratique, permet alors généralement de ne conserver que les coefficients les plus importants.

(ii) le paramètre de parcimonie final est imposé source par source à $\tau_{\sigma} \sigma_i^{\text{grad}}$, σ_i^{grad} étant une estimation du niveau de bruit sur la $i^{\text{ième}}$ ligne du gradient. Cette estimation peut être effectuée au cours de l'algorithme. Une méthode grossière consiste ainsi à estimer le niveau de bruit pour chaque observation en utilisant un MAD (cf. équation (IV.2) et section A), sur le résidu $\mathbf{R} = \mathbf{AS} - \mathbf{Y}$, puis à en déduire une

estimation du niveau de bruit du gradient en remarquant que $\nabla f(\mathbf{S}) = \mathbf{A}^T \mathbf{R}$ et en considérant que le bruit sur chaque observation est indépendant.

(iii) λ est décroissant de manière linéaire entre λ_0 et l'estimée courante de $\tau_\sigma \sigma_i^{\text{grad}}$, ce qui permet de modifier au minimum le problème à résoudre entre deux itérations globales.

(iv) le seuil reste inchangé lors des dernières itérations afin de pouvoir affiner la solution, c'est-à-dire de laisser le temps à l'algorithme de converger vers une solution à un λ donné. Cette convergence fait l'objet de la section B.5.

B.3 Initialisation de l'algorithme

En NMF et plus généralement dans l'ensemble des problèmes non-convexes, l'initialisation revêt une extrême importance. En effet, des initialisations différentes peuvent alors mener à des résultats très différents. Quelques étapes d'ALS [Paatero & Tapper 1994] sur une matrice aléatoire positive sont souvent utilisées en guise d'initialisation. En effet, cet algorithme à l'avantage d'être rapide et peu sensible aux points critiques dus à la contrainte de positivité, puisque cette contrainte n'est appliquée par projection que dans un second temps, après la résolution des sous-problèmes non-contraints.

Dans nGMCA cependant, l'initialisation est d'autant plus cruciale que de celle-ci dépend le choix de λ_0 . Pour que celui-ci ne soit pas trop grand et risque d'inactiver tous les coefficients, il est nécessaire d'être déjà proche d'un minimum en \mathbf{S} à \mathbf{A} fixé, ce qui n'est généralement pas le cas après une itération de ALS. Une deuxième étape est donc rajoutée en plus des itérations d'ALS et consiste en la résolution des problèmes sous contrainte de positivité en \mathbf{A} et \mathbf{S} , sans parcimonie ($\lambda = 0$). En pratique, l'initialisation de nGMCA est effectuée en alternant deux fois entre une étape d'ALS et une de minimisation sous contrainte de positivité, ce qui permet une initialisation rapide et très robuste. Cela semble dû au fait d'utiliser deux algorithmes très différents dans leur conception et chacun sensible à des aspects particuliers du problème.

B.4 Renormalisations des sources et des mélanges

Comparé à nGMCA^{naive}, une nouvelle difficulté apparaît avec cette formulation : la gestion des normes de \mathbf{A} et \mathbf{S} . En effet, comme vu au chapitre III, si $\mathbf{\Delta}$ est une matrice diagonale dont les coefficients diagonaux sont strictement positifs, les solutions $(\mathbf{A}\mathbf{\Delta}, \mathbf{S})$ et $(\mathbf{A}, \mathbf{\Delta}\mathbf{S})$ sont parfaitement équivalentes. Ainsi, les normes de chaque ligne de \mathbf{S} et les normes correspondantes de chaque colonne de \mathbf{A} sont interchangeableables et seul importe le produit des deux. Ainsi, le terme de pondération $\|\mathbf{S}\|_1$ ayant toujours pour conséquence de pousser à une norme de \mathbf{S} minimale, le problème peut converger vers la solution dégénérée $\mathbf{S} = \mathbf{0}$ et $\mathbf{A} = \infty$. L'approche envisagée afin d'éviter cette difficulté se décline en deux points.

Afin d'éviter de converger vers la solution dégénérée, il est possible de renormaliser les colonnes de \mathbf{A} avant la mise-à-jour de \mathbf{S} , et inversement. Renormaliser

permet alors également de reconditionner le problème et d'améliorer la convergence des sous-itérations. Cette approche est utilisée tout au long de l'algorithme.

B.5 Convergence de l'algorithme

Lors de la phase d'affinement de la solution, la renormalisation ne permet pas d'assurer la convergence, celle-ci modifiant la fonction de coût. Afin de gérer le problème de normalisation de façon plus rigoureuse, la mise-à-jour de \mathbf{A} peut être modifiée comme suit, sous le modèle de Zibulevsky & Pearlmutter [Zibulevsky & Pearlmutter 2001] :

$$6 : \quad \mathbf{A}^{(k)} \leftarrow \underset{\mathbf{A} \geq \mathbf{0}, \forall i \|\mathbf{A}_{\cdot, i}\|_2 \leq 1}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}^{(k)}\|_2^2. \quad (\text{IV.9})$$

La fonction de coût complète du problème que l'algorithme vise à minimiser est donc en fait la suivante :

$$F(\mathbf{A}, \mathbf{S}) = \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2 + i_{\|\mathbf{A}_{\cdot, k}\|_2 \leq 1, \forall k}(\mathbf{A}) + i_{\geq \mathbf{0}}(\mathbf{A}) + \|\mathbf{A} \odot \mathbf{S}\|_1 + i_{\geq \mathbf{0}}(\mathbf{S}). \quad (\text{IV.10})$$

Dans le cas de nGMCA, la minimisation de F est effectuée en utilisant une méthode de descente par blocs, c'est-à-dire en minimisant de manière alternative les sous-problèmes en \mathbf{A} et \mathbf{S} . Pour ce type d'algorithmes, Tseng [Tseng 2001] fournit des conditions de convergence quand la fonction de coût n'est pas différentiable.

Une des conditions est que les ensembles de niveau $E_a = \{x, F(x) \leq a\}$ de la fonction de coût soient compacts. C'est le cas pour F grâce à la contrainte que nous avons rajouté, en effet :

- les E_a sont fermés, en tant qu'images inverses d'ensembles fermés par une fonction continue.
- ils sont bornés en \mathbf{S} grâce à la régularisation ℓ_1 , et en \mathbf{A} grâce à la nouvelle contrainte sur la norme des colonnes de \mathbf{A} .

Ainsi, en temps qu'ensembles fermés et bornés d'un espace de dimension fini, ils sont compacts. Aussi, le terme d'attache aux données $(\mathbf{A}, \mathbf{S}) \mapsto \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2$ est différentiable sur son domaine, son domaine est ouvert, F est continue, et le sous-problème en \mathbf{A} est pseudoconvexe (étant donné qu'il est convexe). Cela permet d'utiliser le théorème 4.1(b) de [Tseng 2001] qui prouve que la méthode de descente par blocs converge vers un point critique de F . Plus de détails sont fournis dans l'exemple 6.4 [Tseng 2001] qui s'attaque à un problème très similaire à celui qui nous intéresse ici.

La mise-à-jour de \mathbf{A} peut toujours être effectuée en utilisant le même algorithme que précédemment, mais en modifiant l'opérateur proximal en conséquence. En effet, l'opérateur proximal de :

$$g : x \mapsto i_{\geq \mathbf{0}}(x) + i_{\|\cdot\|_2 \leq 1}(x), \quad (\text{IV.11})$$

est explicite et donné par :

$$\text{prox}_g(x) = \frac{[x]_+}{\max(\| [x]_+ \|_2^2, 1)} \quad (\text{IV.12})$$

La démonstration du calcul de cet opérateur proximal peut être trouvée dans l'annexe A-B (page 129). Dans la pratique, afin de pouvoir continuer à renormaliser les sources et les directions de mélange tel qu'indiqué en B.4, la borne de norme pour chaque colonne de \mathbf{A} peut être imposée indépendamment à une valeur différente de 1.

B.6 Version "seuillage dur"

La résolution exacte des fonctions de coût des mises-à-jour de \mathbf{A} et \mathbf{S} fournit l'algorithme nGMCA décrit précédemment et utilisant du seuillage doux, mais il est également possible d'utiliser du seuillage dur afin de s'approcher d'une régularisation ℓ_0 . On perd cependant toute propriété de convergence dans ce cas là. Dans les expérimentations ci-dessous, cette version sera notée nGMCA^H. La stratégie de seuillage dans ce cas doit aussi être modifiée. Cet algorithme étant un mélange de nGMCA^{naive} et de nGMCA, le seuil est géré comme nGMCA^{naive} au début, afin de ne garder que les plus coefficients les plus significatifs. Elle décroît ensuite jusqu'à être gérée comme dans nGMCA à la fin, le seuillage s'effectuant également sur le gradient.

C Expérimentations sur des données simulées

Dans cette section, nous comparons les différentes versions de nGMCA présentées ci-dessus avec, dans un premier temps, des algorithmes classiques en NMF sur des données non-bruitées, puis, dans un deuxième temps, avec des algorithmes de NMF parcimonieux issus de l'état-de-l'art sur des données bruitées.

C.1 Simulation des données

Les données sont ici simulées afin de pouvoir évaluer les résultats. Les matrices de référence à retrouver seront \mathbf{A}^{ref} et \mathbf{S}^{ref} dont les coefficients sont tirés respectivement comme des variables aléatoires $|B_{p_A} G_{\alpha_A}|$ et $|B_{p_S} G_{\alpha_S}|$ où :

- B_p est une variable de Bernoulli ayant un taux d'activation p , c'est-à-dire valant 1 avec une probabilité p et 0 avec une probabilité $1 - p$.
- G_α est une variable aléatoire Gaussienne généralisée centrée réduite dont le paramètre de forme est α .

En pratique, p et α contrôlent deux formes de parcimonie. Le paramètre de Bernoulli affecte le nombre de vrais zéros dans \mathbf{A} et \mathbf{S} . La variable α permet de choisir à quelle point la distribution des coefficients est piquée en zéro, et contrôle donc une parcimonie approchée. La distribution est en effet proportionnelle à $e^{\frac{-|x-\mu|^\alpha}{\beta}}$ (avec

$\mu = 0$ ici et β dépendant de l'écart-type qui est choisi à 1 dans notre cas). En pratique, dans le cas où $\alpha = 2$, G_α est une variable Gaussienne, et pour $\alpha = 1$ elle est Laplacienne. Quand $\alpha \leq 1$, G_α est considérée comme approximativement parcimonieuse, et l'est d'autant plus que α est petit.

Dans les expérimentations ci-dessous, et à moins qu'il en soit mentionné autrement, les paramètres de base seront $p_A = 1$ et $\alpha_A = 2$ (\mathbf{A} est donc générée comme la valeur absolue d'une matrice aléatoire Gaussienne), et $\alpha_S = 1$ et $p_S = 0.8$ (soit 0 pour 20% des coefficients de \mathbf{S} et la valeur absolue d'une variable Laplacienne sur les autres). Les données sont de tailles $m \times n$ avec $m = 200$ observations et $n = 200$ échantillons dans chaque signal.

C.2 Comparaison avec les algorithmes standards dans le cas non-bruité

Dans cette première partie de l'expérimentation, nGMCA^{naive}, nGMCA^H et nGMCA sont testés sur un jeu de données non-bruitées mais peu parcimonieuses, avec un taux d'activation $p_S = 80\%$. Ces conditions ne sont pas favorables aux algorithmes de type nGMCA et permettent alors de faire ressortir leurs différences de comportement. Puisque le problème se résume à une factorisation exacte lorsque les données ne sont pas bruitées, τ_σ est imposé à 0. Il n'y a donc plus de régularisation à la fin des algorithmes, pendant la phase d'affinement. ALS [Paatero & Tapper 1994], l'algorithme multiplicatif [Lee & Seung 2001] et HALS (sans parcimonie) [Gillis & Glineur 2012] sont également testés afin de servir de références. Le nombre maximum d'itérations est à chaque fois imposé à un nombre suffisamment grand pour assurer la convergence des algorithmes quand ils sont convergents. Plus précisément, il est de 5000 pour HALS, 40 000 pour l'algorithme multiplicatif, et 500 pour ALS et les algorithmes basés sur nGMCA, avec un maximum de 80 sous-itérations pour nGMCA et nGMCA^H. Tous les résultats des tests de performances sont tirés de simulations de Monte-Carlo dont le nombre est fourni dans la légende des figures.

C.2.1 Sommaire des expériences

- Figure [IV.2](#) : ce test de performances montre l'influence du nombre de sources r sur l'estimation de \mathbf{S} . Ce paramètre est bien sûr primordial puisque les sources sont d'autant plus difficiles à séparer qu'elles sont nombreuses. nGMCA^{naive} obtient de meilleurs résultats qu'ALS alors qu'avec $\tau_\sigma = 0$, les dernières itérations (lors de la phase d'affinement) sont identiques pour les deux algorithmes. La stratégie de seuillage utilisée dans la première phase de nGMCA^{naive} montre ainsi son efficacité pour une meilleure séparation des sources. Bien que nGMCA ne soit pas aussi performant que nGMCA^H et nGMCA^{naive} à faible nombre de sources, il est plus robuste qu'eux quand celui-ci devient élevé. Ce constat est analysé plus en détail en [C.2.3](#).
- Figure [IV.3](#) : cette figure montre l'évolution de la fonction de coût $\|\mathbf{Y} - \mathbf{AS}\|_2^2$ au fur et à mesure des itérations des algorithmes de type nGMCA, avec $r = 40$

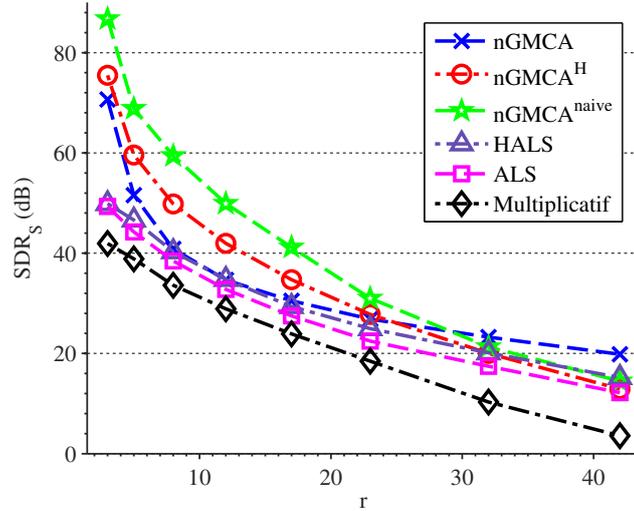


Figure IV.2 – SDR de la reconstruction (SDR_S) en fonction du nombre de sources r ($p_S = 80\%$, non-bruité, moyenne de 48 simulations).

sources. Pendant la phase d’affinement, le paramètre de parcimonie λ vaut 0 afin d’observer si la solution peut alors être améliorée. nGMCA converge dans ce cas à une valeur plus faible que celle où nGMCA^H semble se stabiliser (la convergence n’étant pas garantie) alors que nGMCA^{naive} ne converge pas du tout. Une explication est fournie en C.2.2

- Figure IV.4 : cette évaluation permet d’observer l’influence de la distribution des coefficients de \mathbf{A}^{ref} sur l’estimation de \mathbf{S} . Modifier le paramètre α_A est une façon de rendre sa distribution plus ou moins parcimonieuse. Quand \mathbf{A}^{ref} est plus parcimonieux, ses colonnes sont moins corrélées et le conditionnement du problème est meilleur (voir figure IV.5) ce qui simplifie la séparation. Encore une fois, nGMCA est plus robuste à un problème mal conditionné que les autres algorithmes.

C.2.2 Prise en compte des contraintes

Les différences de performances entre nGMCA et nGMCA^{naive} pour un grand nombre de source ($r = 40$) et un taux d’activation élevé ($p_S = 0.8$) comme sur la figure IV.2 peuvent s’expliquer en regardant l’évolution de la valeur de la fonction de coût pendant les itérations, sur la figure IV.3. On observe en effet que nGMCA^{naive} ne converge pas dans ce cas-là. Cela est dû au fait que cet algorithme prend en compte la contrainte de positivité et l’a priori de parcimonie de manière complètement séparée, et peut ainsi être facilement déstabilisé dans ces conditions difficiles. nGMCA, qui a été conçu spécifiquement pour éviter ce problème, est quant à lui convergent comme expliqué en B.5 et la phase d’affinement permet alors de converger ici vers une valeur bien plus faible de la fonction de coût. Entre les deux, on

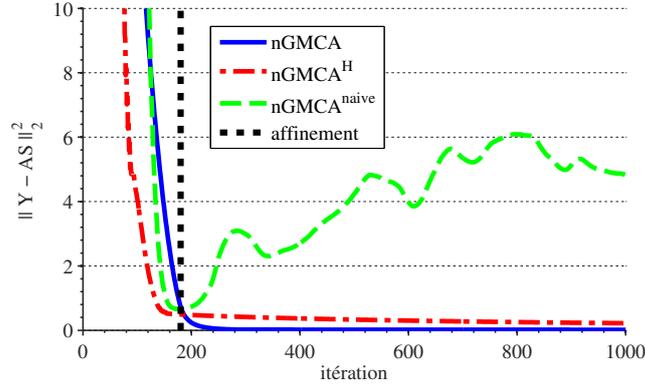


Figure IV.3 – Évolution de la fonction de coût $\|Y - AS\|_2^2$ pendant les itérations, sur un exemple représentatif ($p_S = 80\%$, $r = 40$, non-bruité).

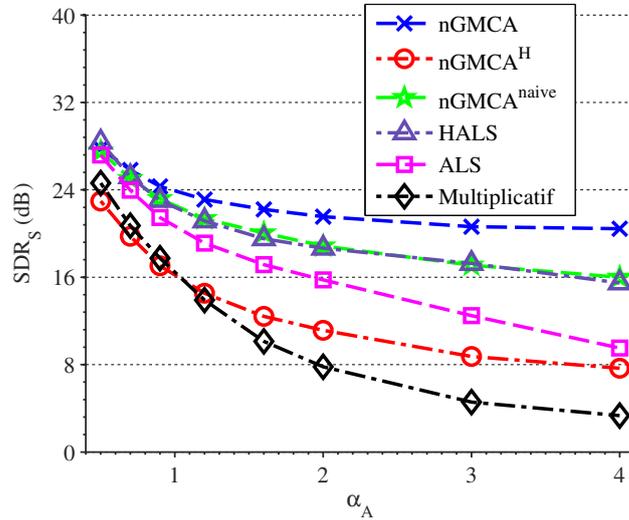


Figure IV.4 – SDR de la reconstruction (SDR_S) en fonction du paramètre de forme α_A ($r = 35$, $p_S = 80\%$, non-bruité, moyenne de 48 simulations).

retrouver nGMCA^H qui semble plus stable que nGMCA^{naive} mais pour sa convergence n'est tout de même pas assurée à cause du seuillage dur.

C.2.3 Seuillage dur et seuillage doux

L'explication des différences entre nGMCA et nGMCA^H vient des propriétés différentes du seuillage doux et du seuillage dur. Ces propriétés sont résumées dans la figure IV.6 qui montre le seuillage de 3 points de dimension 2, de la forme $x = (x_1, x_2)$. Quand un point (une colonne de S) a deux grands coefficients (quand il se trouve dans le cadran), il souffre d'un biais dans le cas du seuillage doux mais reste inchangé avec du seuillage dur. D'un autre côté, le mouvement induit par le seuillage doux augmente le ratio entre le plus grand coefficient de x et son plus petit,

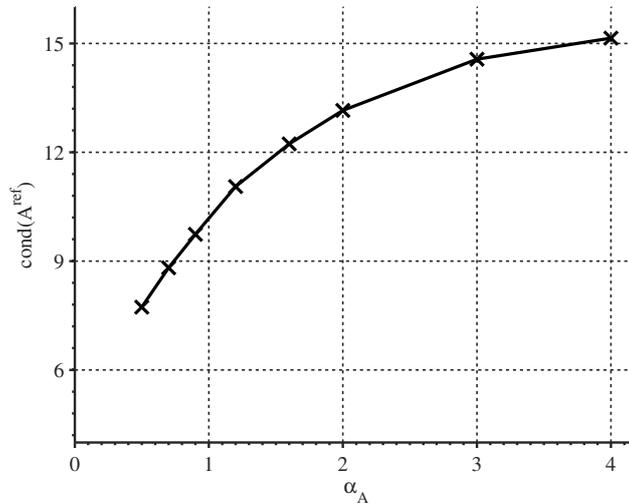


Figure IV.5 – Conditionnement de \mathbf{A}^{ref} en fonction de α_A ($r = 35$, moyenne de 48 simulations).

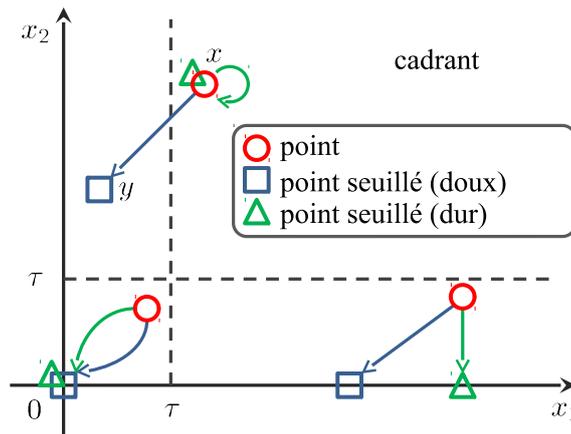


Figure IV.6 – Différences entre le seuillage doux et le seuillage dur (seuil à τ).

ce qui aide à renforcer l'affectation de ce point sur l'axe de son plus grand coefficient (dans le cas présent : x_2). Avec $y = \text{Soft}_\tau(x) = (y_1, y_2)$ le point seuillé, cela veut dire que $\frac{y_2}{y_1} > \frac{x_2}{x_1}$. Alors que le moindre biais du seuillage dur peut permettre une meilleure précision, renforcer l'affectation de chaque point à une direction principale grâce au seuillage dur permet une meilleure séparation des sources.

Lorsqu'il y a peu de sources, nGMCA et nGMCA^H séparent tous les deux de manière correcte, comme indiqué figure IV.2. Le biais est ainsi coûteux pour nGMCA et conduit à des phénomènes de compensation : avec un taux d'activation de 80%, presque tous les coefficients se voient imposés un biais, et une source à tendance à entièrement récupérer ces biais. C'est ce que l'on observe sur la source 5 de la figure IV.7. Ce biais corrèle avec l'ensemble des sources de référence comme cela est visible sur la figure IV.8 : la source estimée #5 a rassemblé l'ensemble des parties seuillées

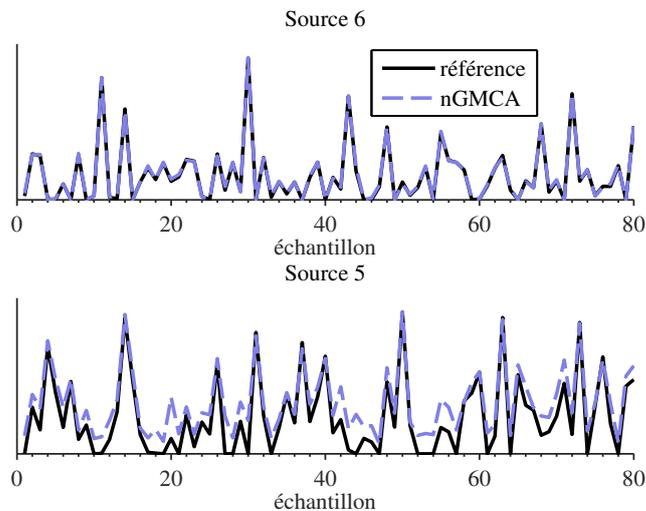


Figure IV.7 – Sources estimées et sources de référence numéros 5 et 6 (nGMCA, $r = 15$, $p_S = 80\%$, non-bruité, échantillons 1 à 80).

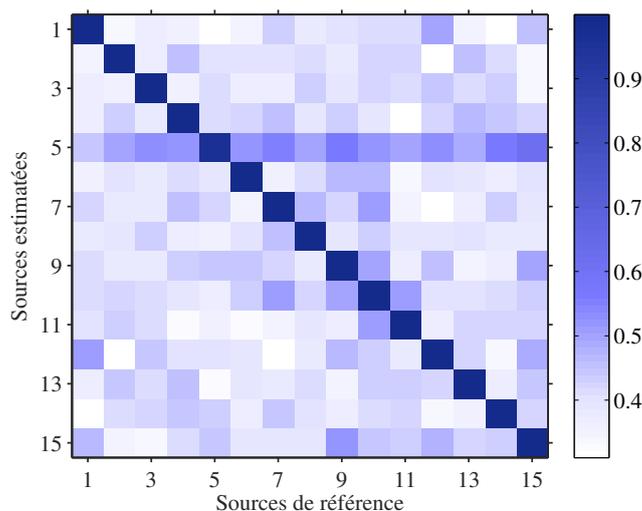


Figure IV.8 – Matrice de corrélation (non-centrée) entre les sources de référence et les sources estimées, sur un exemple représentatif ($S^{\text{est}} S^{\text{ref}T}$ après normalisation, nGMCA, $r = 15$, $p_S = 80\%$, non-bruité).

des coefficients des autres sources et est ainsi contaminée par des interférences.

D'un autre côté, l'effet du seuillage doux sur l'amplitude des coefficients reste essentiel pour la séparation des sources en amplifiant le poids des plus grand coefficients. Cela est particulièrement utile dans le cas où le problème est moins bien conditionné, tel que sur la figure IV.4 ou avec un nombre de sources r élevé (figure IV.2).

C.3 Expérimentations avec du bruit

On a vu jusque là le comportement de nGMCA dans le cas non-bruité. Dans cette partie, du bruit est ajouté aux données et les algorithmes prennent donc en entrée la matrice $\mathbf{Y} = \mathbf{A}^{\text{ref}}\mathbf{S}^{\text{ref}} + \mathbf{Z}$, où \mathbf{Z} est une matrice Gaussienne dont les coefficients sont identiquement distribués et indépendants entre eux ainsi qu'avec \mathbf{A} et \mathbf{S} (bruit additif Gaussien). Dans les expériences, le niveau de bruit est donné en terme de SNR sur \mathbf{Y} , noté SNR_Y . Alors que dans la section précédente les algorithmes présentés ci-dessus étaient comparés à des algorithmes standards n'utilisant pas la parcimonie (le dernier seuil des algorithmes de type nGMCA étant fixé à $\lambda = 0$), les algorithmes que nous utiliserons dans cette partie sont des algorithmes compétitifs, disponibles en ligne et prenant la parcimonie en compte de manières variées (cf. chapitre III pour plus de détails). Les comparaisons seront ainsi effectuées avec l'algorithme avec contrainte de parcimonie de Hoyer [Hoyer 2004], un algorithme de Kim & Park [Kim & Park 2008] et HALS avec un terme de parcimonie [Gillis 2012]. Puisqu'il n'y a pas de méthode directe pour choisir les paramètres de l'algorithme de Kim & Park, les paramètres par défaut sont utilisés. En ce qui concerne l'algorithme de Hoyer, aucun a priori n'est imposé sur \mathbf{A} et le paramètre de parcimonie sur \mathbf{S} est ajusté au ratio de parcimonie moyen des sources de référence, et nécessite donc une connaissance fine des données. De la même manière, le niveau de parcimonie recherché par HALS est calculé à partir des données de référence. Dans l'essentiel de cette section et tant qu'il n'est pas précisé autrement, $\tau_\sigma = 1$ pour les algorithmes de type nGMCA. Ce choix est un compromis entre un débruitage efficace et un biais qui ne détériore pas les résultats.

Les comparaisons comprennent aussi un algorithme "oracle" qui résout le problème des moindres carrés en \mathbf{S} avec régularisation parcimonieuse et contrainte de positivité, en considérant la matrice de mélange \mathbf{A}^{ref} connue :

$$\underset{\mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}^{\text{ref}}\mathbf{S}\|_2^2 + \|\mathbf{A} \odot \mathbf{S}\|_1. \quad (\text{IV.13})$$

Le paramètre de parcimonie est fixé à $\tau_\sigma \sigma_i^{\text{grad}}$ pour chacune des sources, de la même manière que dans l'étape d'affinement de nGMCA. Cet oracle permet d'avoir une idée des meilleures sources \mathbf{S} qu'il serait possible de retrouver avec l'approche nGMCA si les directions de mélange \mathbf{A}^{ref} étaient connues. Bien entendu, puisque les mélanges ne sont en pratique pas connus en SSA, cet oracle obtient des résultats qui sont inatteignables, mais permettent d'observer la marge de progression de nGMCA.

C.3.1 Sommaire des expériences

- Figures *IV.9*, *IV.10* et *IV.11* : ces tests de performances montrent les résultats de reconstructions avec 15 sources et un taux d'activation de $p_S = 10\%$ (figures *IV.9* et *IV.10*) et 30% (figure *IV.11*) en fonction du niveau de bruit sur les données. Comme indiqué précédemment, plus le taux d'activation est bas, plus les données sont parcimonieuses et plus l'a priori parcimonieux est

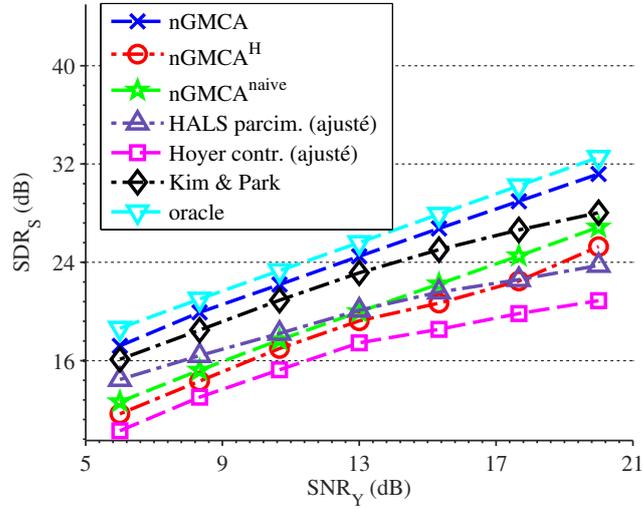


Figure IV.9 – SDR de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($p_S = 10\%$, $r = 15$, moyenné sur 192 simulations)

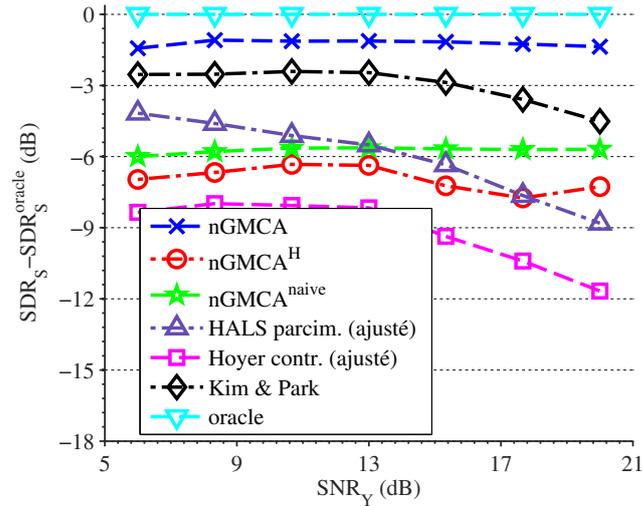


Figure IV.10 – SDR de la reconstruction (SDR_S) moins celui de l'oracle en fonction du niveau de bruit sur les données en dB (SNR_Y) ($p_S = 10\%$, $r = 15$, moyenné sur 192 simulations)

justifié. Les résultats pour les figures *IV.10* et *IV.11* sont donnés sous la forme du nombre de décibels perdus sur le SDR comparé à l'oracle, ceci afin de simplifier leur visualisation. La plage de valeur est identique entre les trois graphiques afin de pouvoir les comparer plus facilement. Dans les deux situations exposées, nGMCA est moins sensible au bruit et obtient des reconstructions de meilleure qualité que les autres algorithmes.

- Figure *IV.12* : cette évaluation permet d'observer la même expérience mais

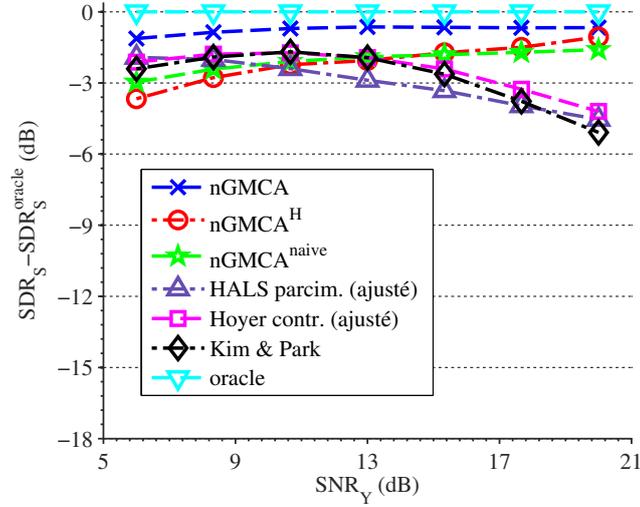


Figure IV.11 – SDR de la reconstruction (SDR_S) moins celui de l’oracle en fonction du niveau de bruit sur les données en dB (SNR_Y) ($p_S = 30\%$, $r = 15$, moyenné sur 192 simulations)

avec un taux d’activation plus élevé ($p_S = 50\%$) et donc moins favorable aux algorithmes de type nGMCA car la parcimonie des sources décroît. nGMCA reste tout de même l’algorithme obtenant les meilleurs résultats dans la plupart des cas. L’algorithme de Hoyer obtient des résultats similaires sur les données les plus bruitées. Il est cependant important de se rappeler que dans ces expériences, les paramètres de l’algorithme de Hoyer et de HALS ont été ajustés grâce aux données de référence, ce qu’il ne serait pas possible de faire en pratique. Quand les données sont moins bruitées, nGMCA^H et nGMCA^{naive} commencent à dépasser nGMCA, ce qui corrobore les observations précédentes sur les données non-bruitées avec un fort taux d’activation et peu de sources (figure IV.2).

- Figure IV.13 : Ce test de performances montre les résultats de reconstruction en fonction du nombre d’observations m sur des données bruitées (15dB), avec 15 sources et un faible taux d’activation ($p_S = 30\%$). Plus le nombre d’observations est faible et plus le problème est difficile, étant donné que la surdétermination du problème peut aider à débruiter et à identifier les sources. Alors que jusqu’ici seul le cas fortement surdéterminé a été étudié, ces nouvelles courbes montrent que les comparaisons sont également favorables à nGMCA quand le nombre d’observations est plus restreint.
- Figure IV.14 : Cette évaluation fournit les résultats de reconstruction en fonction du nombre de sources r pour des données parcimonieuses (taux d’activation de 30%) et bruitée (15dB). La complexité de la séparation augmente bien entendu avec le nombre de sources, mais, dans tous les cas, nGMCA se comporte mieux que les autres algorithmes dans ce cas également.

