Separation of parameterized and delayed sources : application to spectroscopic and multispectral data

Hassan Mortada

ÉCOLE DOCTORALE MATHÉMATIQUES, SCIENCES DE L'INFORMATION ET DE L'INGÉNIEUR

UNIVERSITÉ DE STRASBOURG - ICUBE

THESIS

by

Hassan Mortada

Submitted in partial fulfillment of requirements for the degree of
Doctor of Philosophy

at
Université de Strasbourg

Speciality: signal and image processing

SEPARATION OF PARAMETERIZED AND DELAYED SOURCES: APPLICATION TO SPECTROSCOPIC AND MULTISPECTRAL DATA

December 13, 2018

Pierre-Olivier Amblard	Research director, GIPSA-lab, CNRS	Examiner
Hervé Carfantan	Associate professor, IRAP, Université de Toulouse	Reviewer
Christophe Collet	Professor, ICube, Université de Strasbourg	PhD director
Jean-François Giovannelli	Professor, IMS, Université de Bordeaux	Reviewer
Vincent Mazet	Associate professor, ICube, Université de Strasbourg	PhD supervisor
Charles Soussen	Professor, L2S, CentraleSupélec	PhD co-director

Contents

List of Figures

Acronyms

ALS	**A**lternating **L**east **S**quares
ICM	**I**terated **C**onditional **M**odes
MCMC	**M**onte **C**arlo **M**arkov **C**hain
MMV	**M**ultiple **M**easurment **V**ectors
MSE	**M**ean **S**quared **E**rror
OMP	**O**rthogonal **M**atching **P**ursuit
RJMCMC	**R**eversible **J**ump **M**onte **C**arlo **M**arkov **C**hain
SNR	**S**ignal to **N**oise **R**atio

Symbols

$\widehat{\cdot}$	estimated parameter
$\cdot^*$	ground-truth parameter
$*$	convolution product
λ	wavelength
i	mixture index
j	source index
I	number of mixtures
J	number of sources
a_{ij}	amplitude of source j in mixture i
c_{ij}	delay of source j in mixture i
w_{ij}	shape of source j in mixture i
$x_i(\lambda)$	mixture i sample at wavelength λ
$s(\lambda - c_{ij}; w_{ij})$	source j delayed by c_{ij} and parameterized by w_{ij}
$n_i(\lambda)$	noise sample in mixture i at wavelength λ
σ_n^2	noise variance
N	mixture length
$\boldsymbol{A} \in \mathbb{R}^{I \times J}$	matrix gathering the amplitudes a_{ij}
$\boldsymbol{C} \in \mathbb{R}^{I \times J}$	matrix gathering the delays c_{ij}
$\boldsymbol{W} \in \mathbb{R}^{I \times J}$	matrix gathering the shapes w_{ij}
$\boldsymbol{x}_i \in \mathbb{R}^N$	vector gathering the mixture samples $x(\lambda; w_{ij})$
$\boldsymbol{n}_i \in \mathbb{R}^N$	vector gathering the noise samples $n_i(\lambda)$
$\boldsymbol{s}\big[c_{ij}; w_{ij}\big] \in \mathbb{R}^N$	vector gathering the source samples $s(\lambda - c_{ij}; w_{ij})$
$\left\| \cdot \right\|_p$	ℓ_p norm
$\left\| \cdot \right\|_0$	ℓ_0 pseudo-norm
$\big(\cdot \big)_+$	$\max(\cdot, 0)$
$\boldsymbol{I}$	identity matrix
Δ	delay sampling step
ℓ_{ij}	delay index of source j in mixture i
p_{ij}	shape index of source j in mixture i
$\mathcal{G}$	set of cliques
τ	delay regularization parameter
μ	shape regularization parameter
$\boldsymbol{S}$	block dictionary
$\boldsymbol{S}[j]$	the jth sub-block dictionary
$\boldsymbol{\alpha}_{ij}$	the sub-sparse vector corresponding to $\boldsymbol{S}[j]$

$\mathcal{S}$	sparse vector support
$\mathcal{I}_j$	subset of mixture indices where source j is present
$\boldsymbol{\alpha}$	sparse vector
ρ	ALS stopping parameter
ς	threshold of the OMP-like implementation
κ	threshold of the joint OMP-like implementation
$\mathcal{J}$	set of selected sources
$\mathcal{T}, \mathcal{B}$	sets used for 2D ICM sweeping
d	B-spline degree
$\boldsymbol{k}$	knot vector
M	number of B-spline bases functions
$b_m^d(\cdot)$	the mth B-spline basis function of degree d
$\boldsymbol{b}[\cdot]$	vector gathering the B-spline bases functions
$\beta_m^d(\cdot)$	bivariate B-spline basis function of degree d
$\boldsymbol{\beta}[\cdot]$	vector gathering the bivariate B-spline bases function
ϕ_m^j	the mth amplitude control point of source j
σ_m^j	the mth delay control point of source j
ω_m^j	the mth shape control point of source j
$\boldsymbol{\Phi} \in \mathbb{R}^{M \times J}$	matrix gathering the amplitude control points ϕ_m^j
$\boldsymbol{\Sigma} \in \mathbb{R}^{M \times J}$	matrix gathering the delay control points σ_m^j
$\boldsymbol{\Omega} \in \mathbb{R}^{M \times J}$	matrix gathering the shape control points ω_m^j
$\lambda_{\max}$	the wavelength with the highest energy in the mixture sum
$w_{\min}$	shape parameter lower bound
$w_{\max}$	shape parameter upper bound

Résumé étendu

Contexte et objectifs

Cette thèse fait partie du projet DSIM (Décomposition Spectroscopique en Imagerie Multispectrale, dsim.unistra.fr) financé par l'ANR (Agence Nationale de la Recherche). Ce projet vise à développer des méthodes numériques pour analyser plusieurs signaux spectroscopiques (spectres) obtenus à partir d'un même phénomène physique.

Le sujet de cette thèse est motivé par deux applications. La première est la spectroscopie photoélectronique à résolution temporelle, qui est un outil expérimental pour obtenir la configuration électronique des atomes ou des molécules [SBN04]. Les données consistent en un ensemble de spectres acquis à différents instants pendant une expérience, donnant ainsi une séquence temporelle de spectres. Chaque spectre contient plusieurs raies à partir desquelles la distribution d'électrons par rapport aux niveaux d'énergie peut être déterminée. L'évolution de ces raies dans la séquence temporelle des spectres renseigne sur la dynamique de relaxation de la distribution des électrons. La deuxième application traite de la cinématique des galaxies et de l'étude du mouvement des gaz galactiques [MB81]. Les données sont des images multispectrales où les raies observées dans les pixels sont décalées spectralement. Ceci est une conséquence de l'effet Doppler provoqué par le mouvement du gaz.

Les deux applications considérées proviennent de phénomènes très différents et reposent sur diverses données (une séquence temporelle de spectres pour la spectroscopie de photoélectrons à résolution temporelle et une image multispectrale pour la cinématique des galaxies). Cependant, les deux données présentent un ensemble de spectres ordonnés où les raies portent les informations d'intérêts. De plus, deux spectres voisins sont très similaires car ils sont acquis à des instants très proches où en des positions spatiales contiguës. Par conséquent, les raies montrent un *évolution lente* (temporellement ou spatialement) dans les spectres.

Cette thèse vise à atteindre deux objectifs:

- estimer les caractéristiques de chaque pic dans les données, telles que l'intensité, la position spectrale ou la largeur;
- associer les raies de différents spectres à mesure qu'ils évoluent lentement.

La Figure 1.7 illustre ces deux objectifs dans le cas d'une séquence temporelle de spectres et dans le cas d'une image multispectrale.

Extraire des informations d'intérêts (les caractéristiques des raies et leur évolution) à partir des mesures est un *problème inverse*. Un tel problème est souvent mal posé au sens de Hadamard [Had23], c'est-à-dire que la solution n'existe pas ou n'est pas unique ou n'est pas une fonction continue des

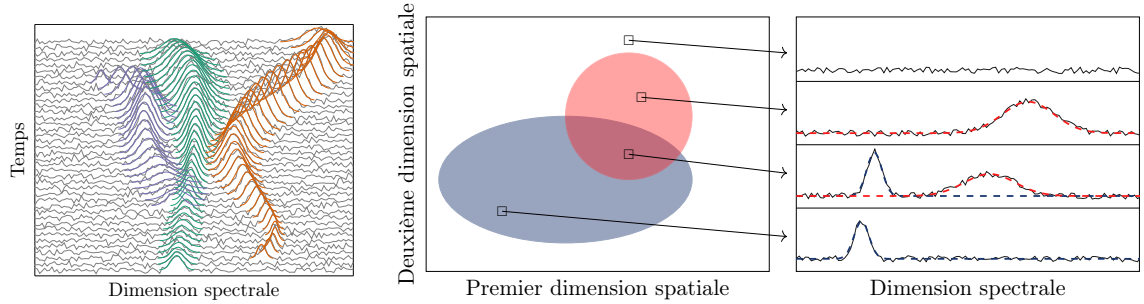


Figure 1: (a) Exemple synthétique d'une séquence temporelle de spectres avec trois groupes de raies en évolution lente indiquées par des couleurs différentes. (b) Illustration d'une image multispectrale avec deux groupes de raies qui se chevauchent spatialement. Des spectres provenant de différentes positions spatiales sont également affichés.

données. Pour obtenir une solution satisfaisante, il faut inclure des connaissances supplémentaires sur le problème. Par conséquent, les spectres sont modélisés comme une somme de fonctions paramétrées décrivant les raies, et la lente évolution de leurs caractéristiques peut être favorisée.

Plusieurs méthodes dans la littérature ont abordé ces objectifs (ou des objectifs similaires). Une première approche [MCV] consiste à estimer les caractéristiques des raies dans les spectres indépendamment des autres, puis à les associer avec un post-traitement. Cependant, une telle approche n'exploite pas les similitudes entre les spectres voisins lors de l'étape d'estimation et peut conduire à une propagation d'erreur. Ainsi, les méthodes qui traitent conjointement les deux objectifs constituent une meilleure solution. Plus précisément, un problème très similaire a été abordé dans le cadre bayésien [Maz+15]. Cette méthode est efficace mais souffre d'un temps de calcul élevé et ne peut être appliquée à des images multispectrales. Un autre cadre conjoint est la séparation des sources [CJ10] qui vise à extraire des sources (raies) de leurs mélanges (spectres). En particulier, la séparation de source retardée est capable de traiter les sources décalées. Les connaissances préalables sont introduites en supposant souvent que les sources sont non corrélées ou statistiquement indépendantes. Malheureusement, ces hypothèses ne sont pas valables dans les applications considérées.

Dans cette thèse, nous adoptons une approche conjointe en utilisant le cadre de séparation des sources. Nous exploitons explicitement la connaissance des sources paramétrées, ce qui nous permet de traiter des sources très corrélées. Les approches proposées peuvent être appliquées à des séquences temporelles de mélanges et d'images multispectrales, tout en étant à la fois efficaces et rapides.

Formulation de problème

Modèle de mélange

Considérons I mélanges. Chaque mélange $\boldsymbol{x}_i \in \mathbb{R}^N$ est modélisé comme étant la somme de J sources paramétriques $\boldsymbol{s}\left[c_{ij}; w_{ij}\right] \in \mathbb{R}^N$ plus un bruit additif $\boldsymbol{n}_i \in \mathbb{R}^N$:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij} \boldsymbol{s}\left[c_{ij}; w_{ij}\right] + \boldsymbol{n}_i. \tag{1}$$

ou $i \in \{1, \ldots, I\}$ est l'indice d'un mélange qui peut représenter un instant d'acquisition ou une position spatiale. Les paramètres a_{ij}, c_{ij} et w_{ij} représentent respectivement les amplitudes, retards (position spectrale) et paramètres de formes (largeur par exemple) de la source j dans le mélange i. Par ailleurs, les amplitudes sont supposées positives au fait des applications considérées.