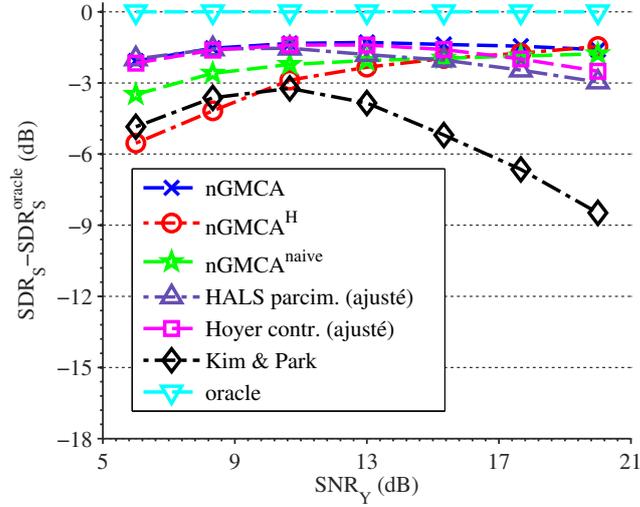


Figure IV.12 – SDR de la reconstruction (SDR_S) moins celui de l'oracle en fonction du niveau de bruit sur les données en dB (SNR_Y) ($p_S = 50\%$, $r = 15$, moyenné sur 192 simulations)

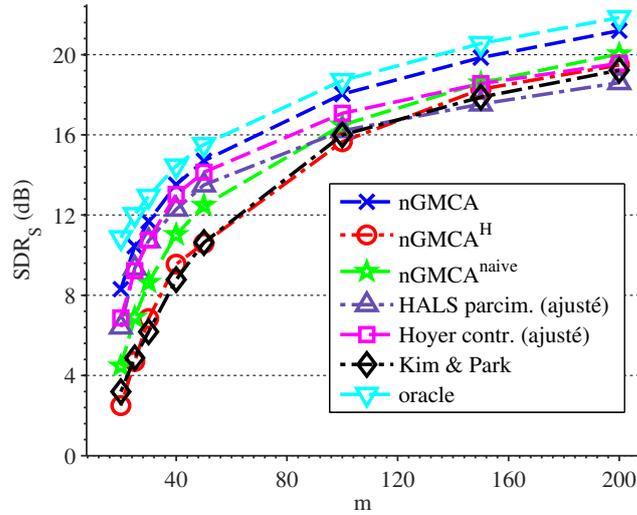


Figure IV.13 – SDR de la reconstruction (SDR_S) en fonction du nombre d'observations m ($SNR_Y=15$ dB, $p_S = 30\%$, $r = 15$, moyenné 72 simulations)

C.3.2 À propos de l'initialisation et de séparation

La figure IV.15 se place exactement dans le même cadre que la figure IV.11 mais compare cette fois nGMCA^{naive} et nGMCA avec des versions de ces algorithmes qui sont initialisées avec \mathbf{A}^{ref} et \mathbf{S}^{ref} et commence ainsi avec la meilleure initialisation possible. La version initialisée de nGMCA est également paramétrée en connaissant

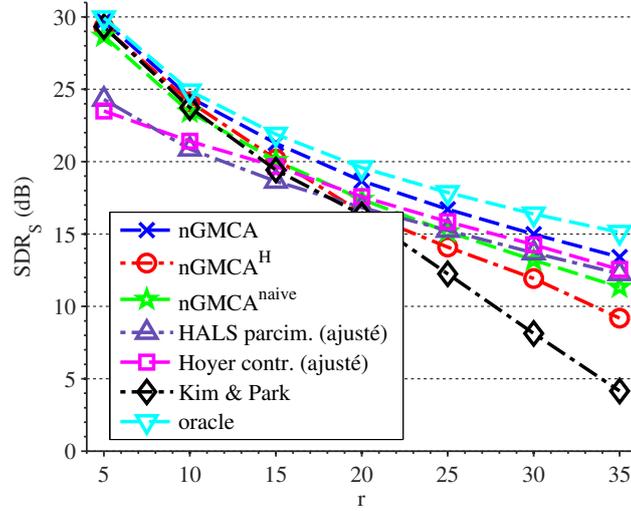


Figure IV.14 – SDR de la reconstruction (SDR_S) en fonction du nombre de sources ($SNR_Y=15\text{dB}$, $p_S = 30\%$, moyenné sur 48 simulations)

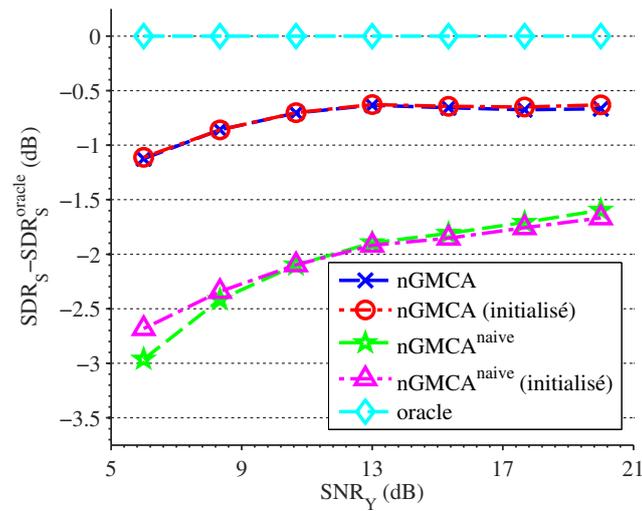


Figure IV.15 – SDR de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($p_S = 30\%$, $r = 15$, moyenné sur 192 simulations).

exactement le niveau de bruit présent dans les données. On peut ainsi observer que la différence en terme de qualité de reconstruction entre les versions initialisées et non-initialisées est extrêmement faible. Cela montre que l'estimation automatique du niveau de bruit par nGMCA parvient à un résultat similaire au vrai niveau de bruit, et surtout que l'initialisation de nGMCA et nGMCA^{naive} est très robuste puisqu'elle permet de trouver une solution proche de celle attendue.

C.3.3 Conclusion des expériences : le compromis

Etant donné que les sources estimées peuvent être décomposées comme suit (cf. [Vincent *et al.* 2006] et section III-E, page 45) :

$$s^{\text{est}} = s_{\text{cible}} + s_{\text{interf}} + s_{\text{bruit}} + s_{\text{artefacts}}, \quad (\text{IV.14})$$

les algorithmes de SSA doivent minimiser dans le même temps l'apparition d'interférences, de bruit et d'artefacts afin d'obtenir de bonnes performances en terme de SDR. Ces termes sont tous les trois affectés de manière différente par l'a priori parcimonieux :

- *Les Interférences* : elles apparaissent quand les sources ne sont pas correctement ou complètement séparées. La parcimonie, en temps que mesure de diversité, peut être d'une grande aide pour obtenir une séparation correcte des sources et permet donc de garder le terme s_{interf} relativement faible. Cependant, il peut également créer des interférences en raison du biais induit par la régularisation, comme cela est visible sur la figure **IV.7**.
- *Le Bruit* : puisque le bruit Gaussien étudié dans cette partie s'étale uniformément sur tous les coefficients alors que les sources concentrent leur énergie dans un faible nombre de coefficients, le seuillage induit par les régularisations ℓ_0 et ℓ_1 débruite grandement les sources estimées. Cela réduit la norme du terme s_{bruit} et permet donc une meilleur reconstruction.
- *Les Artefacts* : le biais introduit par l'opérateur de seuillage doux contribue à créer des artefacts. Ceux-ci sont d'autant plus importants que les sources sont peu parcimonieuses.

Ces différents aspects interagissent entre eux. Comme cela a été montré dans cette section, nGMCA apporte un compromis efficace entre le bruit les interférences et le biais. En effet, grâce aux expériences, nous avons montré que nGMCA surpasse les autres algorithmes dans la plupart des scénarios envisagés en terme du critère SDR, qui prend en compte ces trois sources de détérioration des reconstructions.

Lorsque le taux d'activation est bas (sources très parcimonieuses), nGMCA obtient de meilleures performances que les autres algorithmes sur une grande plage de niveaux de bruit (figures **IV.9** et **IV.11**) tout en donnant des résultats satisfaisants dans le cas non-bruité. Dans ces contextes, le terme de régularisation ℓ_1 joue ses rôles, soient :

- obtenir une séparation correcte des sources en comparaison des autres autres types de régularisation, et en particulier de la régularisation ℓ_0 .
- débruiter les sources.

nGMCA est aussi visiblement robuste aux écarts au modèle parcimonieux : l'algorithme obtient des performances satisfaisantes avec des sources faiblement parcimonieuses (figures **IV.2** et **IV.12**) mais ceci au coût d'un biais plus important des sources estimées (figures **IV.8** et **IV.7**).

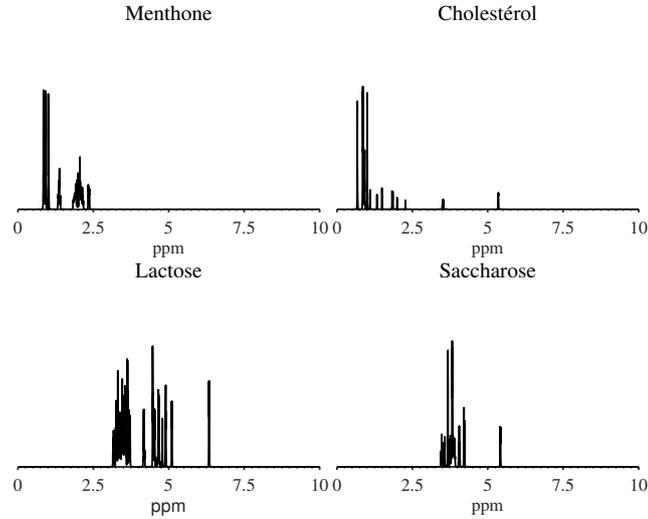


Figure IV.16 – Spectres RMN de 4 composés chimiques.

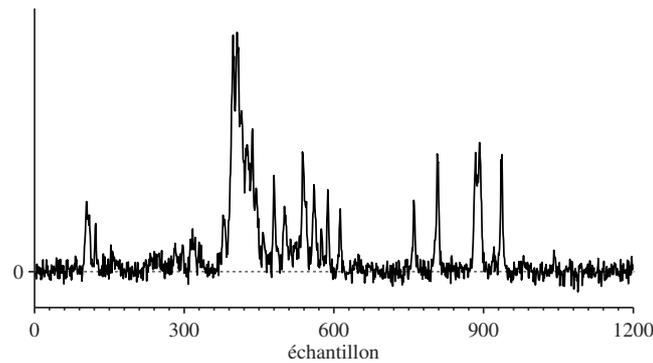


Figure IV.17 – Exemple de mélange (composé principal : lactose, $\text{SNR}_Y = 15\text{dB}$)

Enfin, nGMCA permet également une meilleure séparation que les autres algorithmes quand le nombre de sources est élevé (figure IV.14) ainsi que pour des problèmes mal-conditionnés en raison d'un faible nombre d'observations comme sur la figure IV.13, ou à cause de directions de mélange corrélées comme sur la figure IV.4. Ces résultats peuvent s'expliquer par l'effet séparant de la régularisation ℓ_1 avec une gestion appropriée de son paramètre, mais aussi par la prise en compte simultanée de la positivité et de la parcimonie.

C.4 Application à des données RMN synthétiques

La résonance magnétique nucléaire (RMN) est une technique physique qui permet de caractériser les molécules par un spectre qui leur est spécifique. Dans cette section, nous simulons des données plus réalistes que précédemment, à partir de spectres RMN de molécules réelles. Ces spectres sont bien adaptés à la NMF [Toumi *et al.* 2013], et en particulier au cas étudié ici puisqu'ils sont très parcimonieux et

positifs. Les informations concernant la localisation des pics de ces spectres et leur intensité peuvent être trouvées sur la "Spectral Database for Organic Compounds", SDBS¹. Afin de prendre en compte l'imperfection de l'acquisition, ces listes de pics ont été convolués par une Laplacienne dont la largeur à mi-hauteur est de 3 échantillons. Le nombre d'échantillons est fixé à $n = 1200$. \mathbf{S}^{ref} est fait de $r = 15$ tels spectres. La figure **IV.16** montrent 4 d'entre eux. Le produit scalaire normalisé entre certaines de ces sources peut être élevé et atteint par exemple 0.67 dans le cas du cholestérol et du menthone et peuvent ainsi difficilement être considérées comme indépendantes. Les coefficients de mélange \mathbf{A}^{ref} sont simulés de la même manière que précédemment ($p_A = 1$, $\alpha_A = 2$). Les données observées sont $\mathbf{Y} = \mathbf{A}^{\text{ref}}\mathbf{S}^{\text{ref}} + \mathbf{Z}$, où \mathbf{Z} est une matrice aléatoire Gaussienne dont les coefficients sont indépendants et identiquement distribués. Un exemple d'observation où le lactose est particulièrement présent est montré sur la figure **IV.17**.

Pour le test de performances de la figure **IV.18**, le nombre d'observations a été fixé au même nombre que celui des sources, c'est-à-dire $m = 15$, ce qui peut arriver dans certaines applications. Les courbes y montrent l'influence du bruit sur l'estimation des sources. Avec aussi peu d'observations, débruiter devient plus important que dans les conditions précédentes, alors que dans le même temps, le niveau de bruit est sous-estimé par l'algorithme puisque le problème est moins contraint. Pour compenser ces difficultés, τ_σ est fixé à 2 dans ces tests.

nGMCA^{naive} échoue à obtenir des résultats satisfaisants. En effet, dans ce scénario, le conditionnement du problème est très faible : $\text{cond}(\mathbf{A}^T\mathbf{A}) \approx 10^4$ et nGMCA^{naive} ne réussit pas à s'approcher d'une solution de qualité. À l'inverse, nGMCA obtient des résultats de 3 à 5dB meilleurs que tous les autres algorithmes. Cela montre encore une fois que nGMCA est particulièrement robuste à une grande variété de conditions. Un exemple de reconstruction est fourni figure **IV.19**, où nGMCA permet d'identifier davantage de pics que HALS avec régularisation parcimonieuse. Sa reconstruction n'est toutefois pas complètement débruitée, puisqu'il y a toujours un compromis à trouver entre le débruitage, la séparation, et le biais induit.

Sur la figure **IV.20**, le nombre d'observations varie de 15 à 90. Puisque le conditionnement augmente grandement pour des nombres d'observations plus élevés, les résultats de nGMCA^{naive} s'améliorent rapidement. Néanmoins, bien que nGMCA^{naive} et HALS obtiennent des résultats comparables à nGMCA quand il y a suffisamment d'observations, nGMCA est l'algorithme qui est le plus robuste et obtient de bons résultats dans l'ensemble des scénarios.

D Justification a posteriori de l'approche choisie

Après avoir observé le comportement de nGMCA et avant de conclure, il est intéressant de remettre l'algorithme en perspective en précisant les choix qui ont

1. http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

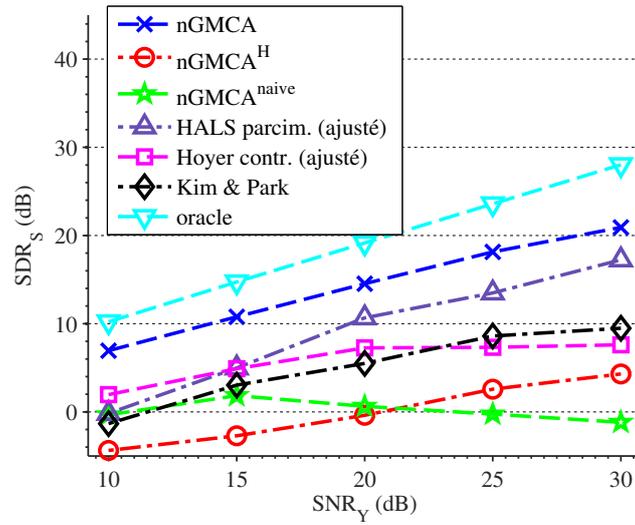


Figure IV.18 – SDR de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($m = 15$, $\tau_\sigma^\infty = 2$, moyenné sur 96 simulations de données RMN)

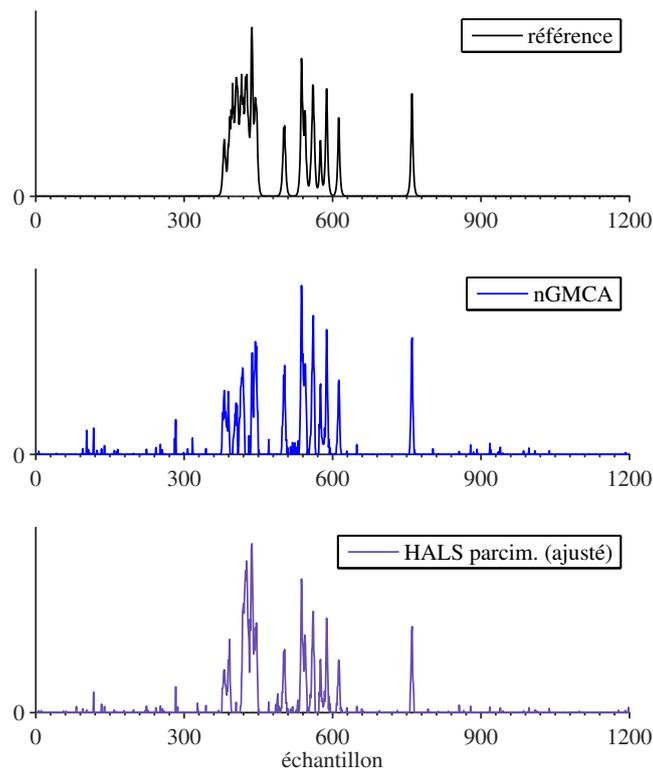


Figure IV.19 – Exemple de reconstruction (lactose, $SNR_Y = 15$ dB)

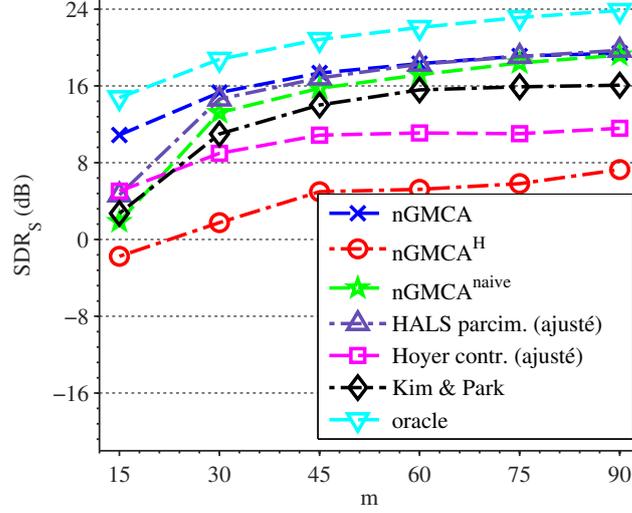


Figure IV.20 – Reconstruction SDR (SDR_S) en fonction du nombre d'observations m ($\text{SNR}_Y = 15\text{dB}$, $\tau_\sigma^\infty = 2$, moyenné sur 36 simulations de données RMN)

été effectués lors de sa conception. Deux choix plus particulièrement n'ont pas été précisés précédemment alors qu'ils revêtent une grande importance : les choix de formulation et de l'approche algorithmique utilisée.

D.1 Choix de la formulation régularisée

Le sous-problème en \mathbf{S} envisagé en (IV.6) peut être réécrit sous la forme sous contrainte suivante lorsque $\lambda > 0$:

$$\underset{\mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \|\mathbf{S}\|_1, \|\mathbf{AS} - \mathbf{Y}\|_2 \leq \varepsilon \quad (\text{IV.15})$$

Les deux formulations sont en effet équivalentes car pour n'importe quel $\lambda > 0$, il existe un ε tel que les deux problèmes ont le même minimum (ou les mêmes minima si le problème est sous-contraint). Le lien entre λ et ε n'est toutefois pas explicite. Cependant, alors qu'ajuster λ est assez complexe comme nous l'avons vu, et que nous ne proposons qu'une méthode grossière pour effectuer ce choix, il existe un moyen plus direct de choisir ε . En effet, si le bruit sur les coefficients des données est Gaussien indépendant et identiquement distribué, d'écart-type σ , alors la norme du résidu attendue peut être bien approximée par :

$$\|\mathbf{A}^{\text{ref}} \mathbf{S}^{\text{ref}} - \mathbf{Y}\|_2 \approx \sqrt{mn} \sigma \quad (\text{IV.16})$$

Deux raisons principales ont mené au choix de la formulation pénalisée (IV.6) plutôt que la formulation sous contrainte (IV.15). La première est la difficulté accrue en terme de minimisation de la fonction de coût. En effet, celle-ci peut toujours s'écrire sous la forme (IV.7) mais avec $f(\mathbf{S}) = i_{\|\mathbf{A} \cdot \mathbf{Y}\|_2 \leq \varepsilon}(\mathbf{S})$ qui n'est cette fois plus

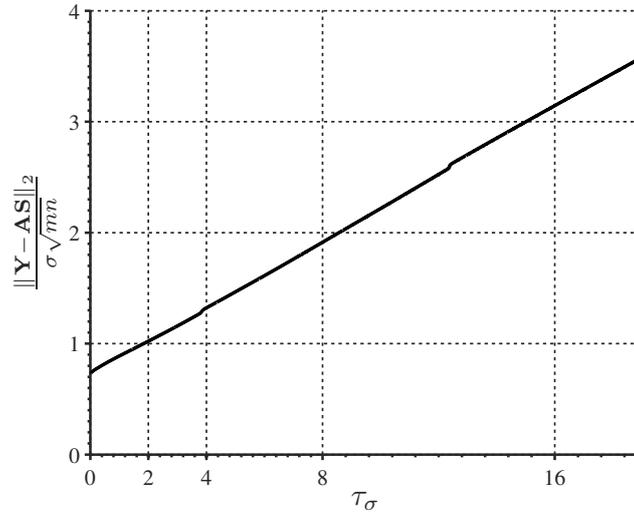


Figure IV.21 – Résidu en fonction de τ_σ (simulation RMN, $m = 15$, $\tau_\sigma^\infty = 2$, $\text{SNR}_Y = 25\text{dB}$).

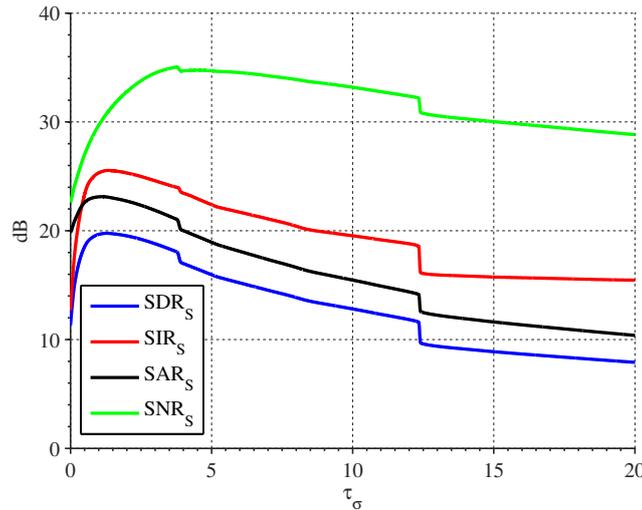


Figure IV.22 – SDR, SIR, SNR et SAR en fonction de τ_σ (simulation RMN, $m = 15$, $\tau_\sigma^\infty = 2$, $\text{SNR}_Y = 25\text{dB}$).

dérivable. Etant donné que f reste convexe, propre et semi-continue inférieurement, elle est toutefois proximable. Il est donc possible d'utiliser l'algorithme de Douglas-Rachford pour effectuer cette minimisation (voir section II-C.3.3, page 22), mais il n'existe pas d'accélération de celui-ci similaire à celle de l'algorithme Forward-Backward. Aussi, le proximal de f n'est pas explicite, et doit être calculé par la méthode de la sécante par exemple, et nécessite donc des sous-itérations.

La deuxième raison tient dans le choix d' ε . Celui-ci est conditionné par le niveau de bruit σ des données qui peut être approximé au cours de l'algorithme comme

expliqué précédemment (cf. B.2). Cependant, dans la formulation sous contrainte, un changement minime d' ε peut avoir une forte importance sur la solution, et donc σ doit être connu avec une précision difficilement atteignable.

Ceci peut être observé à l'aide des figures IV.21 et IV.22. La figure IV.21 montre le résidu normalisé par le bruit $\frac{\|\mathbf{Y}-\mathbf{A}\mathbf{S}\|_2^2}{\sqrt{mn}\sigma}$ en fonction du choix de τ_σ (le paramètre du parcimonie étant choisi à $\tau_\sigma\sigma$ dans nGMCA). Elle a été réalisée en utilisant un très grand nombre d'itérations (20 000) et en faisant décroître τ_σ linéairement vers 0. On remarque ainsi qu'une faible modification de $\varepsilon = \sqrt{mn}\sigma$ peut conduire à une forte modification du τ_σ équivalent, une mauvaise estimation de σ aura ainsi moins d'influence sur la version régularisée que sur la version contrainte.

La figure IV.22 vient compléter ce graphique en montrant les SDR, SIR, SNR et SAR en fonction de τ_σ . L'algorithme partant d'une grande valeur de τ_σ et le faisant décroître vers 0, ce graphique se lit de droite à gauche. On voit ainsi que la solution s'améliore d'abord en faisant décroître τ_σ , certaines sources apparaissant subitement (provoquant la discontinuité à $\tau_\sigma = 12,5$ par exemple). Le bruit commence à entrer dans la solution pour $\tau_\sigma \approx 4$ puisque le SNR commence alors à baisser. Ce résultat est attendu et motivait le choix de prendre τ_σ entre 1 et 3. Le plus faible biais induit par un seuillage plus faible permet cependant d'encore améliorer les autres critères jusqu'à $\tau_\sigma \approx 1,5$, et en particulier du SDR, qui est le critère prenant en compte tous les aspects de la séparation. À partir de cette valeur, tous les critères chutent, le bruit pouvant entraîner un démixage des sources. Encore une fois, un choix de τ_σ entre 1 et 3 semble donc convenable et permet une marge d'erreur (de l'ordre de 50% si l'on se place à $\tau_\sigma = 2$) importante, ce qui est utile en cas de mauvaise estimation de σ . D'après la figure IV.21, cette plage correspond à une marge d'erreur d'environ 5% seulement dans le cas de la formulation sous contrainte.

En résumé, la minimisation sous contrainte demande donc beaucoup plus d'efforts que celle de la formulation régularisée, et le choix de ε est également plus difficile que le choix de λ . La formulation régularisée est ainsi plus efficace et plus souple.

D.2 Approche algorithmique

Le choix d'utiliser des algorithmes proximaux a été dicté par leur adaptabilité à un grand nombre de types d'a priori, ce qui sera utile dans le chapitre suivant. Cependant, au sein de ces algorithmes, de nombreuses approches sont possibles. Plus spécifiquement, nous avons fait le choix ici de minimiser entièrement et alternativement chaque sous-problème, ce qui permet de converger vers un point critique du problème (IV.5) (voir section B.5). Récemment, des algorithmes proximaux ont été conçus pour traiter directement certains problèmes non-convexes comme le problème (IV.5), avec les mêmes propriétés de convergence vers un point critique [Atouch *et al.* 2010, Bolte *et al.* 2013, Xu & Yin 2013, Ochs *et al.* 2014]. Il aurait donc été possible d'utiliser un tel algorithme directement sur le problème global au lieu de se concentrer sur la minimisation alternée des sous-problèmes. Dans [Bolte

Algorithme IV.4 : nGMCA (PALM)**Paramètre(s) d'entrée** : \mathbf{Y} , K

- 1: initialiser $\mathbf{A}^{(0)}$, $\mathbf{S}^{(0)}$
- 2: **Pour** $k \leftarrow 1, K$ **faire**
- 3: **choisir** $\Lambda^{(k)}$ au niveau du bruit.
- 4: **choisir** $\gamma < \frac{1}{L}$ avec $L = \|\mathbf{A}^T \mathbf{A}\|_{\text{ope},2}$.
- 5: $\mathbf{S}^{(k)} \leftarrow \left[\text{Soft}_{\gamma \Lambda^{(k)}} (\mathbf{S}^{(k-1)} - \gamma \mathbf{A}^{(k-1)T} (\mathbf{A}^{(k-1)} \mathbf{S}^{(k-1)} - \mathbf{Y})) \right]_+$
- 6: **choisir** $\gamma < \frac{1}{L}$ avec $L = \|\mathbf{S} \mathbf{S}^T\|_{\text{ope},2}$.
- 7: $\mathbf{A}^{(k)} \leftarrow \text{proj}_{\{\mathbf{X}, \|\mathbf{x}_{\cdot,k}\|_2 \leq 1 \forall k\}} \left(\left[\mathbf{A}^{(k-1)} - \gamma (\mathbf{A}^{(k-1)} \mathbf{S}^{(k)} - \mathbf{Y}) \mathbf{S}^{(k)T} \right]_+ \right)$
- 8: **renvoyer** $\mathbf{A}^{(K)}$, $\mathbf{S}^{(K)}$

et al. 2013, Xu & Yin 2013], les auteurs montrent qu'utiliser une unique étape de gradient proximal pour les mises-à-jour de \mathbf{A} et \mathbf{S} , au lieu d'itérer afin de trouver le minimum global des sous-problèmes, permet toujours de converger vers un point critique du problème.

Cette approche est appelée PALM (proximal alternating linearized minimization) [Bolte *et al.* 2013] ou prox-linéaire [Xu & Yin 2013]. L'utiliser pour résoudre le problème (IV.5) de nGMCA pose toutefois de nouvelles difficultés. La stratégie de décroissance du seuil a ainsi moins d'intérêt. En effet, ce mécanisme se basait sur l'observation qu'un seuil élevé permettait d'obtenir que les plus grands coefficients de \mathbf{S} , davantage susceptibles de n'appartenir qu'à une unique source, et ainsi de pouvoir correctement en déduire les directions de mélanges \mathbf{A} . Avec l'approche PALM, les sous-problèmes ne sont pas résolus entièrement et cette observation ne tient donc plus. Aussi, si la norme des colonnes de \mathbf{A} n'est pas contrainte, la question de la normalisation afin d'éviter de converger vers une solution dégénérée se pose à nouveau, et est plus compliquée à gérer en plein cœur des itérations. Il est donc nécessaire de contraindre la norme des colonnes de \mathbf{A} dès le début de l'algorithme, alors qu'il peut être préférable de ne pas trop contraindre le problème dans un premier temps. L'Algorithme IV.4 montre le pseudo-code de nGMCA modifié afin d'utiliser l'approche PALM.

Le problème (IV.5) n'étant pas convexe, des algorithmes différents peuvent également converger vers des points critiques différents. La figure IV.23 compare donc la version nGMCA standard avec la version explicitée ci-dessus, utilisant l'approche PALM. Elle montre ainsi que la qualité de la reconstruction a une tendance significative à être meilleure dans le cas de la version standard de nGMCA. Ainsi, bien que rien ne permet de trancher dans le cas d'un problème non-convexe en général, il semble que minimiser complètement les sous-problèmes avec la stratégie de seuillage décroissant permet ici de converger vers des points critiques plus proches de la solution recherchée.

Enfin, dans le chapitre suivant, nous traiterons des cas où l'opérateur proximal de la parcimonie et la positivité n'est pas explicite, ce qui rendrait l'utilisation de

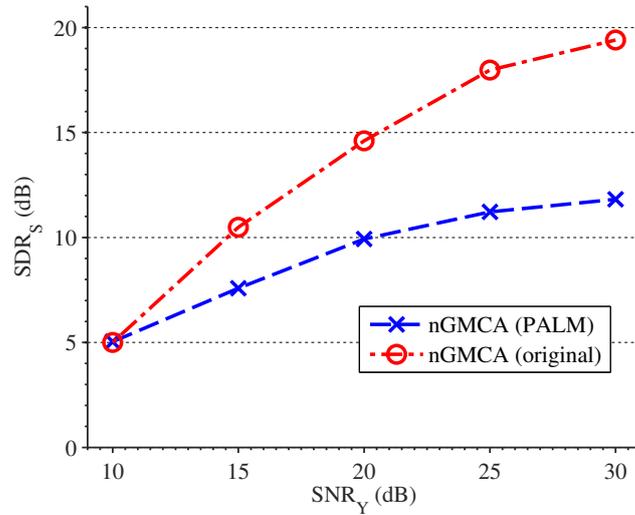


Figure IV.23 – SDR de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($m = 15$, $\tau_\sigma^\infty = 2$, moyenné sur 96 simulations de données RMN)

l'approche PALM plus difficile. Nous verrons cependant qu'à l'inverse, la minimisation globale des sous-problèmes peut encore être mise en œuvre sans l'aide de sous-routines.

E Conclusion

Afin de se conformer à la philosophie de la recherche reproductible [Buckheit & Donoho 1995], les algorithmes présentés dans ce chapitre sont disponibles sur le site web de l'équipe CosmoStat du Service d'Astrophysique du CEA². L'algorithme mis en avant, nGMCA, a montré ici sa robustesse à un grand nombre de situations différentes, allant des données bruitées aux données mal-conditionnées. Ceci a été permis par la prise en compte rigoureuse à la fois de la régularisation parcimonieuse et de la contrainte de positivité. Il sert ainsi de base aux adaptations qui suivent. En particulier, nGMCA n'est jusque-là capable de ne prendre en compte la parcimonie des sources qu'uniquement dans l'espace direct. Le prochain chapitre présente les améliorations nécessaires afin d'utiliser des a priori parcimonieux dans des espaces transformés afin de pouvoir élargir le champ des applications.

2. <http://www.cosmostat.org/GMCA Lab>

nGMCA dans un espace transformé

Sommaire

A	Parcimonie dans un espace transformé	80
A.1	Formulation en synthèse	80
A.2	Formulation en analyse	82
A.3	Complexité numérique	83
B	Expérimentations	83
B.1	Conditions expérimentales	84
B.2	Comparaison des approches	85
B.3	Conclusions des expérimentations	89
C	Mélanges d'images et repondération- ℓ_1	90
C.1	Influence du biais sur la séparation	90
C.2	La repondération comme compromis	92
C.3	Expérimentations	96
D	Conclusion	100

Nous avons pu observer dans le chapitre précédent qu'une prise en compte appropriée de la positivité et de la parcimonie permettait d'obtenir de meilleures séparations. Nous avons également vu comment utiliser la parcimonie afin de gérer une contamination par du bruit additif Gaussien. Cependant, la parcimonie telle que nous l'avons utilisée ne permet pas de prendre en compte toutes nos connaissances sur un signal comme celui montré à la figure [V.1](#). En effet, celui-ci peut être bien approximé par une fonction polynomiale par morceaux. Certains algorithmes tel que celui de Zdunek & Cichocki [[Zdunek & Cichocki 2007a](#)] (cf. chapitre [III](#)) permettent d'ajouter un a priori lisse sur les sources via la régularisation ℓ_2 du gradient par exemple. Cette approche n'est cependant pas complètement satisfaisante non plus, puisqu'elle pourra difficilement s'accommoder de singularités telles que celles observées à l'extrémité des pics de la figure [V.1](#).

Dans ce chapitre, nGMCA est modifié afin de pouvoir gérer la parcimonie dans des espaces transformés, aussi bien avec une formulation en synthèse qu'une formulation en analyse (section [A](#)). Des expérimentations permettent de montrer la puissance de ce type de régularisation (section [B](#)). L'utilisant d'ondelettes redondantes est ainsi particulièrement adaptée à la séparation de signaux semblables à celui de la figure [V.1](#). Cette régularisation peut cependant poser des problèmes pour

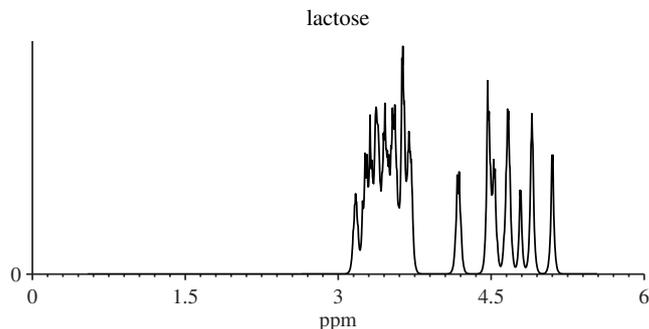


Figure V.1 – Spectre RMN du lactose (synthétique).

des signaux plus complexes en créant un biais important auquel nous remédions à l'aide d'une méthode de repondération- ℓ_1 (section C).

A Parcimonie dans un espace transformé

En séparation de sources en aveugle, la parcimonie dans un espace transformé a montré sa capacité à mettre en évidence les diversités morphologiques entre des sources et ainsi d'en améliorer la séparation [Zibulevsky & Pearlmutter 2001, Li *et al.* 2003, Bobin *et al.* 2007]. Dans cette partie, nous développons une version de nGMCA capable d'utiliser une régularisation parcimonieuse dans un espace transformé en NMF. Cette régularisation permet ainsi de modéliser plus finement des signaux comme celui de la figure V.1 en prenant tout à la fois en compte sa continuité et sa parcimonie dans l'espace direct, sans être mis en difficulté par les singularités aux extrémités des pics. Il faut cependant bien se rendre compte que contrairement au cas du chapitre précédent, la parcimonie des sources et leur positivité ne s'expriment plus dans le même domaine. On doit alors prendre en compte deux a priori différents dans deux espaces différents.

A.1 Formulation en synthèse

Comme vu en section B.3 du chapitre II (page 15), la parcimonie dans un espace transformé peut être exprimée avec deux formulations distinctes : en synthèse et en analyse. En synthèse, le but est de chercher la solution sous la forme de ses coefficients dans l'espace transformé. Il est ainsi nécessaire de modifier nGMCA (Algorithme IV.3, page 55) afin de rechercher non plus directement \mathbf{S} , mais ses coefficients dans l'espace image d'une transformée \mathbf{W} , notés \mathbf{S}_w . Cela mène ainsi à l'Algorithme V.1¹. Le résultat final pour les sources sera alors obtenu en reprojettant les coefficients d'ondelettes dans l'espace direct, via l'équation : $\mathbf{S}^{(K)} = \mathbf{S}_w^{(K)} \mathbf{W}$.

Alors que la mise-à-jour de \mathbf{A} n'est pas modifiée, celle de \mathbf{S}_w pose de nouvelles difficultés par rapport au chapitre IV. En effet, au chapitre IV et en reprenant

1. les problèmes de renormalisation sont ici passés sous silence, mais les mêmes considérations qu'en sections IV-B.4 et IV-B.5 doivent être appliquées (voir à partir de la page 57).

Algorithme V.1 : nGMCA-synthèse**Paramètre(s) d'entrée** : \mathbf{Y} , K

- 1: initialiser $\mathbf{A}^{(0)}$, $\mathbf{S}_w^{(0)}$ et $\mathbf{\Lambda}^{(1)}$
- 2: **Pour** $k \leftarrow 1, K$ **faire**
- 3: $\mathbf{S}_w^{(k)} \leftarrow \underset{\mathbf{S}_w \mathbf{W} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}^{(k-1)} \mathbf{S}_w \mathbf{W}\|_2^2 + \|\mathbf{\Lambda}^{(k)} \odot \mathbf{S}_w\|_1$
- 4: $\mathbf{S}^{(k)} = \mathbf{S}_w^{(k)} \mathbf{W}$
- 5: $\mathbf{A}^{(k)} \leftarrow \underset{\mathbf{A} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A} \mathbf{S}^{(k)}\|_2^2$
- 6: **choisir** $\mathbf{\Lambda}^{(k+1)} \leq \mathbf{\Lambda}^{(k)}$
- 7: **renvoyer** $\mathbf{A}^{(K)}$, $\mathbf{S}^{(K)}$

l'écriture (IV.7), $g(\mathbf{S}) = \lambda \|\mathbf{\Lambda} \odot \mathbf{S}\|_1 + i_{\geq \mathbf{0}}(\mathbf{S})$ avait un opérateur proximal analytique. Cependant, pour la plupart des transformées \mathbf{W} , l'opérateur proximal de $g(\mathbf{S}_w) = \|\mathbf{\Lambda} \odot \mathbf{S}_w\|_1 + i_{\mathbf{W} \geq \mathbf{0}}(\mathbf{S}_w)$ n'a plus d'expression aussi simple. Le calculer demanderait l'usage d'une routine à l'intérieur de l'algorithme, ce qui le ralentirait fortement. Suivant la transformation \mathbf{W} , plusieurs cas peuvent être envisagés afin d'éviter l'usage d'une telle sous-routine :

- (i) le cas d'une transformée sans caractéristiques particulières.
- (ii) le cas d'une transformée orthonormale ou d'une trame ajustée.
- (iii) le cas d'une décomposition positive sur des atomes positifs.

A.1.1 Cas général

La fonction g peut être décomposée comme la somme des fonctions $g_1(\mathbf{S}) = \|\mathbf{\Lambda} \odot \mathbf{S}_w\|_1$ et $g_2(\mathbf{S}) = i_{\mathbf{W} \geq \mathbf{0}}(\mathbf{S}_w)$. Dans le cas où l'opérateur proximal de chacune de ces fonctions peut être facilement calculé, l'algorithme GFB est particulièrement adapté à la résolution du sous-problème en \mathbf{S} (cf. chapitre II, section C.3.2, page 21). L'opérateur proximal de g_1 étant explicite (le seuillage doux), c'est sur l'étude de g_2 qu'il faut s'attarder.

Dans le cas général, l'opérateur proximal de g_2 n'est pas nécessairement explicite. Afin de résoudre le sous-problème en \mathbf{S} sans l'aide de sous-itérations, on peut utiliser l'algorithme de Chambolle-Pock [Chambolle & Pock 2010] présenté section II-C.3.4 (page 23). Ce cas n'est pas étudié ici, mais l'utilisation de cet algorithme est détaillée pour la formulation en analyse dans la section A.2 ci-dessous.

A.1.2 Transformées orthonormales et trames ajustées

En pratique, beaucoup de transformées sont cependant orthonormales, comme la transformée de Fourier ou les ondelettes orthonormales par exemple. L'opérateur proximal de g_2 est alors explicite dans ce cas là. Celui-ci se réduit en effet au proximal #5 du tableau A.1 en page 130. Dans les expériences de la section suivante, nous considérerons cette situation avec une transformée en ondelettes orthonormales sous le nom "nGMCA-orthonormal".

Algorithme V.2 : nGMCA-analyse**Paramètre(s) d'entrée** : \mathbf{Y} , K

- 1: initialiser $\mathbf{A}^{(0)}$, $\mathbf{S}^{(0)}$ et $\mathbf{\Lambda}^{(1)}$
- 2: **Pour** $k \leftarrow 1, K$ faire
- 3: $\mathbf{S}^{(k)} \leftarrow \underset{\mathbf{S} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}^{(k-1)}\mathbf{S}\|_2^2 + \|\mathbf{\Lambda}^{(k)} \odot (\mathbf{S}\mathbf{W}^T)\|_1$
- 4: $\mathbf{A}^{(k)} \leftarrow \underset{\mathbf{A} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}^{(k)}\|_2^2$
- 5: choisir $\mathbf{\Lambda}^{(k+1)} \leq \mathbf{\Lambda}^{(k)}$
- 6: renvoyer $\mathbf{A}^{(K)}$, $\mathbf{S}^{(K)}$

Comme mentionné précédemment (cf. chapitre II), l'usage de transformée redondantes peut souvent améliorer la reconstruction des signaux. Dans le cas particulier d'une trame ajustée, c'est-à-dire une transformée $\mathbf{W} \in \mathbb{R}^{p \times n}$ telle que $\mathbf{W}^T \mathbf{W} = \mathbf{I}_n$, l'opérateur proximal est toujours analytique et généralise le cas des transformées orthonormales. Il est fourni en tant qu'opérateur proximal #6 du tableau page 130, et son calcul est explicité dans l'annexe A-C. Dans la partie expérimentale qui suit, ces conditions sont utilisées avec des ondelettes redondantes pour l'algorithme appelé "nGMCA-synthèse", et la mise-à-jour est détaillée dans l'annexe B-A (page 133).

A.1.3 Dictionnaire d'atomes positifs

Le cas où les atomes de la transformée sont positifs mérite une attention particulière. En effet, il peut alors être intéressant de vouloir synthétiser les signaux comme une combinaison linéaire positive et parcimonieuse des atomes de \mathbf{W} , qui sont donc positifs. Ce modèle, plus restrictif que la positivité des coefficients de \mathbf{S} , prend la forme suivante :

$$\mathbf{S}^{(k)} \leftarrow \underset{\mathbf{S}_w}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}^{(k-1)}\mathbf{S}_w\mathbf{W}\|_2^2 + \|\mathbf{\Lambda}^{(k)} \odot \mathbf{S}_w\|_1 + i_{\geq \mathbf{0}}(\mathbf{S}_w). \quad (\text{V.1})$$

L'avantage de ce cas particulier est que la positivité et la parcimonie sont cette fois à nouveau exprimée dans le même domaine (le domaine transformé), c'est-à-dire que l'on utilise la contrainte $\mathbf{S}_w \geq \mathbf{0}$ au lieu de $\mathbf{S}_w\mathbf{W} \geq \mathbf{0}$ dans les cas précédents. La fonction $g(\mathbf{S}_w) = \|\mathbf{\Lambda} \odot \mathbf{S}_w\|_1 + i_{\geq \mathbf{0}}(\mathbf{S}_w)$ a alors comme opérateur proximal l'opérateur #3 du tableau A.1 (page 130), comme au chapitre IV, et l'algorithme FB peut être utilisé. Dans les expérimentations qui suivent, nous utilisons ce cas de figure avec une transformée convolutive pour l'algorithme nommé "nGMCA-convolutif".

A.2 Formulation en analyse

Alors que la minimisation s'effectuait dans l'espace transformé avec la formulation en synthèse, la formulation en analyse vise à rechercher les sources \mathbf{S} dans l'espace direct tout en faisant en sorte que leurs transformées $\mathbf{S}\mathbf{W}^T$ soient parcimonieuses. Les grandes étapes de nGMCA avec une régularisation en analyse sont

présentées dans l’**Algorithme V.2**. La seule différence est l’ajout de la transformée \mathbf{W} dans le terme de régularisation ℓ_1 . Mais comme dans le cas de la formulation en synthèse, celui-ci complexifie la résolution de la mise-à-jour de \mathbf{S} . De la même manière que précédemment, il est possible de décomposer le problème en :

$$f(\mathbf{S}) = \frac{1}{2} \|\mathbf{Y} - \mathbf{AS}\|_2^2, \quad (\text{V.2})$$

$$g_1(\mathbf{S}) = \|\mathbf{A} \odot (\mathbf{SW}^T)\|_1, \quad (\text{V.3})$$

$$g_2(\mathbf{S}) = i_{\geq \mathbf{0}}(\mathbf{S}). \quad (\text{V.4})$$

On peut alors résoudre le sous-problème en \mathbf{S} en utilisant l’algorithme GFB comme dans [Rapin *et al.* 2013b, Rapin *et al.* 2013c]. Contrairement au cas de la synthèse cependant, pour des transformations redondantes et même dans le cas de trame ajustées, l’opérateur proximal de g_1 n’est pas explicite et nécessite l’usage d’une sous-routine. Le calcul de cet opérateur est détaillé dans l’annexe A-D et le résultat est fourni en tant qu’opérateur #7 du tableau A.1 (page 130). Afin d’éviter les sous-itérations induites par l’usage de cet opérateur proximal, il est possible d’utiliser l’algorithme de Chambolle-Pock [Chambolle & Pock 2010] pour mener à bien la minimisation. L’annexe B-B (page 134) explique comment se placer dans le cadre de cet algorithme. Cette approche est utilisée dans la partie expérimentale de ce chapitre sous le nom "nGMCA-analyse" et utilisera des ondelettes redondantes comme transformée.

A.3 Complexité numérique

Chaque itération des mises-à-jour de \mathbf{S} implique une multiplication matricielle, une transformée avant et arrière, et des étapes linéaires telles que du seuillage, des projections sur l’orthant positif et des combinaisons linéaires. La multiplication matricielle s’effectue entre une matrice de taille $r \times r$ et \mathbf{S} , et a donc une complexité en $\mathcal{O}(r^2n)$. En notant $\phi(n)$ la complexité de la transformée d’un signal de longueur n , la complexité de l’étape impliquant les transformées en $\mathcal{O}(r\phi(n))$, ce qui donne une complexité totale de l’ordre de $\mathcal{O}(r\phi(n) + r^2n)$.

Dans ce chapitre, nous utilisons principalement une transformée en ondelettes redondante, qui a une complexité $\phi(n) = n \log_2(n)$ [Coifman & Donoho 1995]. La complexité numérique de la mise-à-jour de \mathbf{S} est donc linéarithmique en son nombre d’échantillons n à cause de la transformée, et quadratique en son nombre de sources r à cause de la multiplication matricielle.

Dans tous les cas, le nombre d’observations m n’intervient pas dans la complexité de la mise-à-jour de \mathbf{S} . Le nombre d’observations apparaît seulement (linéairement) dans la mise-à-jour de \mathbf{A} , qui a une complexité en $\mathcal{O}(r^2m)$ à cause de la multiplication matricielle.

B Expérimentations

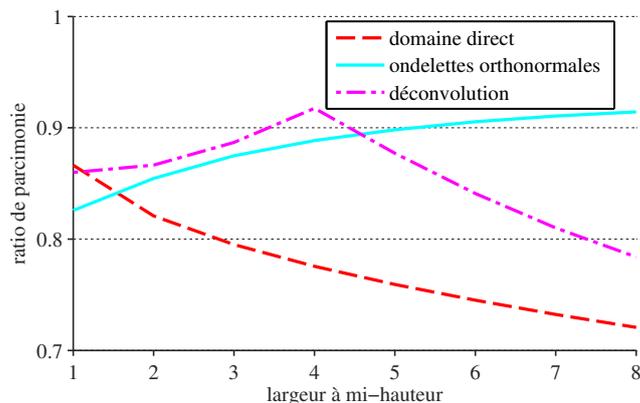


Figure V.2 – Ratio de parcimonie moyen des sources RMN sources dans les différents domaines.

B.1 Conditions expérimentales

Afin de comparer ces algorithmes, des données RMN sont simulées de la même manière que dans la section C.4 du chapitre précédent. Le nombre d'observations est choisi à $m = 32$, et le nombre d'échantillons à $n = 1024$. Le filtre convolutif permettant la génération des données est un filtre Laplacien de largeur à mi-hauteur de 4 échantillons, sauf quand il en est spécifié autrement, et est ainsi légèrement plus large que dans le chapitre précédent afin de mettre en valeur l'intérêt de la parcimonie en ondelettes. La source de la figure V.1 au début de ce chapitre a été synthétisée de cette manière.

Les sources générées de cette façon sont déjà naturellement parcimonieuses dans l'espace direct mais le sont davantage en ondelettes, comme indiqué à la figure V.2. Cette figure montre le ratio de parcimonie (cf. équation (III.15)) des sources dans l'espace direct, l'espace des ondelettes, et après déconvolution par un filtre Laplacien de largeur 4, en fonction de la largeur du filtre utilisé pour synthétiser les données. Dans le cas actuel, pour une largeur de 4 échantillons, on voit donc que les sources sont plus parcimonieuses (i.e. ont un ratio de parcimonie plus élevé) en ondelettes que dans l'espace direct. Ces sources sont en effet proches d'être polynomiales par morceaux. Nous reviendrons sur cette figure avec plus de détails à la fin de la prochaine section. Dans la suite \mathbf{W} est une transformée en ondelettes (symmlets) à 3 échelles pour nGMCA-orthonormal, nGMCA-analyse, et nGMCA-synthèse. La version orthonormale utilise Wavelab² et la version redondante pour l'analyse et la synthèse est basée sur une réimplémentation de la "Rice wavelet toolbox"³. Ces deux versions satisfont à la condition $\mathbf{W}^T \mathbf{W} = \mathbf{I}_n$ après renormalisation si nécessaire, ce qui en fait des trames ajustées. Avec ces configurations, le ratio de parcimonie moyen est de 0,78 dans le domaine direct et monte à 0,89 dans le domaine des ondelettes orthonormales. τ_σ est choisi à 1 pour la version standard de

2. <http://www-stat.stanford.edu/~wavelab/>

3. <http://dsp.rice.edu/software/rice-wavelet-toolbox>

nGMCA et nGMCA-orthonormal, et à 2 pour les versions en analyse et en synthèse.

Les sources utilisées pour les expérimentations sont simulées à partir d'impulsions positives convoluées avec un filtre Laplacien. En tant que telles, ces sources s'écrivent donc comme combinaisons linéaires positives et parcimonieuses d'atomes de la transformée convolutive ayant ce même filtre Laplacien. La version convolutive de nGMCA (nGMCA-convolutif) utilisera donc une transformée convolutive \mathbf{W} telle que pour un vecteur colonne x , $\mathbf{W}x = f * x$. f sera le noyau de convolution Laplacien de largeur à mi-hauteur de 4 échantillons utilisé pour générer l'essentiel des données. En ce sens, nGMCA-convolutif agira comme un oracle puisqu'il a accès au filtre ayant généré les données, ce qui n'est pas envisageable sur des données réelles. Pour cette version, on utilise $\tau_\sigma = 3$.

Comme dans le chapitre précédent, ces algorithmes sont comparés avec une version accélérée et parcimonieuse de HALS [Gillis 2012] dont le paramètre est ajusté en fonction du niveau de parcimonie des sources. Afin de comparer l'a priori en ondelettes avec un terme de lissage, les tests incluent aussi la NMF lissée [Zdunek & Cichocki 2007a] introduite au chapitre III. Puisqu'il n'y a pas de méthode directe permettant d'ajuster le paramètre de lissage α , nous gardons le meilleur résultat obtenu à partir de 12 valeurs de α distribuées logarithmiquement entre 0.00001 et 1. De cette façon, ces deux algorithmes ont des paramètres proches des paramètres optimaux alors qu'aucun ajustement n'est effectué pour les algorithmes basés sur nGMCA. Comme pour le chapitre précédent, l'évaluation des résultats s'effectue à l'aide des critères présentés dans [Vincent *et al.* 2006].

B.2 Comparaison des approches

Les figures V.3, V.4 et V.5 permettent d'observer les performances des algorithmes basés sur nGMCA comparées à celles des autres méthodes, respectivement en fonction du niveau de bruit, du nombre d'observations m et du nombre de sources r . Les résultats montrent une tendance commune sur ces trois figures. La NMF lissée n'obtient pas de bons résultats pour ce type de sources. Les méthodes utilisant la parcimonie dans l'espace direct (nGMCA, HALS parcimonieux) se comportent quant à elles légèrement mieux. Enfin, l'a priori parcimonieux dans l'espace transformé améliore significativement les performances. nGMCA-orthonormal est ainsi globalement 1 à 3dB au-dessus de nGMCA et HALS parcimonieux, et les ondelettes redondantes améliorent encore les résultats de 1 à 2dB aussi bien en analyse qu'en synthèse. nGMCA-convolutif obtient quant à lui encore 5dB supplémentaires en moyenne.

Il est important de garder à l'esprit que le paramètre de régularisation dans le cas des algorithmes basés sur nGMCA est géré de façon automatique. Ce n'est pas le cas des autres algorithmes pour lesquels le choix des paramètres n'est pas toujours aisé. Il est ainsi particulièrement difficile de savoir quel paramètre convient le mieux dans le cas de la NMF lissée. Dans nos expérimentations, ce paramètre est ajusté au mieux afin d'obtenir les meilleurs résultats, évalués en considérant les sources connues. Dans HALS parcimonieux, ce paramètre est ajusté à partir du niveau de

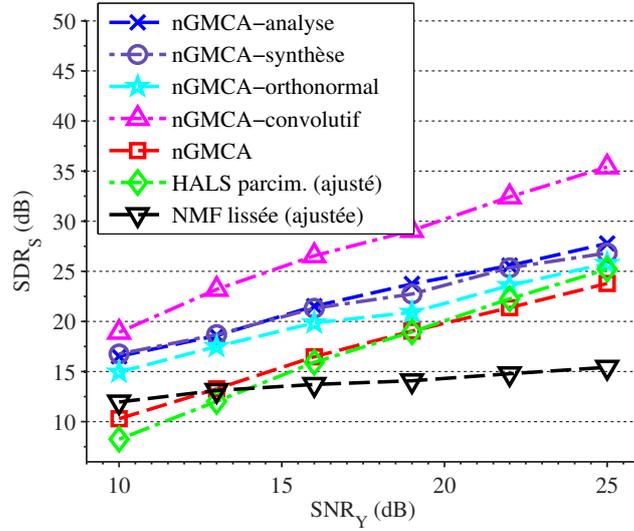


Figure V.3 – SDR moyen de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($r = 12$, $m = 32$, moyenné sur 36 simulations de données RMN)

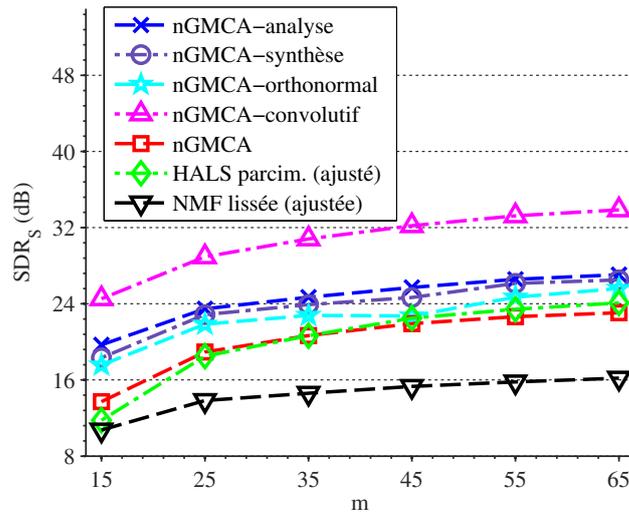


Figure V.4 – SDR moyen de la reconstruction (SDR_S) en fonction du nombre d'observations m ($r = 12$, $SNR_Y = 20$ dB, moyenné sur 36 simulations de données RMN)

parcimonie des vraies sources. Bien évidemment, cela ne pourrait pas être le cas avec des données non-simulées.

Le fait que nGMCA-convolutif obtienne les meilleurs résultats uniformément sur ces trois figures était attendu car le modèle convolutif qu'il utilise est exact, ayant servi à générer les données. Cette information n'est cependant pas connue et

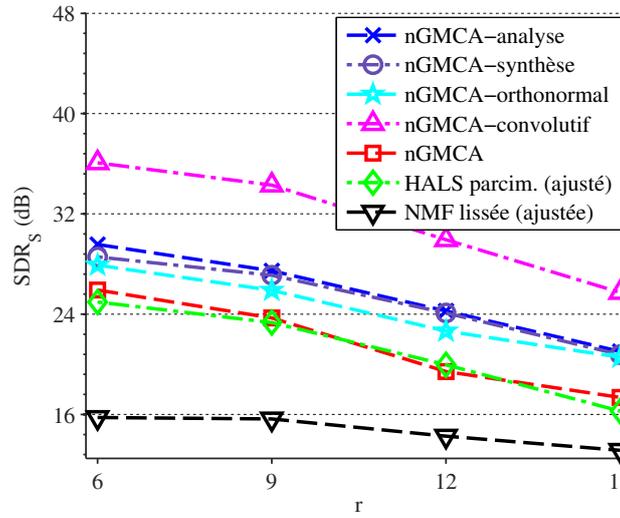


Figure V.5 – SDR moyen de la reconstruction (SDR_S) en fonction du nombre de sources r ($SNR_Y = 20\text{dB}$, $m = 32$, moyenné sur 36 simulations de données RMN)

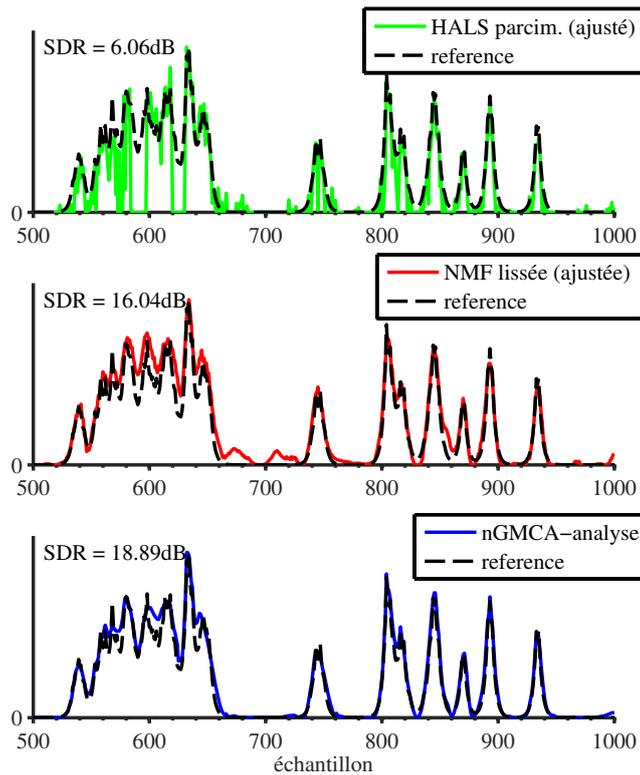


Figure V.6 – Exemples de reconstruction ($r = 12$, $m = 32$, $SNR_Y = 10\text{dB}$)

ne serait même pas fiable en pratique. Cela sera le propos d'une expérience dans la suite de cette section. La NMF lissée n'est pas capable de bien identifier les

sources. En effet, alors qu'elle est bien adaptée pour retrouver des structures géométriques lisses, elle n'est pas appropriée pour retrouver les singularités des signaux qui nous intéressent ici. nGMCA et HALS parcimonieux obtiennent des résultats légèrement meilleurs puisque les sources sont plutôt parcimonieuses dans l'espace direct. Comme mentionné précédemment, elles sont cependant plus parcimonieuses une fois exprimées avec des ondelettes orthonormales (ratio de parcimonie de 0.89 comparé à 0.78 dans l'espace direct), ce qui explique les 1 à 3dB supplémentaires obtenus par nGMCA-orthonormal. L'invariance par translation apportée par les ondelettes redondantes est aussi très bénéfique comme le montre les résultats de nGMCA-analyse et nGMCA-synthèse qui sont encore 1 à 2dB meilleurs que les ondelettes orthonormales, avec un net avantage pour la formulation en analyse. Cela corrobore les résultats de [Elad *et al.* 2007, Selesnick & Figueiredo 2009] pour des signaux naturels qui ne peuvent pas être synthétisés exactement à partir d'un faible nombre d'atomes de la transformée en ondelettes. Dans ce cas-là, la régularisation en analyse apporte généralement une meilleure robustesse et obtient des meilleurs résultats de reconstruction.

La figure **V.6** montre un exemple de reconstruction pour nGMCA-analyse, HALS parcimonieux et la NMF lissée, et confirme notre analyse. En effet, l'effet de lissage de la NMF lissée est très clairement visible, avec des pics qui ont été arrondis. Elle souligne aussi le fait que la parcimonie dans l'espace direct n'est pas complètement adaptée puisqu'elle ne capture par la structure continue de ces sources ; elle tend essentiellement à biaiser les sources estimées vers 0. Les ondelettes ont la capacité de représenter aussi bien les singularités locales (la pointe des pics) que les structures lisses à plus grande échelle. Imposer la parcimonie dans le domaine des ondelettes, comme avec nGMCA-analyse, permet alors une meilleure estimation à la fois des pics et de la continuité de ces signaux.

Il est aussi très instructif de regarder l'influence de la largeur des pics des sources. Sur la figure **V.7**, la largeur à mi-hauteur des pics devient un paramètre et est ainsi modifiée par rapport aux cas précédents où elle était fixée à 4 échantillons. Cela a un impact important sur le modèle puisque plus la largeur est faible, plus les pics sont pointus et les données parcimonieuses, et à l'inverse, plus la largeur est grande, plus prendre en compte la continuité est important. Sans surprise, la NMF lissée réussit d'autant moins à reconstruire les sources que les pics deviennent étroits. À l'inverse, la version de base de nGMCA obtient de meilleurs résultats quand les pics sont étroits. Ses performances baissent cependant comparé à nGMCA-synthèse et nGMCA-analyse quand la largeur des pics augmente. Il est finalement intéressant de remarquer que dans cette expérience, le noyau de convolution utilisé par nGMCA-convolutif est laissé inchangé. Il n'est alors plus parfaitement adapté que pour une largeur à mi-hauteur de 4 échantillons, et les performances de nGMCA-convolutif baissent vite quand la largeur du filtre ayant servi à générer les données s'éloigne de cette valeur. Cet algorithme est ainsi très sensible aux écarts au modèle utilisé.

Ces résultats sont à mettre en regard avec les courbes de la figure **V.2** qui fournissent le ratio de parcimonie dans les domaines direct, des ondelettes, ainsi

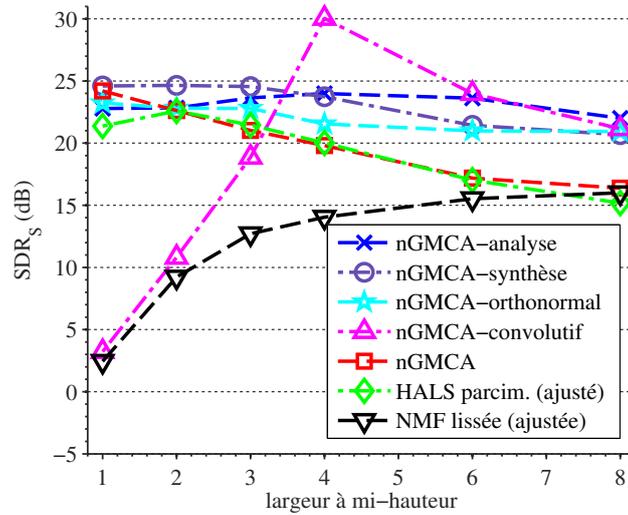


Figure V.7 – SDR moyen de la reconstruction (SDR_S) en fonction de la largeur des pics ($SNR_Y = 20\text{dB}$, $r = 12$, moyenné sur 36 simulations de données RMN)

qu’après une inversion du filtre Laplacien de largeur 4, en fonction de la largeur du filtre ayant généré les données. Les résultats de la figure V.7 présentent en effet sensiblement les mêmes tendances. Ainsi, les résultats de la version standard de nGMCA baissent quand la largeur augmente car la parcimonie dans l’espace direct baisse. De la même manière, les versions de nGMCA utilisant la parcimonie en ondelettes sont moins affectées par la largeur des pics car celle-ci modifie peu le ratio de parcimonie des sources dans ce domaine. Enfin, nGMCA-convolutif apparaît comme étant bien plus sensible que les autres algorithmes à l’exactitude du modèle génératif qu’il met en œuvre, avec des résultats fortement dégradés quand la largeur de son filtre convolutif n’est pas la même que celle des pics des données. Bien qu’il présente tout de même de bons résultats ici, cet algorithme ne serait pas envisageable en pratique, les pics n’étant pas uniformes dans la réalité et aucun filtre convolutif ou banque de filtres convolutifs comportant des largeurs et des formes différentes ne pouvant donc les modéliser parfaitement.

B.3 Conclusions des expérimentations

Au travers de ces expériences, nous avons vu que les informations structurelles apportées par la transformée s’avèrent utiles pour la reconstruction de sources comportant des structures géométriques complexes qui ne peuvent pas être correctement prises en compte par la NMF lissée ou pas les méthodes standards de NMF utilisant la parcimonie dans l’espace direct. Cela est particulièrement vrai dans des conditions difficiles comme lorsque le nombre d’observations est faible ou quand le nombre de sources est élevé. Comparée au lissage, la parcimonie dans un domaine bien choisi (ici dans le domaine des ondelettes) a aussi l’avantage de préserver les pics tout en reconstruisant correctement les structures lisses à plus grande échelle. Enfin,

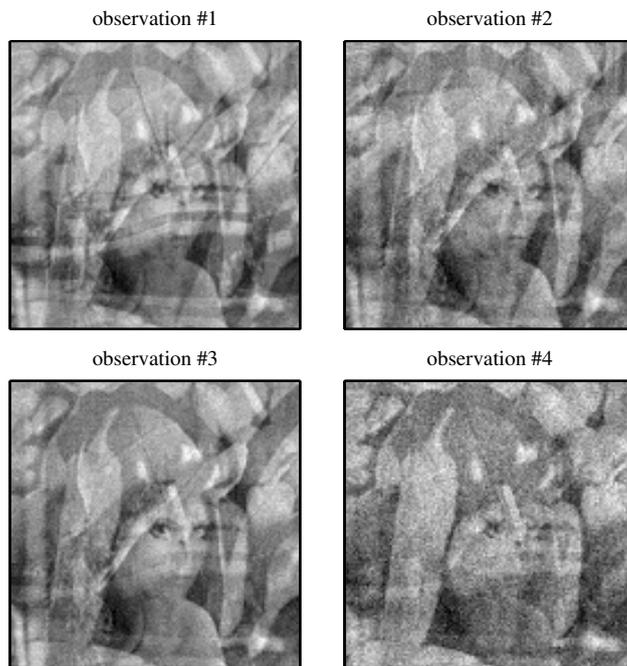


Figure V.8 – 4 mélanges bruités sur 32 (25dB).

la stratégie de seuillage proposée ici est aisée à mettre en place, ne nécessitant pas d’ajustement fin du paramètre de parcimonie, et est efficace dans une grande variété de contextes.

À notre connaissance, ces expérimentations présentent également la première comparaison entre les formulations en analyse et en synthèse dans le domaine de la SSA et plus particulièrement de la NMF. Comme il a déjà été souligné dans le cas des problèmes inverses [Elad *et al.* 2007, Selesnick & Figueiredo 2009], les problèmes formulés en analyse sont généralement plus robustes à des écarts au modèle puisque, contrairement à ceux exprimés en synthèse, ils ne supposent pas que les signaux à retrouver peuvent nécessairement se décomposer avec un faible nombre d’éléments non nuls dans l’espace transformé.

C Mélanges d’images et repondération- ℓ_1

On peut directement étendre nGMCA à la séparation de sources 2D telles que des images, en modifiant la transformée utilisée en conséquence. On observe cependant des résultats qui ne sont pas à la hauteur de ceux obtenus en 1D. Cette section explique pourquoi et apporte une solution au problème observé.

C.1 Influence du biais sur la séparation

Au chapitre IV, nous avons discuté des avantages et inconvénients de l’usage de la régularisation ℓ_1 dans le contexte de la SSA positive. En pratique, alors que cette



Figure V.9 – Vérité terrain des 4 sources 2D (128×128 pixels).

régularisation peut parvenir à des performances de séparation supérieures, elle est également connue pour introduire un biais qui peut gêner cette même séparation. Si le biais induit par la régularisation ℓ_1 peut être négligé pour des sources très parcimonieuses (c'est-à-dire ayant un faible nombre d'entrées non nulles), ce n'est pas le cas pour des données plus complexes qui ne sont qu'approximativement parcimonieuses. Dans ce dernier cas, toutes les entrées sont non nulles mais seul un faible nombre ont une amplitude vraiment significative. Les images naturelles sont ainsi un bon exemple de signaux approximativement parcimonieux dans le domaine des ondelettes. Dans ces conditions, la régularisation ℓ_1 peut ne pas être aussi efficace.

La figure **V.8** montre 4 mélanges bruités sur 32 réalisés des 4 sources présentées figure **V.9** (respectivement : Léna, des poivrons, un bateau et Barbara), les coefficients de mélange étant tirés comme les valeurs absolues des entrées d'une matrice aléatoire Gaussienne. La figure **V.10a** montre la reconstruction de Léna après avoir résolu le problème inverse de mise-à-jour de \mathbf{S} à partir de la matrice de mélange \mathbf{A}^{ref} ayant servie à générer les données :

$$\operatorname{argmin}_{\mathbf{S} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}^{\text{ref}} \mathbf{S}\|_2^2 + \|\mathbf{\Lambda}_\mu \odot (\mathbf{S} \mathbf{W}^T)\|_1, \quad (\text{V.5})$$

où les coefficients de la matrice $\mathbf{\Lambda}_\mu$ valent 0 sur l'échelle grossière et μ sur les échelles fines. On peut ainsi observer de fortes interférences : des fragments des poivrons et de la tête de Barbara sont visibles sur l'image. Les figures **V.10b**, **V.10c** et **V.10d** montrent respectivement l'intensité des contaminations par les interférences, le bruit et les artefacts (le blanc signifiant une valeur de 0, et la dynamique étant amplifiée

4 fois par rapport à [V.10a](#) afin d'améliorer la visualisation), telles que définies dans Vincent et al. [[Vincent et al. 2006](#)] et introduites section III-E (page 45). Il ressort ainsi que les artefacts sont particulièrement présents, et localisés sur les contours de l'image. Ceci est dû au seuillage des grands coefficients d'ondelettes sur ces contours. Les figures en [V.11](#) montrent quant à elles la reconstruction en prenant $\mu = 0$ et donc $\mathbf{\Lambda}_\mu = \mathbf{0}$, c'est-à-dire en résolvant le problème des moindres carrés positifs. Cette fois, il n'y a pas d'interférences et d'artefacts visibles puisque pas de régularisation (figures [V.11b](#) et [V.11d](#)), mais la contamination par le bruit est importante (figure [V.11c](#)).

Dans le cadre des problèmes inverses, la régularisation ℓ_1 est souvent remplacée par une régularisation ℓ_0 qui produit généralement un moindre biais. Dans le contexte de la NMF parcimonieuse, nous avons observé au chapitre IV qu'une pénalisation ℓ_0 ne produit pas ce genre d'interférences, mais ne se comporte pas aussi bien que la régularisation ℓ_1 pour séparer les sources. C'est particulièrement vrai dans des conditions de séparation difficiles, comme lorsque le nombre de sources r est élevé par exemple. Afin d'obtenir une séparation plus précise de signaux approximativement parcimonieux, comme des images naturelles, un traitement spécifique doit être appliqué afin de ne garder que le meilleur des pénalisations ℓ_0 et ℓ_1 .

C.2 La repondération comme compromis

Dans le contexte des problèmes inverses linéaires régularisés, plusieurs techniques ont été conçues afin d'éviter l'apparition du biais généralement induit par la régularisation ℓ_1 . Dans [[Gao & Bruce 1997](#)], les auteurs ont introduit un mélange entre les seuillages doux et dur qu'ils ont appelé seuillage ferme. Dans [[Voronin & Woerdeman 2013](#)], les auteurs ont proposé de convertir itération après itération le seuillage doux en seuillage ferme. Cette stratégie profite ainsi de la convexité de la régularisation ℓ_1 au début de l'algorithme pour approcher une solution unique, avant de profiter du biais moindre du seuillage ferme.

Ici, nous optons pour un choix alternatif appelé repondération- ℓ_1 [[Candès et al. 2007](#)]. Cette méthode consiste à choisir les coefficients de la matrice de régularisation $\mathbf{\Lambda}$ en fonction de l'amplitude des coefficients correspondants de la transformée de \mathbf{S} . Les entrées de la transformée de \mathbf{S} ayant les plus grandes amplitudes sont alors moins pénalisées que celles ayant une plus faible amplitude, plus susceptibles de n'être dues qu'à du bruit. Cela est généralement effectué en résolvant de manière itérative une série de problèmes régularisés avec du ℓ_1 repondéré. À chaque étape de la série, les coefficients de $\mathbf{\Lambda}$ sont modifiés en considérant l'estimée courante des sources \mathbf{S} . L'avantage principal de cette méthode est que, bien qu'elle tente de résoudre un problème non-convexe, chaque étape reste tout de même convexe.