Le problème de séparation de sources revient à estimer les paramètres a_{ij}, c_{ij} et w_{ij}. Trois méthodes ont été proposées. Les deux premières utilisent en particulier les approches d'approximation parcimonieuses. La deuxième méthode diffère de la première par la considération de l'évolution lente des retards à travers les mélanges en utilisant une régularisation. La troisième méthode considère que les paramètres sont modélisés par des fonctions B-splines afin de garantir leur évolutions lentes et de réduire le nombre d'inconnus.

Sources paramétriques

En spectroscopie, il est habituel de considérer les sources $\boldsymbol{s}\left[c; w\right]$ comme des fonctions paramétrées par w [Pav+08]. De nombreux facteurs conduisent à ce que l'on appelle l'élargissement des raies [Pea81] tels que l'effet Doppler et la pression. Par conséquent, les sources sont rarement modélisées par des impulsions de Dirac, mais plutôt par des fonctions en forme de cloche. De plus, les fonctions sont généralement symétriques par rapport à la position spectrale de leur maximum. Les fonctions gaussiennes et lorentziennes [Hol04; Maz05; VC14] sont très utilisées en spectroscopie et ont un seul paramètre de forme w caractérisant leur largeur. Les méthodes proposées dans cette thèse ne sont pas limitées à une fonction spécifique. Sans perte de généralité, les méthodes seront testées avec des fonctions gaussiennes.

Modèle du bruit

Dans cette thèse, on considère un bruit additif, gaussien, de moyenne zéro et de variance σ_n^2:

$$\forall i, \qquad \boldsymbol{n}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}), \tag{2}$$

où $\boldsymbol{I}$ est la matrice d'identité.

Méthodes proposées

Séparation de sources paramétriques et retardées

Dans la première méthode proposée nous faisons l'hypothèse que les retards sont discrétisés sur une grille avec un pas d'échantillonnage Δ. Les retards s'écrivent alors $c_{ij} = \ell_{ij}\Delta$ où $\ell_{ij} \in \mathbb{N}$ est l'indice d'échantillonnage. En plus on fait l'hypothèse que les paramètres de forme pour une source j ne varient pas dans les mélanges. Cela conduit au modèle suivant:

$$\forall\, i,\ \ \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; w_j\right] + \boldsymbol{n}_i. \tag{3}$$

La séparation de sources retardées et paramétriques est adressé comme un problème de moindres carrés. L'estimateur du maximum de vraisemblance est alors obtenu en minimisant le critère suivant:

$$\min_{\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}} E(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}) = \min_{\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}} \sum_{i=1}^{I} \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; w_j\right] \right\|_2^2 \tag{4}$$

où $\|.\|_2$ est la norme ℓ_2. Les matrices $\boldsymbol{A} \in \mathbb{R}_+^{I\times J}$ et $\boldsymbol{L} \in \mathbb{N}^{I\times J}$ et le vecteur $\boldsymbol{w} \in \mathbb{R}^J$ regroupent respectivement les amplitudes, retards et paramètres de forme.

On propose de résoudre ce problème complexe d'optimisation en utilisant un schéma de moindres carrés alterné en deux étapes. La première étape consiste à estimer les paramètres de formes $\boldsymbol{w}$: pour cela on utilise un solveur de moindres carrés non-linéaire (l'algorithme de Levenberg-Marquardt). La deuxième étape consiste à estimer les amplitudes et les retards. D'après (4), il apparaît que le critère E est séparable par rapport à $\boldsymbol{A}$ et $\boldsymbol{C}$, *i.e.*,

$$\min_{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{L}} E(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}) \Leftrightarrow \forall i, \qquad \min_{\boldsymbol{A}_{i:}\geq\boldsymbol{0},\boldsymbol{L}_{i:}} \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; w_j\right] \right\|_2^2. \tag{5}$$

On propose d'utiliser un algorithme d'approximation parcimonieuse glouton inspiré d'OMP [PRK93] pour estimer les amplitudes et les retards dans chaque mélange séparément. Cela revient à approximer chaque mélange comme le produit d'un dictionnaire avec un vecteur parcimonieux. Le dictionnaire est construit à partir de J blocs contenant toutes les versions retardées d'une source. Par ailleurs, le vecteur parcimonieux à estimer est contraint d'avoir au plus un élément non-nul pour chaque bloc comme illustre la Figure 2. Cette contrainte assure qu'une source peut apparaître au plus une fois dans un mélange et permet d'associer les sources dans les mélanges simultanément avec l'estimation de leurs paramètres.

Séparation de sources paramétriques avec évolution lente des retards

Dans la deuxième méthode proposée, on considère l'évolution lente des retards d'un mélange à un autre. L'ajout de cette connaissance est justifié par les applications considérées et permet de mieux séparer les sources surtout quand elles sont très corrélées. On propose donc d'ajouter une régularisation au terme d'attache aux données présentées dans la méthode précédente. Cette régularisation a comme objectif de mesurer les distances entre les retards des mélanges et ceux des

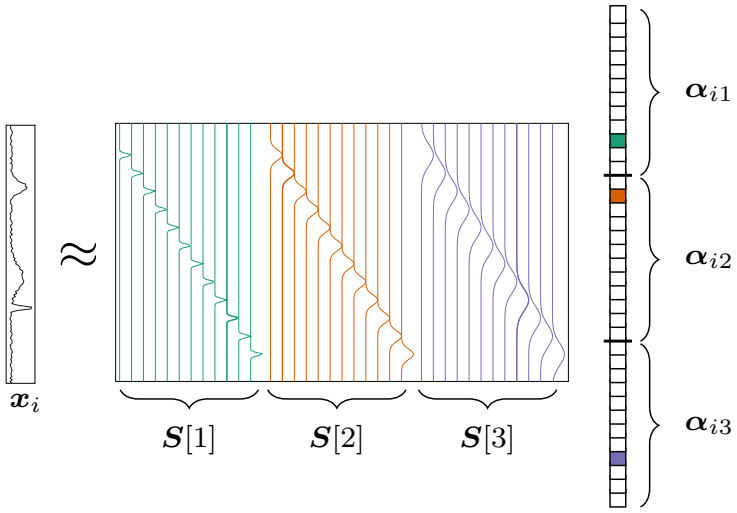


Figure 2: Un exemple de l'approximation parcimonieuse d'un mélange $\boldsymbol{x}_i$ avec $J = 3$ sources. Le support du vecteur parcimonieux (les indices des carrées colorées) indique les retards estimés, alors que les valeurs non-nulles correspondent aux amplitudes.

mélanges voisins. Ainsi, le critère composé s'écrit:

$$F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) = E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) + \tau\Delta^2 \sum_{j=1}^{J} \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}\right)^2, \tag{6}$$

où τ est le terme de régularisation et $\mathcal{G}$ contient tous les cliques d'ordre deux, c'est-à-dire toutes les couples i et i' de deux mélanges voisins.

De nouveau, un schéma ALS est proposé pour minimiser ce critère et estimer les amplitudes, les retards et les paramètres de formes. La première étape qui consiste à estimer $\boldsymbol{w}$ reste identique à la première méthode car le terme de régularisation ne dépend pas des paramètres de forme. Par contre, la deuxième étape est modifiée puisque le critère n'est plus séparable par rapport à $\boldsymbol{A}$ et $\boldsymbol{L}$. Un algorithme glouton conjoint est proposé (Figure 3). A chaque itération la contribution de la "meilleure" source est estimée dans tous les mélanges en utilisant une stratégie d'estimation séquentielle des retards et des amplitudes inspirée par l'algorithme ICM [Bes86]. La meilleure source est celle qui modélise au mieux les données (en minimisant le terme d'attache aux données) et avec l'évolution la plus lente possible des retard (en minimisant le terme de régularisation).

Modélisation des paramètres par des B-splines

Une troisième méthode proposée consiste à modéliser l'évolution lente des amplitudes, retards et paramètres de formes par des B-splines [DB72]. Les fonctions B-splines sont multipliées par des scalaires nommés poids. Évidemment, cette modélisation permet de prendre en compte l'évolution lente des paramètres car les B-splines sont obtenues par combinaison linéaire des fonctions polynomiales (on utilise des B-spline cubiques, donc les degrées des polynômes est trois) définis par un ensemble des nœuds. La Figure 4 présente un exemple génératif d'une séquence de mélanges en illustrant les fonctions B-spline pondérées. Les modélisations B-spline des amplitudes,

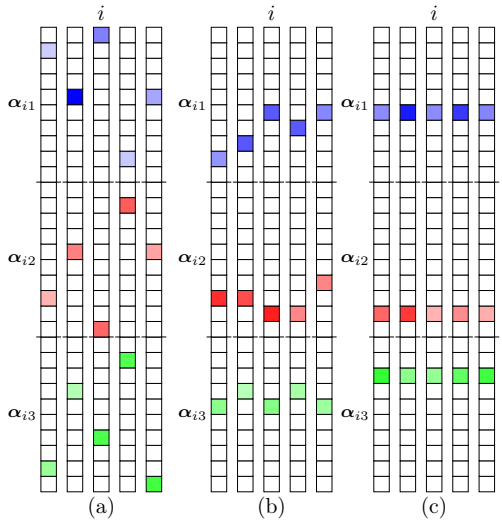


Figure 3: Exemple d'approximation parcimonieuse de $J = 3$ sources et $I = 6$ mélanges voisins. (a) Cas ou il n'y a pas de régularisation sur les retards ($\tau = 0$). (b) cas avec régularisation modérés. (c) cas avec régularisation très forte.

retards et paramètres de formes s'écrivent comme suit:

$$\forall\, j, \qquad a_{ij} = \sum_{m=1}^{M} \gamma_m^j b_m(i), \;\; c_{ij} = \sum_{m=1}^{M} \sigma_m^j b_m(i), \;\; w_{ij} = \sum_{m=1}^{M} \omega_m^j b_m(i) \tag{7}$$

où $b_m(i)$ est la $m^{\text{ème}}$ fonction B-spline de degré trois évalué en i. Les nouvelles inconnues sont γ_m^j,σ_m^j et ω_m^j et représentent respectivement les poids des B-splines pour les amplitudes, retards et paramètres de formes. Notons que cette modélisation réduit largement le nombre des inconnues à estimer ($M < I$). En remplaçant ces paramètres dans le modèle (1), on obtient:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} \boldsymbol{b}[i]^T \boldsymbol{\phi}_j \boldsymbol{s}\left[\boldsymbol{b}[i]^T \boldsymbol{\sigma}_j; \boldsymbol{b}[i]^T \boldsymbol{\omega}_j\right] + \boldsymbol{n}_i, \tag{8}$$

ou $\boldsymbol{b}[i] \in \mathbb{R}^M$ est un vecteur qui regroupe les fonctions B-splines $b_m(i)$. Les poids des B-splines pour les amplitudes(γ_m^j), retards (σ_m^j) et paramètres de formes (ω_m^j) pour la source j sont respectivement regroupés dans $\boldsymbol{\phi}_j$, $\boldsymbol{\sigma}_j$ et $\boldsymbol{\omega}_j$. Le critère à minimiser devient:

$$\mathcal{L}(\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} \boldsymbol{b}[i]^T \boldsymbol{\phi}_j \boldsymbol{s}\left[\boldsymbol{b}[i]^T \boldsymbol{\sigma}_j; \boldsymbol{b}[i]^T \boldsymbol{\omega}_j\right] \right\|_2^2. \tag{9}$$

Les matrices $\boldsymbol{\Phi} \in \mathbb{R}_+^{M \times J}$ et $\boldsymbol{\Sigma} \in \mathbb{R}^{M \times J}$ et $\boldsymbol{\Omega} \in \mathbb{R}^{M \times J}$ regroupent respectivement les poids des amplitudes, retards et paramètres de formes. Pour minimiser ce critère on utilise un algorithme de type SQP (*Sequential Quadratic Programming*) [NW99] qui permet de prendre en compte les contraintes (par exemple la positivité des amplitudes).