Plusieurs stratégies de repondération ont été proposées [[Candès et al. 2007](#)]. Dans les expérimentations qui suivent, la matrice $\mathbf{\Lambda}$ sera mise à jour comme suit :

$$\mathbf{\Lambda}^{\text{rew.}} = \mathbf{\Lambda} \oslash \left(\mathbf{1} + (\mathbf{S}_{\text{inv}|w} \oslash \mathbf{\Sigma})^2 \right), \quad (\text{V.6})$$

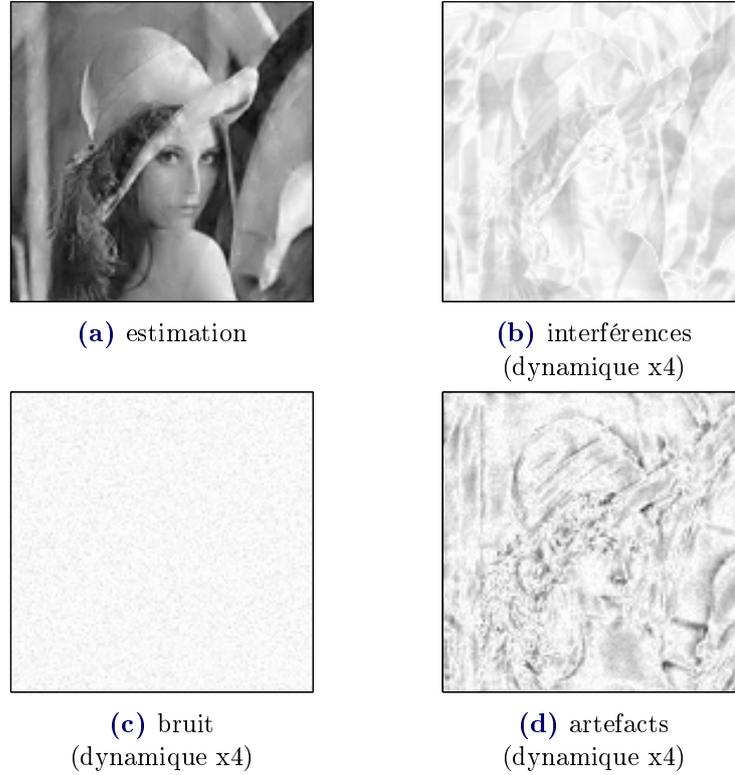


Figure V.10 – Exemple des composantes de reconstruction pour Léna (régularisation ℓ_1 standard, $\mu = 0.05$, SDR = 21, 77dB pour cette estimation de Léna).

où le carré est appliqué élément par élément, et $\mathbf{S}_{w|inv}$ est une estimée des sources faite à partir du problème des moindres carrés utilisant l'estimée courante de \mathbf{A} :

$$\mathbf{S}_{w|inv} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} \mathbf{W}^T. \quad (\text{V.7})$$

La matrice Σ contient l'écart-type du bruit contaminant les sources dans le domaine transformé. En pratique, celui-ci est évalué à partir de l'estimateur robuste de l'écart-type du bruit qu'est le MAD de $\mathbf{S}_{w|inv}$, que nous utilisons également section A.

Cette repondération du terme de régularisation ℓ_1 peut être utilisée sur le cas précédent des mélanges bruités d'images naturelles avec \mathbf{A}_μ , connaissant la matrice de mélange \mathbf{A}^{ref} . Le résultat pour la reconstruction de Léna est exposé figure V.12a. On observe alors que cela fait largement baisser l'importance des artefacts (figure V.12a) et fait quasiment disparaître les interférences (figure V.12b) tout en gardant un niveau de bruit très bas (figure V.12c)

De façon plus quantitative, les critères d'évaluations de Vincent et al. [Vincent et al. 2006] sont résumés dans le tableau V.1 pour les 3 cas exposés. (aucune régularisation, régularisation ℓ_1 standard, et régularisation ℓ_1 repondérée). En comparaison d'un simple moindres carrés, l'usage d'une régularisation ℓ_1 aide visiblement à

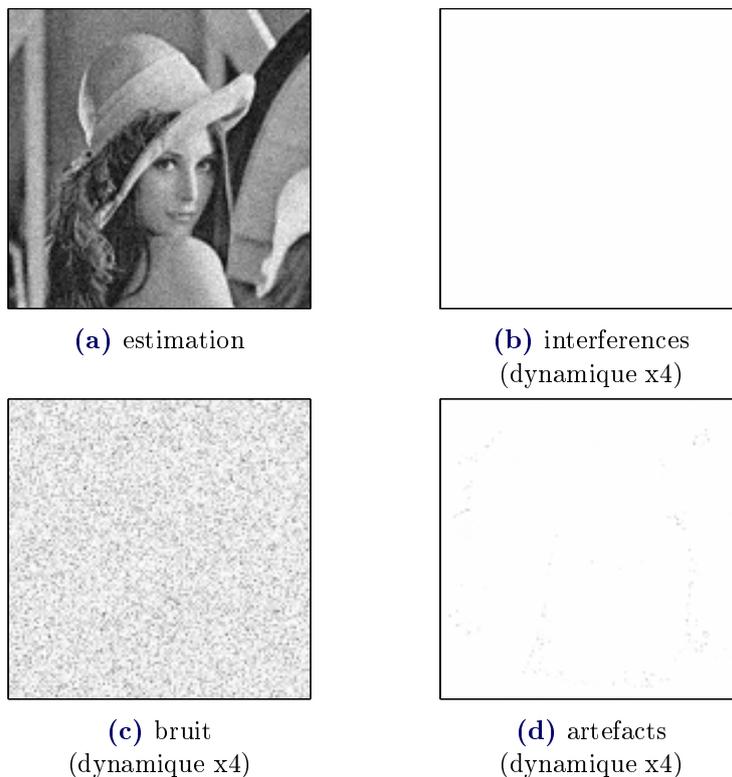


Figure V.11 – Exemple des composantes de reconstruction pour Léna (aucune régularisation, $\text{SDR} = 24, 10\text{dB}$ pour cette estimation de Léna).

critères	aucune régularisation	ℓ_1 standard	ℓ_1 repondéré
SDR_S	23.5	20.9	25.8
SIR_S	64.9	25.7	35.4
SNR_S	23.5	37.0	30.6
SAR_S	51.1	22.9	28.5

Tableau V.1 – Critères de séparations pour une inversion avec les seuillages standards et repondérés ($\text{SNR}_Y = 25\text{dB}$, $\mu = 0, 05$).

réduire la contamination par le bruit (augmentant ainsi le SNR), mais au coût de plus d'interférences et d'artefacts (faisant baisser le SIR et le SAR). Dans la dernière colonne, la pénalisation ℓ_1 repondérée offre un bon équilibre entre le débruitage et une séparation efficace, avec un SNR plus faible mais un SIR bien plus élevé.

On pourrait néanmoins avancer que la différence de résultats serait due uniquement au fait que le μ dans Λ_μ^{rew} et dans Λ_μ n'ont pas la même signification, et qu'à un μ différent, on pourrait trouver des résultats contradictoires. La figure **V.13** apporte cependant des éléments consolidant nos observations. En effet, elles montrent la qualité des reconstructions en terme de SDR des versions standards et repondérées de l'inversion en analyse, en fonction du paramètre μ et en fonction du résidu $\|\mathbf{Y} - \mathbf{AS}\|_2$.

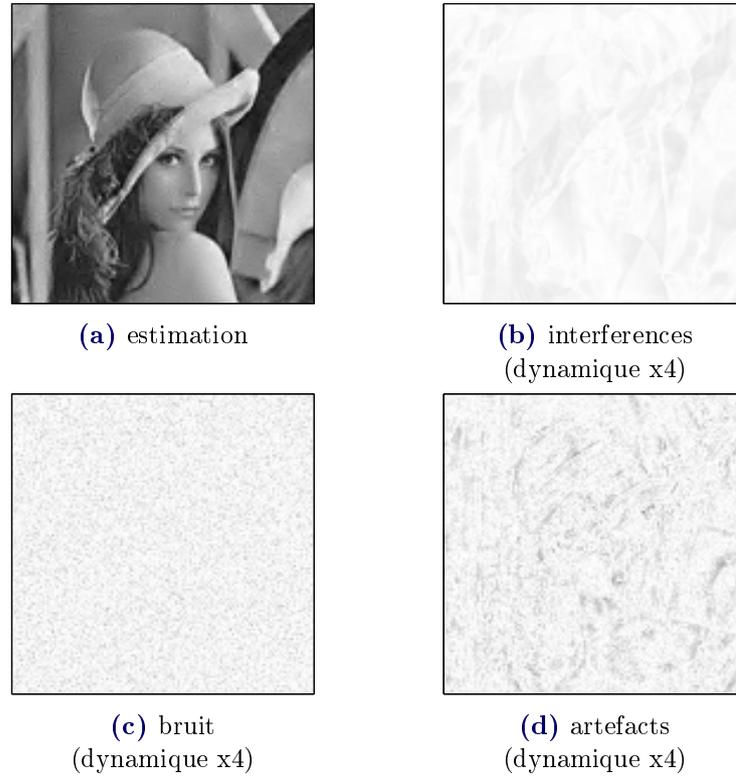


Figure V.12 – Exemple des composantes de reconstruction pour Léna (régularisation ℓ_1 repondérée, $\mu = 0.05$, SDR = 26,62dB pour cette estimation de Léna).

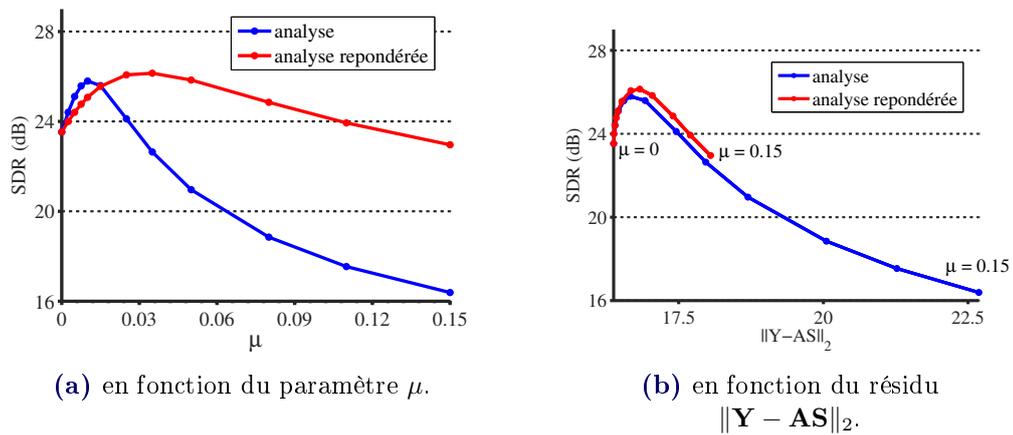


Figure V.13 – Évolution du SDR sur un exemple représentatif ($\text{SNR}_Y = 25\text{dB}$).

En gardant à l'esprit que les deux courbes la figure **V.13a** ne peuvent pas vraiment être comparées à un μ fixé, ce premier tracé permet tout de même de constater que le processus de repondération permet d'atteindre de meilleures reconstructions et surtout, apporte une robustesse bien plus grande puisqu'il permet d'augmenter

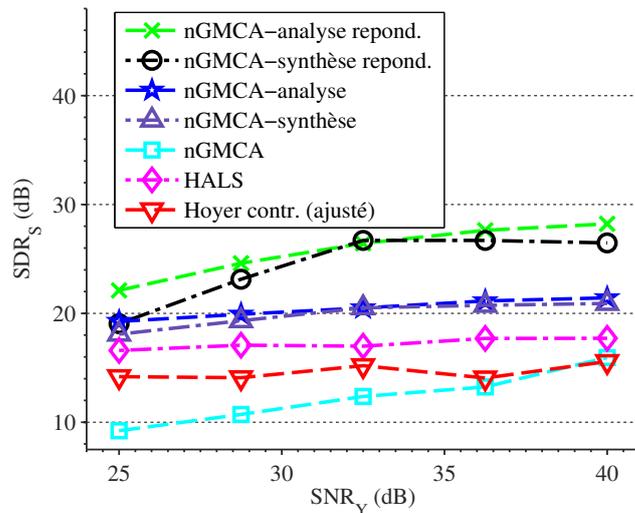


Figure V.14 – SDR moyen de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($r = 4$, $m = 16$, moyenné sur 24 simulations de mélanges d’images).

le SDR sur une plage de valeurs de μ bien plus large. Ainsi, une moindre précision est nécessaire pour l’ajustement de μ lorsque l’on utilise la repondération. L’expérience précédente utilisant $\mu = 0,05$, on observe effectivement que la régularisation non-repondérée détériore alors la solution quand l’utilisation de la repondération l’améliore par rapport à la solution non-régularisée ($\mu = 0$).

Le deuxième tracé en figure V.13b montre exactement les mêmes points, mais cette fois en fonction de la norme résiduelle obtenue $\|\mathbf{Y} - \mathbf{AS}\|_2$. Cela permet une comparaison plus juste puisque ce terme d’attache aux données est identique dans les deux cas, seul le terme de régularisation étant modifié. Lorsque μ augmente, la norme résiduelle augmente également car la régularisation biaise la solution, et on retrouve ainsi le cas sans régularisation $\mu = 0$ à l’extrémité gauche des deux courbes, et le cas $\mu = 0,15$ à leur extrémité droite. Le fait que la courbe non-repondérée parte rapidement vers de plus grandes valeurs de résidu est un point remarquable. Cela indique en effet qu’un léger changement de μ biaise très vite la solution obtenue, et donc, encore une fois, que la version repondérée, à l’inverse, est plus robuste. Il est aussi intéressant de noter qu’à une norme résiduelle donnée, la version standard n’obtient jamais une reconstruction de meilleure qualité que la version repondérée, ce qui conforte l’observation que cette régularisation semble mieux adaptée.

C.3 Expérimentations

C.3.1 Séparation aveugle d’images

Dans ce paragraphe, des expérimentations sont effectuées afin de comparer les régularisations en analyse et en synthèse μ , ainsi que leurs pendants repondérés, dans

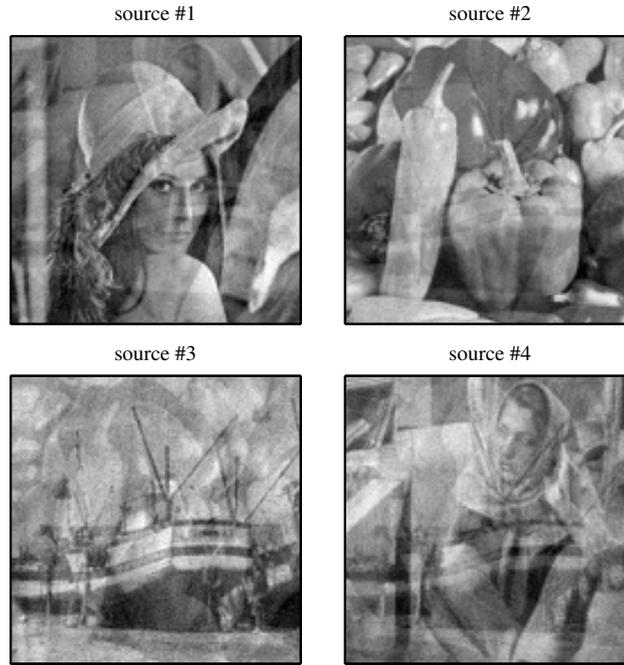


Figure V.15 – Exemple de reconstruction avec HALS (16 observations, $\text{SNR}_Y = 30\text{dB}$, obtenant un SDR_S de 14.79dB).

le contexte de l'imagerie. Dans ces conditions, il est courant de ne pas pénaliser l'échelle grossière des ondelettes, qui contient les valeurs moyennes locales du signal et n'est donc pas parcimonieuse. Ainsi, les coefficients de \mathbf{A} qui correspondent à cette échelle sont mis à 0. De la même manière, cette échelle ne contient généralement pas d'information discriminante. Il est donc usuel de ne pas en prendre compte lors de l'estimation de \mathbf{A} . Les ondelettes utilisées dans ces expérimentations sont des Daubechies-4 avec 3 échelles.

À part en ce qui concerne les sources, qui sont maintenant les 4 images de la figure V.9, les simulations sont conduites de la même manière qu'à la section précédente. La figure V.14 montre le SDR moyen de chacune des reconstructions pour des mélanges bruités avec les formulations en analyse et en synthèse, repondérées ou non, ainsi qu'avec le nGMCA de base et HALS. Dans ces conditions, nGMCA de base donne de mauvais résultats puisque les sources recherchées ne sont clairement pas parcimonieuses dans l'espace direct. Il pourrait en effet être préférable de ne pas régulariser le problème plutôt que d'utiliser une régularisation mal adaptée. Pour cette raison, HALS est testé ici sans régularisation parcimonieuse. Il obtient de meilleurs résultats que nGMCA mais n'est pas capable de débruiter les sources ni de les séparer convenablement, comme le montre les interférences visibles sur la figure V.15.

Sur la figure V.16, la solution obtenue à l'aide de nGMCA-analyse sans la repondération est très peu bruitée mais présente de fortes interférences également,



Figure V.16 – Exemple de reconstruction avec du ℓ_1 standard (nGMCA-analyse, 16 observations, $\text{SNR}_Y = 30\text{dB}$, obtenant un SDR_S de 18.64dB).



Figure V.17 – Exemple de reconstruction avec du ℓ_1 repondéré (nGMCA-analyse repond., 16 observations, $\text{SNR}_Y = 30\text{dB}$, obtenant un SDR_S de 21.07dB).

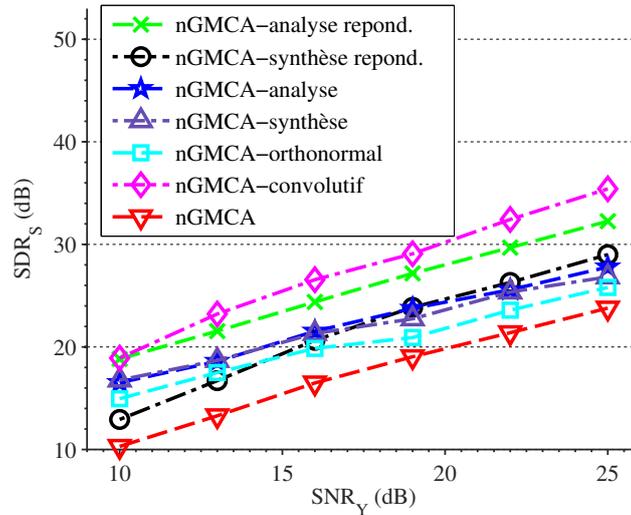


Figure V.18 – SDR moyen de la reconstruction (SDR_S) en fonction du niveau de bruit sur les données en dB (SNR_Y) ($r = 12$, $m = 32$, moyenné sur 36 simulations de données RMN)

ce qui est particulièrement visible sur la source #4. Enfin, la figure [V.17](#) montre une reconstruction effectuée avec le même algorithme mais cette fois en utilisant la repondération. Le procédé de repondération proposé permet d'obtenir des interférences bien plus réduites tant en conservant une faible contamination par le bruit. À des fins de cohérence, la même repondération est appliquée à la formulation en synthèse, alors qu'elle n'est pas aussi bien adaptée à ce cas. En effet, la formulation en synthèse est bien moins contrainte et pourrait ainsi bénéficier d'un choix de repondération plus dur.

Ces expériences corroborent nos observations précédentes sur des données bien plus complexes. La formulation en analyse se montre plus robuste que toutes les autres régularisations parcimonieuses dans le cas d'images naturelles qui ne sont pas parfaitement parcimonieuses en ondelettes.

C.3.2 Retour sur les résultats 1D

Le procédé de repondération proposé dans cette section peut également être utilisé sur les données RMN simulées introduites précédemment (section [B.1](#)). Les figures [V.18](#) et [V.19](#) sont ainsi des mises-à-jour des figures [V.3](#) et [V.7](#) sur lesquelles ont été ajoutées les courbes pour les formulations en analyse et en synthèse repondérées dans le domaine des ondelettes. Comme attendu, le procédé de repondération permet d'obtenir de claires améliorations, avec un gain de 4 à 5dB en analyse aussi bien sur la figure [V.18](#) que la figure [V.19](#). En ce qui concerne la formulation en synthèse, il apparaît encore une fois sur la figure [V.18](#) que la repondération permet d'obtenir une légère amélioration pour cette formulation quand il y a peu de bruit, mais peut gêner la reconstruction aux régimes les plus bruités. Dans ce cas, elle pourrait en

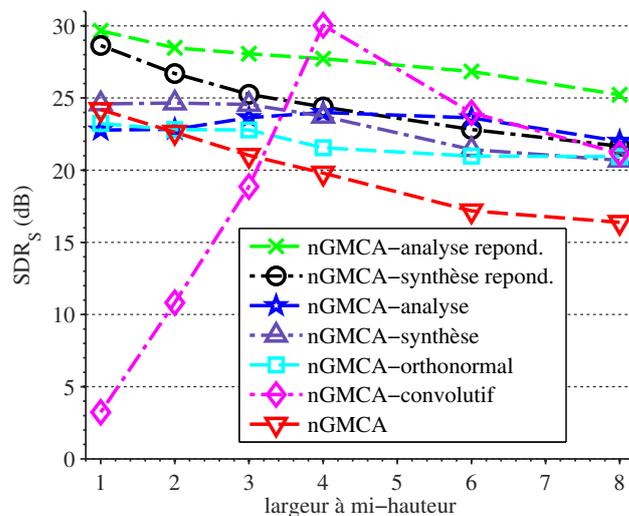


Figure V.19 – SDR moyen de la reconstruction (SDR_S) en fonction de la largeur des pics ($\text{SNR}_Y = 20\text{dB}$, $r = 12$, $m = 32$, moyenné sur 36 simulations de données RMN)

effet ne pas pénaliser suffisamment les petits coefficients en ondelettes qui ont de grandes chances d’être dus au bruit.

D Conclusion

Ce chapitre a étendu nGMCA aux cas de régularisations parcimonieuses. À notre connaissance un seul algorithme de NMF avait été conçu afin de gérer la parcimonie dans un espace transformé et cela sans résoudre parfaitement les sous-problèmes [Jiang & Yin 2012] (cf. section II-C.2, 40). Nous avons montré ici que résoudre ces sous-problèmes était possible à l’aide d’algorithmes proximaux. De nombreux cas ont pu alors être explorés, en utilisant une transformée orthonormale, ou une transformée redondante exprimée en analyse ou en synthèse. Nous avons aussi proposé de considérer une représentation utilisant des combinaisons positives d’atome positifs.

Comme attendu, l’usage de transformées bien adaptée aux signaux et en particulier de transformées redondantes s’est montré extrêmement efficace sur des signaux 1D. Le même procédé sur des images a montré l’apparition de fortes interférences, dues au biais de la régularisation parcimonieuses. L’usage d’une méthode de repondération a alors permis de largement diminuer ces interférences.

De la même manière que pour le chapitre précédent, le code des algorithmes et des scripts ayant servi à créer les figures est disponible en ligne.

Application aux données LC/MS

Sommaire

A	Présentation des données	101
A.1	Spectres de masse et profils d'élution	102
A.2	À propos du bruit et du modèle de mélange	103
A.3	Chaîne de traitement des données	105
A.4	Création de données synthétiques	106
B	Adaptations de nGMCA aux données LC/MS	108
B.1	Formulation avec du bruit non-stationnaire	108
B.2	Validation de l'approche non-stationnaire	110
B.3	Mises-à-jour stochastiques	112
C	Comparaison des algorithmes	113
C.1	Premiers résultats sur les mélanges synthétiques	113
C.2	Étude qualitative sur les données réelles - Cas de nGMCA	115
C.3	Étude quantitative sur les données réelles	119
D	Perspectives d'amélioration du modèle	121
D.1	Modèle de bruit	122
D.2	Modèle de mélange	122
D.3	Modélisation des spectres et des profils d'élution	123

La NMF est particulièrement bien adaptée à l'étude des données de chromatographie liquide/spectrométrie de masse (LC/MS). En effet, celles-ci sont composées de mélanges positifs de spectres de masse qui sont également positifs. La première section de ce chapitre décrit les différents aspects de ces données comme la structure des spectres de masse et des profils de mélange, ainsi que le bruit multiplicatif qui les contamine. Nous proposerons ensuite d'adapter nGMCA aux spécificités de ces données en utilisant un modèle de bruit non-stationnaire et une approche stochastique afin d'être moins sensible aux points stationnaires induits par la structure très piquée des données. Nous évaluerons enfin plusieurs approches de la NMF d'abord sur des données LC/MS synthétiques, puis sur un jeu de données réelles annotées afin d'en tirer des conclusions sur la manière la plus efficace d'en extraire les spectres de masse recherchés.

A Présentation des données

La LC/MS permet d'analyser un fluide afin de trouver les composés chimiques qui le constitue. Le processus d'acquisition est résumé à la figure [VI.1](#). Le fluide

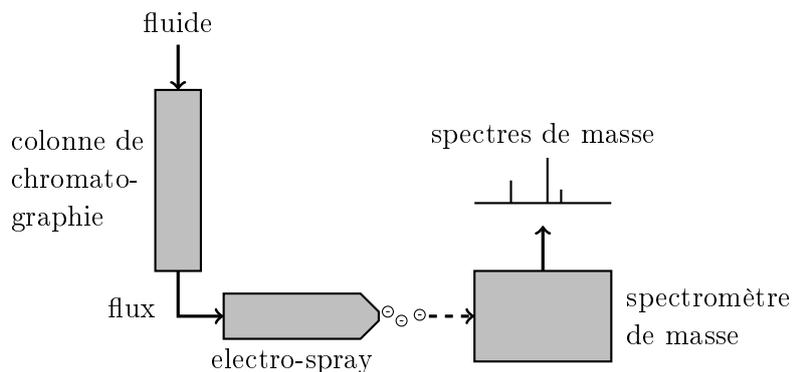


Figure VI.1 – Principe de la chromatographie en phase liquide-spectrométrie de masse.

Le fluide passe d'abord dans une colonne de chromatographie, à travers laquelle les composés circulent à des vitesses différentes. Ainsi, les composés du fluide sortent de la colonne de chromatographie à des temps différents. À chaque instant t , les composés sortants de la colonne sont réduits sous forme de spray et ionisés, puis analysés par un spectromètre de masse. Les composés ayant des spectres de masse différents, ce processus fournit ainsi une première double séparation imparfaite : en temps et en masse. On obtient alors des données 2D comme celles de la figure III.2 du chapitre III (page 27). Celles-ci seront notées \mathbf{Y} et chaque ligne de cette matrice est un spectre de masse à un temps donné.

A.1 Spectres de masse et profils d'élution

La figure VI.2a montre un spectre de masse typique à un instant t donné, c'est-à-dire $\mathbf{Y}_{t,\cdot}$. Toutes les observations comme celle-ci sont constituées d'un mélange de spectres élémentaires qui sont caractéristiques des composés du mélange. Les démêler nous permet ainsi d'identifier les composés. Les composés observés dans les spectres de masse sont des molécules. Celles-ci sont cassées afin de former des ions, appelés adduits. Chacun des pics de ces spectres élémentaires correspond au rapport de la masse sur la charge d'un adduit. L'unité utilisée pour le rapport masse-sur-charge est généralement le Thomson (Th), qui correspond à la masse exprimée en Dalton (Da), divisée par la charge exprimée en nombre de charges élémentaires. En pratique, les pics observés sur ces données proviennent d'ions ayant une charge de $-1e$, et ils correspondent donc directement à la masse des adduits exprimés en Dalton, 1Da valant approximativement la masse d'un nucléon. Les atomes de ces adduits pouvant avoir des isotopes, ceux-ci apparaissent également dans les spectres de masse. En particulier, un zoom sur le pic principal (figure VI.2b) permet d'observer un pic plus petit situé à +1Da qui est typique de la présence du carbone-13 ^{13}C (qui a un neutron supplémentaire par rapport au carbone-12 ^{12}C , le plus courant dans la nature, et est près de 100 fois moins abondant que ce dernier). Cette figure met aussi en évidence la grande gamme d'intensités à extraire

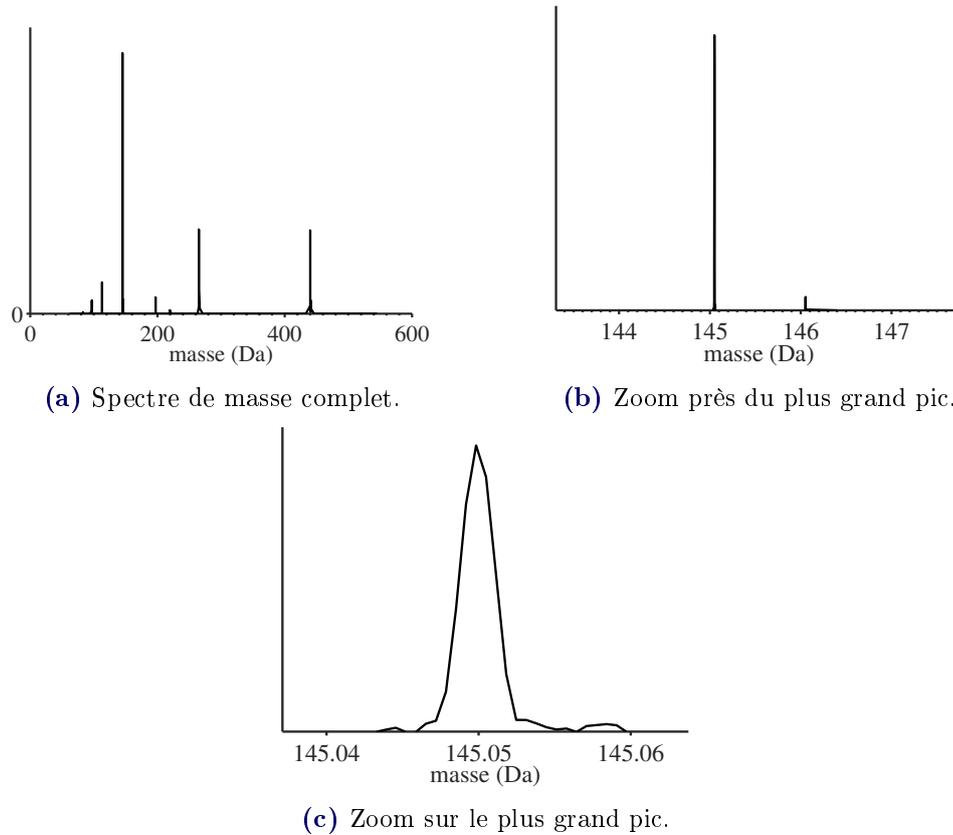


Figure VI.2 – Exemple de spectre de masse à un temps t .

des données. Zoomer encore au plus près du pic principal (figure VI.2c) montre la grande précision du spectromètre Orbitrap qui a été utilisé pour acquérir ces données [Makarov 2000, Makarov *et al.* 2006]. En effet, bien qu'à première vue le pic ne correspond qu'à un coefficient, il est en fait larges d'une dizaine d'échantillons.

La figure VI.3 montre quant à elle un exemple représentatif de profil temporel, c'est-à-dire d'intensité en fonction du temps, à une masse m donnée, $\mathbf{Y}_{.,m}$. Ces profils sont également appelés profils d'éluion. Ceux-ci ressemblent typiquement à des Gaussiennes d'une durée inférieure à une minute, avec une queue allongée après leur valeur maximale. Une caractéristique des données est clairement visible sur ce graphique : plus la valeur des coefficients est grande, plus ceux-ci sont bruités et présentent de grandes variations autour d'une valeur moyenne. Les données semblent ainsi contaminées par du bruit multiplicatif, c'est-à-dire que son écart-type est proportionnel à la valeur du coefficient du signal.

A.2 À propos du bruit et du modèle de mélange

Afin de corroborer l'observation précédente en ce qui concerne le bruit sur les données, on peut appliquer la procédure suivante :

- (i) extraire les profils temporels les plus énergétiques des données.

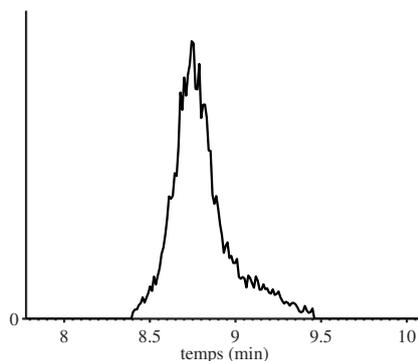


Figure VI.3 – Exemple de profil temporel dans les données LC/MS.

(ii) lisser ceux-ci à l'aide d'un filtre gaussien afin d'en éliminer une partie du bruit.

(iii) calculer une moyenne locale et estimer un écart-type local tout le long des profils sur les morceaux non-uniformément nuls, en comparant les profils bruités et filtrés.

Le nuage de point affiché à la figure [VI.4a](#) correspond alors aux écarts-types en fonction de l'amplitude moyenne des données ainsi créées. Celui-ci fait apparaître une corrélation entre l'amplitude et l'écart-type. De nombreuses études sont parvenues à des conclusions similaires [[Anderle et al. 2004](#), [Du et al. 2008](#)]. En considérant que le bruit est dû à un mélange d'une composante additive et d'une composante multiplicative, on obtient le modèle suivant pour l'écart-type de chaque coefficient des données :

$$\Sigma_{ij} = \sqrt{\sigma_{\text{add}}^2 + \sigma_{\text{mult}}^2 \mathbf{Y}_{ij}^2}, \quad (\text{VI.1})$$

où σ_{add} est l'écart-type du bruit additif, et σ_{mult} le coefficient de bruit multiplicatif. Le modèle est montré sur la figure précédente (ligne noire) les valeurs obtenues, c'est-à-dire $\sigma_{\text{mult}} = 0,074$ et $\sigma_{\text{add}} = 15\,000$. La figure [VI.4b](#) montre les mêmes données avec une échelle logarithmique. Le bruit multiplicatif y est davantage visible, avec une pente de coefficient directeur 1 pour les grandes valeurs. On remarque cependant sur ce graphique que le bruit additif est surévalué à cause des points aberrants à faible moyenne mais grand écart-type. Dans la suite, nous utiliserons tout de mêmes les valeurs surévaluées trouvées de cette manière, afin d'être légèrement moins sensible à ces points. Les biologistes utilisant ces données estiment de plus que les pics d'intensité plus faible que 10 000 en ordre de grandeur sont peu significatifs, ce qui vient conforter ce choix.

En plus des points aberrants et du bruit multiplicatif, un dernier aspect rend également ces données difficiles à traiter : les non-linéarités. En effet, le modèle de mélange n'est pas parfaitement vérifié sur ces données, ce qui peut entraîner des erreurs. On peut ainsi trouver dans les données des masses qui "dévient" comme sur la figure [VI.5](#) où deux composés semblent changer de masse au cours du temps (entre les échantillons 7600 et 7700). Ces non-linéarités peuvent être dues à des parasites

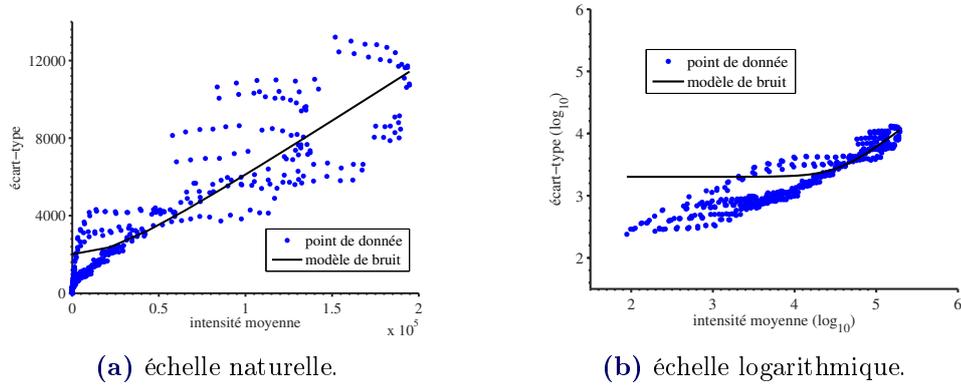


Figure VI.4 – Ecart-types en fonction des intensités, calculés à partir des plus profils temporels les plus énergétiques.

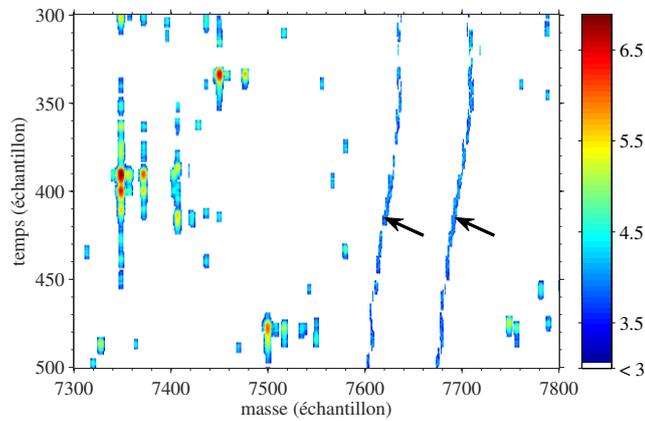


Figure VI.5 – Zoom sur une non-linéarité (masses "variables" indiquées par des flèches entre les échantillons 7600 et 7800, données non-annotées, intensités en échelle logarithmique).

électromagnétiques. Durant certaines acquisitions, certains composés chimiques réagissent également entre eux, ce qui n'est pas non plus modélisable à l'aide d'un mélange linéaire.

A.3 Chaîne de traitement des données

Les données que nous considérons dans ce chapitre proviennent d'un mélange de composés commerciaux dont les spectres de masse sont connus. Cet échantillon a été réalisé à des fins d'évaluation au sein du CEA-Saclay [Xu *et al.* 2010], et a été analysé à l'aide d'un spectromètre Orbitrap. Les données obtenues sont annotées, et la position en temps et en masse de chacun des pics des composés mélangés est ainsi connue. La plage de temps va de 2 à 18 minutes, et les masses de 69 à 544 Da. Deux simulations des spectres de masse des composés du mélange sont visibles à la figure III.4. Un exemple de spectre de masse observé est également donné figure

VI.11b et la liste complète des composés est affichée sur la figure **VI.6** avec une estimation de leur profil d'élution.