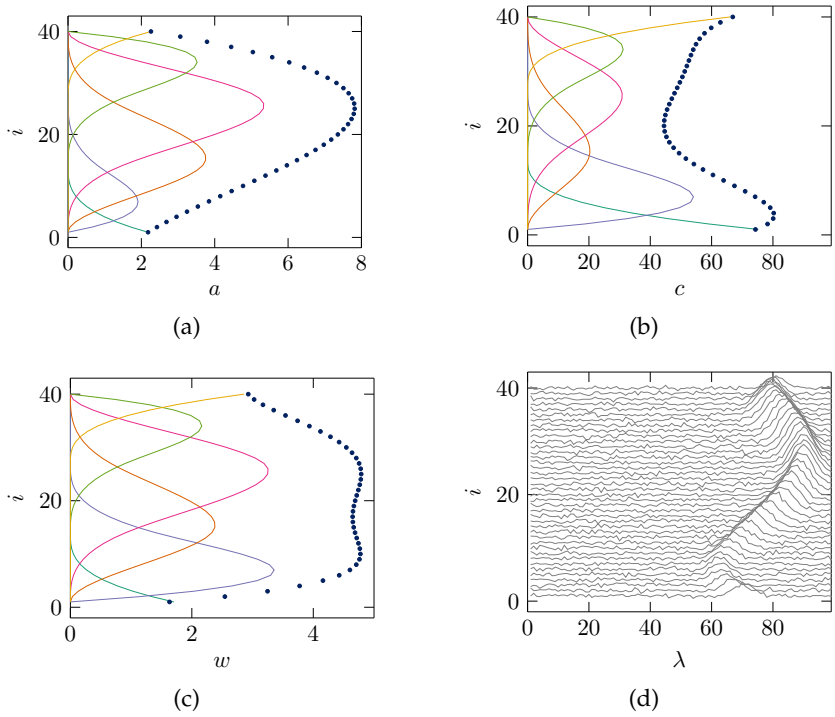


Figure 4: Une exemple générative.(a)–(c) Les amplitudes, retards et paramètres (les points bleus) d'une source on été généré avec des B-spline qui sont également affiché. (d) Les mélanges résultants.

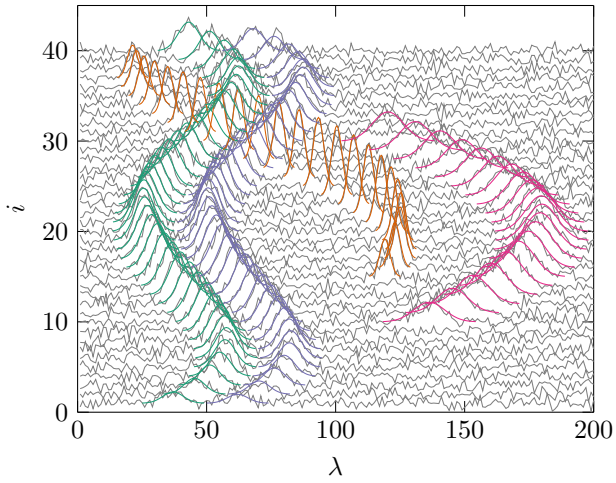


Figure 5: $I = 40$ mélanges avec $J = 4$ sources et un RSB = 15 dB. Chaque source est représentée par une couleur particulière.

Résultats

Dans cette partie nous présentons des résultats obtenues avec les trois méthodes proposées, notamment la méthode non-régularisée , la méthode régularisée et la méthode des B-splines . En outre, une comparaison est effectué avec la méthode de [Maz+15] qui est une méthode bayésienne adaptée au problème de décomposition des spectres qui tient compte de l'évolution lente des amplitudes, retards et largeurs. Les résultats sont obtenus sur une séquence temporelle de mélanges synthétiques, des données réelles de spectroscopie de photo-électrons et une image multispectrale de galaxie.

Données synthétiques

D'abord, on applique les méthodes proposées et la méthode de [Maz+15] sur une séquence synthétique de $I = 40$ mélanges avec $J = 3$ sources modélisées par des gaussiennes de largeur w. Les paramètres réels sont générés grâce à des polynôme de degré deux, trois ou quatre. la Figure 5 affiche les mélanges et les vérités terrains des sources. Les données présentent plusieurs difficultés. D'abord, deux sources partagent exactement la même largeur. Ensuite, les mélanges présentent plusieurs chevauchements spectraux entre les sources. Finalement, les sources n'apparaissent pas dans tous les mélanges.

Les reconstructions des mélanges et les paramètres estimés sont affichées Figure 6. Les résultats montrent que toutes les méthodes réussissent à reconstruire les sources correctement. Concernant les paramètres estimés, on peut remarquer que l'estimation des retards obtenus avec la méthode non-régularisée n'est pas satisfaisante surtout pour les deux sources avec largeurs identiques (la séparation devient ambigue dans ce cas). En revanche, les méthodes qui considèrent l'évolution lente des paramètres donnent des meilleurs résultats. L'avantage principal des méthodes proposées par rapport à la méthode de [Maz+15] c'est le temps de calcul; les méthodes non-régularisée et régularisée sont au moins vingt fois plus rapide, et la méthode des B-splines est quatre fois plus rapide.

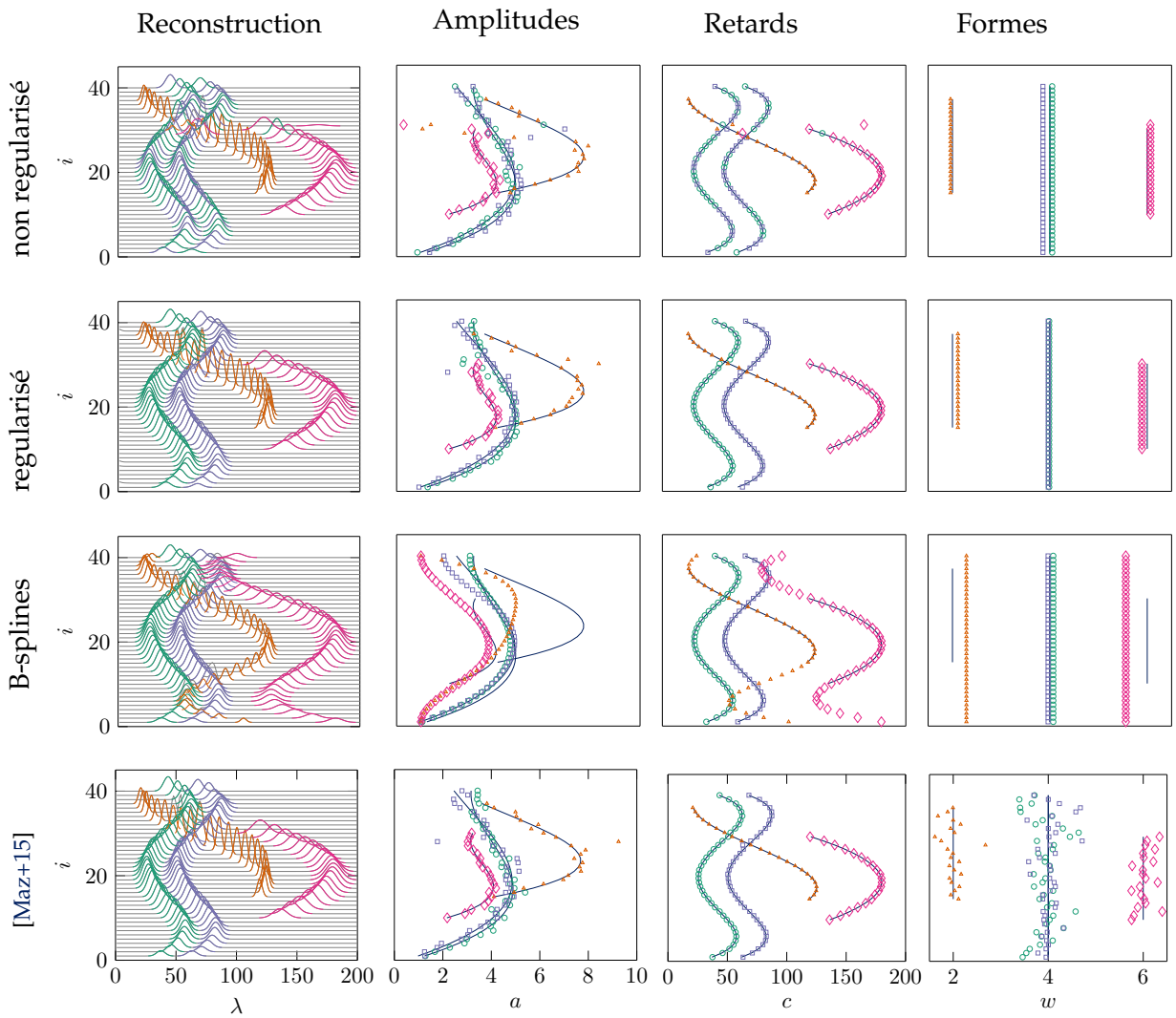


Figure 6: Reconstructions et sources estimées sur les données de Figure 5.

Photoélectrons

Les méthodes sont appliquées sur des données réelles provenant de la spectroscopie de photoélectron; il s'agit d'une séquence temporelle de $I = 44$ mélanges (Figure 7). Le nombre de sources a été fixé à $J = 6$ pour toutes les méthodes.

Les reconstructions, les paramètres estimés et les vérités terrains sont affichés Figure 8. De nouveau, les résultats montrent qu'il est indispensable de considérer l'évolution lente des paramètres, en particulier pour les retards et que les méthodes proposées sont beaucoup plus rapides que la méthode de [Maz+15]. Les résultats obtenus avec la méthode régularisée et la méthode des B-splines ont été validés par un expert.

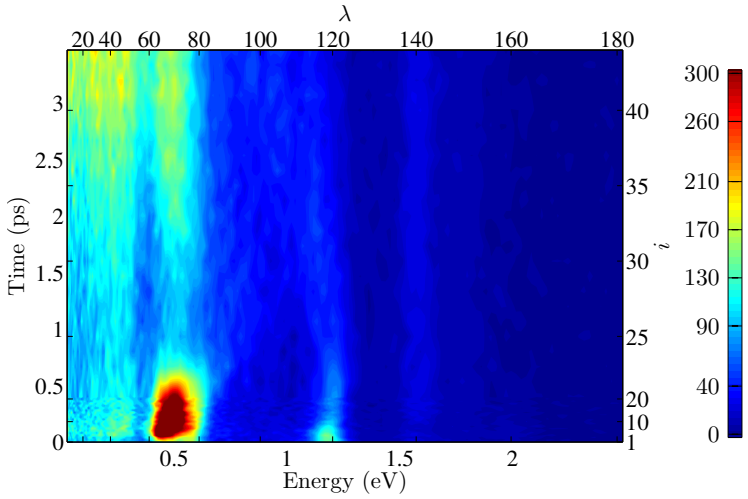


Figure 7: Une séquence temporelle de spectre de photoélectons ($I = 44$ mélanges) représentée comme une carte d'intensité de dimension $I \times N$.