Le traitement des données LC/MS se fait généralement à l'aide de logiciels tels que XCMS [Benton *et al.* 2008] et MZmine [Pluskal *et al.* 2010], qui recherchent les pics 2D un par un avant de les regrouper. Comme indiqué dans [Dubroca *et al.* 2012], l'utilisation de la NMF en LC/MS permettrait de combiner ces deux étapes en une seule. Nous nous proposons donc ici de comparer les résultats de plusieurs approches en NMF sur ces données, avec la chaîne de traitement suivante :

1. Mise sous forme matricielle : chaque spectre acquis par le spectromètre de masse a une grille de masses spécifique et différente de celle des autres spectres. Ainsi, une grille de masse commune est créée sur laquelle placer l'ensemble données. Cela permet de les mettre sous forme matricielle.
2. Sélection de masses : les masses n'apparaissant qu'à un faible nombre de temps consécutifs ou ayant une intensité trop faible sont jugées non-significatives et sont généralement retirées des données. De la même manière, les masses présentes à chaque temps, comme celles correspondants aux diluants, ne sont pas jugés intéressantes et sont enlevées des données. On obtient alors des données \mathbf{Y} nettoyées.
3. Sous-échantillonnage : afin de réduire davantage la taille des données, \mathbf{Y} est sous-échantillonné le long de sa dimension masse, ce qui permet d'obtenir une matrice \mathbf{Y}_{sub} de plus faible dimension. Cela rend les calculs plus rapides au coût d'une moindre précision. Cette moindre précision peut cependant être un avantage puisqu'elle diminue le niveau de bruit des données. Ainsi, on trouve alors $\sigma_{\text{add}} = 12\,000$ et $\sigma_{\text{mult}} = 0,023$, ce qui est significativement plus bas que les niveaux estimés en section A.2 sur l'ensemble des données. De plus, la précision d'origine peut être partiellement retrouvée via une étape d'inversion présentée en dernière étape de cette chaîne de traitement.
4. Factorisation : les données sont décomposées sous la forme $\mathbf{Y}_{\text{sub}} \approx \mathbf{A}\mathbf{S}_{\text{sub}}$ avec \mathbf{A} et \mathbf{S} des matrices positives, en utilisant des algorithmes de NMF.
5. Inversion : on se ramène à la grille de masse d'origine en résolvant le problème inverse suivant :

$$\operatorname{argmin}_{\mathbf{S} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2, \quad (\text{VI.2})$$

où la matrice de mélange \mathbf{A} est celle trouvée à l'étape précédente, et où \mathbf{Y} est la matrice de données avec la grille de masse commune créée à l'étape 1 (c'est-à-dire les données non-sous-échantillonnées).

A.4 Création de données synthétiques

Afin d'obtenir une estimation des profils temporels qui devraient être observés pour les sources annotées, les annotations sont utilisées afin de créer une matrice de

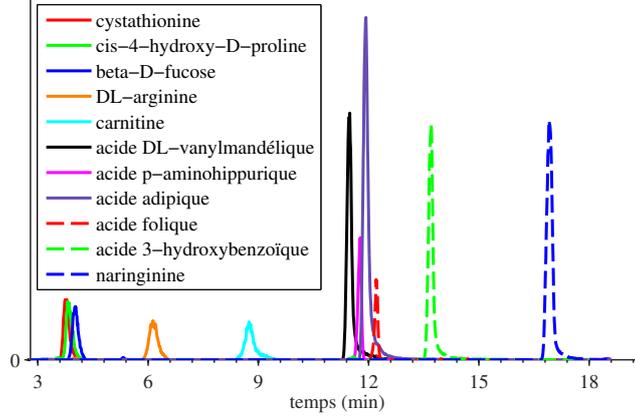


Figure VI.6 – Profils d'élution $\mathbf{A}^{\text{annot.}}$ des sources annotées, estimés par inversion (norme infinie des sources partagée entre $\mathbf{A}^{\text{annot.}}$ et $\mathbf{S}_{\text{sub}}^{\text{annot.}}$).

sources de référence $\mathbf{S}_{\text{sub}}^{\text{annot.}}$. On peut alors calculer les profils temporels correspondants en résolvant le problème inverse suivant :

$$\mathbf{A}^{\text{annot.}} = \underset{\mathbf{A} \geq \mathbf{0}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y}_{\text{sub}} - \mathbf{A} \mathbf{S}_{\text{sub}}^{\text{annot.}}\|_2^2. \quad (\text{VI.3})$$

Ces profils sont montrés sur la figure VI.6 en répartissant équitablement l'échelle sur $\mathbf{A}^{\text{annot.}}$ et sur $\mathbf{S}_{\text{sub}}^{\text{annot.}}$ afin que les plus petites sources soient suffisamment visibles par rapport aux sources les plus énergétiques. En effet, les dynamiques des sources sont très différentes les unes des autres.

À partir de là, il est possible de synthétiser une matrice de données de la manière suivante :

$$\mathbf{Y}_{\text{synth.}} = \mathbf{A}^{\text{annot.}} \mathbf{S}_{\text{sub}}^{\text{annot.}} + \mathbf{Z}, \quad (\text{VI.4})$$

où \mathbf{Z} modélise une contamination par un mélange de bruit additif non-stationnaire. Dans la suite, on considère que chaque coefficient \mathbf{Z}_{ij} est une variable aléatoire Gaussienne, indépendante des autres, de moyenne nulle et d'écart type $\Sigma_{ij} = \sqrt{\sigma_{\text{add}}^2 + \sigma_{\text{mult}}^2 \mathbf{Y}_{ij}^2}$. Ce bruit s'approche ainsi du bruit contaminant les données réelles, avec une composante additive et une composante multiplicative. Une différence subsiste cependant : on considère ici que tous les coefficients sont contaminés par du bruit alors que l'analyse du bruit sur les données n'avaient été effectuées que sur les sections non nulles des données. Le bruit des données est ainsi plus parcimonieux que celui utilisé dans ces simulations. Dans les expérimentations des sections qui suivent, σ_{add} est fixé à 12 000 et le paramètre contrôlant le bruit multiplicatif σ_{mult} varie entre 0 et 0,1 afin d'évaluer la sensibilité des algorithmes au bruit multiplicatif sur ces données synthétisées.

B Adaptations de nGMCA aux données LC/MS

Les données LC/MS sont différentes des données simulées des chapitres précédents. En effet, ils possèdent une très grande dynamique et des profils temporels et spectraux peu corrélés. Aussi, ces données sont contaminées par un bruit qui n'est pas simplement Gaussien. Cette section vise à adapter nGMCA à ces conditions.

B.1 Formulation avec du bruit non-stationnaire

Dans [Dubroca *et al.* 2012], Dubroca et al. adaptèrent l'algorithme multiplicatif de Lee & Seung ([Lee & Seung 1999], section III-B.2.2, page 36) afin de tenir compte du bruit multiplicatif présent dans les données LC/MS. Pour ce faire, ils modifièrent le terme d'attache aux données en le pondérant par l'écart-type du bruit à chaque coefficient, en considérant que chaque coefficient de bruit est indépendant des autres. L'écart-type est estimé au début de l'algorithme via l'équation (VI.1), directement à partir des observations et des coefficients σ_{add} et σ_{mult} que l'on suppose connus. Le même principe peut être utilisé avec nGMCA, qui, comme nous l'avons vu au chapitre IV est plus efficace que l'algorithme de mise-à-jour multiplicative. La fonction de coût devient donc :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \frac{1}{2} \|\boldsymbol{\Sigma}^{-1} \odot (\mathbf{Y} - \mathbf{AS})\|_2^2 + \|\mathbf{A} \odot \mathbf{S}\|_1, \quad (\text{VI.5})$$

où l'on considère les exposants (et donc ici l'exposant -1 sur $\boldsymbol{\Sigma}$) élément par élément. Ce type de repondération aurait été plus difficile à adapter avec des méthodes de type ALS [Paatero & Tapper 1994] ou HALS [Cichocki *et al.* 2008, Cichocki *et al.* 2009] car celles-ci utilisent des propriétés de séparabilité du problème de NMF qui ne sont pas conservées par le produit terme à terme avec $\boldsymbol{\Sigma}^{-1}$. La version de nGMCA conçue pour résoudre le problème (VI.5) sera nommée "nGMCA non-stationnaire" car elle vise à prendre un compte un bruit qui n'est pas identique sur tous les coefficients. Les paragraphes qui suivent fournissent les détails de sa mise en œuvre.

B.1.1 Gestion du seuillage

Une attention toute particulière doit être accordée à la gestion des seuils \mathbf{A} . En gardant l'approche utilisée au chapitre IV, les seuils sont fixés au niveau de l'écart-type du bruit estimé sur le gradient. Le gradient du terme d'attache aux données f s'écrit :

$$\nabla f(\mathbf{S}) = \mathbf{A}^T (\boldsymbol{\Sigma}^{-2} \odot (\mathbf{AS} - \mathbf{Y})). \quad (\text{VI.6})$$

À la fin de l'algorithme, on souhaiterait trouver un résidu $\mathbf{R} = \mathbf{AS} - \mathbf{Y}$ qui soit compatible statistiquement avec du bruit, et donc que sa matrice de variance élément par élément soit donnée par $\boldsymbol{\Sigma}^2$. On fait ainsi le choix simplificateur de négliger les covariances, étant donné qu'elles sont extrêmement difficiles à modéliser et à prendre

en compte. On peut ainsi en déduire une estimation de l'écart-type du bruit attendu sur le gradient, ce qui mène au choix suivant pour $\mathbf{\Lambda}$:

$$\mathbf{\Lambda} = \sqrt{(\mathbf{A}^T)^2 \mathbf{\Sigma}^{-2}}, \quad (\text{VI.7})$$

où la racine carrée est appliquée élément par élément, comme les exposants. Alors que $\mathbf{\Lambda}$ est décroissant dans la version originale de nGMCA, ce principe ne sera pas appliqué pour nGMCA non-stationnaire. En effet, l'ajustement du seuillage est déjà complexe et est modifiée à chaque nouvelle estimation de \mathbf{A} . Sans seuil décroissant, la contrainte de norme, envisagée pour la convergence en section IV-B.5 (page 58), est appliquée sur \mathbf{A} à chaque itération. Celle-ci n'assure cependant pas la convergence de l'algorithme ici, puisque le seuil $\mathbf{\Lambda}$ évolue à chaque itération. L'**Algorithme VI.1** résume ce nouvel algorithme (en prenant $\sigma_D^{(k)} = 0$, ce dernier terme étant expliqué par la suite).

B.1.2 Reconditionnement du gradient

En raison du terme $\mathbf{\Sigma}^{-2}$ dans le calcul de ∇f (VI.6), le gradient peut être très mal conditionné. En effet, dans le cas des données LC/MS, on a :

$$\frac{\max(\mathbf{\Sigma}^{-2})}{\min(\mathbf{\Sigma}^{-2})} \approx 10^3. \quad (\text{VI.8})$$

Pire, puisque les coefficients Σ_{ij} sont d'autant plus grands que les \mathbf{Y}_{ij} le sont, les pas de gradient dans ces directions sont plus petits, alors que ce sont les coefficients les plus importants à estimer.

Il est possible de reconditionner le gradient afin d'accélérer la convergence des algorithmes proximaux [Pock & Chambolle 2011, Becker & Fadili 2012]. Le cas spécifique du reconditionnement diagonal pour l'algorithme FB avec du seuillage doux est détaillé dans l'annexe B-D (page 135). Le reconditionnement peut ainsi être utilisé pour contrer l'effet néfaste de $\mathbf{\Sigma}$ sur le pas du gradient. En pratique, nous reconditionnons ainsi le gradient par un produit terme-à-terme avec $\mathbb{1} \odot \sqrt{(\mathbf{A}^T)^2 \mathbf{\Sigma}^{-2}}$. Ce choix permet de nettement accélérer la convergence de l'algorithme.

B.1.3 Formulation en analyse via synthèse contrainte

Afin de pouvoir utiliser le même type de terme non-stationnaire d'attache aux données avec un a priori parcimonieux en analyse, il faut pouvoir résoudre l'étape de mise-à-jour de \mathbf{S} :

$$\operatorname{argmin}_{\mathbf{S} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{\Sigma}^{-1} \odot (\mathbf{Y} - \mathbf{AS})\|_2^2 + \|\mathbf{\Lambda} \odot (\mathbf{SW}^T)\|_1. \quad (\text{VI.9})$$

L'approche de résolution de la formulation en analyse présentée précédemment utilise l'algorithme de Chambolle-Pock [Chambolle & Pock 2010]. Celui-ci n'est pas aussi aisé à utiliser avec le terme de bruit non-stationnaire. En effet, cet algorithme nécessite l'utilisation de l'opérateur proximal du terme d'attache aux données

$\mathbf{S} \mapsto \frac{1}{2} \|\boldsymbol{\Sigma}^{-1} \odot (\mathbf{Y} - \mathbf{A}\mathbf{S})\|_2^2$ qui n'est alors plus explicite à cause de la matrice de poids $\boldsymbol{\Sigma}^{-1}$ qui détruit la structure séparable du problème. Son calcul nécessiterait donc l'usage d'une sous-routine qui ralentirait d'autant l'ensemble de la minimisation. On pourrait aussi utiliser l'algorithme GFB ([Raguet *et al.* 2013], présenté section II-C.3.2, page 21) qui ne nécessite que le calcul du gradient du terme d'attache aux données, mais nécessiterait l'opérateur proximal de $\mathbf{S} \mapsto \|\boldsymbol{\Lambda} \odot (\mathbf{S}\mathbf{W}^T)\|_1$ (ce qui a été fait dans [Rapin *et al.* 2013b, Rapin *et al.* 2013c], sans $\boldsymbol{\Sigma}$). Celui-ci n'est cependant pas explicite non plus.

Dans l'annexe B-C (page 134), nous montrons que résoudre le problème (VI.9) revient à résoudre le problème suivant lorsque \mathbf{W} est une trame ajustée telle que $\mathbf{W}^T \mathbf{W} = \mathbf{I}_n$:

$$\operatorname{argmin}_{\mathbf{S}_w} \frac{1}{2} \|\boldsymbol{\Sigma}^{-1} \odot (\mathbf{Y} - \mathbf{A}\mathbf{S}_w \mathbf{W})\|_2^2 + \|\boldsymbol{\Lambda} \odot (\mathbf{S}_w)\|_1 + i_{\mathbf{w} \geq \mathbf{0}}(\mathbf{S}_w) + i_{\operatorname{Im}^T(\mathbf{W})}(\mathbf{S}_w), \quad (\text{VI.10})$$

avec la notation $\operatorname{Im}^T(\mathbf{W}) = \{\mathbf{S}_w, \exists \mathbf{S} \mid \mathbf{S}_w = \mathbf{S}\mathbf{W}^T\}$. Cette contrainte revient à se placer dans l'espace image de la transformée \mathbf{W} pour les signaux \mathbf{S} . On retrouve ensuite la solution dans l'espace direct par $\hat{\mathbf{S}} = \hat{\mathbf{S}}_w \mathbf{W}$. L'avantage de cette formulation est qu'elle peut être facilement résolue à l'aide de l'algorithme GFB. En effet, on prend alors :

$$f(S_w) = \frac{1}{2} \|\boldsymbol{\Sigma}^{-1} \odot (\mathbf{Y} - \mathbf{A}\mathbf{S}_w \mathbf{W})\|_2^2, \quad (\text{VI.11})$$

$$g_1(S_w) = \|S_w\|_1, \quad (\text{VI.12})$$

$$g_2(S_w) = i_{\mathbf{w} \geq \mathbf{0}}(\mathbf{S}_w) + i_{\operatorname{Im}^T(\mathbf{W})}(\mathbf{S}_w), \quad (\text{VI.12})$$

où l'on note $\operatorname{Im}^T(\mathbf{W}) = \{\mathbf{S}_w, \exists \mathbf{S} \mid \mathbf{S}_w = \mathbf{S}\mathbf{W}^T\}$. Le gradient de f est simple à calculer, l'opérateur proximal de g_1 est le seuillage doux, et celui de g_2 est $\mathbf{S} \mapsto [\mathbf{S}_w \mathbf{W}]_+ \mathbf{W}^T$ comme cela est montré dans l'annexe A-E (page 132).

Le choix du seuillage se fait de la même manière que dans le cas de la parcimonie dans l'espace direct, et l'on trouve alors $\boldsymbol{\Lambda} = \sqrt{(\mathbf{A}^T)^2 \boldsymbol{\Sigma}^{-2} (\mathbf{W}^T)^2}$. Le reconditionnement se généralise de façon directe dans ce cas là, étant donné que celui-ci n'affecte pas les projections. En pratique, nous utiliserons $\mathbf{1} \oslash \left((\mathbf{A}^T)^2 \boldsymbol{\Sigma}^{-2} (\mathbf{W}^T)^2 \right)$ qui se révèle très efficace pour la formulation en analyse.

B.2 Validation de l'approche non-stationnaire

La figure VI.7 résume les résultats pour les différentes versions de nGMCA envisagées sur les données synthétiques conçues en section A.4. Bien que les versions standards de nGMCA peuvent estimer le niveau de bruit au cours de l'algorithme, il sera ici fixé manuellement au niveau du bruit sur les données, c'est-à-dire $\sigma_{\text{add}} = 12\,000$. Les versions non-stationnaires prennent quant à elles à la fois σ_{add} et σ_{mult} utilisés en entrée.

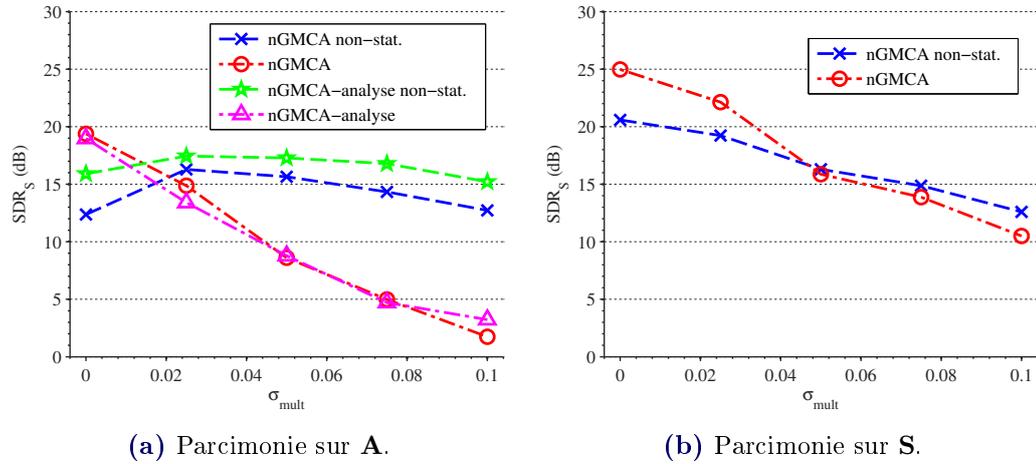


Figure VI.7 – SDR moyen de la reconstruction en fonction du niveau de bruit multiplicatif σ_{mult} (bruit additif à 40dB, données LC/MS synthétisées, moyenné sur 100 simulations de bruit).

Les résultats sont séparés en deux parties. En effet, la régularisation parcimonieuse peut être appliquées sur les spectres **S** ou sur les profils d'éluion **A** qui sont également parcimonieux. Même si les spectres sont bien entendu beaucoup plus parcimonieux que les profils d'éluion, savoir quel a priori est le plus adapté à la séparation de ces sources n'est pas nécessairement facile à première vue. La figure **VI.7a** montre les résultats lorsque la parcimonie est mise sur **A**, **VI.7b** lorsqu'elle est mise sur **S**. Dans ce dernier cas, les versions en analyse ne sont pas utilisées, étant donné que les spectres sont trop parcimonieux dans l'espace direct pour qu'elle soit véritablement utile, comme nous l'avons vu au chapitre précédent (cf. section V-B.2, page 85).

Ces figures permettent d'observer plusieurs comportements intéressants :

- Versions non-stationnaires : lorsque l'on considère la parcimonie des profils d'éluion, sur la figure **VI.7a**, les versions non-stationnaires se montrent bien plus robustes au bruit multiplicatif que les versions standards de nGMCA, avec un gain dépassant les 10dB lorsque σ_{mult} est élevé. Cette distinction n'existe cependant pas sur la figure **VI.7b**, c'est-à-dire dans le cas où la régularisation est mise sur **S**. Cela laisse penser que, dans ce dernier cas, les algorithmes sont limités par un autre problème, sur lequel nous reviendrons.
- Domaine de parcimonie : sur la figure **VI.7a**, la parcimonie en ondelettes des profils d'éluion apporte une amélioration de 1 à 3dB pour les versions non-stationnaires de nGMCA. Cette amélioration est cependant légère comparée au gain de la non-stationnarité.
- Parcimonie sur A ou sur S : en comparant les figures **VI.7a** et **VI.7b**, on peut s'apercevoir que la parcimonie sur **S** permet d'obtenir 2 à 5dB supplémentaires par rapport à la parcimonie sur **A** à faible bruit multiplicatif, et à peine

légèrement moins bonne dans des conditions de bruit plus élevé.

- Cas sans bruit multiplicatif : le premier point des deux graphiques, à $\sigma_{\text{mult}} = 0$, correspond au cas sans bruit multiplicatif. En ce point, on observe que les versions non-stationnaires sont à chaque fois défavorisées par rapport au versions standards, de 2 à 6dB. On notera pourtant que, dans ce cas là, le seuil final est identique pour les deux approches. Ainsi, la seule différence entre celles-ci vient du fait que le seuil est constant pour les versions non-stationnaires alors que les versions standards appliquent la stratégie de seuillage décroissant. Cette stratégie est donc extrêmement utile à la séparation sur ces données.

Au vu de ces résultats, les expérimentations suivantes ne porteront plus que sur l'impact de la parcimonie sur \mathbf{S} , qui peut semble pouvoir apporter davantage que la parcimonie sur \mathbf{A} , que ce soit dans l'espace direct ou en ondelettes. On a cependant mentionné que l'apport de la version non-stationnaire est inexistant dans ce cas là, ce qui suscite des interrogations au vu des résultats de la figure VI.7a, où les versions non-stationnaires sont bien plus robustes.

B.3 Mises-à-jour stochastiques

Le seuil décroissant était visiblement très bénéfique sur les données synthétiques. On peut donc supposer que les algorithmes peuvent facilement rester bloqués par des points stationnaires, et plus spécifiquement ce qui semble être des minima locaux. Ce fait peut provenir de la structure des données, ayant des spectres très piqués et peu de pics similaires, ainsi que des profils d'élution qui sont pour certains très éloignés des autres. Cela est amplifié par le bruit multiplicatif. Ainsi, l'estimation des sources peut se retrouver dans des cuvettes convexes desquelles il est impossible de sortir pour ce type d'algorithmes qui font nécessairement décroître la fonction de coût.

On se propose ici d'ajouter aux données un terme stochastique $\sigma_D^{(k)} \mathbf{D}^{(k)}$ à chaque itération k de l'algorithme global. L'**Algorithme VI.1** détaille les étapes de nGMCA non-stationnaire avec ce terme stochastique. La matrice $\mathbf{D}^{(k)}$ est choisie comme une matrice Gaussienne centrée réduite, et $\sigma_D^{(k)}$ est choisi au niveau du bruit additif σ_{add} au début de l'algorithme, puis décroissant linéairement jusqu'à 0 pour la phase d'affinement.

La figure VI.8 montre la différence obtenue avec ou sans le terme stochastique pour nGMCA standard et non-stationnaire avec une régularisation parcimonieuse sur \mathbf{S} . L'amélioration est flagrante pour la version non-stationnaire, avec un gain allant jusqu'à 10dB dans le cas le plus bruité. La version standard est également améliorée mais seulement d'environ 1dB supplémentaire. Cela corrobore ainsi deux observations précédentes : les versions non-stationnaires apportent une très nette amélioration, et la structure particulière des données LC/MS amène à se retrouver facilement limité par ce qui semble être des minima locaux. Bien que l'usage de ce terme stochastique ne soit pas optimal et nécessiterait davantage d'études, il semble permettre efficacement de sortir les algorithmes de ces cuvettes convexes.

Algorithme VI.1 : nGMCA non-stationnaire (stochastique)

Paramètre(s) d'entrée : \mathbf{Y} , K , σ_{add} , σ_{mult} .

- 1: initialiser $\mathbf{A}^{(0)}$ et $\mathbf{S}^{(0)}$, $\sigma_D^{(1)} = \sigma_{\text{add}}$, $\boldsymbol{\Sigma} = \sqrt{\sigma_{\text{add}}^2 + \sigma_{\text{mult}}^2} \mathbf{Y}^2$.
 - 2: **Pour** $k \leftarrow 1, K$ **faire**
 - 3: tirer $\mathbf{D}^{(k)}$ comme matrice aléatoire Gaussienne centrée réduite.
 - 4: $\boldsymbol{\Lambda}^{(k)} = \sqrt{(\mathbf{A}^{(k-1)T})^2 \boldsymbol{\Sigma}^{-2}}$
 - 5: $\mathbf{S}^{(k)} \leftarrow \underset{\mathbf{S} \geq \mathbf{0}}{\text{argmin}} \frac{1}{2} \|\boldsymbol{\Sigma}^{-1} \odot (\mathbf{Y} + \sigma_D^{(k)} \mathbf{D}^{(k)} - \mathbf{A}^{(k-1)} \mathbf{S})\|_2^2 + \|\boldsymbol{\Lambda}^{(k)} \odot \mathbf{S}\|_1$
 - 6: $\mathbf{A}^{(k)} \leftarrow \underset{\mathbf{A} \geq \mathbf{0}, \forall i \|\mathbf{A}_{\cdot, i}\|_2 \leq 1}{\text{argmin}} \frac{1}{2} \|\boldsymbol{\Sigma}^{-1} \odot (\mathbf{Y} + \sigma_D^{(k)} \mathbf{D}^{(k)} - \mathbf{A} \mathbf{S}^{(k)})\|_2^2$
 - 7: choisir $\sigma_D^{(k+1)} \leq \sigma_D^{(k)}$
 - 8: renvoyer $\mathbf{A}^{(K)}$, $\mathbf{S}^{(K)}$
-

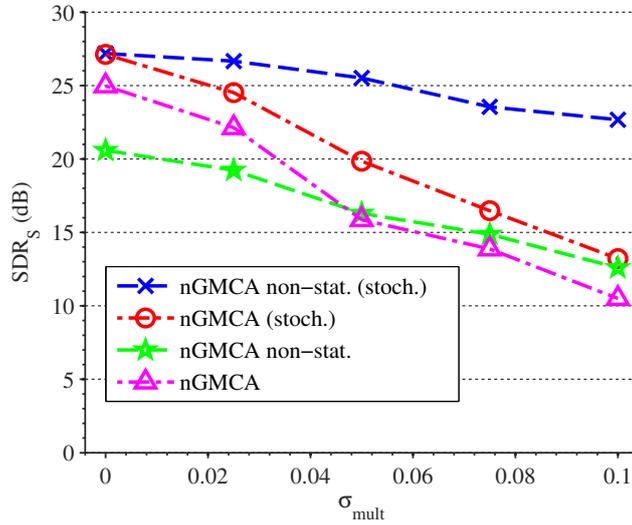


Figure VI.8 – SDR moyen de la reconstruction en fonction du niveau de bruit multiplicatif σ_{mult} (bruit additif à 40dB, données LC/MS synthétisées, moyenné sur 100 simulations de bruit, parcimonie sur \mathbf{S}).

C Comparaison des algorithmes

C.1 Premiers résultats sur les mélanges synthétiques

Afin d'effectuer des comparaisons plus complètes, on teste également sur ces données synthétiques l'algorithme multiplicatif de Lee & Seung [Lee & Seung 2001] avec divergence ℓ_2 ("mult. (ℓ_2)") et de Kullback-Leibler ("mult. (KL)"), l'algorithme multiplicatif de Dubroca et al. adapté au bruit multiplicatif des données LC/MS ("mult. (non-stat.)"), HALS, l'algorithme de Hoyer et l'algorithme de Kim & Park. Le niveau de parcimonie nécessaire à HALS et le ratio de parcimonie nécessaire à l'algorithme de Hoyer sont calculés sur les observations \mathbf{Y} . En effet, les sources étant

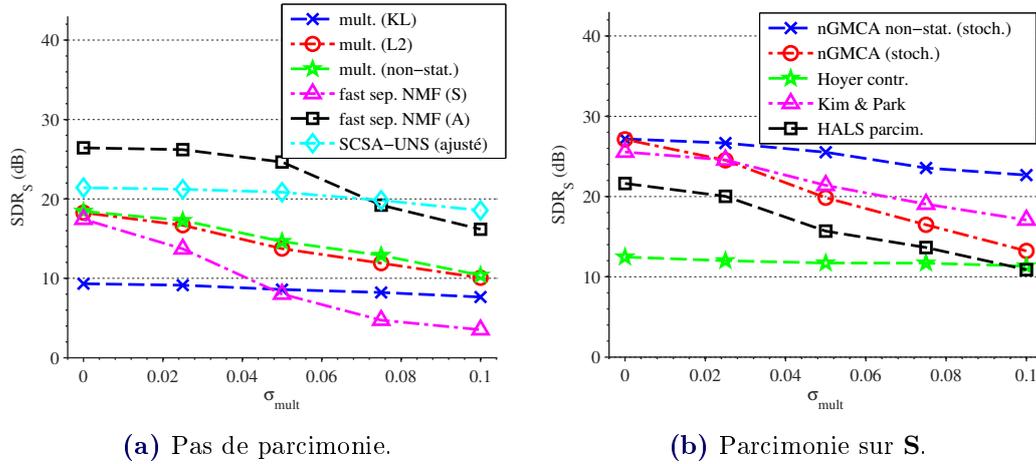


Figure VI.9 – SDR moyen de la reconstruction en fonction du niveau de bruit multiplicatif σ_{mult} (bruit additif à 40dB, données LC/MS synthétisées, moyenné sur 100 simulations de bruit).

peu mélangées sur ces données, cette estimation devrait être proche des vraies valeurs, qui ne sont pas connues à l'avance en pratique. Les paramètres de l'algorithme de Kim & Park sont quant à eux ceux proposés par défaut.

Puisque les spectres sont très peu corrélés entre eux et les profils d'éluion se chevauchent peu, les comparaisons incluent également deux approches géométriques, celles-ci étant bien adaptées à ce type de situation (cf. section III-B, page 32). La première, appelée "Fast separable NMF" [Gillis & Vavasis 2014] et dont le code est disponible sur le site de l'auteur¹, vise à retrouver les lignes de \mathbf{S} parmi les lignes de \mathbf{Y} ou les colonnes de \mathbf{A} parmi les colonnes de \mathbf{Y} . Ces algorithmes sont notés respectivement "Fast sep. NMF (S)" et "Fast sep. NMF (A)" dans les expérimentations qui suivent, et sont dits déterministes car ils ne dépendent pas de l'initialisation. La seconde, appelée SCSA-UNS [Ouedraogo *et al.* 2012], vise à retrouver le cône de volume minimal contenant les données. Ce dernier algorithme nécessite un paramètre visant à gérer le bruit, qui est ici ajusté manuellement sur les données simulées afin d'obtenir le meilleur SDR possible à $\sigma_{\text{mult}} = 0,05$.

Les résultats sont partagés entre deux figures afin de simplifier leur visualisation. Dans l'ensemble, tous les algorithmes sont fortement sensibles au bruit multiplicatif. La figure VI.9a rassemble les résultats des algorithmes n'utilisant pas de parcimonie. On observe que Fast sep. NMF (A) obtient d'excellents résultats, suivi par SCSA-UNS, qui le dépasse même à haut niveau de bruit. À l'inverse, Fast sep. NMF (S) obtient les moins bons résultats à haut niveau de bruit multiplicatif parmi tous les algorithmes testés, près de 10dB en dessous de Fast sep. NMF (A). Ceci est un des inconvénients de ces techniques : il est difficile de prévoir si elles obtiendront de bons résultats ou non. Ces algorithmes sont cependant extrêmement rapides

1. <https://sites.google.com/site/nicolasgillis/code>

et peuvent être une source d'information à confronter avec d'autres approches. On remarquera finalement sur ce graphique les résultats globalement médiocres obtenus par les algorithmes multiplicatifs, avec cependant un léger avantage pour la version non-stationnaire.

En ce qui concerne les algorithmes utilisant la parcimonie sur \mathbf{S} de la figure **VI.9b**, la version non-stationnaire et stochastique de nGMCA obtient les meilleurs résultats. Cet algorithme obtient également de meilleurs résultats que les algorithmes de la figure précédente. L'algorithme de Kim & Park est globalement 5dB en deçà de cette version de nGMCA, mais près de 5dB au-dessus de tous les autres algorithmes lorsqu'il y a du bruit multiplicatif. On se rappellera au passage que les paramètres standards de l'algorithme ont été utilisés et qu'ils ne sont donc pas nécessairement optimaux. Cela met en lumière le fait que comme le modèle de bruit, une régularisation différente, et ainsi un modèle de sources mieux adapté, peut être bénéfique. En effet, la régularisation utilisée dans cet algorithme favorise les sources ne partageant pas de coefficients en commun. Ce modèle est bien vérifié en pratique sur ces données.

C.2 Étude qualitative sur les données réelles - Cas de nGMCA

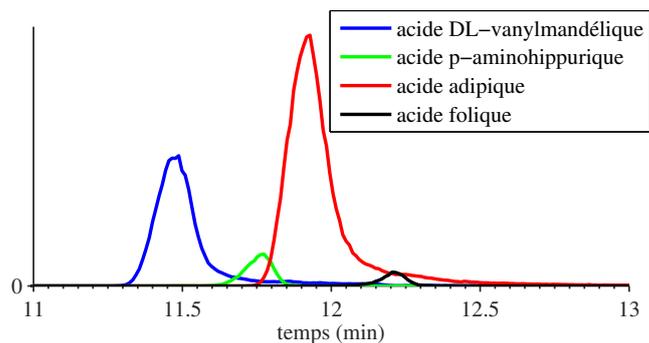
Les observations précédentes ont été effectuées sur des données simulées proches des données réelles. L'évaluation directe des résultats sur les données réelles est cependant plus difficile, bien que celles-ci soient annotées. En effet, les annotations ne fournissent que la localisation des principaux pics de chaque source, mais pas exactement les sources à retrouver. Dans cette section, nous développons une méthodologie permettant de qualifier les résultats obtenus par les algorithmes de NMF.

Pour le traitement des données réelles, le réglage des seuils finaux des algorithmes nGMCA s'effectuera manuellement avec $\sigma_{\text{add}} = 12\,000$, et $\sigma_{\text{mult}} = 0,05$ pour les versions non-stationnaire, soit légèrement au dessus du niveau estimé. Les paramètres des autres algorithmes sont choisis comme précédemment. Pour tous les algorithmes on considérera la recherche de 18 sources, parmi lesquelles on souhaite retrouver les 11 sources annotées.

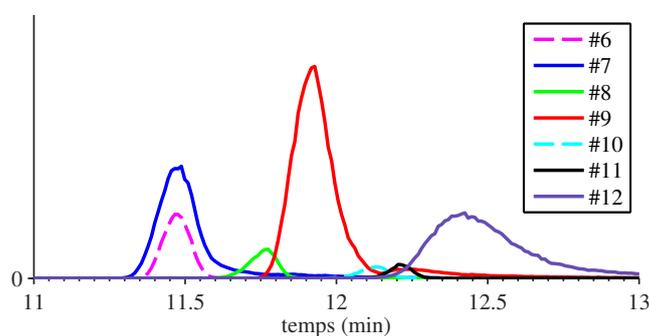
Les profils temporels $\mathbf{A}^{\text{annot.}}$ de la figure **VI.10a** peuvent être comparés à ceux retrouvés à l'aide de nGMCA et de nGMCA non-stat. (stoch.) sur les données réelles (respectivement figure **VI.10b** et figure **VI.10c**) sur la plage temporelle allant de 10 à 13 minutes, où plusieurs sources se superposent. On peut alors faire les observations suivantes :

(i) la base de données des pics ne contient pas tous les composés présents dans le mélange : nGMCA et nGMCA non-stat. (stoch.) trouvent tous les deux une source très prédominante (ligne violette à 12,4min, source #12 et #14 respectivement) qui n'est pas annotée mais dont le spectre de masse comporte une structure typique de la présence de certains adduits.

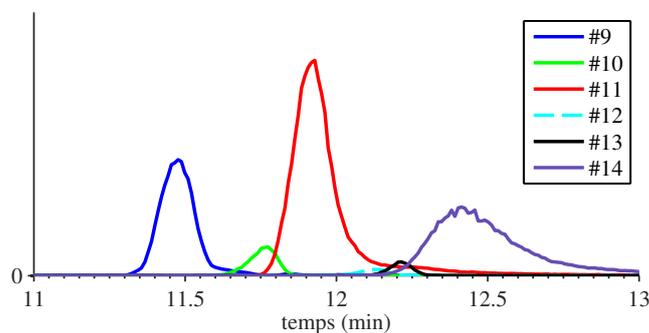
(ii) la source #6 obtenue par nGMCA (ligne pointillée magenta à 11,45min) interfère avec l'acide DL-vanylmandélique (source #7, ligne bleue à 11,45min), et la



(a) inversion des spectres annotés.



(b) NMF avec nGMCA.

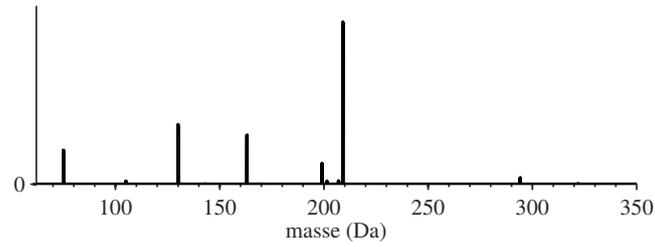


(c) NMF avec nGMCA non-stat. (stoch.).

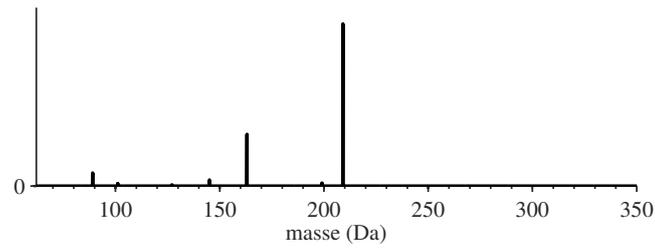
Figure VI.10 – Zoom sur les profils temporels entre 11 et 13 minutes (données réelles).

source #10 (ligne pointillée cyan à 12,1min) avec l'acide adipique (source #9, ligne rouge à 11,9min). Ces interférences sont inexistantes avec nGMCA non-stat. (stoch.) dans le cas de l'acide DL-vanylmandélique (source #9), et plus faible dans le cas de l'acide adipique, ce qui tend à montrer la meilleure qualité des reconstructions avec cet algorithme.

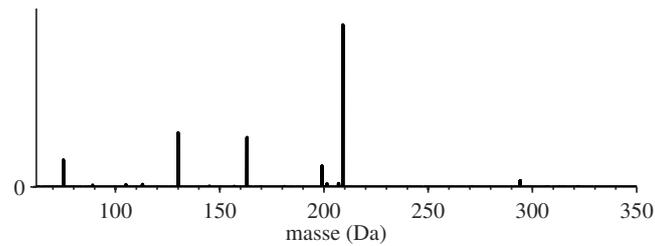
Un exemple de spectre retrouvé par nGMCA est visible sur la figure VI.11a, correspondant au beta-D-fucose. La figure VI.11b montre le spectre tel qu'il est indiqué dans les annotations une fois mis sur la grille de masse des données. On



(a) source obtenue à l'aide de nGMCA (#3).



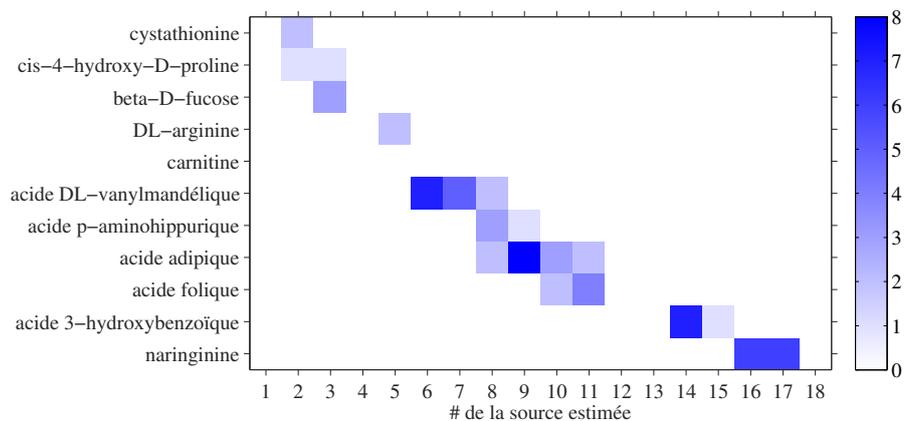
(b) spectre annoté.

(c) réestimation du spectre à partir de $\mathbf{A}^{\text{annot.}}$.**Figure VI.11** – Spectre du beta-D-fucose (données réelles).

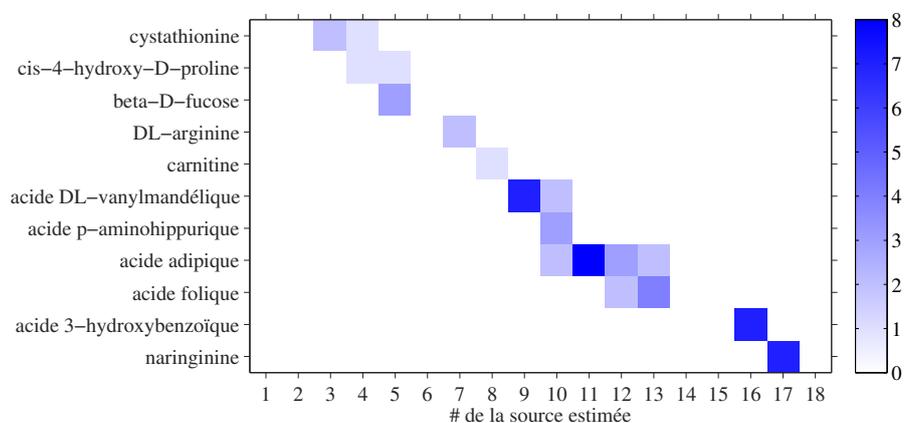
remarque en particulier deux pics proéminents à environ 163 et 209 Da qui sont identiques sur les deux spectres. Une grande partie des pics est cependant différente. La figure [VI.11c](#) montre la source estimée en inversant $\mathbf{A}^{\text{annot.}}$ via l'équation (VI.2). Cela revient à réestimer les sources à partir des profils d'éluion obtenus en inversant les annotations. Le spectre obtenu est alors très similaire à celui obtenu par nGMCA, ce qui montre donc qu'en pratique les données sont plus complexes que ce qui est indiqué dans les annotations.

Les figures [VI.12a](#) et [VI.12b](#) montrent le nombre de pics estimés avec respectivement nGMCA et nGMCA non-stat. (stoch.) qui sont compatibles avec les annotations à la fois en temps et en masse. Il est difficile d'évaluer quantitativement ces résultats sur les données réelles. On peut cependant faire un certain nombre de remarques sur la qualité de l'identification des sources recherchées :

(i) toutes les sources annotées ne sont pas toujours nécessairement retrouvées. Sur la figure correspondant à nGMCA, la ligne correspondant à la carnitine est vide et ainsi aucun pic n'a été identifié pour cette source. nGMCA non-stat. (stoch.) récupère quant à lui toutes les sources sur cet exemple.



(a) avec nGMCA.



(b) avec nGMCA non-stat. (stoch.).

Figure VI.12 – Nombre de pics des composants recherchés retrouvés (données réelles).

(ii) certaines sources estimées contiennent des pics correspondants à plusieurs composés annotés (colonnes ayant plus d'un unique coefficient non nul). Ces pics sont des interférences qui doivent être évitées. Au vu des figures [VI.12a](#) et [VI.12b](#), les interférences semblent ainsi moins courantes avec nGMCA non-stat. (stoch.) qu'avec la version standard de nGMCA.

(iii) certains composés annotés sont retrouvés dans plusieurs sources estimées (lignes comportant plus d'un unique coefficient non nul). Cela signifie que le composé recherché a été scindé en plusieurs sources. C'est par exemple ce que l'on a observé sur la reconstruction de nGMCA à la figure [VI.10b](#) avec les sources #6 et #7, et les sources #9 et #10. Ces composés scindés dans plusieurs sources semblent encore une fois moins fréquentes avec nGMCA non-stat. (stoch.).

C.3 Étude quantitative sur les données réelles

En considérant l'exemple qualitatif précédent on peut concevoir des critères qualitatifs. Pour ce faire on extrait les 200 plus grands pics des sources retrouvées. Ce choix d'un nombre prédéfini de pics extraits permet de ne pas privilégier l'usage de seuillages très sélectifs qui pourraient éviter artificiellement des interférences. Comme précédemment, ces pics sont affectés à un composé lorsqu'ils sont compatibles en temps et en masse avec les annotations. Étant donné que l'on estime 18 sources afin de retrouver les 11 composés annotés, on définit la matrice $\mathbf{E} \in \mathbb{R}^{11 \times 18}$ dont l'élément \mathbf{E}_{ij} est la somme des énergies (au sens du carré de l'amplitude) des pics de la $j^{\text{ième}}$ source estimée qui sont affectés au $i^{\text{ième}}$ composé. On définit également le vecteur $e \in \mathbb{R}^{18}$ où l'élément e_j est la somme des énergies des pics de la $j^{\text{ième}}$ source estimée qui ne sont affectés à aucun composé. On affecte ensuite chaque composé à une unique source estimée en choisissant une permutation des colonnes de \mathbf{E} maximisant $\sum_{i=1}^{11} \mathbf{E}_{ii}$. Ainsi le $i^{\text{ième}}$ composé est affecté à la $i^{\text{ième}}$ source estimée. On définit alors les critères quantitatifs suivants :

(i) ratio d'énergie correcte : le ratio moyen d'énergie d'une source estimée qui appartient au composé auquel la source est affectée. Ceci correspond ainsi à :

$$\text{corr.} = \frac{1}{11} \sum_{j=1}^{11} \frac{\mathbf{E}_{jj}}{e_j + \sum_{i=1}^{11} \mathbf{E}_{ij}}. \quad (\text{VI.13})$$

(ii) ratio d'énergie des interférences : le ratio moyen d'énergie des pics des sources estimées qui appartient à des composés annotés différents de celui auquel la source est affectée, rapporté à l'énergie totale de la source estimée (ratio d'énergie sur une colonne de la matrice \mathbf{E} , en prenant en compte les pics non-identifiés) :

$$\text{interf.} = \frac{1}{11} \sum_{j=1}^{11} \left(1 - \frac{\mathbf{E}_{jj} - e_j}{e_j + \sum_{i=1}^{11} \mathbf{E}_{ij}} \right). \quad (\text{VI.14})$$

(iii) ratio d'énergie des scissions : l'énergie moyenne des pics d'un composé qui n'est pas affecté à ce composé, rapporté à l'énergie totale des pics de ce composé (ratio d'énergie sur les lignes de la matrice d'identification de la figure VI.12a). Ce critère vise à évaluer la tendance des algorithmes à scinder les composés en plusieurs sources estimées, alors que chacun de ceux-ci devrait dans l'idéal être rassemblé dans une unique source. Celui-ci se calcule de la manière suivante :

$$\text{div.} = \frac{1}{11} \sum_{i=1}^{11} \left(1 - \frac{\mathbf{E}_{ii}}{\sum_{j=1}^{18} \mathbf{E}_{ij}} \right). \quad (\text{VI.15})$$

(iv) ratio de composés identifiés : le ratio du nombre de composés pour lesquels au moins un pic a été retrouvé, rapporté au nombre de composés annotés (c'est-à-dire à 11).

algorithm \ critère	corr. (%)	interf. (%)	sciss. (%)	iden. (%)	récup. (%)
Fast Sep. NMF (A)	86,9 (N/D)	9,2 (N/D)	4,9 (N/D)	90,9 (N/D)	N/D
Fast Sep. NMF (S)	86,7 (N/D)	4,6 (N/D)	5,8 (N/D)	100 (N/D)	N/D
SCSA-UNS (ajusté)	53,3 (+/-5,7)	10,9 (+/-2,6)	20,7 (+/-2,3)	99,5 (+/-2,2)	95,5
mult. (non-stat.)	74,6 (+/-6,0)	13,9 (+/-7,0)	14,7 (+/-5,3)	94,4 (+/-5,4)	44
mult. (KL)	69,2 (+/-5,6)	14,6 (+/-5,9)	15,2 (+/-4,8)	94,0 (+/-5,9)	43,5
mult. (L2)	74,7 (+/-5,6)	13,9 (+/-6,8)	13,9 (+/-4,8)	93,6 (+/-5,5)	37,5
Hoyer contr.	66,4 (+/-7,6)	16,9 (+/-7,4)	16,9 (+/-6,7)	90,0 (+/-7,5)	21
Kim & Park	78,1 (+/-5,1)	9,6 (+/-7,4)	10,8 (+/-3,8)	95,9 (+/-4,5)	54,6
HALS parcim.	75,3 (+/-4,1)	12,4 (+/-6,2)	13,4 (+/-3,7)	93,5 (+/-4,3)	29,5
nGMCA	77,2 (+/-4,9)	11,2 (+/-5,2)	11,0 (+/-3,5)	93,8 (+/-5,0)	36
nGMCA (stoch.)	79,4 (+/-2,8)	5,7 (+/-3,7)	10,6 (+/-1,8)	98,8 (+/-3,1)	87
nGMCA non-stat.	77,0 (+/-4,2)	9,9 (+/-5,7)	13,3 (+/-4,3)	96,9 (+/-4,6)	67,5
nGMCA non-stat. (stoch.)	80,3 (+/-3,4)	4,5 (+/-2,4)	10,3 (+/-3,1)	99,7 (+/-1,6)	97

Tableau VI.1 – Résultats moyens (et écart-types) obtenus à partir de 200 différentes initialisations.

(v) récupération : on considérera que toute les sources ont été retrouvées si au moins un pic de chaque composé annoté est identifié. Cela ne signifie pas que la factorisation est parfaite, mais en est cependant une condition nécessaire.

Tous les critères sont moyennés sur un grand nombre de réalisations, correspondant à des initialisations différentes des algorithmes. Les résultats sont rassemblés dans le tableau **VI.1** avec les valeurs moyennes pour chacun des critères et leur écart-type entre parenthèse. Ces tableaux apportent de nombreuses informations qui diffèrent sur certains points des expérimentations sur les données synthétiques, mettant en lumière la complexité des données réelles :

(i) résultats des algorithmes déterministes : Fast Sep. NMF (**S**) et Fast Sep. NMF (**A**) obtiennent peu d'interférences et de scissions. La version recherchant les lignes de **S** obtient d'excellents résultats puisqu'elle retrouve toutes les sources, alors que ce n'est pas le cas de la version recherchant les lignes de **A**. On remarquera que les résultats sont ainsi inversés par rapport aux données synthétiques. Ces algorithmes sont ainsi très difficiles à manipuler puisqu'ils est difficile de prévoir dans quelles conditions ils réussiront à retrouver les sources ou non.

(ii) sensibilité à l'initialisation : les algorithmes déterministes ne dépendant pas de l'initialisation, la variabilité de les résultats est bien entendu nulle. C'est pour cela qu'elle est notée comme N/D (non-définie) dans le tableau. La variabilité des algorithmes non-déterministes en fonction de l'initialisation est quant à elle très élevée (dans le cas de mult. (non-stat.) par exemple : 74.6% d'énergie correcte avec un écart-type de 6.0%). L'initialisation est ainsi extrêmement importante. Ceci tend à confirmer encore une fois la présence de nombreux points critiques auxquels ces algorithmes sont très sensibles. Le terme stochastique de nGMCA non-stationnaire permet cependant de baisser cette variabilité de plus d'un point.

(iii) résultats des algorithmes non-déterministes : la plupart des critères montrent des résultats similaires à l'incertitude près. Le critère le plus discrimi-

nant est le critère de "récupération" qui va de 21% à 97%. Une grande partie des algorithmes obtiennent dans les 20 à 45% sur ce critère. L'algorithme de Kim & Park obtient cependant toujours des résultats dépassant la plupart des autres algorithmes avec un taux de récupération de 54,6%. Seuls SCSA-UNS (ajusté) ainsi que les versions non-stationnaires et/ou stochastiques de nGMCA parviennent à des taux de récupération plus élevés. SCSA-UNS (ajusté) obtient ainsi un taux de récupération de 95,5% mais obtient le plus haut taux scissions (20,6%) et le plus bas taux d'énergie correcte. La version non-stationnaire et stochastique de nGMCA obtient quant à elle un taux de récupération de 97%, avec un taux d'interférence parmi les plus faibles et un taux d'énergie correcte parmi les plus élevés.

(iv) attache aux données : l'a priori non-stationnaire est très efficace pour nGMCA et améliore également légèrement ceux de "mult. (non-stat.)" par rapport à "mult. (L2)". La divergence de Kullback-Leibler permet d'obtenir un taux de récupération comparable à celui de "mult. (non-stat.)" (43,5%) mais au prix d'une détérioration des autres critères, et en particulier du ratio d'énergie correcte qui baisse de près de 5 points par rapport aux autres algorithmes multiplicatif. L'utilisation d'un terme d'attache aux données particulier n'est ainsi pas neutre est mène à des résultats différents. Sur ces données, l'intégration complète de l'a priori de bruit non-stationnaire dans nGMCA, avec l'usage d'un terme de parcimonie adapté en conséquence, semble ainsi l'approche la plus appropriée.

On voit ainsi qu'en résumé, l'algorithme de Kim & Park est efficace et robuste, grâce à un terme de régularisation qui paraît particulièrement approprié à ces données. L'adaptation de nGMCA permet cependant une très nette amélioration à l'aide de la prise en compte du bruit multiplicatif et du terme stochastique le rendant moins sensible aux points critiques. Il est indéniable que les approches déterministes peuvent dans certains cas être encore meilleures, et ceci de manière extrêmement rapide. Les algorithmes géométriques et l'algorithme de Kim & Park profitent ici du fait que les données sont peu mélangées, avec des sources n'ayant pas beaucoup de pics en commun, et des profils d'élution se chevauchant peu. Ceux-ci ne seraient cependant pas aussi performants dans le cas contraire, car l'hypothèse des mélanges purs ne seraient alors plus vérifiée pour Fast Sep. NMF, et l'hypothèse de source dominante non plus pour l'algorithme de Kim & Park. Dans des jeux de données plus complexes, avec un plus grand nombre de composés chimique, cela est susceptible de poser problème. SCSA-UNS pourrait s'avérer plus robuste, ne nécessitant pas strictement l'hypothèse de mélanges purs, mais cet algorithme nécessite l'utilisation d'un paramètre qui n'est pas aussi direct à régler que celui de nGMCA. Les approches géométriques sont également moins souples en terme d'améliorations possibles.

D Perspectives d'amélioration du modèle

Ces expérimentations fournissent des résultats prometteurs pour l'usage de la NMF sur les données LC/MS. On peut cependant en extraire de nombreuses pos-

sibilités d'améliorations qui pourraient se révéler utiles afin d'obtenir une meilleure identification de spectres avec les algorithmes non-déterministes.

D.1 Modèle de bruit

Le bruit n'est pas encore complètement compris sur les données LC/MS. En plus du bruit multiplicatif sur les plus grands coefficients, des études ont ainsi suggéré de modéliser le bruit sur les plus petits coefficients par du bruit de Poisson, au lieu du bruit Gaussien considéré ici [Anderle *et al.* 2004]. Dans ce papier, nous avons comparé des algorithmes utilisant plusieurs types de terme d'attache aux données, et en particulier la divergence de Kullback-Leibler, et les distances euclidiennes standard et repondérées. La divergence KL semble être plus adéquate que la distance euclidienne standard, et l'usage de la norme euclidienne repondérée peut également nettement améliorer les résultats. Une divergence \mathcal{D} mieux adaptée au bruit pourrait vraisemblablement améliorer les résultats. Celle-ci pourrait cependant rendre beaucoup plus difficiles les mises-à-jours de \mathbf{A} et \mathbf{S} , en particulier par rapport à un terme d'attache aux données quadratique qui a l'avantage d'avoir un gradient Lipschitzien.

D.2 Modèle de mélange

Le modèle de mélange, qui stipule que chaque observation est une combinaison linéaire des sources, n'est pas parfait en pratique comme nous l'avons vu à la figure VI.5 par exemple. Encore une fois, si l'on a accès à un modèle plus précis, il pourrait être utile de modifier le terme d'attache aux données en conséquence mais un tel modèle peut se révéler également difficile à mettre en œuvre.

Ces problèmes de non-linéarités et écarts au modèle peuvent également être pris en compte via l'ajout d'un terme robuste, sur le modèle de [Candès *et al.* 2011]. En gardant une régularisation parcimonieuse sur \mathbf{S} , cela revient à considérer le problème suivant :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}, \mathbf{E}} \mathcal{D}(\mathbf{Y} || \mathbf{AS} + \mathbf{E}) + \|\mathbf{\Lambda}_S \odot \mathbf{S}\|_1 + \|\mathbf{\Lambda}_E \odot \mathbf{E}\|_1. \quad (\text{VI.16})$$

Le cadre utilisé dans nGMCA peut facilement s'adapter à l'ajout de la minimisation de \mathbf{E} , qui pourrait être effectuée à chaque itération après les étapes de mise-à-jour de \mathbf{A} et de \mathbf{S} . Cette variable capturerait les éléments énergétiques des données qui sont mal modélisés par une matrice de faible rang, ou par des interactions positives. Ainsi, ces éléments ne risqueraient plus de contaminer les sources \mathbf{S} et les mélanges \mathbf{A} . Ce modèle peut ainsi s'apparenter au modèle d'analyse en composantes morphologiques présenté en II-B.4 (page 16), avec un terme parcimonieux dans l'espace direct, \mathbf{E} , et une somme de termes de rang 1 et positifs, \mathbf{AS} .

D.3 Modélisation des spectres et des profils d'éluion

L'utilisation d'une régularisation sur \mathbf{S} a été bénéfique, comme l'atteste la figure [VI.9b](#). Parmi celles-ci, le terme $\sum_{t=1}^n \|\mathbf{S}_{\cdot,t}\|_1^2$ envisagé dans l'algorithme de Kim & Park s'est montré extrêmement efficace, autant sur les données synthétiques que sur les données réelles. Ceci peut s'expliquer par le fait que cette régularisation privilégie les spectres présentant peu de coefficients en commun, avec une source dominante à chaque échantillon de masse, ce qui est le cas sur ce jeu de données.

L'utilisation d'une régularisation sur \mathbf{A} a également permis d'obtenir de nettes améliorations. En effet, les profils d'éluion étant lisses, ils ont bénéficié de la parcimonie en ondelettes de nGMCA non-stationnaire en analyse comme le montre la figure [VI.7a](#), même si ce bénéfice est moindre qu'avec l'usage de la parcimonie sur \mathbf{S} . Il serait alors intéressant d'étudier l'usage simultané de la parcimonie sur \mathbf{S} et de la parcimonie en ondelettes sur \mathbf{A} , ce qui s'exprimerait comme suit :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \mathcal{D}(\mathbf{Y}|\mathbf{A}\mathbf{S}) + \|\mathbf{\Lambda}_S \odot \mathbf{S}\|_1 + \|\mathbf{\Lambda}_A \odot (\mathbf{W}\mathbf{A})\|_1, \quad (\text{VI.17})$$

Ce type de modèle avec une double parcimonie a déjà été utilisée dans [\[Moudeden & Bobin 2011\]](#) par exemple. Il présente cependant une nouvelle difficulté pour nGMCA : alors que le choix des seuils y était effectué au niveau du bruit, il faudra également tenir compte dans ce cas là de la dynamique des paramètres $\mathbf{\Lambda}_A$ et $\mathbf{\Lambda}_S$. En effet, ceux-ci peuvent influencer l'un sur l'autre, $\mathbf{\Lambda}_A$ étant calculé à partir du résidu et de \mathbf{S} , qui est influencé par $\mathbf{\Lambda}_S$ qui est dépendant de \mathbf{A} . Il faut ainsi éviter que l'énergie ne se concentre uniquement dans \mathbf{A} ou uniquement dans \mathbf{S} afin de ne pas obtenir des situations instables.

Conclusion

Principaux enseignements

La conception de l'algorithme nGMCA nous a permis d'aborder de nombreuses problématiques allant des formulations mathématiques (parcimonie en ondelettes, bruit non-stationnaire) aux problèmes algorithmiques (sensibilités aux points stationnaires, gestion des paramètres). Les points suivants ont ainsi été abondamment développés dans cette thèse :

Formulations de la parcimonie : Nous avons mis en place des algorithmes permettant de régulariser le problème de NMF avec des a priori parcimonieux dans l'espace direct (chapitre IV) et des espaces transformés (chapitre V), via l'utilisation d'algorithmes proximaux. La parcimonie dans un espace transformé n'a été que très rarement étudiée dans la littérature en NMF, étant plus difficile à gérer que la parcimonie dans l'espace direct. En effet, manipuler la contrainte de positivité dans l'espace direct et la parcimonie dans un autre espace nécessite l'usage d'algorithmes plus complexes. Lorsque la transformation est redondante, à la fois la formulation en synthèse (section V-A.1, page 80) et la formulation en analyse (section V-A.2, page 82) ont été envisagées et ont obtenu de très bons résultats. Avec la formulation en synthèse, le cas d'un modèle convolutif avec un filtre positif est également intéressant. La formulation en analyse semble cependant plus robuste à des écarts au modèle (section V-B.2, page 85) .

Gestion du bruit : Les paramètres de parcimonie sont difficiles à régler, des approches de type validation croisée n'étant pas toujours envisageables, par manque de temps ou d'une base de données permettant de tester les paramètres. Dans les algorithmes de type GMCA, le niveau de seuillage est choisi en fonction du niveau de bruit. Cette stratégie, aisée à mettre en place de façon automatique, est relativement efficace comme l'atteste par exemple les expérimentations de la section IV-C.3 (page 65). Plusieurs stratégies doivent cependant être considérées suivant si la minimisation des sous-problèmes se fait de manière itérative ou non, et si le seuillage utilisé est doux ou dur (voir IV-B, page 54). Alors que les chapitres IV et V ont uniquement considéré un bruit additif Gaussien uniforme, le traitement des données LC/MS au chapitre VI a nécessité l'usage d'un a priori de bruit non-stationnaire afin de pouvoir prendre en compte un bruit multiplicatif sur les observations, avec des résultats tout aussi encourageants (cf. section VI-B.2, page 110).

Points critiques : Dans GMCA, la stratégie de seuillage a également pour but de déstabiliser les points critiques. En effet, les problèmes de SSA étant non-convexes, les algorithmes peuvent converger vers des points critiques éloignés de la solution recherchée. Ainsi, le seuil est fixé à un niveau élevé au début des algorithmes et baisse ensuite petit à petit afin d'affiner la solution. Encore une fois, plusieurs cas doivent être envisagés suivant l'approche de minimisation utilisée (itérative ou non) et le type de seuillage (dur ou doux). Cette stratégie est efficace, comme ceci est particulièrement visible sur les données LC/MS (cf. section VI-B.2, page 110). Elle peut être cependant délicate à mettre en place suivant les algorithmes utilisés (voir sections IV-D.2 page 76 et VI-B.1 page 108). Ainsi une autre approche a été envisagée, consistant à rajouter un terme stochastique à la minimisation afin de déstabiliser les points critiques (section VI-B.3, page 112). Elle a alors montré son efficacité sur les données LC/MS.

Biais dû au seuillage : La question du biais provoqué par le seuillage doux est revenue de manière récurrente au cours de cette thèse. Son effet est ambigu. En effet, il mène d'un côté à des bénéfices en terme de séparation (voir section IV-C.2.3, page 62) et permet d'assurer dans certains cas la convergence de l'algorithme (section IV-B.5, page 58). D'un autre côté, il crée un biais pouvant être extrêmement néfaste. Celui-ci s'observe au travers des artefacts et interférences qu'il engendre sur les figures IV.7 (page 64) et V.10 (page 93) par exemple. L'équilibre entre ces deux aspects est périlleux. La technique de repondération- ℓ_1 que nous avons mise en application section V-C (page 90) a cependant montré son intérêt, permettant de séparer convenablement les sources recherchées tout en évitant le biais et ainsi les interférences qui peuvent découler du seuillage.

Étude des données LC/MS : L'étude de données réelles s'est montrée extrêmement instructive, ceci autant du point de vue méthodologie qu'algorithmique. Du point de vue méthodologique tout d'abord, l'évaluation des résultats sur des données réelles est beaucoup plus difficile. Nous avons eu l'opportunité d'avoir accès à des données LC/MS annotées, ce qui nous a permis de concevoir des critères ad hoc pour cette évaluation (section VI-C.3, page 119). Du point de vue algorithmique, le traitement de ces données a nécessité des adaptations particulières comme l'usage d'un bruit non-stationnaire ainsi que d'un terme permettant déstabiliser les nombreux points critiques qu'elles engendraient (section VI-B, page 108). La prise en compte efficace de ces contraintes via nGMCA a permis d'obtenir d'excellents résultats.

Publications et logiciels

Ces travaux ont mené aux publications suivantes dans des journaux et conférences à comité de lecture :

- Jérémy Rapin, Jérôme Bobin, Anthony Larue, Jean-Luc Starck. *Sparse Regularizations and Non-negativity in BSS*. SPARS, Lausanne, Suisse, 8-11 Juillet

2013.

- Jérémy Rapin, Jérôme Bobin, Anthony Larue, Jean-Luc Starck. *Sparse Redundant Formulations and Non-negativity in Blind Source Separation*. EUSIPCO, Marrakech, Maroc, 9-13 Septembre 2013.
- Jérémy Rapin, Jérôme Bobin, Anthony Larue, Jean-Luc Starck. *Sparse and Non-negative BSS for Noisy Data*. IEEE Transactions on Signal Processing, vol. 61, issue 22, p. 5620-5632, 2013.
- Jérémy Rapin, Jérôme Bobin, Anthony Larue, Jean-Luc Starck. *NMF with Sparse Regularizations in Transformed Domains*, SIAM Journal on Imaging Sciences (accepté), 2014.

Des travaux annexes ont également donné lieu à une publication dans la conférence ICIP [Bobin *et al.* 2014] (voir annexe C, page 137).

En plus de ces publications et afin de se conformer à la philosophie de la recherche reproductible [Buckheit & Donoho 1995], les algorithmes présentés dans les articles [Rapin *et al.* 2013a] et [Rapin *et al.* 2014] sont disponibles en Matlab et en Python sur le site web de l'équipe CosmoStat du Service d'Astrophysique du CEA¹.

Perspectives

Après ces travaux, de nombreuses pistes restent à explorer. La dernière section du chapitre précédent (section VI-D, page 121) en fournit un certain nombre qui pourraient permettre d'améliorer le traitement des données LC/MS. Les trois approches qui suivent nous apparaissent comme étant les plus prometteuses :

Approches robustes : Le modèle de mélange linéaire envisagé en SSA est rarement parfaitement vérifié. L'exemple des données LC/MS est ici flagrant, celles-ci présentant des non-linéarités visibles à la figure VI.5 (page 105). On pourrait aussi citer les interactions chimiques entre composés en LC/MS et les changements d'illumination dus aux reliefs sur les données d'imagerie hyperspectrale aérienne. Quand ces non-linéarités apparaissent de manière parcimonieuse dans les données, elles peuvent être modélisées à l'aide d'un terme robuste. Ceci prendrait typiquement la forme suivante :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}, \mathbf{E}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AS} - \mathbf{E}\|_2^2 + \|\mathbf{\Lambda}_S \odot (\mathbf{WS})\|_1 + \|\mathbf{\Lambda}_E \odot \mathbf{E}\|_1, \quad (\text{VII.1})$$

Le terme robuste \mathbf{E} a ainsi pour but de récupérer ces non-linéarités qui sont mal modélisées par le modèle linéaires, et ainsi éviter qu'elles ne contaminent les sources \mathbf{S} .

1. <http://www.cosmostat.org/GMCALab>

Double parcimonie : Beaucoup de données multispectrales présentent un certain type de parcimonie à la fois sur leur dimension spectrale et sur les profils de mélange. C'est ainsi le cas des données LC/MS comme nous l'avons vu au chapitre précédent, les spectres étant très parcimonieux dans l'espace direct, et les profils de mélange étant lisses, ils sont parcimonieux dans l'espace des ondelettes par exemple. De même, les données Mars Express de la figure III.1 (page 26) présentent le même type de propriété, étant donné qu'à la fois les spectres infrarouges et les cartes de concentration pourraient être modélisés comme parcimonieux en ondelettes, respectivement 1D et 2D, comme cela est visible sur la figure III.3 (page 28). Ainsi, utiliser la parcimonie simultanément sur \mathbf{A} et \mathbf{S} pourrait améliorer l'identification des sources. Cela s'exprimerait par exemple de la manière suivante :

$$\operatorname{argmin}_{\mathbf{A} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2 + \|\mathbf{\Lambda}_S \odot (\mathbf{W}_S \mathbf{S})\|_1 + \|\mathbf{\Lambda}_A \odot (\mathbf{W}_A \mathbf{A})\|_1, \quad (\text{VII.2})$$

où \mathbf{W}_A et \mathbf{W}_S sont des transformées rendant respectivement parcimonieux \mathbf{A} et \mathbf{S} . La difficulté de ce type de formulation vient de la manière de gérer les paramètres $\mathbf{\Lambda}_A$ et $\mathbf{\Lambda}_S$. En effet, suivre la méthode de régularisation en fonction du bruit utilisée par nGMCA pourrait alors mener à des instabilités, le paramètre $\mathbf{\Lambda}_S$ étant choisi en considérant \mathbf{A} , qui dépend du choix de $\mathbf{\Lambda}_A$ qui lui-même dépendrait de la valeur de \mathbf{S} . On pourrait ainsi se retrouver dans la situation dégénérée où \mathbf{A} tend vers l'infini et \mathbf{S} vers $\mathbf{0}$.

Considérations algorithmiques : Les deux points précédents correspondent à des problématiques de modélisation. Les considérations algorithmiques ont cependant un intérêt extrêmement important sous plusieurs aspects. Ainsi, la sensibilité aux minima locaux est un problème qui est peu étudié en NMF. Le mécanisme le plus utilisé, et utilisé en particulier dans nGMCA, consiste à faire décroître le paramètre de régularisation au fur et à mesure des itérations. D'autres mécanismes peuvent cependant être mis en place, comme le terme stochastique envisagé en section VI-B.3 (page 112). Au vu des résultats du chapitre VI, ces mécanismes peuvent avoir une très grande influence, et mériteraient d'être davantage étudiés. Les autres considérations algorithmiques dépendent des applications visées. Ainsi, les algorithmes utilisés pourraient être efficacement codés sur GPU afin d'en obtenir des versions plus rapides et peut-être de pouvoir envisager éventuellement des problématiques temps réel. Les applications en grande dimension requièrent également des approches spécifiques. Les algorithmes utilisant des méthodes de déflation comme HALS, estimant une source à la fois, sont alors particulièrement bien adaptés mais sont plus difficiles à mettre en place lorsque l'on utilise la parcimonie dans un espace transformé, ou des modèles de bruit non-stationnaire.

Opérateurs proximaux

Sommaire

A	Récapitulatif des opérateurs utilisés	129
B	Positivé et contrainte de norme	129
C	Positivité et trames ajustées	130
D	Parcimonie dans un espace transformé	131
E	Contrainte de sous-espace et de positivité	132

Cette annexe rassemble les informations sur les opérateurs proximaux pouvant être utiles pour la lecture de ce manuscrit. La première section rappelle la définition et contient un tableau récapitulatif des opérateurs proximaux utilisés. Les sections qui suivent développent les calculs des opérateurs proximaux pour certaines fonctions non-triviales.

A Récapitulatif des opérateurs utilisés

L'opérateur proximal d'une fonction g convexe propre et semi-continue à gauche est défini comme suit :

$$\text{prox}_g(x) = \underset{y}{\text{argmin}} \frac{1}{2} \|y - x\|_2^2 + g(y) \quad (\text{A.1})$$

Les opérateurs proximaux utilisés dans cette thèse sont donnés dans le tableau **A.1**. Les opérateurs proximaux #4, #6 et #7 sont détaillés respectivement en section B, C et D.

B Positivé et contrainte de norme

Le but est de trouver \hat{y} tel que :

$$\hat{y} = \text{prox}_{\geq 0, \|\cdot\|_2 \leq 1}(x). \quad (\text{A.2})$$

Cet opérateur est bien défini, comme l'opérateur proximal d'une fonction convexe propre semi-continue à gauche puisque c'est une fonction caractéristique d'une intersection non-vide d'ensembles convexes. Le Lagrangien du problème d'optimisation correspondant est donné par :

$$\mathcal{L}(y, u, \lambda) = \frac{1}{2} \|y - x\|_2^2 - \langle u, y \rangle + \frac{\lambda}{2} (\|y\|_2^2 - 1). \quad (\text{A.3})$$

#	Fonction	Opérateur proximal
1	$i_{\geq 0}(x)$	$[x]_+ = \max(x, 0)$
2	$\ \lambda \odot x\ _1$	$\text{Soft}_\lambda(x) = \text{sign}(x) \odot [x - \lambda]_+$
3	$\ \lambda \odot x\ _1 + i_{\geq 0}(x)$	$[\text{Soft}_\lambda(x)]_+$
4	$i_{\geq 0, \ \cdot\ _2 \leq 1}(x)$	$[x]_+ / \max(\ [x]_+\ _2, 1)$
5	$i_{(\mathbf{R}^T \cdot) \geq 0}(x)$	$\mathbf{R}[\mathbf{R}^T x]_+$
6	$i_{(\mathbf{W}^T \cdot) \geq 0}(x)$	$x + \mathbf{W}[-\mathbf{W}^T x]_+$
7	$\ \lambda \odot (\mathbf{W}x)\ _1$	$x - \mathbf{W}^T \left(\underset{ u_w _i \leq \lambda_i, \forall i}{\text{argmin}} \ x - \mathbf{W}^T u_w\ _2^2 \right)$
8	$i_{ \cdot _i \leq \lambda_i, \forall i}(x)$	$x - \text{Soft}_\lambda(x)$

Tableau A.1 – Quelques fonctions et leurs opérateurs proximaux, avec x un vecteur colonne, \mathbf{R} une transformée orthonormale, et \mathbf{W} une transformée telle que $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ (trame ajustée).