Galaxie NGC-4254

Dans cette section des résultats obtenus sur l'image multispectrale de la galaxie NGC-4254 de dimension $140 \times 140 \times 42$ (Figure 9). On montre juste les résultats obtenus avec la méthode régularisée car les autres méthodes donnent des résultats moins satisfaisants et la méthode de [Maz+15] n'est pas applicable à des images multispectrales. Le nombre de sources est fixé à trois. La reconstruction de la galaxie est correcte comme illustre la Figure 9. Les résultats de l'estimation des paramètres qui montrent les amplitudes et les retards de chaque source sont présentés Figure 10. La méthode permet de bien identifier les différentes structures dans la galaxie. En particulier, la structure principale représentant le bras de la galaxie (la source 1 pour les deux méthodes), une source artificielle avec retards et largeurs fixes (la source 2), et une troisième source très atténuée en intensité (source 3), mais malgré ça la méthode a réussi à l'estimer. Ces résultats montrent que les méthodes proposées permettent de traiter de grandes données et de bien identifier les différentes structures de gaz. Aussi, l'évolution des retards est douce comme attendu.

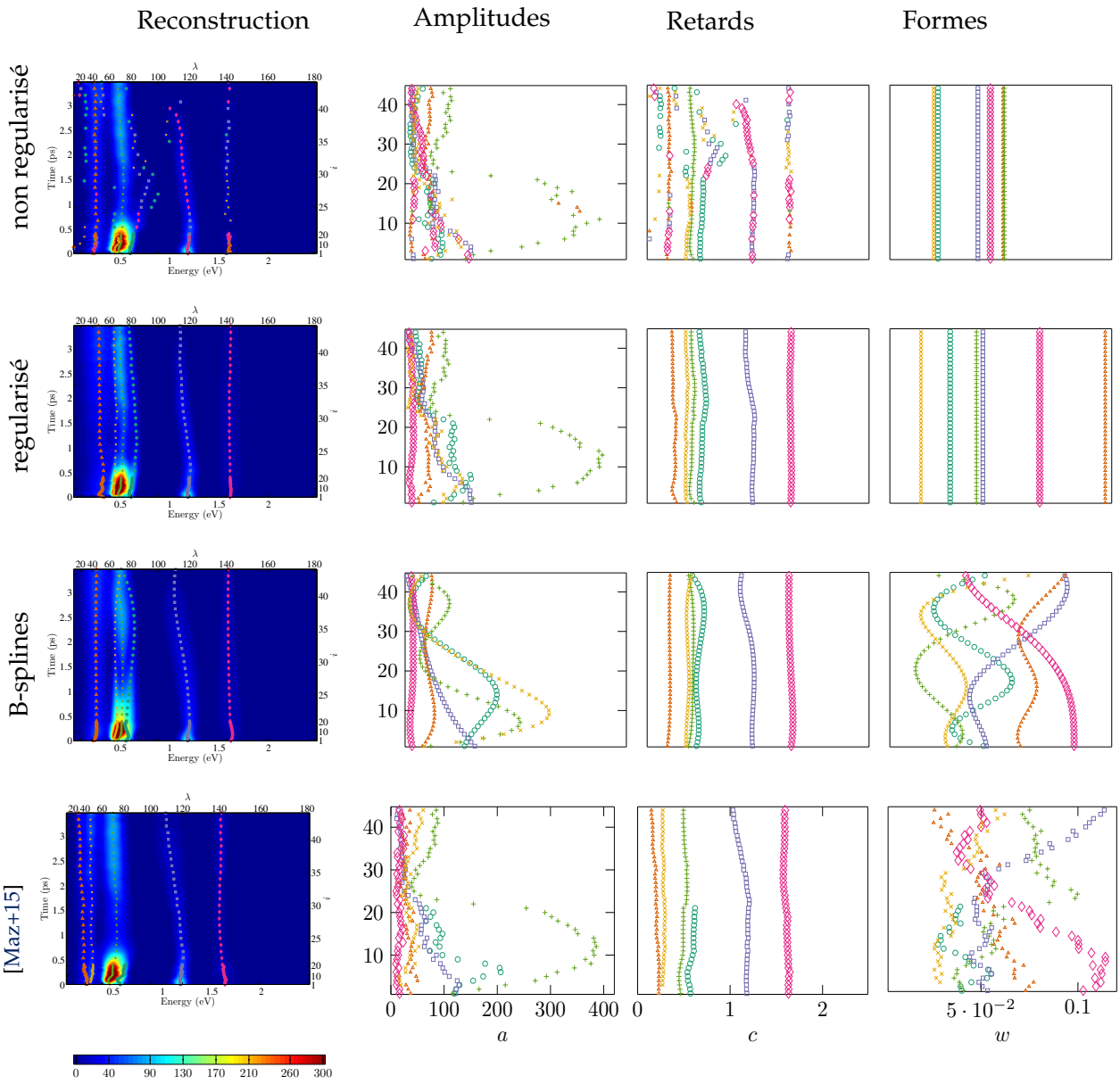


Figure 8: Premier colonne: la reconstruction des mélanges surimposé par les retards estimés. La deuxième, troisième et quatrième colonnes représentent respectivement les amplitudes, retards et largeurs estimés.

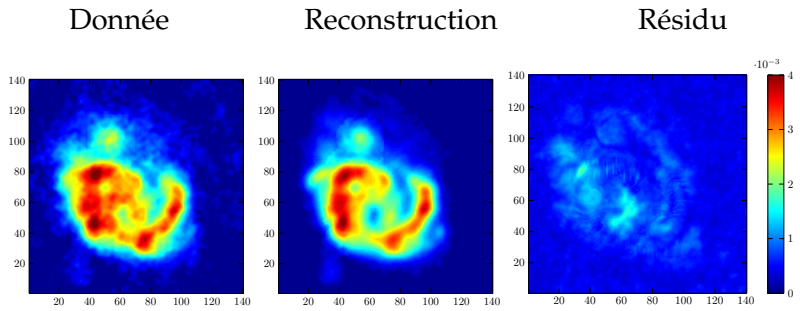


Figure 9: L'image moyenne de la galaxie NGC-4254, sa reconstruction et le résidu.

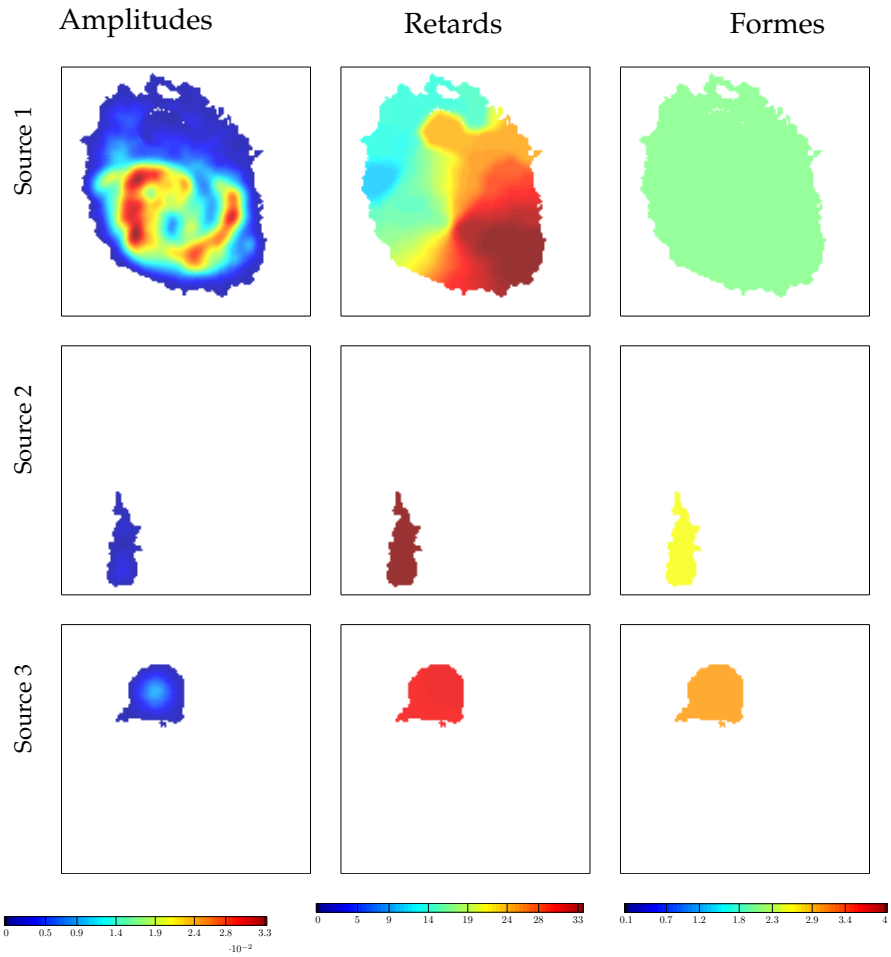


Figure 10: Les cartes des paramètres estimées pour $J = 3$ sources en utilisant la méthode régularisée.

Introduction

This thesis takes part in the project DSIM[1] funded by the ANR (the French National Research Agency). The project aims at developing numerical methods to analyze multiple spectroscopic signals (spectra) obtained from the same physical phenomenon.

The work in this thesis is motivated by two applications. The first one is time-resolved photoelectron spectroscopy, which is an experimental tool to obtain the electronic configuration of atoms or molecules [SBN04]. Data consist of a set of spectra acquired at different time instants during an experiment, thus giving a temporal sequence of spectra. Each spectrum shows several peaks from which the electron distribution with respect to the energy levels can be determined. The evolution of these peaks within the temporal sequence of spectra informs about the relaxation dynamics of the electron distribution. The second application deals with galaxy kinematics and the study of galactic gas motion [MB81]. Data are multispectral images where the peaks observed in the pixels are spectrally shifted. This is a consequence of the Doppler effect caused by gas motion.

The two considered applications come from very different phenomena and rely on various data (a temporal sequence of spectra for time-resolved photoelectron spectroscopy and a multispectral image for galaxy kinematics). However, both data are a set of ordered spectra where the peaks carry the information of interest. Moreover, two neighboring spectra are very similar, therefore the peaks show a (temporally or spatially) *slow evolution* within the spectra. Hence, this thesis aims at achieving two goals:

- estimate the characteristics of each peak in the data, such as intensity, spectral position or width;
- associate the peaks from different spectra as they evolve slowly.

Extracting information of interest (the peak characteristics and their evolution) from measurements is an *inverse problem*. Such problem is often ill-posed in the sense of Hadamard [Had23], that is the solution does not exist, or is not unique, or is not a continuous function of the data. In order to obtain a satisfying solution, one must include additional knowledge about the problem. Therefore, the spectra are modeled as a sum of *parameterized* functions describing the peaks, and the slow evolution of their characteristics may be favored within the spectra.

Multiple methods in the literature have addressed these goals (or similar ones). A first approach is to estimate the peak characteristics in the spectra independently of the others, and then associate them in a post-processing. However, such an approach does not exploit the similarities between neighboring spectra during the estimation step and may lead to error propagation. Thus, methods that jointly process the two goals are a better solution. Specifically, a very similar problem has

[1] DSIM is the French acronym of Décomposition Spectroscopique en Imagerie Multispectrale, dsim.unistra.fr.

been addressed in the Bayesian framework [Maz+15]. This method is effective but suffers from high computation time and cannot be applied to multispectral images. Another joint framework is source separation [CJ10] that aims at extracting sources (peaks) from their mixtures (spectra). In particular, the delayed source separation is able to deal with shifted sources. Prior knowledge is introduced by often assuming the sources to be non-correlated or statistically independent. Unfortunately, these assumptions are not valid in the considered applications.