À l'optimum $(\hat{y}, \hat{u}, \hat{\lambda})$, les conditions de Karush-Kuhn-Tucker (KKT) permettent d'obtenir que :

- (i) $\hat{y} \geq 0, \|\hat{y}\|_2^2 \leq 1,$
- (ii) $\hat{u} \geq 0, \hat{\lambda} \geq 0,$
- (iii) $\hat{y} - x - \hat{u} + \hat{\lambda} \hat{y} = 0,$
- (iv) $\hat{u} \odot \hat{y} = 0, \hat{\lambda} (\|\hat{y}\|_2^2 - 1) = 0.$

En rassemblant toutes ces conditions, on obtient que :

$$\text{prox}_{\geq 0, \|\cdot\|_2 \leq 1}(x) = \frac{[x]_+}{\max(\|[x]_+\|_2, 1)}. \quad (\text{A.4})$$

L'opérateur proximal (A.2) se réduit ainsi à une projection sur l'ensemble des vecteurs à coefficients positifs, suivie par une projection sur la boule unité ℓ_2 .

C Positivité et trames ajustées

Le but est de calculer :

$$\hat{y}_w = \text{prox}_{\mathbf{W}^T \cdot \geq 0}(x_w), \quad (\text{A.5})$$

pour un vecteur colonne x avec \mathbf{W} une opérateur tel que $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ (trame ajustée). Cet opérateur proximal est bien défini puisque $i_{\mathbf{W}^T \cdot \geq 0}$ est une fonction convexe propre et semi-continue à gauche. Puisque $\max_{u \leq 0} \langle u, \mathbf{W}^T y_w \rangle = i_{\geq 0}(\mathbf{W}^T y_w)$, calculer cet opérateur est équivalent à résoudre :

$$\underset{y_w}{\text{argmin}} \max_{u \leq 0} \frac{1}{2} \|y_w - x_w\|_2^2 + \langle u, \mathbf{W}^T y_w \rangle. \quad (\text{A.6})$$

En considérant le théorème min-max de Von Neumann, les opérateurs min et max peuvent être intervertis. Trouver le \hat{y}_w optimal qui minimise la fonction de coût pour un u donné produit la relation : $\hat{y}_w = x_w - \mathbf{W}u$. Remplacer y_w par sa valeur optimale et réarranger les termes mène au problème d'optimisation suivant :

$$\operatorname{argmin}_{u \leq 0} \frac{1}{2} \|\mathbf{W}u - x_w\|_2^2. \quad (\text{A.7})$$

Puisque $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ par hypothèse, ce problème est équivalent à :

$$\operatorname{argmin}_{u \leq 0} \frac{1}{2} \|u - \mathbf{W}^T x_w\|_2^2, \quad (\text{A.8})$$

qui est directement la projection de $\mathbf{W}^T x_w$ sur l'ensemble des matrices à coefficients négatifs : $\hat{u} = -[-\mathbf{W}^T x_w]_+$.

Ainsi, rassembler toutes les relations permet d'obtenir la solution analytique suivante pour le proximal opérateur recherché :

$$\operatorname{prox}_{i, \geq 0}(\mathbf{W}^T \cdot)(x_w) = x_w + \mathbf{W}[-\mathbf{W}^T x_w]_+. \quad (\text{A.9})$$

D Parcimonie dans un espace transformé

Le but est de trouver \hat{y} tel que :

$$\hat{y} = \operatorname{prox}_{\|\lambda \odot \mathbf{W} \cdot\|_1}(x), \quad (\text{A.10})$$

avec x un vecteur colonne x et \mathbf{W} une transformée linéaire sur x . Cet opérateur proximal est bien défini puisque $\|\lambda \odot \mathbf{W} \cdot\|_1$ est une fonction convexe propre et semi-continue à gauche. En utilisant le fait que :

$$\max_{|u_w|_i \leq \lambda_i} \langle u, \mathbf{W}y \rangle = \|\lambda \odot \mathbf{W}y\|_1, \quad (\text{A.11})$$

et en suivant les mêmes étapes que dans la section précédente C jusqu'à l'équation (A.7), on obtient la relation $\hat{y} = x - \mathbf{W}^T u_w$, et que le calcul de (A.10) est équivalent au problème suivant :

$$\operatorname{argmin}_{|u_w|_i \leq \lambda_i} \frac{1}{2} \|\mathbf{W}^T u_w - x\|_2^2, \quad (\text{A.12})$$

Cependant, contrairement au cas précédent, $\mathbf{W}\mathbf{W}^T \neq \mathbf{I}$ et le problème (A.12) n'a alors pas de solution analytique. Celle-ci peut néanmoins être calculée en utilisant l'algorithme FB (cf. section II-C.3.1, page 20), avec comme gradient $u_w \mapsto \mathbf{W}(\mathbf{W}^T u_w - x)$ et comme opérateur proximal $P_\lambda^\infty : u_w \mapsto u_w - \operatorname{Soft}_\lambda(u_w)$ (projection sur la boule ℓ_∞).

Finalement, on obtient :

$$\operatorname{prox}_{\|\lambda \odot \mathbf{W} \cdot\|_1}(x) = x - \mathbf{W}^T \operatorname{argmin}_{|u_w|_i \leq \lambda_i} \frac{1}{2} \|\mathbf{W}^T u_w - x\|_2^2. \quad (\text{A.13})$$

E Contrainte de sous-espace et de positivité

On cherche ici à trouver l'opérateur proximal de la fonction $i_{\text{Im}(\mathbf{W})}(x_w) + i_{\cdot, \mathbf{W}^T \geq \mathbf{0}}$, avec $\text{Im}(\mathbf{W}) = \{y \in \mathbb{R}^p, \exists x \in \mathbb{R}^n \mid y = \mathbf{W}x\}$. Cela revient donc à résoudre le problème suivant :

$$\underset{x_w}{\text{argmin}} \frac{1}{2} \|y_w - x_w\|_2^2 + i_{\text{Im}(\mathbf{W})}(x_w) + i_{\cdot, \geq \mathbf{0}}(\mathbf{W}^T x_w).$$

Ce problème est équivalent par bijection (voir annexe B-C, page 134) à la minimisation suivante dans le domaine direct :

$$\underset{x}{\text{argmin}} \frac{1}{2} \|\mathbf{W}x - y_w\|_2^2 + i_{\cdot, \geq \mathbf{0}}(x),$$

qui est la projection de $\mathbf{W}^T y_w$ sur la contrainte de positivité puisque $\mathbf{W}^T \mathbf{W} = \mathbf{I}_n$. L'opérateur proximal est alors :

$$\text{prox}_{\text{Im}(\mathbf{W}), \cdot, \geq \mathbf{0}}(x) = \mathbf{W} [\mathbf{W}^T x_w]_+ \quad (\text{A.14})$$

On remarque alors que cet opérateur est très proche de l'opérateur développé en C. L'annexe B-C montre ainsi que remplacer l'opérateur proximal de la positivité dans l'algorithme de mise-à jour en synthèse de l'annexe B-A (page 133) par l'opérateur de l'équation (A.14) revient à résoudre le problème en analyse.

Détails des algorithmes

Sommaire

A	Mise-à-jour positive régularisée en synthèse	133
B	Mise-à-jour positive régularisée en analyse	134
C	Analyse formulée en synthèse contrainte	134
D	Reconditionnement élément par élément de l'algorithme FB avec une régularization ℓ_1	135

Les différentes versions de nGMCA nécessitent de pouvoir résoudre les sous-problèmes de mises-à-jour des mélanges \mathbf{A} et des sources \mathbf{S} . Cette annexe fournit les détails des instances des algorithmes proximaux utilisées pour la résolution de ces sous-problèmes.

A Mise-à-jour positive régularisée en synthèse

L'**Algorithme B.1** ci-dessous implémente la mise-à-jour positive de \mathbf{S} en synthèse :

$$\operatorname{argmin}_{\mathbf{S}_w \mathbf{W} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A} \mathbf{S}_w \mathbf{W}\|_2^2 + \|\mathbf{\Lambda}^{(k)} \odot \mathbf{S}_w\|_1. \quad (\text{B.1})$$

Il converge pour $\gamma < \frac{2}{L}$ et $\mu \in]0, \min\left(\frac{3}{2}, \frac{1+2/(L\gamma)}{2}\right)[$ où $L = \|\mathbf{A}^T \mathbf{A}\|_{s,2}$, et utilise l'algorithme GFB [Raguet *et al.* 2013] vu en section II-C.3.2 (page 21).

Algorithme B.1 : Mise-à-jour en synthèse

- 1: **initialiser** $\mathbf{S}_w^{(0)}, \mathbf{U}_w^{(0)}, \mathbf{V}_w^{(0)}$ $\gamma > 0, \omega_u, \omega_v \in]0, 1[$ tels que $\omega_u + \omega_v = 1, k = 0$.
- 2: **tant que** (non-convergé) **faire**
- 3: $\mathbf{G}_w = \mathbf{A}^T (\mathbf{A} \mathbf{S}_w^{(k)} \mathbf{W} - \mathbf{Y}) \mathbf{W}^T$ %calcul du gradient
- 4: $\mathbf{U}_w^{(k+1)} = \mathbf{U}_w^{(k)} - \mu \mathbf{S}_w^{(k)} + \mu \operatorname{Soft}_{\frac{\gamma}{\omega_u} \mathbf{\Lambda}} (2\mathbf{S}_w^{(k)} - \mathbf{U}_w^{(k)} - \gamma \mathbf{G}_w)$
- 5: $\mathbf{V}_w^{(k+1)} = \mathbf{V}_w^{(k)} - \mu \mathbf{S}_w^{(k)} + \mu \operatorname{prox}_{i \geq 0}(\mathbf{W}^T) (2\mathbf{S}_w^{(k)} - \mathbf{V}_w^{(k)} - \gamma \mathbf{G}_w)$
- 6: $\mathbf{S}_w^{(k+1)} = \omega_u \mathbf{U}_w^{(k+1)} + \omega_v \mathbf{V}_w^{(k+1)}$
- 7: $k = k + 1$
- 8: **renvoyer** $\mathbf{S}_w^{(k)}$

B Mise-à-jour positive régularisée en analyse

Le problème :

$$\min_{\mathbf{S} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2 + \|\mathbf{A} \odot (\mathbf{S}\mathbf{W}^T)\|_1, \quad (\text{B.2})$$

peut être mis sous la forme utilisée dans [Chambolle & Pock 2010] (cf. II-C.3.4 (page 23)) :

$$\min_{\mathbf{S}} \max_{|\mathbf{U}_w|_{i,j} \leq \mathbf{A}_{i,j}} \max_{\mathbf{V} \leq \mathbf{0}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2 + \langle \mathbf{S}\mathbf{W}^T, \mathbf{U}_w \rangle + \langle \mathbf{S}, \mathbf{V} \rangle. \quad (\text{B.3})$$

Ainsi, il peut s'écrire :

$$\min_{\mathbf{S}} \max_{\mathbf{U}_w, \mathbf{V}} G(\mathbf{S}) + F_1^*(\mathbf{U}_w) + F_2^*(\mathbf{V}) + \langle \mathbf{S}\mathbf{W}^T, \mathbf{U}_w \rangle + \langle \mathbf{S}, \mathbf{V} \rangle, \quad (\text{B.4})$$

avec :

$$\begin{aligned} G(\mathbf{S}) &= \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_2^2. \\ F_1^*(\mathbf{U}_w) &= i_{|\mathbf{U}_w|_{i,j} \leq \mathbf{A}_{i,j}}(\mathbf{U}_w). \\ F_2^*(\mathbf{V}) &= i_{\mathbf{V} \leq \mathbf{0}}(\mathbf{V}). \end{aligned}$$

On remarquera au passage que la variable duale a été séparée en deux parties par simplicité. Elle pourrait facilement être rassemblées en une unique variable.

L'algorithme nécessite alors la connaissance de l'opérateur proximal de G qui est directement donné par :

$$\begin{aligned} \text{prox}_{\tau G}(\mathbf{S}) &= \underset{\mathbf{R}}{\text{argmin}} \frac{\tau}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{R}\|_2^2 + \frac{1}{2} \|\mathbf{R} - \mathbf{S}\|_2^2, \\ &= (\mathbf{I} + \tau \mathbf{A}^T \mathbf{A})^{-1} (\mathbf{S} + \tau \mathbf{A}^T \mathbf{Y}), \end{aligned} \quad (\text{B.5})$$

et il nécessite également les opérateurs proximaux de σF_1 and σF_2 qui sont respectivement les projections sur la boule ℓ_∞ , notée P_Λ^∞ (opérateur proximal #8 du tableau page 130), et sur la contrainte de négativité (opérateur proximal #1).

L'**Algorithme B.2** converge alors vers la solution du problème (B.2) si $\tau \sigma L^2 < 1$, où $L = \sqrt{1 + \|\mathbf{W}\|_{s,2}^2}$. Si \mathbf{W} est une trame ajustée avec $\mathbf{W}\mathbf{W} = \mathbf{I}_n$, ce qui est le cas dans nos expérimentations, $L = \sqrt{2}$.

C Analyse formulée en synthèse contrainte

Le but de cette annexe est de montrer que la formulation en analyse (B.2) peut être mise sous la forme d'un problème en synthèse, avec une contrainte supplémentaire.

Algorithme B.2 : Mise-à-jour en analyse

- 1: initialiser $\mathbf{S}^{(0)}, \mathbf{U}_w^{(0)}, \mathbf{V}^{(0)}, \bar{\mathbf{S}}^{(0)} = \mathbf{S}^{(0)}, \tau > 0, \sigma > 0, k = 0$.
 - 2: tant que (non-convergé) faire
 - 3: $\mathbf{U}_w^{(k+1)} = P_{\Lambda}^{\infty} \left(\mathbf{U}_w^{(k)} + \sigma \bar{\mathbf{S}}^{(k)} \mathbf{W}^T \right)$
 - 4: $\mathbf{V}^{(k+1)} = - \left[-\mathbf{V}^{(k)} - \sigma \bar{\mathbf{S}}^{(k)} \right]_+$
 - 5: $\mathbf{S}^{(k+1)} = (\mathbf{I} + \tau \mathbf{A}^T \mathbf{A})^{-1} \left(\mathbf{S}^{(k)} - \tau (\mathbf{U}_w \mathbf{W} + \mathbf{V} - \mathbf{A}^T \mathbf{Y}) \right)$
 - 6: $\bar{\mathbf{S}}^{(k+1)} = 2\mathbf{S}^{(k+1)} - \mathbf{S}^{(k)}$
 - 7: $k = k + 1$
 - 8: renvoyer $\mathbf{S}^{(k)}$
-

On considère pour commencer une application linéaire injective $\mathbf{W} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ et son pseudo-inverse \mathbf{W}^\dagger de telle manière que l'on a : $\mathbf{W}^\dagger \mathbf{W} = \mathbf{I}$. On remarquera alors que la fonction f définie par :

$$f : \mathbb{R}^n \rightarrow \text{Im}(\mathbf{W}) \subset \mathbb{R}^p \quad (\text{B.6})$$

$$x \mapsto \mathbf{W}x$$

est une bijection et $f^{-1}(y) = \mathbf{W}^\dagger y$.

En effet, soit $y \in \text{Im}(\mathbf{W})$, il existe alors par définition un $x \in \mathbb{R}^n$ tel que $y = \mathbf{W}x$ et f est donc surjective. Puisque elle est également injective par définition de \mathbf{W} , f est une bijection. Avec $y \in \text{Im}(\mathbf{W})$ et $x \in \mathbb{R}^n$ tels que $y = \mathbf{W}x$, $\mathbf{W}^\dagger y = \mathbf{W}^\dagger \mathbf{W}x = x$ puisque $\mathbf{W}^\dagger \mathbf{W} = \mathbf{I}$, et l'on a donc $f^{-1}(y) = \mathbf{W}^\dagger y$.

Ceci peut être utilisé pour la résolution des problèmes de type :

$$\min_x f(x) + g(\mathbf{W}x), \quad (\text{B.7})$$

puisque par bijection, celui-ci peut être réécrit sous la forme :

$$\min_y f(\mathbf{W}^\dagger y_w) + g(y_w), \quad y_w \in \text{Im}(\mathbf{W}) \quad (\text{B.8})$$

ou de manière équivalente :

$$\min_y f(\mathbf{W}^\dagger y_w) + g(y_w) + i_{\text{Im}(\mathbf{W})}(y_w) \quad (\text{B.9})$$

Cela permet en particulier de simplifier la résolution quand f n'est pas différentiable alors que f l'est. En pratique, on utilise ainsi une formulation en synthèse avec une contrainte supplémentaire afin de rester dans l'image de \mathbf{W} .

D Reconditionnement élément par élément de l'algorithme FB avec une régularisation ℓ_1

La résolution du problème suivant :

$$\min_x f(x) + \|\lambda \odot x\|_1, \quad (\text{B.10})$$

avec f est parfaitement équivalente à la résolution du problème après changement de variable à $x = b \odot z$ où b est un vecteur à coefficients strictement positifs :

$$\min_z f(b \odot z) + \|\lambda \odot b \odot z\|_1. \quad (\text{B.11})$$

Lorsque f est convexe et dérivable, ce problème peut être résolu en utilisant l'algorithme FB, dont les itérations sont données par :

$$z^{k+1} \leftarrow \text{Soft}_{\tau\lambda \odot b} \left(z^k - \tau b \odot \nabla f(b \odot z^k) \right) \quad (\text{B.12})$$

En revenant à la variable $x = b \odot z$, et en notant que :

$$b \odot \text{Soft}_\lambda (b^{-1} \odot \cdot) = \text{Soft}_{b \odot \lambda} (\cdot), \quad (\text{B.13})$$

on retrouve que ces itérations sont équivalentes aux suivantes :

$$x^{k+1} \leftarrow \text{Soft}_{\tau\lambda \odot b^2} \left(x^k - \tau b^2 \odot \nabla f(x^k) \right). \quad (\text{B.14})$$

À partir de là, on peut observer qu'il est possible de reconditionner le gradient de f avec n'importe quel vecteur strictement positif, du moment que le seuillage est adapté en conséquence. Dans ce cas là, le pas τ doit être pris de telle sorte que $\frac{1}{\tau}$ soit inférieur à la constante de Lipschitz de $x \mapsto b \odot \nabla f(b \odot x)$. Si ∇f est une application linéaire, on peut facilement trouver cette constante via la méthode de la puissance.

Séparation de sources partiellement corrélées

Des travaux annexes ont été effectués sur la séparation de sources lorsque celles-ci sont partiellement corrélées. Ce cas se présente en effet couramment en spectrométrie de masse par exemple, deux isomères partageant un grand nombre de pics. Il en a découlé un article pour la conférence ICIP [Bobin *et al.* 2014] dont les résultats ne sont pas décrits dans le cadre de ce manuscrit afin d'en conserver la cohérence.

L'algorithme présenté dans [Bobin *et al.* 2014], appelé Adaptive Morphological Component Analysis (AMCA) est une adaptation de GMCA (cf. section IV-A, page 52). Il introduit un procédé de repondération du terme d'attache aux données visant à davantage utiliser l'information provenant des échantillons les plus discriminants afin d'effectuer la séparation des sources. La recherche de ces échantillons discriminants y est effectuée de manière automatique tout au long des itérations.

Des expérimentations montrent qu'AMCA est très significativement plus robuste que les autres algorithmes de SSA lorsque les sources recherchées partagent certains grands coefficients et ne satisfont donc pas l'hypothèse de diversité morphologique telle que présentée en III-A.2.2 (page 31). Le lecteur intéressé pourra trouver davantage de détails dans la reproduction ci-dessous de l'article présentant AMCA.

SPARSE BLIND SOURCE SEPARATION FOR PARTIALLY CORRELATED SOURCES

*J. Bobin, J. Starck**

CEA Saclay
IRFU/Sap-SEDI
Orme des merisiers, Bat. 709
91191 Gif-sur-Yvette, France

J. Rapin, A. Larue

CEA Saclay
LIST/LADIS
Bat. 536
91191 Gif-sur-Yvette, France

ABSTRACT

Blind source separation (BSS) is a very popular technique to analyze data which can be modeled as linear mixtures of elementary sources. Standard approaches generally make the assumption that such sources are statistically independent or at least uncorrelated. However, this is barely the case for real-world sources which are very often partially correlated. We present a new sparsity-enforcing BSS method coined Adaptive Morphological Component Analysis (AMCA) designed to retrieve sparse and partially correlated sources based on an adaptive weighting scheme. Numerical experiments have been carried out which show that the proposed method is robust to the partial correlation of the sources while standard BSS techniques fail. The performances of the proposed algorithm are further illustrated with simulations in the context of astrophysics.

Index Terms— Sparsity, wavelets, blind source separation, morphological diversity

1. INTRODUCTION

Blind source separation is a classical problem in statistics and signal/image processing for the efficient processing of multi-channel or multi-wavelength data. In this context, each observation $\{x_i\}_{i=1,\dots,m}$ is modeled as the linear combination of $n \leq m$ sources $\{s_j\}_{j=1,\dots,n}$:

$$\forall i = 1, \dots, m; \quad x_i = \sum_{j=1}^n a_{ij}s_j + n_i$$

where a_{ij} models quantifies the contribution of source j in the observation x_i . Each datum is assumed to have t entries. The term n_i models for additive noise or model imperfections. The well-known linear mixture model is customarily recast in the following matrix form:

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \quad (1)$$

*Supported by the French National Agency for Research (ANR) 11-ASTR-034-02-MultiID. Contact : name.lastname@cea.fr

The goal of blind source separation is to blindly estimate both \mathbf{A} and \mathbf{S} from the knowledge of \mathbf{X} only. However, there exists an infinite number of couples (\mathbf{A}, \mathbf{S}) which satisfy $\mathbf{X} = \mathbf{AS}$. Disentangling between the sources then requires using prior information that emphasize some discriminative information or diversity between the sources.

In the last two decades Independent Component Analysis (ICA - [1, 2] references therein) has taken the lion share. In brief, ICA enforces the statistical independence of the sources as a discriminative measure of diversity between the sources. More recently, sparked by advances in harmonic analysis and applied mathematics [3], sparse blind source separation has emerged as an effective separation technique (see [4, 5]). In the framework of ICA, sparse and independent sources have been studied in [6, 7].

However, real-life sources are seldom perfectly independent nor decorrelated. In such cases, neither statistical independence nor morphological diversity hold. In this case, most classical source separation methods generally yield erroneous separation results.

In some applications like spectroscopy, the sources of interest are often partially correlated sources. In this particular field of research, the development of dedicated source separation techniques has already been considered assuming strong assumptions about the signal [8, 9]. However these methods generally make strong assumptions about the signals : i) the sources and the mixing matrix are non-negative, ii) for each source there exists one non-zero sample that vanishes for the other sources and iii) sparsity is modeled only in the direct domain.

2. SPARSE BSS FOR PARTIALLY CORRELATED SOURCES

2.1. Sparse BSS

Sparse blind source separation has emerged as an effective separation technique (see [4, 5]). In this setting, the sources are assumed to be sparse; this means that there exists some basis (or waveform dictionary) Φ so that each source can

be decomposed as follows $s = \alpha\Phi$ where α has few entries with significant amplitudes and most of them are close or equal to zero. The compressibility of the sources in Φ is particularly appealing as it greatly helps enhancing the contrast between the sources. More precisely, in [5] the sources are assumed not to share their most significant entries Φ . In this setting, the most discriminative data samples are the ones which are related to the high amplitude entries of the sources. Such property has further been called morphological diversity. Building on this concept, Generalized Morphological Component Analysis (GMCA) has been showed to be an effective BSS method and it proved to be robust to additive noise [5]. In the GMCA algorithm, the sources and the mixing matrix are estimated by solving the following optimization problem:

$$\min_{\mathbf{A}, \mathbf{S}} \sum_{i=1}^n \lambda_i \|s_i \Phi^T\|_1 + \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 \quad (2)$$

where the first term penalizes non-sparse solutions and the quadratic term assesses the fidelity to the data. The Frobenius norm is defined as follows: $\|\mathbf{Y}\|_F^2 = \text{Trace}(\mathbf{Y}\mathbf{Y}^T)$. The sparsity level is measured by the ℓ_1 norm of the sources which is defined as the sum of the absolute values of the sources' entries. This problem is solved by using an iterative alternate projected least-square estimation of \mathbf{S} and \mathbf{A} . It is well established that the problem in Equation 2 is non-convex. However, fixing the mixing matrix \mathbf{A} (*resp.* the sources \mathbf{S}) makes each sub-problem convex. Assuming that the sources are sparse in Φ , the GMCA algorithm can be decomposed into the following two steps :

- *Step 1 : Estimating \mathbf{S} for fixed \mathbf{A} :*

$$\min_{\mathbf{S}} \sum_{i=1}^n \lambda_i \|s_i \Phi^T\|_{\ell_1} + \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$$

- *Step 2 : Estimating \mathbf{A} for fixed \mathbf{S} :*

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$$

Unless \mathbf{A} and Φ are orthogonal matrices, the solution to the problem in Step 1 does not have a closed-form expression. This solution could be further computed using an iterative thresholding algorithm such as the Forward-Backward splitting algorithm[10] but at the cost of highing the computational burden of this step. In the GMCA algorithm, this step is approximated with a projected least-square step :

$$\forall i = 1, \dots, n; \quad s_i = \mathcal{S}_{\lambda_i} ([\mathbf{A}^+ \mathbf{X}]_i \Phi^T) \Phi$$

where $\mathbf{A} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ is the pseudo-inverse of the matrix \mathbf{A} and $[\mathbf{Y}]_i$ stands for the i -th row of the matrix \mathbf{Y} . The operator \mathcal{S}_{λ_i} is the well-known soft-thresholding operator with threshold λ_i .

The second step of the GMCA algorithm is a least-square problem which admits a solution with closed-form expression : $\mathbf{A} = \mathbf{X}\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1}$.

It has been emphasized in [11] that the choice of the thresholds $\{\lambda_i\}_{i=1, \dots, n}$ plays an important role. Their values are lowered at each iteration starting for high values and decreasing towards a final value that is function of the noise level of the sources. The role played by the thresholds is twofold: i) a coarse-to-fine estimation process : According to the concept of morphological diversity, the most discriminative entries of the sources are those with the highest amplitudes. Therefore, the thresholds are first fixed to high values to only select the entries with the most significant amplitudes. ii) the thresholding procedure in Step 1 of the algorithm allows for the rejection of small entries which are likely related to noise thus making the algorithm more robust to noise. The final value of each threshold is function of the noise level. In practice, they are fixed to $3 \text{MAD}([\mathbf{A}^+ \mathbf{X}]_i)$ where the MAD (median absolute deviation) is an empirical estimator of the noise standard deviation.

2.2. The AMCA algorithm

Whether relying on statistical independence or morphological diversity, standard sparse BSS methods are not well suited to estimate partially correlated sources. However, the partial correlation of signals can be found in rather distinct cases in real-world signal. In the finite sample case, chance correlations between correlation can arise as single realizations of theoretically independent or uncorrelated sources are likely to have non-zero correlation. In imaging, the sparse modeling of signals is quite often done in a multiscale signal representation like the wavelets or the curvelets (see [12, 5]); in these domains, the expansion coefficients which are related to the low-frequency content of images (*e.g.* the wavelet coarse scale) have generally the most energetic coefficients. However, on these scales, the effective correlation between two images is generally non-zero. This is especially true in the low number of samples case.

It is important to note that, in the setting of sparse BSS, the most discriminative samples are the coefficients which have the most significant amplitudes. In the case of partially correlated sources, the most discriminative samples should be the ones that are not common to more than one sources or at least mainly significant in source. In contrast to the morphological diversity principle defined in Section 2.1, the most discriminative coefficients may not be the large amplitude entries of the sources.

The gist of the forthcoming method relies on the definition of some measure of samples discriminativeness to penalize non-discriminative sample and favor the others. As stated earlier, a good way to assess the discriminant power of some sample is to check whether this sample is mainly significant in only a single source or shared between different sources. A straight-

forward approach consists, for each sample t in the sparse domain Φ , in computing the ℓ_1 norm of this sample across the sources : $q^{(t)} = \|\mathbf{S}\Phi^T\|_{\ell_1}^{(t)}$ where the operator $[\mathbf{Y}]^{(t)}$ extracts the t -th column of \mathbf{Y} . Small (*resp.* large) values for $q^{(t)}$ indicate that the t -th sample is mainly active in few (*resp.* many) sources. This quantity is then used to define a diagonal weighting matrix $\mathbf{W} = \text{diag}(1/q^{(t)^2})$ to further favor discriminative samples. The weights are squared to give W the units of a covariance matrix. Step 2 of the GMCA algorithm is then substituted with the following update rule :

$$\min_{\mathbf{A}} \text{Trace}(\mathbf{X} - \mathbf{A}\mathbf{S}) \mathbf{W} (\mathbf{X} - \mathbf{A}\mathbf{S})^T \quad (3)$$

The ‘‘optimal’’ weighting matrix \mathbf{W} – which would be defined from the true sources – is not known and has further to be learnt as well. The weighting matrix is therefore updated at each iteration (k) of the algorithm from the current estimate of the sources $\mathbf{S}^{(k)}$. We will therefore call this algorithm Adaptive Morphological Component Analysis (AMCA) as it adapts itself to learn the discriminative level of the samples. The AMCA algorithm is detailed below:

1. Initialize $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}, \{\lambda_i^{(0)}\}_{i=1, \dots, n}$ and $p^{(0)} = 1$
2. While $k < P_{\max}$,
 - Updating \mathbf{S} assuming \mathbf{A} is fixed :

$$\bar{\mathbf{S}} = \left(\mathbf{A}^{(k)T} \mathbf{A}^{(k)} \right)^{-1} \mathbf{A}^{(k)T} \mathbf{X}$$

$$\forall i = 1, \dots, n; \quad [\mathbf{S}^{(k+1)}]_i = \mathcal{S}_{\lambda_i^{(k)}}(\bar{\mathbf{S}}\Phi^T) \Phi$$
 - Updating the weight matrix :

$$\mathbf{W}^{(k)} = \text{diag} \left(\frac{1}{\|\mathbf{S}^{(k+1)}\Phi^T\|_{\ell_1}^{(t)}} \right)$$
 - Updating \mathbf{A} assuming \mathbf{S} is fixed :

$$\mathbf{A}^{(k+1)} = \mathbf{X}\Phi \mathbf{W}^{(k)} \mathbf{S}_{\Phi}^{(k+1)T} \left(\mathbf{S}_{\Phi}^{(k+1)} \mathbf{W}^{(k)} \mathbf{S}_{\Phi}^{(k+1)T} \right)^{-1}$$
 where $\mathbf{X}_{\Phi} = \mathbf{X}\Phi^T$ and $\mathbf{S}_{\Phi} = \mathbf{S}\Phi^T$.
 - Decrease each threshold $\{\lambda_i^{(k)}\}_{i=1, \dots, n}$.
3. Stop when $k = P_{\max}$.

The number of iterations P_{\max} is generally of the order of 10^2 ; in the forthcoming numerical experiments, it was typically set to $P_{\max} = 500$. The thresholding strategy is the same as in GMCA – see Section 2.1.

3. NUMERICAL RESULTS

3.1. Simulations

In this section, we evaluate the performances of the AMCA algorithm with respect to standard sparse BSS method : i) the GMCA algorithm, ii) RNA (Relative Newton Algorithm) [6] and iii) EFICA (Efficient FastICA) [7]. Each of the 4096 entries of the sources are drawn at random so that : i) only

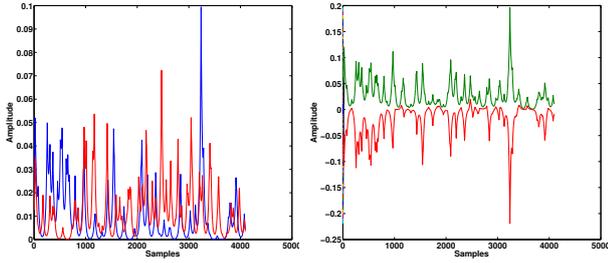
2% of the entries of each source is active among which a certain ratio c – the so-called coherence level – is shared by all the sources, ii) active entries are distributed according to a Gaussian with variance 1. The coherence level c controls the ratio of spikes which are common to all the sources; c takes its values between 0 (the entries of the sources are independently distributed) and 1 (the sources have all the same active entries). The sources are then convolved with a Laplacian kernel of full width at half maximum (fwhm) equal 15 samples. This model will tend to mimic spectroscopic data. The sparse representation Φ is a tight frame of redundant (translation invariant) 1D wavelet. The mixing matrix will be picked at random from a Gaussian distribution. Examples of simulated 2 sources (*resp.* 2 mixtures) are shown on the left (*resp.* right) at the top of Figure 3.1.

The experimental results will be evaluated with two different criteria : the SDR (Signal-to-Distortion Rate – see [13]) and the Mixing matrix criterion defined as $C_A = \|\mathbf{P}\hat{\mathbf{A}}^+ \mathbf{A}^*\|_{\ell_1}$ where $\hat{\mathbf{A}}^+$ is the pseudo-inverse of the estimated mixing matrix corrected – through \mathbf{P} – for the scale/permutation indeterminacies; \mathbf{A}^* is the true mixing matrix. The lower C_A is, the better the separation is. In this experiment, the number of sources and the number of channels are both fixed to 10. The noise level is fixed to 120dB.

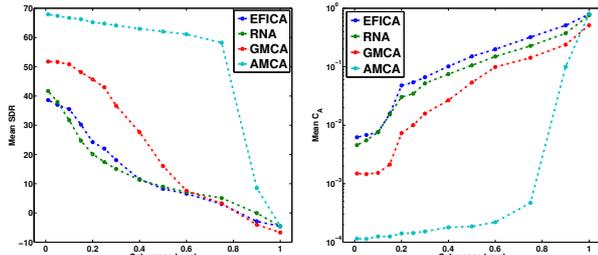
The evolution of the SDR as a function of the coherence level c is displayed in the bottom-left panel of Figure 3.1. First, it is interesting to notice that the performances of the standard BSS algorithms (GMCA, EFICA and RNA) are quickly hampered by the partial correlation between the sources. The GMCA algorithm seems to be less sensitive to the correlation for $c < 0.2$. When more than 20% of the entries are shared by all the sources, these algorithms equally behave badly. Interestingly, the AMCA algorithm already shows a much higher SDR even when only a few entries are correlated. When less than 80% of the entries are correlated, the SDR of the AMCA algorithm is higher than 60 dB. Its performances start degrading for $c > 0.8$ to eventually behave similarly to the standard BSS algorithm when c tends towards 1. The evolution of the mixing matrix criterion are shown in the bottom-right panel of Figure 3.1. One can notice that the mixing matrix criterion with the AMCA algorithm is consistently more than 1 order of magnitudes lower than for the other standard BSS methods.

3.2. Illustration in astrophysics

In astrophysics, source separation has turned to be crucial for the analysis of microwave surveys [14, 15, 16]. In this context, the observations are modeled as linear mixtures of various physical emissions such as the cosmological microwave background (CMB) and the foreground galactic emissions. In this section, we focus on disentangling between the different galactic sources: free-free emission, spinning dust, synchrotron and thermal. Simulations of these sources using the



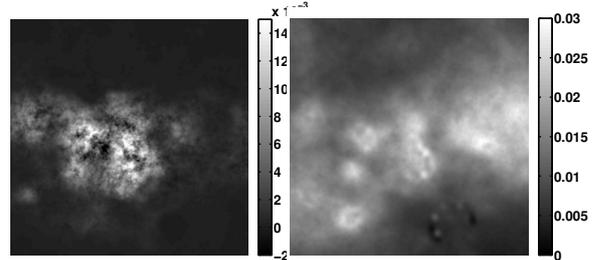
Left: Example of 2 sources. **Right:** Mixtures.



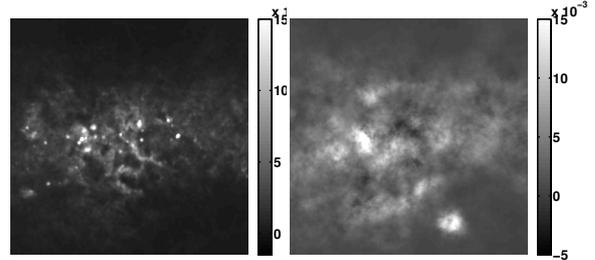
Left: SDR. **Right:** Mixing matrix criterion.

PSM (Planck sky model¹) are displayed in Figure 3.2. We will further specifically evaluate the standard sparse BSS algorithms for the estimation of these 4 sources from PSM simulations of the 5 WMAP observations; for more details about WMAP, we invite the interested reader to refer to [14]. In these simulations, the linear mixture model (1) holds true. The sources come from various galactic emissions; however they are dominant in the same galactic regions of the sky as can be seen in Figure 3.2. This makes them particularly correlated at large scale (low frequency). In the numerical experiments, the separation performances are evaluated using the SDR and the mixing matrix criterion C_A for various settings. The sparse BSS algorithms are applied to the same data at different resolutions or sample size ranging from 64×64 to 256×256 pixels; at low sample size, the effect of chance correlations at large scales should hamper the separation performances. The noise level is set to the nominal WMAP noise level. The panel on the left of Figure 3.2 shows the SDR of the reconstructed sources as a function of the resolution; the right panel displays the evolution of the mixing criterion. Clearly, the performances of all methods decrease when the resolution decrease; indeed, when the resolution decreases, the large scales become more prominent. However, the AMCA algorithm outperforms all other algorithms with a SDR gain of about 7 dB. These preliminary results show that the AMCA algorithm is more robust to partial correlations than existing sparse BSS methods.

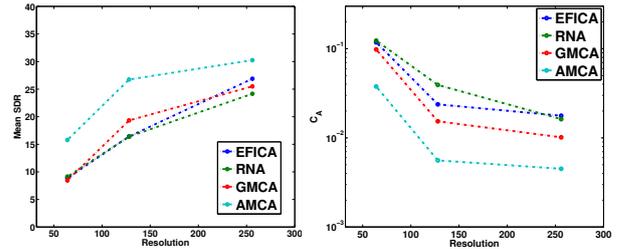
¹<http://www.apc.univ-paris7.fr/delabrou/PSM/psm.html>



Astrophysical sources - left: free-free. **Right:** synchrotron.



Astrophysical sources - left: dust. **Right:** spinning dust.



Varying resolution - left: SDR. **Right:** criterion C_A .

4. CONCLUSION

We introduce a novel blind source separation algorithm (BSS) to tackle the separation of sparse and partially correlation sources. In contrast to standard sparse BSS methods, the key ingredient for the separation of partially correlated sources is the estimation of discriminative samples to be favored in the separation process. For that purpose, the proposed AMCA algorithm (Adaptive Morphological Component Analysis) relies on a weighting scheme which iteratively evaluates the discriminative power of the data sample based on the sparse distribution of the estimated sources. Numerical results are provided which demonstrate the robustness of the AMCA to the partial correlation of the sought-after sources. The good performances of the AMCA algorithm are illustrated in the context of astrophysics where the sources to be estimated are highly correlated at large scale.

5. REFERENCES

- [1] P. Comon and C. Jutten, *Handbook of blind source separation*, Elsevier, 2010.
- [2] S. Choi, A. Cichocki, H. Park, and S. Lee, "Blind source separation and independent component analysis: a review," *Neural Inform. Process. Lett. Rev.*, vol. 6, pp. 1–57, 2005.
- [3] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.
- [4] Y. Li, S. Amari, A. Cichocki, and C. Guan, "Underdetermined blind source separation based on sparse representation," *IEEE Transactions on information theory*, vol. 52, pp. 3139–3152, 2006.
- [5] J. Bobin, J.-L. Starck, M. J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, 2007, 2662–2674.
- [6] M. Zibulevski, "Blind source separation with relative Newton method," *Proceedings of ICA, Independent Component Analysis, 2003*, 2003, 897-902.
- [7] P. Tichavsky Z. Koldovsky and E. Oja, "Efficient variant of algorithm fastica for independent component analysis attaining the cramer-rao lower bound," *IEEE Transactions on neural networks*, vol. 17, pp. 1265–1277, 2006.
- [8] W. Naanaa and J-M Nuzillard, "Blind source separation of positive and partially correlated data," *Signal Processing*, vol. 85, pp. 1711–1722, 2005.
- [9] Y. Sun, C. Ridge, F. del Rio, A.J. Shaka, and J. Xin, "Postprocessing and sparse blind source separation of positive and partially overlapped data," *Signal Processing*, vol. 91, pp. 1838–1851, 2011.
- [10] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [11] J. Bobin, J.-L. Starck, Y. Moudden, and M. J. Fadili, "Blind source separation: The sparsity revolution," *Advances in Imaging and Electron Physics*, vol. 152, 2008, 221–306.
- [12] J.-L. Starck, F. Murtagh, and J.M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge, 2010.
- [13] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [14] Bennett C. L. and et.al., "Nine-year wilkinson microwave anisotropy probe (wmap) observations: Final maps and results," *ApJS*, vol. 208, no. 20B, 2013.
- [15] J. Bobin, F. Sureau, P. Paykari, A. Rassat, S. Basak, and J.-L. Starck, "Wmap nine-year cmb estimation using sparsity," *A&A*, vol. 553, pp. L4, 2013.
- [16] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., "Planck 2013 results. XII. Component separation," *ArXiv e-prints 1303.5072*, Mar. 2013.

Bibliographie

- [Aharon *et al.* 2006] M. Aharon, Michael Elad et Alfred M. Bruckstein. K-SVD : An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pages 4311–4322, Novembre 2006. (cité en pages 14 et 17)
- [Anderle *et al.* 2004] Markus Anderle, Sushmita Roy, Hua Lin, Christopher Becker et Keith Joho. Quantifying reproducibility for differential proteomics : noise analysis for protein liquid chromatography-mass spectrometry of human serum. *Bioinformatics*, vol. 20, no. 18, pages 3575–3582, 2004. (cité en pages 104 et 122)
- [Arora *et al.* 2012] Sanjeev Arora, Rong Ge, Ravindran Kannan et Ankur Moitra. Computing a nonnegative matrix factorization - provably. In *Proceedings of STOC*, pages 145–162, 2012. (cité en page 32)
- [Attouch *et al.* 2010] Hédÿ Attouch, Jérôme Bolte, Patrick Redont et Antoine Soubeyran. Proximal Alternating Minimization and Projection Methods for Nonconvex Problems : An Approach Based on the Kurdyka-Łojasiewicz Inequality. *Mathematics of Operations Research*, vol. 35, no. 2, pages 438–457, Mai 2010. (cité en page 76)
- [Babaie-Zadeh *et al.* 2004] Massoud Babaie-Zadeh, Ali Mansour, Christian Jutten et Farrokh Marvasti. A Geometric Approach for Separating Several Speech Signals. In *Proceedings of ICA*, pages 798–806, 2004. (cité en page 32)
- [Barthélemy *et al.* 2012] Quentin Barthélemy, Anthony Larue, Aurélien Mayoue, David Mercier et Jérôme I. Mars. Shift & 2D Rotation Invariant Sparse Coding for Multivariate Signals. *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pages 1597–1611, 2012. (cité en page 17)
- [Beck & Teboulle 2009] Amir Beck et Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pages 183–202, 2009. (cité en pages 21 et 38)
- [Becker & Fadili 2012] Stephen Becker et Jalal Fadili. A quasi-Newton proximal splitting method. In *Proceedings of NIPS*, pages 2627–2635, 2012. (cité en page 109)
- [Benton *et al.* 2008] H. Paul Benton, Dawn M. Wong, Sunia A. Trauger et Gary Siuzdak. XCMS2 : processing tandem mass spectrometry data for metabolite identification and structural characterization. *Analytical Chemistry*, vol. 80, no. 16, pages 6382–6389, Août 2008. (cité en page 106)
- [Berry & Browne 2005] Michael W. Berry et Murray Browne. Email Surveillance Using Non-negative Matrix Factorization. *Computational & Mathematical Organization Theory*, vol. 11, no. 3, pages 249–264, 2005. (cité en page 32)

- [Berry *et al.* 2007] Michael W. Berry, Murray Browne, Amy N. Langville, V. Paul Pauca et Robert J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, vol. 52, no. 1, pages 155–173, 2007. (cité en page 36)
- [Bittorf *et al.* 2012] Victor Bittorf, Benjamin Recht, Christopher Re et Joel A. Tropp. Factoring nonnegative matrices with linear programs. In *Proceedings of NIPS*, volume abs/1206.1270, pages 1223–1231, 2012. (cité en page 32)
- [Bobin *et al.* 2007] Jérôme Bobin, Jean-Luc Starck, Jalal Fadili et Yassir Moudden. Sparsity and morphological diversity in blind source separation. *IEEE Transactions on Image Processing*, vol. 16, no. 11, pages 2662–2674, 2007. (cité en pages 31, 51 et 80)
- [Bobin *et al.* 2008] Jérôme Bobin, Jean-Luc Starck, Yassir Moudden et Jalal Fadili. Blind Source Separation : the Sparsity Revolution. *Advances in Imaging and Electron Physics*, vol. 152, no. January, pages 221–298, 2008. (cité en pages 31 et 51)
- [Bobin *et al.* 2014] Jérôme Bobin, Jean-Luc Starck, Jérémy Rapin et Anthony Larué. Sparse blind source separation for partially correlated sources. In *ICIP*, 2014. (cité en pages 127 et 137)
- [Bolte *et al.* 2013] Jérôme Bolte, Shoham Sabach et Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, pages 1–36, 2013. (cité en pages 76 et 77)
- [Buckheit & Donoho 1995] Jonathan Buckheit et David Donoho. WaveLab and Reproducible Research. In *Wavelets and Statistics*, pages 55–81. Springer-Verlag, 1995. (cité en pages 78 et 127)
- [Candès & Tao 2006] Emmanuel Candès et Terence Tao. Near-Optimal Signal Recovery From Random Projections : Universal Encoding Strategies ? *IEEE Transactions on Information Theory*, vol. 52, no. 12, pages 5406–5425, 2006. (cité en page 13)
- [Candès *et al.* 2007] Emmanuel J. Candès, Michael B. Wakin et Stephen P. Boyd. Enhancing Sparsity by Reweighted ℓ_1 Minimization. *Journal of Fourier Analysis and Applications*, vol. 14, pages 877–905, 2007. (cité en page 92)
- [Candès *et al.* 2011] Emmanuel J. Candès, Xiaodong Li, Yi Ma et John Wright. Robust principal component analysis ? *Journal of the ACM*, vol. 58, no. 3, page 11, 2011. (cité en page 122)
- [Cardoso & Donoho 1999] Jean-François Cardoso et Dave L. Donoho. Some experiments on independent component analysis of non-Gaussian processes. In *Proceedings of Spw-Hos*, pages 74–77, 1999. (cité en page 29)
- [Cardoso 1999] Jean-François Cardoso. High-order Contrasts for Independent Component Analysis. *Neural Computation*, vol. 11, no. 1, pages 157–192, Janvier 1999. (cité en page 30)

- [Chambolle & Pock 2010] Antonin Chambolle et Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pages 120–145, 2010. (cité en pages 23, 81, 83, 109 et 134)
- [Chan *et al.* 2008] Tsung-Han Chan, Wing-Kin Ma, Chong-Yung Chi et Yue Wang. A Convex Analysis Framework for Blind Separation of Non-Negative Sources. *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pages 5120–5134, Octobre 2008. (cité en page 32)
- [Chen *et al.* 2001] Scott Shaobing Chen, David L. Donoho et Michael A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Review*, vol. 43, no. 1, pages 129–159, Janvier 2001. (cité en page 11)
- [Chu *et al.* 2004] Moody Chu, Fasma Diele, Robert J. Plemmons et Stefania Ragni. Optimality, computation, and interpretation of nonnegative matrix factorizations. Rapport technique, Departments of Mathematics and Computer Science, Wake Forest University, USA, 2004. (cité en pages 37 et 44)
- [Cichocki & Zdunek 2007] Andrzej Cichocki et Rafal Zdunek. Regularized Alternating Least Squares Algorithms for Non-negative Matrix/Tensor Factorization. In *Proceedings of ISNN*, pages 793–802, 2007. (cité en pages 41, 44 et 52)
- [Cichocki *et al.* 2006] Andrzej Cichocki, Rafal Zdunek et Shun-ichi Amari. Csiszárs divergences for non-negative matrix factorization : Family of new algorithms. *Proceedings of ICA*, vol. 3889, no. 1, pages 32–39, 2006. (cité en pages 34, 37 et 44)
- [Cichocki *et al.* 2007] Andrzej Cichocki, Rafal Zdunek et Shun-ichi Amari. Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization. In *Proceedings of ICA*, pages 169–176. Springer, 2007. (cité en pages 34, 38, 41 et 44)
- [Cichocki *et al.* 2008] Andrzej Cichocki, Anh Huy Phan et Cesar Caiafa. Flexible HALS algorithms for sparse non-negative matrix/tensor factorization. *IEEE Workshop on Machine Learning for Signal Processing*, vol. 2, no. 4, pages 73–78, 2008. (cité en page 108)
- [Cichocki *et al.* 2009] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan et Shun-ichi Amari. Nonnegative Matrix and Tensor Factorizations : Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. John Wiley & Sons, Ltd, 2009. (cité en pages 38, 41 et 108)
- [Coifman & Donoho 1995] Ronald R. Coifman et Dave L. Donoho. Translation-invariant de-noising. In G. Oppenheim A. Antoniadis, editeur, *Wavelets and Statistics*, pages 125–150. Springer-Verlag, 1995. (cité en pages 11, 14, 16 et 83)
- [Combettes & Pesquet 2009] Patrick L. Combettes et Jean-Christophe Pesquet. Proximal Splitting Methods in Signal Processing. *Recherche*, vol. 49, no. 8, pages 1–25, 2009. (cité en pages 20 et 22)

- [Combettes & Wajs 2005] Patrick L. Combettes et Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling Simulation*, vol. 4, no. 4, pages 1168–1200, 2005. (cité en page 20)
- [Comon 1994] Pierre Comon. Independent Component Analysis, a New Concept? *Signal Processing*, vol. 36, no. 3, pages 287–314, Avril 1994. (cité en page 30)
- [Davies 2004] Mike Davies. Identifiability Issues in Noisy ICA. *IEEE Signal Processing Letters*, vol. 11, no. 5, pages 470–473, 2004. (cité en page 30)
- [DeVore 1998] Ronald A. DeVore. Nonlinear Approximation. *Acta Numerica*, vol. 7, pages 51–150, 1998. (cité en page 5)
- [Dhillon & Sra 2005] Inderjit S. Dhillon et Suvrit Sra. Generalized Nonnegative Matrix Approximations with Bregman Divergences. In *Proceedings of NIPS*, 2005. (cité en pages 34, 37 et 44)
- [Donoho & Huo 2006] David L. Donoho et Xiaoming Huo. Uncertainty Principles and Ideal Atomic Decomposition. *IEEE Transactions on Information Theory*, vol. 47, no. 7, pages 2845–2862, Septembre 2006. (cité en page 11)
- [Donoho & Johnstone 1994] David L. Donoho et Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, vol. 81, pages 425–455, 1994. (cité en page 14)
- [Donoho & Stodden 2003] David Donoho et Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts? *Earth*, vol. 16, pages 1141–1148, 2003. (cité en pages 31, 32 et 39)
- [Donoho *et al.* 1998] David L. Donoho, Martin Vetterli, Ronald A. DeVore et Ingrid Daubechies. Data Compression and Harmonic Analysis. *IEEE Transactions on Information Theory*, vol. 44, no. 6, pages 2435–2476, 1998. (cité en page 4)
- [Donoho 2006] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, vol. 52, no. 4, pages 1289–1306, 2006. (cité en page 13)
- [Douglas & Rachford 1956] Jim Douglas et Henry H. Rachford. On the numerical solution of heat conduction problems in two or three space variables. *Transaction of the American Mathematical Society*, vol. 82, pages 421–439, 1956. (cité en page 22)
- [Du *et al.* 2008] Peicheng Du, Gustavo Stolovitzky, Peter Horvatovich, Rainer Bischoff, Jihyeon Lim et Frank Suits. A noise model for mass spectrometry based proteomics. *Bioinformatics*, vol. 24, no. 8, pages 1070–1077, 2008. (cité en page 104)
- [Dubroca *et al.* 2012] Rémi Dubroca, Christophe Junot et Antoine Souloumiac. Weighted NMF for high-resolution mass spectrometry analysis. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1806–1810, Août 2012. (cité en pages 32, 106 et 108)
- [Eggert & Körner 2004] Julian Eggert et Edgar Körner. Sparse coding and NMF. *Proceedings of IJCNN*, vol. 4, pages 2529–2533, 2004. (cité en page 40)

- [Elad *et al.* 2007] Michael Elad, Peyman Milanfar et Ron Rubinstein. Analysis Versus Synthesis in Signal Priors. *Inverse Problems*, vol. 23, no. 3, pages 947–968, Juin 2007. (cité en pages 16, 88 et 90)
- [Eriksson & Koivunen 2004] J. Eriksson et V. Koivunen. Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Processing Letters*, vol. 11, no. 7, pages 601–604, July 2004. (cité en page 30)
- [Essid & Fevotte 2013] Slim Essid et Cédric Févotte. Smooth Nonnegative Matrix Factorization for Unsupervised Audiovisual Document Structuring. *IEEE Transactions on Multimedia*, vol. 15, no. 2, pages 415–425, Février 2013. (cité en pages 37, 40 et 44)
- [Févotte *et al.* 2009] Cédric Févotte, Nancy Bertin et Jean-Louis Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence : With application to music analysis. *Neural Computation*, vol. 21, no. 3, pages 793–830, 2009. (cité en pages 32, 34 et 44)
- [Févotte 2011] Cédric Févotte. Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization. In *Proceedings of ICASSP*, pages 1980–1983, 2011. (cité en pages 37 et 40)
- [Gao & Bruce 1997] Hong-Ye Gao et Andrew G. Bruce. Wave shrink with firm shrinkage. *Statistica Sinica*, vol. 7, pages 855–874, 1997. (cité en page 92)
- [Georgiev *et al.* 2005] Pando Georgiev, Fabian Theis et Andrzej Cichocki. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pages 992–996, July 2005. (cité en page 31)
- [Georgiev *et al.* 2007] Pando Georgiev, Fabian Theis, Andrzej Cichocki et Hovagim Bakardjian. Sparse Component Analysis : a New Tool for Data Mining. In PanosM. Pardalos, VladimirL. Boginski et Alkis Vazacopoulos, éditeurs, *Data Mining in Biomedicine*, volume 7 of Springer Optimization and Its Applications, pages 91–116. Springer, 2007. (cité en page 31)
- [Gillis & Glineur 2012] Nicolas Gillis et François Glineur. Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization. *Neural Computation*, vol. 24, no. 4, pages 1085–1105, 2012. (cité en pages 41 et 60)
- [Gillis & Vavasis 2014] Nicolas Gillis et Stephen A. Vavasis. Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pages 698–714, Apr 2014. (cité en pages 32 et 114)
- [Gillis 2012] Nicolas Gillis. Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing. *Journal of Machine Learning Research*, vol. 13, pages 3349–3386, 2012. (cité en pages 41, 65 et 85)
- [Golbabaee *et al.* 2013] Mohammad Golbabaee, Simon Arberet et Pierre Vandergheynst. Compressive Source Separation : Theory and Methods for

- Hyperspectral Imaging. IEEE Transactions on Image Processing, vol. 22, no. 12, pages 5096–5110, Dec 2013. (cité en page 16)
- [Gonzalez & Zhang 2005] Edward F. Gonzalez et Yin Zhang. Accelerating the Lee-Seung Algorithm for Nonnegative Matrix Factorization. Rapport technique TR05-02, Dept. Comput. Appl. Math., Rice University, 2005. (cité en page 36)
- [Guan *et al.* 2012] Naiyang Guan, Dacheng Tao, Zhigang Luo et Bo Yuan. NeNMF : An Optimal Gradient Method for Nonnegative Matrix Factorization. IEEE Transactions on Signal Processing, vol. 60, no. 6, pages 2882–2898, 2012. (cité en pages 38, 41 et 44)
- [Hale *et al.* 2007] Elaine Hale, Wotao Yin et Yin Zhang. A Fixed-Point Continuation Method for l_1 -Regularized Minimization with Applications to Compressed Sensing. Rice University CAAM Technical Report, vol. TR07-07, 2007. (cité en page 53)
- [Ho 2008] Ngoc-Diep Ho. Non-negative matrix Factorization Algorithms and Applications. PhD thesis, Université Catholique de Louvain, 2008. (cité en page 29)
- [Hoyer 2002] Patrik O. Hoyer. Non-negative sparse coding. In Proceedings of Neural Networks for Signal Processing, pages 557–565, 2002. (cité en pages 34, 40, 41 et 44)
- [Hoyer 2004] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research, vol. 5, no. 5, pages 1457–1469, 2004. (cité en pages 41, 44 et 65)
- [Hyvärinen *et al.* 2001] Aapo Hyvärinen, Juha Karhunen et Erkki Oja. Independent component analysis. Adaptive and learning systems for signal processing, communications, and control. John Wiley & Sons, inc., 2001. (cité en page 30)
- [Hyvärinen 1999] Aapo Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks, vol. 10, no. 3, pages 626–634, 1999. (cité en page 30)
- [Jia & Qian 2009] Sen Jia et Yuntao Qian. Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing. IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 1, pages 161–173, 2009. (cité en page 32)
- [Jiang & Yin 2012] Lingling Jiang et Haiqing Yin. Bregman iteration algorithm for sparse nonnegative matrix factorizations via alternating L_1 -norm minimization. Multidimensional Systems and Signal Processing, vol. 23, pages 315–328, 2012. (cité en pages 43, 44 et 100)
- [Jutten & Herault 1991] Christian Jutten et Jeanny Herault. Blind separation of sources, part I : An adaptive algorithm based on neuromimetic architecture. Signal Processing, vol. 24, no. 1, pages 1–10, Juillet 1991. (cité en page 30)

- [Kim & Park 2008] Jingu Kim et Haesun Park. Sparse Nonnegative Matrix Factorization for Clustering. Rapport technique GT-CSE-08-01, Georgia Institute of Technology, 2008. (cité en pages 32, 40, 42, 44 et 65)
- [Lee & Seung 1999] Daniel D. Lee et H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, vol. 401, no. 6755, pages 788–791, 1999. (cité en pages 31, 34, 36 et 108)
- [Lee & Seung 2001] Daniel D. Lee et H. Sebastian Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, vol. 13, no. 1, pages 556–562, 2001. (cité en pages 34, 36, 37, 44, 60 et 113)
- [Lee *et al.* 2000] Te-Won Lee, Mark Girolami, Anthony J. Bell et Terrence J. Sejnowski. A unifying information-theoretic framework for independent component analysis. *Computers & Mathematics with Applications*, vol. 39, no. 11, pages 1–21, 2000. (cité en page 30)
- [Li *et al.* 2003] Yuanqing Li, Andrzej Cichocki, Shun ichi Amari, Sergei L. Shishkin, Jianting Cao et Fanji Gu. Sparse Representation and Its Applications in Blind Source Separation. In *Proceedings of NIPS*, pages 241–248, Vancouver, BC, Canada, Décembre 2003. (cité en pages 31 et 80)
- [Lin 2007a] Chih-Jen Lin. On the Convergence of Multiplicative Update Algorithms for Nonnegative Matrix Factorization. *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pages 1589–1596, Novembre 2007. (cité en page 36)
- [Lin 2007b] Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural Computation*, vol. 19, no. 10, pages 2756–2779, 2007. (cité en pages 38 et 44)
- [Liu & Aeron 2013] Jason Gejie Liu et Shuchin Aeron. Robust large-scale non-negative matrix factorization using Proximal Point algorithm. In *Proceedings of GlobalSIP*, pages 1127–1130, Dec 2013. (cité en page 32)
- [Ma *et al.* 2014] Wing-Kin Ma, José M. Bioucas-Dias, Tsung-Han Chan, Nicolas Gillis, Paul Gader, Antonio J. Plaza, ArulMurugan Ambikapathi et Chong-Yung Chi. A Signal Processing Perspective on Hyperspectral Unmixing : Insights from Remote Sensing. *IEEE Signal Processing Magazine*, vol. 31, no. 1, pages 67–81, Jan 2014. (cité en page 31)
- [Mailhe *et al.* 2009] Boris Mailhe, Rémi Gribonval, Frédéric Bimbot, Matthieu Lemaire, Pierre Vandergheynst et Jean-Marc Vesin. Dictionary learning for the sparse modelling of atrial fibrillation in ECG signals. In *Proceedings of ICASSP*, pages 465–468, Avril 2009. (cité en page 17)
- [Makarov *et al.* 2006] Alexander Makarov, Eduard Denisov, Alexander Kholomeev, Wilko Balschun, Oliver Lange, Kerstin Strupat et Stevan Horning. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Analytical Chemistry*, vol. 78, no. 7, pages 2113–2120, 2006. (cité en page 103)

- [Makarov 2000] A. Makarov. Electrostatic axially harmonic orbital trapping : a high-performance technique of mass analysis. *Analytical Chemistry*, vol. 72, no. 6, pages 1156–62, 2000. (cité en page 103)
- [Mallat & Zhang 1993] S. Mallat et Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pages 3397–3415, Décembre 1993. (cité en page 18)
- [Mallat 2008] Stéphane Mallat. A Wavelet Tour of Signal Processing : The Sparse Way. Academic Press, 3^{ième} édition, Décembre 2008. (cité en pages 8 et 9)
- [Merritt & Zhang 2005] Michael Merritt et Yin Zhang. Interior-Point Gradient Method for Large-Scale Totally Nonnegative Least Squares Problems. *Journal of Optimization Theory and Applications*, vol. 126, no. 1, pages 191–202, 2005. (cité en pages 34, 37 et 44)
- [Moreau 1962] Jean-Jacques Moreau. Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires. *Comptes Rendus de l'Académie des Sciences*, vol. 255, pages 238–240, 1962. (cité en pages 18 et 23)
- [Moudden & Bobin 2011] Yassir Moudden et Jérôme Bobin. Hyperspectral BSS Using GMCA With Spatio-Spectral Sparsity Constraints. *IEEE Transactions on Image Processing*, vol. 20, no. 3, pages 872–879, 2011. (cité en pages 31 et 123)
- [Moussaoui *et al.* 2006] Saïd Moussaoui, David Brie, Ali Mohammad-Djafari et Cédric Carteret. Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling. *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pages 4133–4145, Nov 2006. (cité en page 32)
- [Moussaoui *et al.* 2008] Saïd Moussaoui, Hafrun Hauksdóttir, Frédéric Schmidt, Christian Jutten, Jocelyn Chanussot, David Brie, Sylvain Douté et Jon Atli Benediktsson. On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation. *Neurocomputing*, vol. 71, no. 10-12, pages 2194–2208, 2008. (cité en pages 30 et 31)
- [Nam *et al.* 2013] Sangnam Nam, Mike E. Davies, Michael Elad et Rémi Gribonval. The Cosparsity Analysis Model and Algorithms. *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pages 30–56, 2013. (cité en page 16)
- [Nesterov 2004] Yurii Nesterov. Introductory Lectures on Convex Optimization : A Basic Course. Applied Optimization. Kluwer Academic Publishers, 2004. (cité en page 19)
- [Nesterov 2013] Yurii Nesterov. Gradient methods for minimizing composite objective function. *Mathematical Programming*, vol. 140, no. 1, pages 125–161, Août 2013. (cité en pages 21 et 38)
- [Ochs *et al.* 2014] Peter Ochs, Yunjin Chen, Thomas Brox et Thomas Pock. iPiano : Inertial Proximal Algorithm for Non-convex Optimization. *SIAM Journal on Imaging Sciences*, vol. 7, pages 1388–1419, 2014. (cité en page 76)

- [Olshausen & Field 1996] B. Olshausen et D. Field. Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature*, vol. 381, pages 607–609, 1996. (cité en page 17)
- [Ouedraogo *et al.* 2012] Wendyam S.B. Ouedraogo, Antoine Souloumiac, Meriem Jaïdane et Christian Jutten. Geometrical method using simplicial cones for overdetermined nonnegative blind source separation : application to real PET images. In *Proceedings of LVA/ICA*, pages 494–501, 2012. (cité en pages 32 et 114)
- [Ouedraogo 2012] Wendyam Serge Boris Ouedraogo. Méthode géométrique de séparation de sources non-négatives : Applications à l'imagerie dynamique TEP et à la spectrométrie de masse. PhD thesis, Université de Grenoble et Université de Tunis El Manar, 2012. (cité en page 32)
- [Paatero & Tapper 1994] Pentti Paatero et Unto Tapper. Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, vol. 5, no. 2, pages 111–126, 1994. (cité en pages 31, 34, 35, 44, 57, 60 et 108)
- [Peyré & Fadili 2011] Gabriel Peyré et Jalal Fadili. Learning Analysis Sparsity Priors. In *Proceedings of SampTA*, 2011. (cité en page 17)
- [Plumbley & Oja 2004] Mark D. Plumbley et Erkki Oja. A "Nonnegative PCA" Algorithm for Independent Component Analysis. *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pages 66–76, 2004. (cité en page 32)
- [Pluskal *et al.* 2010] Tomáš Pluskal, Sandra Castillo, Alejandro Villar-Briones et Matej Oresic. MZmine 2 : Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, vol. 11 :395, pages 1–11, 2010. (cité en page 106)
- [Pock & Chambolle 2011] Thomas Pock et Antonin Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *Proceedings of ICCV*, pages 1762–1769, 2011. (cité en page 109)
- [Raguet *et al.* 2013] Hugo Raguet, Jalal Fadili et Gabriel Peyré. A Generalized Forward-Backward Splitting. *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pages 1199–1226, 2013. (cité en pages 22, 110 et 133)
- [Raj *et al.* 2010] Bhiksha Raj, Kevin W. Wilson, Alexander Krueger et Reinhold Haeb-Umbach. Ungrounded independent non-negative factor analysis. In *INTERSPEECH*, pages 330–333, 2010. (cité en page 32)
- [Rapin *et al.* 2013a] Jérémy Rapin, Jérôme Bobin, Anthony Larue et Jean-Luc Starck. Sparse and Non-Negative BSS for Noisy Data. *IEEE Transactions on Signal Processing*, vol. 61, pages 5620–5632, 2013. (cité en pages 43, 44 et 127)
- [Rapin *et al.* 2013b] Jérémy Rapin, Jérôme Bobin, Anthony Larue et Jean-Luc Starck. Sparse Redundant Formulations and Non-negativity in Blind Source Separation. In *Proceedings of EUSIPCO*, Marrakech, Morocco, Septembre 2013. (cité en pages 83 et 110)

- [Rapin *et al.* 2013c] Jérémy Rapin, Jérôme Bobin, Anthony Larue et Jean-Luc Starck. Sparse Regularizations and Non-negativity in BSS. In Proceedings of SPARS, Lausanne, Switzerland, Juillet 2013. (cité en pages 83 et 110)
- [Rapin *et al.* 2014] Jérémy Rapin, Jérôme Bobin, Anthony Larue et Jean-Luc Starck. NMF with Sparse Regularizations in Transformed Domains. SIAM Journal on Imaging Sciences (accepted), vol. 7, no. 4, pages 2020–2047, 2014. SIAM J. Imaging Sci., 7(4), 2020–2047. (28 pages). (cité en pages 44, 45 et 127)
- [Recht *et al.* 2012] Ben Recht, Christopher Ré, Joel A. Tropp et Victor Bittorf. Factoring nonnegative matrices with linear programs. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou et Kilian Q. Weinberger, éditeurs, Proceedings of NIPS, pages 1223–1231, 2012. (cité en page 32)
- [Rivet *et al.* 2007] Bertrand Rivet, Laurent Girin et Christian Jutten. Visual voice activity detection as a help for speech source separation from convolutive mixtures. Speech Communication, vol. 49, no. 7-8, pages 667–677, 2007. (cité en page 31)
- [Rubinstein *et al.* 2013] Ron Rubinstein, Tomer Peleg et Michael Elad. Analysis K-SVD : A Dictionary-Learning Algorithm for the Analysis Sparse Model. IEEE Transactions on Signal Processing, vol. 61, no. 3, pages 661–677, 2013. (cité en page 17)
- [Sardy *et al.* 2000] Sylvain Sardy, Andrew G. Bruce et Paul Tseng. Block coordinate relaxation methods for nonparametric signal denoising with wavelet dictionaries. Journal of Computational and Graphical Statistics, vol. 9, pages 361–379, 2000. (cité en page 34)
- [Schmidt *et al.* 2009] Mikkel N. Schmidt, Ole Winther et Lars Kai Hansen. Bayesian Non-negative Matrix Factorization. In Proceedings of ICA, volume 5441, pages 540–547. Springer Berlin Heidelberg, 2009. (cité en page 32)
- [Schobben *et al.* 1999] Daniel Schobben, Kari Torkkola et Paris Smaragdis. Evaluation Of Blind Signal Separation Methods. In Proceedings of ICA, pages 261–266, january 1999. (cité en page 45)
- [Selesnick & Figueiredo 2009] Ivan W. Selesnick et Mario A. T. Figueiredo. Signal Restoration with Overcomplete Wavelet Transforms : Comparison of Analysis and Synthesis Priors. In Proceedings of SPIE, volume 7446 (Wavelets XIII), Août 2009. (cité en pages 16, 88 et 90)
- [Snyder *et al.* 2008] David A. Snyder, Fengli Zhang, Steven L. Robinette, Lei Bruschweiler-Li et Rafael Brüschweiler. Non-negative matrix factorization of two-dimensional NMR spectra : Application to complex mixture analysis. The Journal of Chemical Physics, vol. 128, no. 5, pages 1–4, 2008. (cité en page 31)
- [Starck *et al.* 2004] Jean-Luc Starck, Michael Elad et David L. Donoho. Redundant Multiscale Transforms and their Application for Morphological Component

- Analysis. *Advances in Imaging and Electron Physics*, vol. 132, pages 287–348, 2004. (cité en page 16)
- [Starck *et al.* 2010] Jean-Luc Starck, Fionn Murtagh et Jalal Fadili. Sparse Image and Signal Processing - Wavelets, Curvelets, Morphological Diversity. Cambridge University Press, 2010. (cité en pages 12 et 13)
- [Takigawa *et al.* 2004] Ichigaku Takigawa, Mineichi Kudo et Jun Toyama. Performance analysis of minimum L1-norm solutions for underdetermined source separation. *IEEE Transactions on Signal Processing*, vol. 52, no. 3, pages 582–591, Mars 2004. (cité en page 16)
- [Tandon & Sra 2010] Rashish Tandon et Suvrit Sra. Sparse nonnegative matrix approximation : new formulations and algorithms. Rapport technique 193, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, 9 2010. (cité en pages 38, 42 et 44)
- [Tibshirani 1994] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B*, vol. 58, pages 267–288, 1994. (cité en page 13)
- [Toumi *et al.* 2013] Ichrak Toumi, Bruno Torrèsani et Stefano Caldarelli. Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation. *Analytical Chemistry*, vol. 85, no. 23, pages 11344–11351, 2013. (cité en pages 31, 32 et 71)
- [Tseng 2001] P Tseng. Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization. *Optimization*, vol. 109, no. 3, pages 475–494, 2001. (cité en page 58)
- [Vavasis 2009] Stephen A. Vavasis. On the Complexity of Nonnegative Matrix Factorization. *SIAM Journal on Optimization*, vol. 20, no. 3, pages 1364–1377, 2009. (cité en pages 32 et 39)
- [Vincent *et al.* 2006] Emmanuel Vincent, Rémi Gribonval et Cédric Févotte. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech & Language Processing*, vol. 14, no. 4, pages 1462–1469, 2006. (cité en pages 45, 70, 85, 92 et 93)
- [Virtanen 2007] Tuomas Virtanen. Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria. *IEEE Transactions on Audio, Speech & Language Processing*, vol. 15, no. 3, pages 1066–1074, 2007. (cité en pages 34, 40 et 44)
- [Voronin & Woerdeman 2013] Sergey Voronin et Hugo J. Woerdeman. A new iterative firm-thresholding algorithm for inverse problems with sparsity constraints. *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pages 151–164, 2013. (cité en page 92)
- [Wilson *et al.* 2008] Kevin W. Wilson, Bhiksha Raj et Paris Smaragdis. Regularized non-negative matrix factorization with temporal dependencies for speech denoising. In *Proceedings of INTERSPEECH*, pages 411–414, 2008. (cité en page 40)

- [Xu & Yin 2013] Yangyang Xu et Wotao Yin. A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion. *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pages 1758–1789, 2013. (cité en pages 76 et 77)
- [Xu *et al.* 2010] Ying Xu, Jean-François Heilier, Geoffrey Madalinski, Eric Genin, Eric Ezan, Jean-Claude Tabet et Christophe Junot. Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-orbitrap mass spectrometer for further metabolomics database building. *Analytical Chemistry*, vol. 82, no. 13, pages 5490–5501, 2010. (cité en page 105)
- [Xu 2011] Yangyang Xu. Alternating proximal gradient method for nonnegative matrix factorization. arXiv :1112.5407v1, Décembre 2011. (cité en pages 37 et 44)
- [Zdunek & Cichocki 2007a] Rafal Zdunek et Andrzej Cichocki. Blind Image Separation Using Nonnegative Matrix Factorization with Gibbs Smoothing. In *Proceedings of ICONIP*, pages 519–528, 2007. (cité en pages 39, 44, 79 et 85)
- [Zdunek & Cichocki 2007b] Rafal Zdunek et Andrzej Cichocki. Gibbs Regularized Nonnegative Matrix Factorization for Blind Separation of Locally Smooth Signals. In *Proceedings of NDES*, pages 317–320, 2007. (cité en page 39)
- [Zdunek & Cichocki 2007c] Rafal Zdunek et Andrzej Cichocki. Nonnegative matrix factorization with constrained second-order optimization. *Signal Processing*, vol. 87, no. 8, pages 1904–1916, 2007. (cité en pages 34, 37, 42, 44 et 52)
- [Zibulevsky & Pearlmutter 2001] Michael Zibulevsky et Barak A. Pearlmutter. Blind Source Separation by Sparse Decomposition in a Signal Dictionary. *Neural Computation*, vol. 13, no. 4, pages 863–882, 2001. (cité en pages 31, 45, 52, 58 et 80)