In this thesis, we adopt a joint approach by using the source separation framework. We explicitly exploit the knowledge of parameterized sources allowing us to deal with highly correlated sources. The proposed approach can be applied to temporal sets of mixtures and multispectral images while being both effective and efficient.

The thesis is organized into six chapters.

Chapter 1 details the considered spectroscopic applications (time-resolved photoelectron spectroscopy and galaxy kinematics) and justify the need for numerical methods. In addition, we define the goal of the thesis from a signal processing point of view and we present the mathematical model of the mixtures, sources and noise. Finally, we highlight the main difficulties related to this model and data, *i.e.*, the indeterminacies inherent in the model, the high correlation between the sources and the unknown number of sources in each mixture.

Chapter 2 presents a survey on methods of the literature which are related to the thesis goals. First, we present methods that estimate the source parameters in a single mixture. Then we present methods that allow to associate the sources of different mixtures. Finally, methods that jointly estimate and associate the sources are presented.

Chapter 3 considers the problem of delayed and parameterized source separation. It is solved by minimizing a data-fit criterion using an Alternating Least Squares (ALS) scheme; the shape parameters are estimated using the Levenberg-Marquardt algorithm, while the amplitudes and delays in each mixture are estimated using a sparse approximation method.

Chapter 4 gives the first solution to take into account the slow delay evolution of each source. To do so, a regularization term is added to the data-fit criterion of Chapter 3. This implies a new sparse approximation strategy to estimate the amplitudes and delays jointly in all the mixtures.

Chapter 5 proposes an alternative model where the shape parameters vary within the mixtures similarly to the amplitudes and delays. The slow evolution of the amplitudes, delays and shape parameters is modeled using B-splines. This modeling significantly reduces the number of unknowns since the source parameters are replaced by the B-spline control points. The problem is then addressed as a constrained optimization problem which is solved using a Sequential Quadratic Programming (SQM) algorithm.

Chapter 6 presents simulations where the three proposed methods are compared both qualitatively and quantitatively with state-of-the-art methods. Comparisons are done in terms of the ability to reconstruct mixtures and accurately recover the sources, their amplitudes and delays. Results on real photoelectron spectra and galaxy multispectral images are also presented. We show that the proposed methods are both effective and efficient, in particular, the knowledge of parameterized sources is crucial.

The methods presented in Chapters 3 and 4 are detailed in a journal paper:

- [Mor+19] H. Mortada, V. Mazet, C. Soussen , C. Collet and L. Poisson. “Parameterized source separation for delayed spectroscopic signals”, Signal Processing, vol. 158, May 2019, p. 48-60.

Two international conference papers with proceedings cover the methods of Chapters 3 and 5:

- [Mor+17b] H. Mortada, V. Mazet, C. Soussen, and C. Collet. "Separation of delayed parameterized sources". EUSIPCO, Aug 2017, pages 1080–1084, Kos, Greece.
- [Mor+18] H. Mortada, V. Mazet, C. Soussen and C. Collet. "Spectroscopic decomposition of astronomical multispectral images using B-splines". WHISPERS, Sep 2018, Amsterdam, The Netherlands.

One national conference paper (the French version of [Mor+17b]):

- [Mor+17a] H. Mortada, V. Mazet, C. Soussen and C. Collet. "Séparation de sources retardées paramétriques". GRETSI, Sep 2017, Juan-Les-Pins, France.

One oral presentation in the scientific day "*Inversion et problèmes multi-**" organized by GDR-ISIS, March 2019, Institut d'Astrophysique de Paris (IAP), Paris, France.

1 Sequential spectra modeling

1.1 Introduction

Spectroscopy deals with measuring and analyzing photons obtained from the interaction of electromagnetic radiation with matter [Pav+08]. The resulting signal is called spectrum and shows emission and absorption peaks that carry important information. So, the determination of peak characteristics such as intensity, spectral position and width is crucial for the practitioner to analyze and understand the studied object. The applications are numerous and can be found in almost every field of science. For instance, in astrophysics [Ten11], the peaks determine the chemical composition of stellar objects or galaxies. In medicine [BH09], the peaks in magnetic resonance spectroscopy identify the molecular structure of organic compounds. To preserve painting artworks [Ang01], the peaks may be used to identify the pigments.

In some applications, numerous spectra of the same object are measured. The order in which the spectra have been acquired is important since it brings precious information about the temporal or spatial evolution of the studied phenomenon. We will refer to such data as *sequential* data. Spectroscopic measurements may be acquired at discrete consecutive time instants, leading to a time series of spectra such as in time-resolved photoelectron spectroscopy [SBN04]. Also, spectra may represent a spatial set of measurements leading to multispectral or hyperspectral images where each pixel represents a spectrum; such images are used for example in remote sensing [Goe+85] or astronomical imaging [MB81]. In other frameworks, such as in chemometrics [Mal02], different spectra are measured when a physical parameter (*e.g.*, temperature or viscosity) variates.

Generally, two neighboring spectra in a sequence, while different, are similar. In other words, the peaks are typically characterized by a *slow evolution* within the sequential spectra; their intensities, spectral positions and widths evolve slowly from one spectrum to its sequentially adjacent ones. This can be the consequence of a fine resolution in time between two acquisition instants in time-resolved photoelectron spectroscopy, or in space between neighboring pixels in astronomical

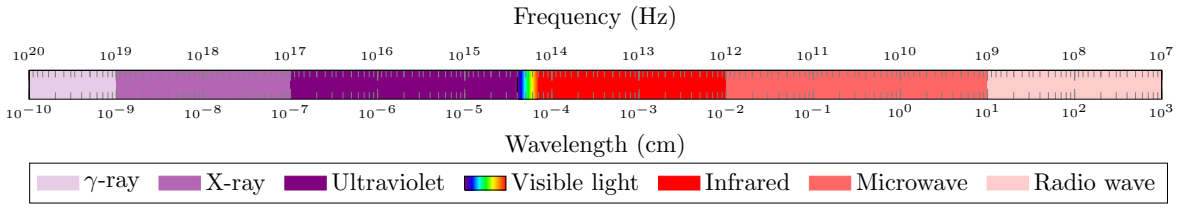


Figure 1.1: Electromagnetic spectrum.

imaging.

In practice, the visual analysis is impossible for accurate peak detection in an acceptable time due to the noise induced by the measurement process, the relatively large number of spectra and the need to account the slow peak evolution. In consequence, numerical methods are required to process the measured spectra.

The numerical methods developed in this thesis are applied to two spectroscopic applications: time-resolved photoelectron spectroscopy and galaxy kinematics, which are detailed in Section 1.2 in addition to a brief introduction to spectroscopy. Then, the goal is stated from a signal processing point of view: the problem is addressed as a parameterized source separation in Section 1.3, and the difficulties that can be encountered in such a model are given in Section 1.4.

1.2 Spectroscopic applications

1.2.1 Light and electromagnetic spectrum

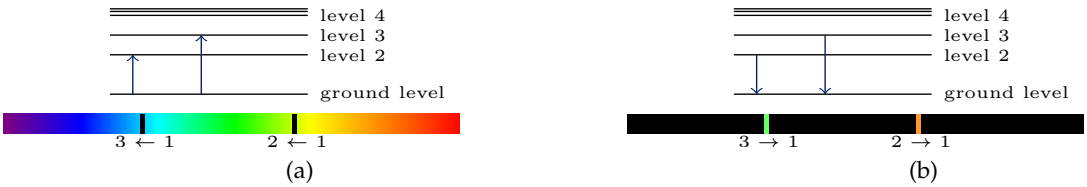


Figure 1.2: (a) Absorption and (b) emission spectra.

Light is composed of photons that have both particle and wave-like properties and carry discrete amounts of energy called quanta that can be transferred to atoms and molecules [Pav+08]. Besides, light is a part of a spectrum of electromagnetic energy that includes the following spectral bands (from lower to higher frequencies): radio waves, microwaves, infrared radiation, visible light, ultraviolet, x-rays, and gamma rays as shown in Figure 1.1. Light in the visible region ($4 \cdot 10^{-5}$ cm to $7 \cdot 10^{-5}$ cm) makes up only a small region of the entire electromagnetic spectrum. The energy E in Joules (J) of a photon is proportional to its frequency ν in Hertz (Hz) and inversely proportional to its wavelength λ in centimeters (cm), following the relation:

$$E = h\nu = h\frac{c}{\lambda}, \tag{1.1}$$

where $h \approx 6.62 \times 10^{-34}$ J·s is the Planck constant and $c \approx 3 \times 10^{10}$ cm·s^{-1} is the light speed in vacuum. The energy can be equivalently expressed in electron-volt (eV) where 1 eV $\approx 1.6 \times 10^{-19}$ J.

The peaks allow one to identify the chemical composition of the studied sample since the spectrum of a sample is unique; the same atoms or molecules will always have same spectra. Two kinds of spectroscopy exist. On the one hand, absorption spectroscopy is the study of light absorbed by a sample. When light passes through a sample, the electrons of the sample absorb some quanta and pass from the "ground" level (the lowest energy level) to a higher "excited" level. The levels refer to atomic of molecular orbitals. The resulting spectrum contains peaks with a negative intensity at these wavelengths. Figure 1.2a shows an absorption spectrum and the corresponding transitions between the energy levels.

On the other hand, emission spectroscopy happens after an absorption when an atom, element or molecule in an excited level returns to a configuration of lower energy level. An emission peak (a peak with positive intensity) appears in a spectrum if the sample emits specific wavelengths of radiation. The energy of the emitted light is equal to the difference between the higher and lower energy levels. Figure 1.2b illustrates an emission spectrum and the corresponding transitions between the energy levels. In this thesis, the proposed methods focus on spectra with emission peaks due to considered applications detailed in the sequel.

1.2.2 Time-resolved photoelectron spectroscopy

A first application that interest us is time-resolved photoelectron spectroscopy since measurements consist in a temporal sequence of spectra with slow evolving peaks.

Photoelectron spectroscopy is an experimental tool that aims at determining the electronic configuration of atoms or molecules *i.e.*, the distribution of electrons according to their energy in atomic or molecular orbitals [SBN04]. Photoelectron spectrometers ionize samples by bombarding them with high-energy radiation resulting in electron ejection; the ejected electrons are called photoelectrons. This reaction occurs on a femtosecond to picosecond time-scale and the measurement of this reaction has been made possible with the pump-probe spectroscopy method developed by Zewail [Zew88] and for which he awarded the chemistry Nobel prize in 1999. This was the key to real-time observations of ultrafast changes in chemical reactions, the so-called femtochemistry. The pump-probe method uses a first ultrafast laser pulse (the pump pulse) to excite an atom or a molecule. The subsequent relaxation is monitored by an interaction with a second ultrafast laser pulse (the probe pulse) which is delayed at a certain time from the pump pulse. The probe pulse lifts the electrons above the final energy level. Then, the electrons are analyzed using an appropriate spectrometer. The latter measures the photoelectron kinetic energy resulting from the absorption of photons with an energy $h\nu$. Figure 1.3 illustrates the photoelectron experiment and the energy transitions after the pump-probe ionizations.

The measured spectrum gives the photoelectron distribution with respect to the kinetic energy. The peak characteristics *i.e.*, their intensity, spectral position and width are crucial for the practitioner for the following reasons:

- the spectral position indicates the energy level from which photoemission is occurring;
- the area under each peak (which depends on both the intensity and width) corresponds to the relative number of electrons.

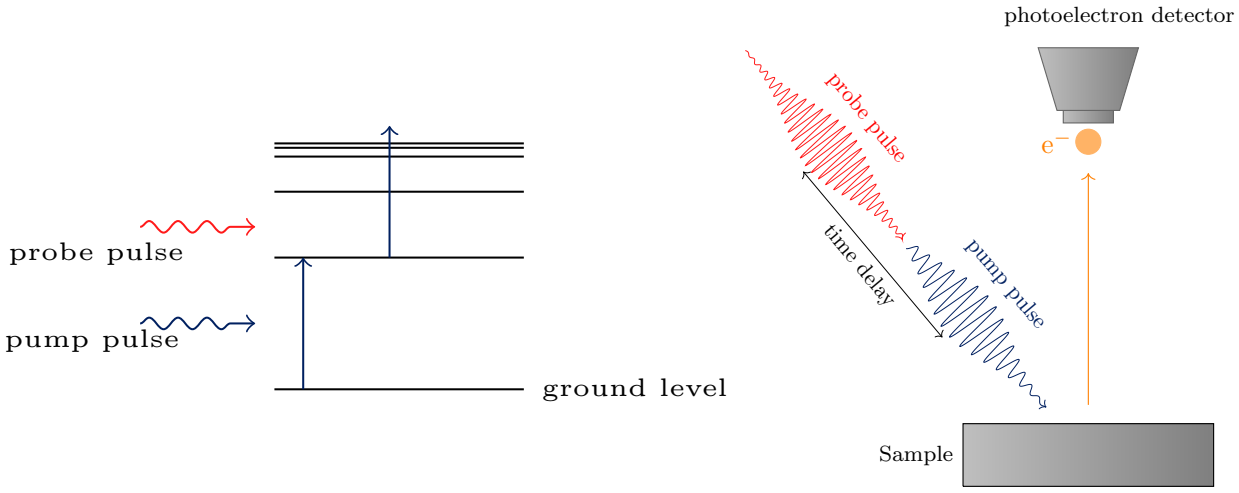
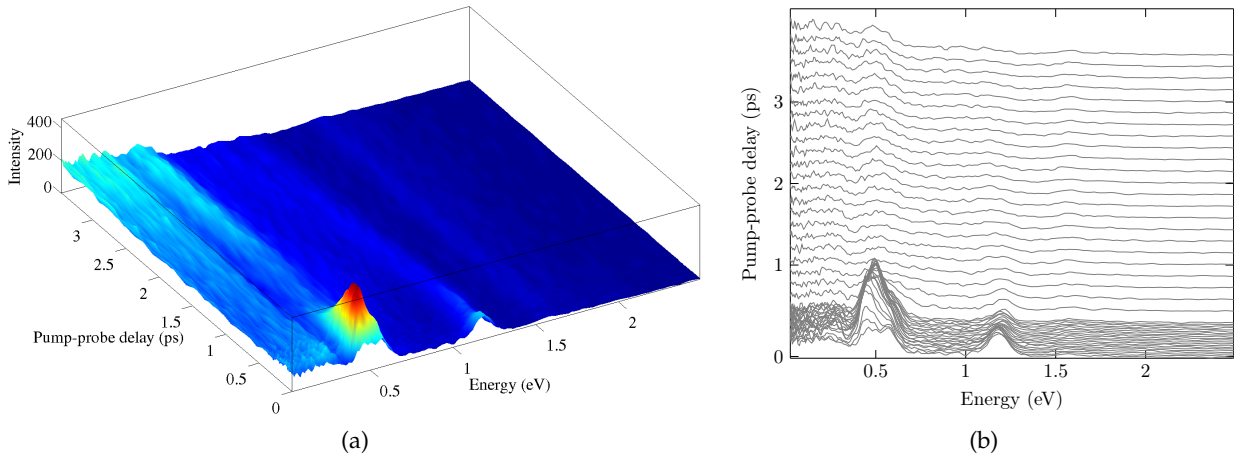


Figure 1.3: The photoelectron spectroscopy experiment. A first pulse (pump pulse) hits the sample to eject electrons from the ground level to an intermediate energy. Then, a second delayed pulse (probe pulse) hits the excited electrons to eject them out of the sample. The ejected electrons are called photoelectrons and are captured and analyzed using an appropriate spectrometer.



(a) (b)

Figure 1.4: (a) a 3D representation of 44 real photoelectron spectra at different pump-probe delay times with respect to the binding energy. (b) Another representation of the spectra.

In time-resolved photoelectron spectroscopy, multiple spectra are obtained for different time delays between pump and probe pulses allowing one to record the relaxation dynamics of the different energy levels [SBN04]. The time delay between the pump and probe pulses is slightly incremented or decremented from one measurement to another. Consequently, the peak characteristics are subject to a slow variation from one spectrum to its adjacent ones.

Figure 1.4 shows a sequence of 44 photoelectron spectra acquired at 44 pump-probe delays. The corresponding experiment studies the relaxation of an atom of barium on an argon droplet;

to get more details of the experiment, the reader could refer to [Mas+10]. For each spectrum, the energy varies from fractions of eV to 3.5 eV (180 energy values). The acquisitions are not regular for the pump-probe delays: for the first 20 spectra, the pump-probe delay step between two adjacent spectra is incremented by 0.02 ps while for the rest of spectra the delay step is incremented by 0.13 ps. Similarly, the energy distribution sampling is not regular and it can be assimilated to a decreasing exponential function. Moreover, a continuum (background) is observed in the spectra and must be corrected to estimate the peak characteristics. Finally, one can observe the slow evolution of multiple groups of peaks within the spectra (the most obvious group of peaks are around 0.5 and 1.2 eV).

1.2.3 Galaxy kinematics

Galaxy kinematics is the second application we focus on and correspond to a different physical phenomenon than the previous one. Data consists in an ordered spatial set of spectra with slowly evolving peaks.

A galaxy is composed of a huge number of stars, as well as gas and interstellar dust that are held together by gravitational attraction [KJ98]. One of the most exciting question in astrophysics is the process of formation and evolution of galaxies. The theory states that clouds of hydrogen and helium have been created in the first few minutes of the Universe. Then, the clouds collapsed due to gravity, resulting in stars. This process of collapse and star formation leads to galaxies around 1 billion years after the beginning of the Universe. When gas clouds are excited by a source of energy (*e.g.*, stars, galactic nucleus), the electrons are excited and as they go back to ground level, they emit photons of a specific wavelength. When analyzing the photons emitted from a given spatial position, a spectrum is obtained and its peaks carry information such as the chemical composition.

The ability to acquire multispectral images provides critical answers like the composition, star birth rates and the rotational velocity of the different part of the galaxies. Telescopes today routinely provide high-resolution multispectral images in the visible, infrared or radio bands. A multispectral image is a 3D data cube where two dimensions correspond to the angular coordinates and the third dimension is the wavelength. Therefore, each pixel of the multispectral image is an electromagnetic spectrum; the length of this spectrum equals the number of bands in the image. Usually, there are tens of spectral bands, such as in VIMOS, FLAMES, SINFONI, VLA or IRAM Interferometer [1] , but very recent instruments now provide a high number of bands (from a few hundreds to four thousands, *e.g.*, CALIFA, KMOS or MUSE) [2].

The spectra in galaxies are characterized by a shift [MB81]. This is a manifestation of the Doppler effect. On the first hand, observers looking at an object that is moving away from them, see a light with higher wavelengths than the emitted ones (a redshift). On the other hand, observers

[1] VIMOS (VIsible Multi-Object Spectrograph), FLAMES (Fibre Large Array Multi Element Spectrograph), and SINFONI (Spectrograph for INtegral Field Observations in the Near Infrared) are integral field spectrographs installed at the Very Large Telescope in Chile in the last decade. The VLA (Very Large Array) and the IRAM Interferometer are radio interferometers respectively located in New Mexico, USA and the French Alps.

[2] CALIFA (Calar Alto Legacy Integral Field Area) is installed on the telescope of Calar Alto Observatory located in Andalucía, Spain; KMOS (K-band Multi-Object Spectrograph) and MUSE (Multi-Unit Spectroscopic Explorer) are second-generation integral field spectrographs installed at the Very Large Telescope in Chile in 2012 and 2014.

looking at an approaching object see a light that is shifted to lower wavelengths (a blueshift). We encounter at least three kinds of shifts in the Universe:

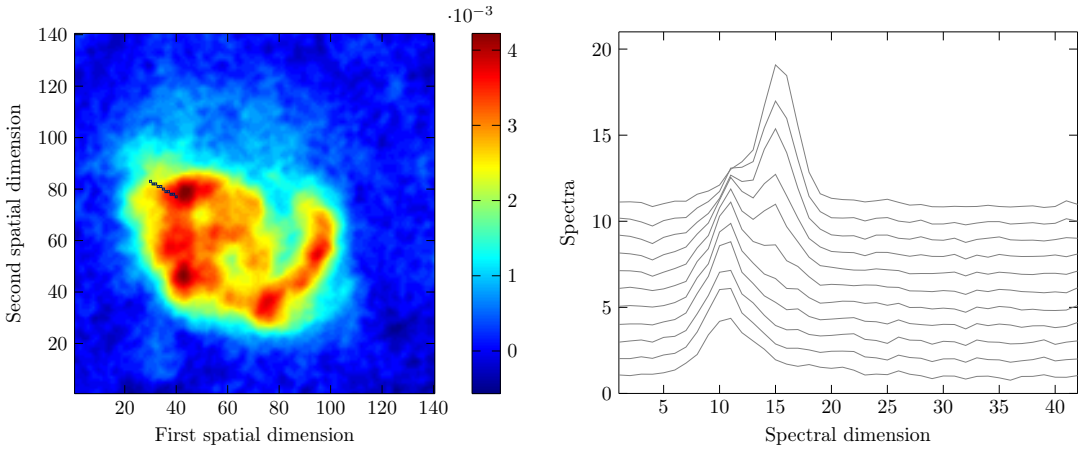
- Cosmological redshift: has been discovered by Edwin Hubble and is a consequence of the Universe expansion; the first piece of the Big Bang theory. By measuring how far the pattern of peaks moves from where it is supposed to be, astronomers can calculate the distance of astronomical objects such as stars and galaxies relative to Earth. Cosmological redshift only becomes noticeable for very far objects (a distance of hundreds of megaparsecs) and it is defined such that:
$$z = \frac{\lambda_o - \lambda_e}{\lambda_e}, \tag{1.2}$$
where λ_o is the observed wavelength and λ_e is the expected wavelength of the gas.
- Gravitational redshift: since photons always travel at the speed of light, the only place where an energy loss can show up, when light tries to escape from a gravitational field, is in a change of frequency. This phenomenon was predicted by Einstein and has been confirmed in laboratory experiments carried out by Pound and Rebka [PRJ59].
- Doppler shift: this is the spectral shift resulting from the inner gas motion of an astronomical object. It shows us how fast each part of the galaxy is moving away from us. Both redshift and blueshift are thus observed: light from parts of the galaxy rotating away from us is shifted towards the red part of the spectrum, while light from parts of the galaxy rotating towards us is shifted towards the blue end of the spectrum [MB81].

The measurement of galaxy kinematics is possible due to the Doppler shift. Indeed, by detecting the peak spectral positions in a spectrum (pixel), one can deduce the velocity of the corresponding kinematic structure since the spectra are shifted proportionally to the gas speed. Different spectroscopic bands allow one to study galaxy kinematics. For instance, in optical spectroscopy and for nearby galaxies, the Hα spectral peak (the first atomic transition in the hydrogen Balmer series) is used [Phi+86]. For distant galaxies, the spectral peak of oxygen [RPL13] is used. In radio spectroscopy, the 21 cm peak of hydrogen is often used to measure the rotation of the galaxy [Bos81] since those radio waves can easily pass through the Earth's atmosphere and can penetrate the massive clouds of interstellar cosmic dust with little interference.

In addition to the gas velocity obtained from the spectral positions of peaks, the peak intensities and widths yield the flux (the amount of energy that crosses a unit area per unit time) and velocity dispersion of gas. Hence, astronomers are interested in obtaining spatial maps[3] of the intensities, positions and widths of the peaks. Traditionally, astronomers analyze galaxy kinematics by visual inspection of the spectra. Such a solution is only feasible for small and simple galaxy spectra. Alternatively, a fast method is to calculate the multispectral image moments which only rely on the observed spectra and do not explicitly estimate the peak characteristics. Namely, the moments 0, 1 and 2 can respectively be interpreted as the flux, velocity field and velocity dispersion maps and are calculated at each spatial position using the following equations:

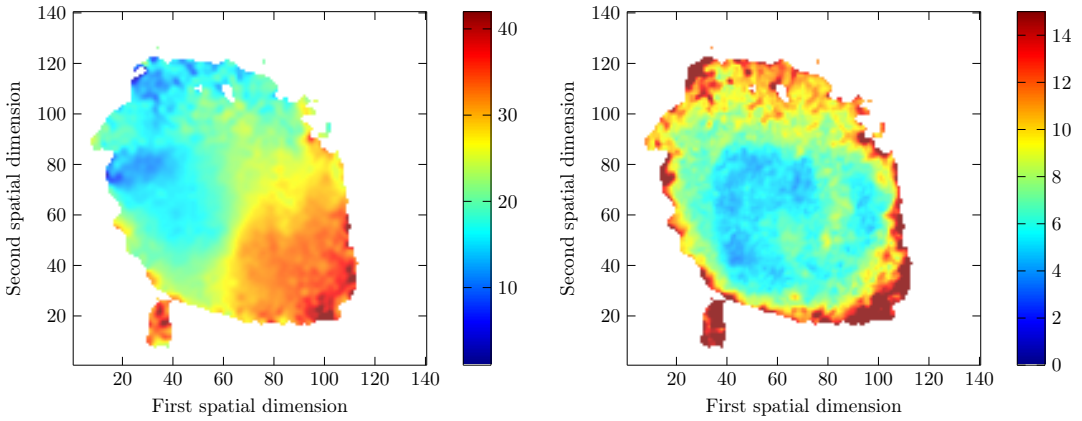
$$M_0 = \sum_{\lambda} X(\lambda), \quad M_1 = \frac{\sum_{\lambda} X(\lambda)\lambda}{M_0}, \quad M_2 = \sqrt{\frac{\sum_{\lambda}(\lambda - M_1)^2 X(\lambda)}{M_0}}, \tag{1.3}$$

[3] A map refers to an image whose colors represent the value of a parameter.



(a) (b)

Figure 1.5: (a) NGC 4254 white image of dimension 140×140 (equivalent to the galaxy moment 0). The pixels indicated in blue squares in (a) are plotted in (b) with respect to the spectral dimension.



(a) Moment 1 (b) Moment 2

Figure 1.6: The moments 1 and 2 of galaxy NGC-4254 calculated only for pixels with high moment 0.

where $X(\lambda)$ represents a pixel spectrum and λ represents the wavelengths. However, this method is very sensitive to noise and only yields a coarse approximation of the maps.

Figure 1.5 shows the white image [4] (which is equivalent to the moment 0 map) of the galaxy NGC-4254 [PVM93] acquired in the radio bands, as well as a sequence of spectra from different spatial positions. This real example clearly shows the slow spatial evolution of the peak characteristics. Besides, multiple peaks can be observed for some pixels, this is the result of multiple kinematic structures with different velocities in the same spatial position. The moments 1 and 2 of the galaxy NGC-4254 are displayed in Figure 1.6.

1.3 Problem statement

The two considered applications represent different physical phenomena with very different scales (from molecules to galaxies). In the first application, a sequence of time-resolved photoelectron spectra is obtained for different time delays. In the second one, the measurement of galaxy kinematics is considered by exploiting the Doppler shift in the galaxy multispectral image. The two applications have the following common points:

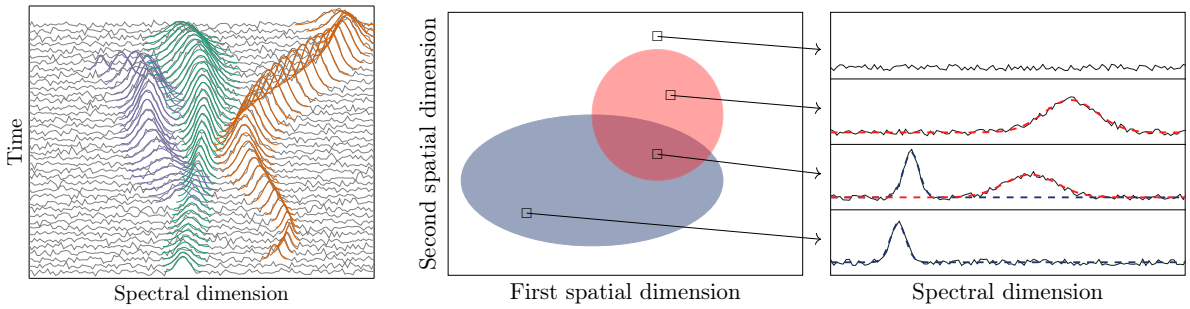
- the measurement of multiple spectra issuing from the same object;
- the spectra are ordered in a given dimension: a temporal dimension for the photoelectron spectra and spatial dimensions for galaxies;
- a slow evolution of the peak characteristics (intensity, spectral position and width) from one spectrum to its adjacent ones;
- the spectra contain only bell-shaped peaks with non-negative intensities;
- the number of peaks is unknown in each spectrum.

From a signal processing point of view, the two applications are similar except for the dimension of the data. The processing of these data aims at achieving two tasks:

1. the detection of an unknown number of peaks and the estimation of their characteristics in each spectrum;
2. the association of the peaks as they evolve slowly within the spectra. This task can be assimilated to a classification problem where:
    - each class is represented by a unique peak in each spectrum;
    - a class does not necessarily appear in all the spectra;
    - a class is contiguous; meaning that all the peaks sharing the same label appear in a (temporally or spatially) sequentially connected set of spectra.

Figure 1.7 clarifies the two tasks on two synthetic examples for both one and two sets of spectra.

[4] A white image refers to an averaged multispectral image along the spectral dimension.



(a) (b)

Figure 1.7: (a) a synthetic example of noisy spectra time-series with 3 evolving class of peaks indicated by different colors. The peak characteristics within each class are slowly evolving. (b) A synthetic illustration of a multispectral image with two evolving classes of peaks. The two classes are spatially overlapping and each is a set of spatially connected pixels. Spectra from different spatial positions are sketched as well.

1.3.1 Notations

In this thesis, we address the peak characteristic (parameter) estimation and association as a source separation problem: each source corresponds to a peak with a given label and each mixture corresponds to a spectrum and is the sum of the sources. Therefore, the terms mixture and source will be used in the sequel. The reader is referred to Section 2.5 for more details about the source separation framework. Also, the terms amplitude, delay and shape are respectively used to indicate the source parameters: intensity, spectral position and width.

The mathematical notations used in this thesis are as follows. Bold and lowercase variables correspond to vectors. Bold and uppercase variables correspond to matrices. The ith row and jth column of a matrix $\boldsymbol{M}$ are respectively denoted as $\boldsymbol{M}_{i:}$ and $\boldsymbol{M}_{:j}$, whereas a matrix element is denoted as m_{ij} as shown in Figure 1.8. A set of columns in $\boldsymbol{M}$ with indices gathered in $\mathcal{L}$ are denoted as $\boldsymbol{M}_{:\mathcal{L}}$.

1.3.2 Mathematical modeling

Let us consider I mixtures. Each mixture $x_i(\lambda)$ is modeled as the noisy sum of J parameterized sources $s(\lambda; w_{ij})$:

$$\forall\, i, \qquad x_i(\lambda) = \sum_{j=1}^{J} a_{ij} s(\lambda - c_{ij}; w_{ij}) + n_i(\lambda), \tag{1.4}$$

where:

- $\lambda \in \{1, \ldots, N\}$ represents a mixture sample (or wavelength) and N is the number of samples. Without loss of generality, the sampling step is supposed to be one;
- $i \in \{1, \ldots, I\}$ represents the mixture index. We distinguish two cases:

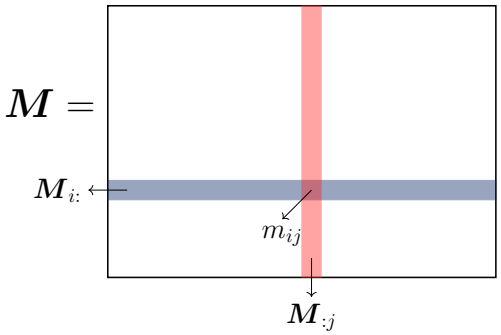


Figure 1.8: Mathematical notation for the elements of a matrix $\boldsymbol{M}$: a row $\boldsymbol{M}_{i:}$, a column $\boldsymbol{M}_{:j}$ and an element m_{ij}.

1. in case of a temporal sequence of mixtures with dimension $N \times I$ such as the photoelectron data, i is the index of the discrete acquisition instants. We refer to this case as *1D set of mixtures*;
2. in case of a multispectral image of dimension $U \times V \times N$, the index $i \triangleq [u\ v]$ gathers the spatial coordinates of a mixture. Also, the number of mixtures I is equal to $U \times V$. We refer to this case as *2D sets of mixtures*;

- the index j refers to the jth source;
- $a_{ij} \in \mathbb{R}_+$ is the amplitude of source j in mixture i. The positivity is considered on the amplitudes since we are only interested in emission peaks;
- $c_{ij} \in [1, N]$ is the delay of source j in mixture i;
- $w_{ij} \in \mathbb{R}$ is the shape parameter of source j in mixture i. Additional constraints like positivity can be added to the shape parameters. For the sake of clarity, we consider w_{ij} to be scalar, but the extension to a multidimensional shape parameter is straightforward;
- $n_i(\lambda)$ is the noise in mixture i. More details about the noise are given in Section 1.3.4.

Further, we introduce $\boldsymbol{x}_i$ as the vector gathering the N samples of a mixture:

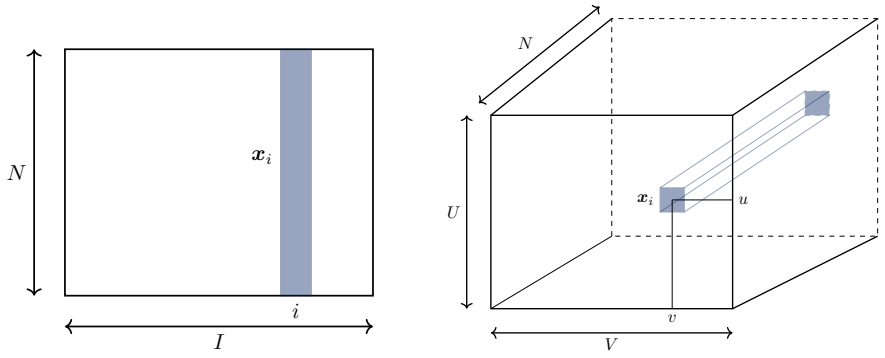
$$\boldsymbol{x}_i = \begin{bmatrix} x_i(1) & x_i(2) & \dots & x_i(N) \end{bmatrix}^T \in \mathbb{R}^N, \tag{1.5}$$

and $\boldsymbol{n}_i$ gathers the noise samples:

$$\boldsymbol{n}_i = \begin{bmatrix} n_i(1) & n_i(2) & \dots & n_i(N) \end{bmatrix}^T \in \mathbb{R}^N. \tag{1.6}$$

Similarly, the samples of each delayed and parameterized source are gathered into a vector $\boldsymbol{s}[c_{ij}; w_{ij}]$ that reads:

$$\boldsymbol{s}\left[c_{ij}; w_{ij}\right] \triangleq \begin{bmatrix} s(1 - c_{ij}; w_{ij}) & s(2 - c_{ij}; w_{ij}) & \dots & s(N - c_{ij}; w_{ij}) \end{bmatrix}^T \in \mathbb{R}^N. \tag{1.7}$$



(a) 1D set of mixtures (b) 2D sets of mixtures

Figure 1.9: 1D and 2D representations of I mixtures.
The two considered 1D and 2D representations of I mixtures.

Finally, using vector notation, equation (1.4) now reads:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[c_{ij}; w_{ij}\right] + \boldsymbol{n}_i. \tag{1.8}$$

We introduce the matrices $\boldsymbol{A}$, $\boldsymbol{C}$ and $\boldsymbol{W}$ of dimension $I \times J$ that respectively gather the amplitudes a_{ij}, delays c_{ij} and shape parameters w_{ij}:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \dots & a_{1J} \\ \vdots & \ddots & \vdots \\ a_{I1} & \dots & a_{IJ} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} c_{11} & \dots & c_{1J} \\ \vdots & \ddots & \vdots \\ c_{I1} & \dots & c_{IJ} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} w_{11} & \dots & w_{1J} \\ \vdots & \ddots & \vdots \\ w_{I1} & \dots & w_{IJ} \end{bmatrix}. \tag{1.9}$$

Indeed, in case of 2D sets of mixtures the number of rows is equal to $U \times V$.

A special case of this model is considered in Chapters 3 and 4, where the shape parameter of a source is supposed to be constant within the mixtures, thus the model writes:

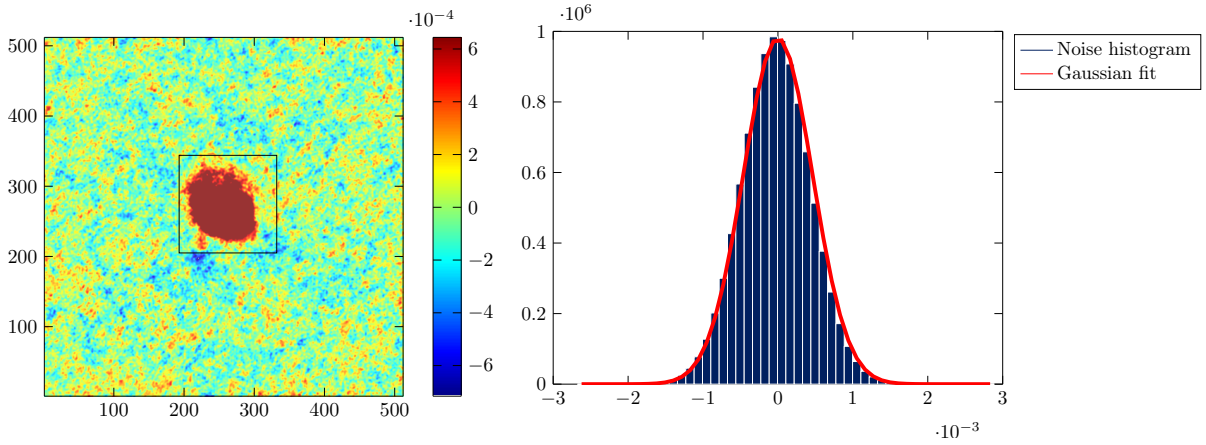
$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[c_{ij}; w_j\right] + \boldsymbol{n}_i. \tag{1.10}$$

In this case, the index i vanishes for w and the shape parameters are gathered in the vector $\boldsymbol{w}$:

$$\boldsymbol{w} = \begin{bmatrix} w_1 & \dots & w_J \end{bmatrix}. \tag{1.11}$$

1.3.3 Parameterized source

It is usual in spectroscopy to consider the sources $s(\lambda; w)$ as functions parameterized by w [Pav+08]. The sources are not infinitely sharp: numerous factors lead to the so-called broadening of peaks [Pea81] such as the Doppler effect, the pressure or the Point Spread Function (PSF) of the spectrometer. Therefore, the sources are rarely modeled as Dirac functions but rather as bell-shaped functions. Also, the functions are mostly symmetric about the spectral position of their maximum. The



(a) (b)

Figure 1.10: (a) The white image of the Galaxy NGC-4254. The pixels outside of the black box are supposed to contain only noise. The pixel inside the box are represented in Figure 1.5a. (b) The histogram of the noise pixels superimposed by a Gaussian function fit with a mean equals to 0 and a standard-deviation equals to $4.6 \cdot 10^{-4}$.

Gaussian and Lorentzian [Hol04; Maz05; VC14] functions are highly used in spectroscopy and have a single shape parameter w characterizing their width. A more complicated function with two shape parameters is the Voigt function [OL77] (convolution of Gaussian and Lorentzian functions).

The methods proposed in this thesis are not restricted to a specific function. Nevertheless, we consider for the sake of clarity, that all the sources are modeled with the same parameterized function which is parameterized by a scalar shape parameter w. Without loss of generality, the methods will be tested with Gaussian functions.

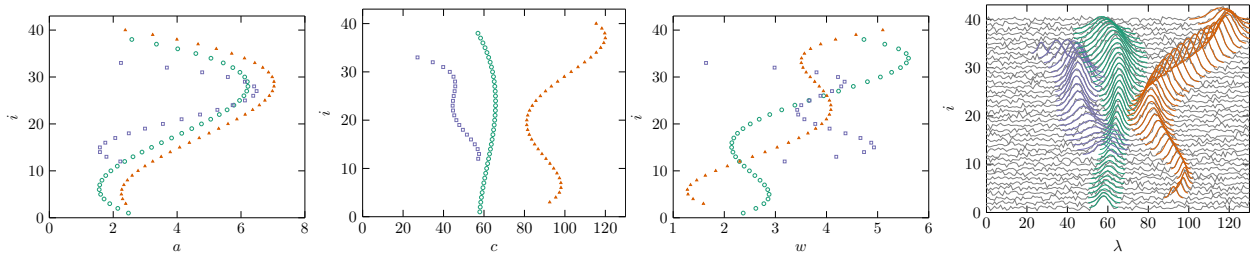
1.3.4 Noise model

The noise $\boldsymbol{n}_i$ in a mixture mainly models:

- the physical noise resulting from the acquisition process;
- the imperfection of the mathematical model.

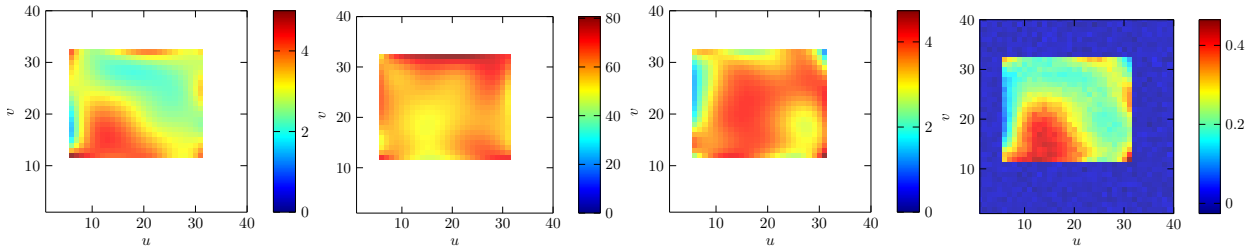
For the first element, it is known that the acquisitions which are obtained from photon counting, as for spectroscopic data, induce a Poisson noise (or shot noise). Moreover, the Poisson probability distribution is correlated with the intensity of each measurement (not additive) making it challenging in the processing. Fortunately, when the number of photons is significant, the Gaussian model for the noise is a good approximation. This is confirmed by the central limit theorem, which states that the sum of a large number of independent random variables tends to a Gaussian, independently of their probability density function [Pet95].

For photoelectron spectroscopy, the number of photons is high; therefore the Gaussian approximation is justified. For the galaxy multispectral images, we are not able to confirm this assumption since the data are acquired using a complex procedure including fast Fourier transform spectrometer. However, the galaxy images considered in this thesis are acquired in the radio



(a) amplitudes (b) delays (c) shapes (d) mixtures

Figure 1.11: Synthetic 1D set of mixtures. (a)–(c) the amplitudes, delays and shapes of $J = 3$ sources. The parameters of each source are plotted in different colors. (d) The resulting mixtures using the generated parameters.



(a) amplitudes $\boldsymbol{A}^*_{:1}$ (b) delays $\boldsymbol{C}^*_{:1}$ (c) shapes $\boldsymbol{W}^*_{:1}$ (d) white image

Figure 1.12: A generative example of 2D mixtures and $J = 1$ source. (a)–(c) the amplitude, delay and shape maps. (d) The resulting averaged image following the spectral dimension.

band (to detect the 21 peaks of hydrogen) which are characterized by a small noise level since, as mentioned earlier, those waves can penetrate the massive clouds of interstellar cosmic dust with little interference. Furthermore, Figure 1.10 represents the histogram of the pixels in the Galaxy NGC-4254 observation which are supposed to only contain noise. This histogram can be correctly fitted by a Gaussian distribution with zero mean as it is shown in Figure 1.10. Thus, in this thesis, the noise is additive, white, zero mean and Gaussian with variance σ_n^2:

$$\forall i, \qquad \boldsymbol{n}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}), \tag{1.12}$$

where $\boldsymbol{I}$ is the identity matrix.

1.3.5 Synthetic data generation

To generate synthetic 1D or 2D sets of mixtures, which are used to evaluate and compare the different methods[5], the model (1.8) is used with Gaussian sources, *i.e.*, $s(\lambda; w) = \exp(-\lambda^2/2w^2)$. A zero mean Gaussian noise with variance σ_n^2 is added to the mixtures. The noise level is measured

[5] All the simulations are done using Matlab on a computer with Intel Core i7, 3.6 GHz CPU and 16 GB RAM.

by the Signal to Noise Ration (SNR) in dB, it is defined as:

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_i \left\| \sum_{j=1}^{J} a_{ij}^* \boldsymbol{s}\left[c_{ij}^*; w_{ij}^*\right] \right\|_2^2}{N \cdot I \cdot \sigma_n^2}, \tag{1.13}$$

where $\|\cdot\|_2$ is the ℓ_2 norm [6].

In case of a 1D set of mixtures, the ground-truth parameters a_{ij}^*, c_{ij}^* and w_{ij}^* [7] are generated for each source j within the mixture using polynomial functions with random coefficients. The degree d of the polynomial allows us to control the evolution smoothness. An example of synthetic data is displayed in Figure 1.11 with $I = 40$ mixtures, $J = 3$ sources, $N = 130$ samples and SNR = 15 dB. The amplitudes, delays and shapes are respectively generated using polynomials of degree 3, 4 and 5.

In case of 2D sets of mixtures, the parameters for each source are represented by maps and are generated using B-spline surfaces [DB72] with random control points thus ensuring a smooth spatial evolution. An example of synthetic multispectral image is displayed in Figure 1.12 with $I = 40 \times 40$ mixtures, $J = 1$ sources, $N = 100$ samples and SNR = 15 dB. The parameters are only generated for a rectangle zone in the image. The amplitude $\boldsymbol{A}_{:1}^*$, delay $\boldsymbol{C}_{:1}^*$ and shape $\boldsymbol{W}_{:1}^*$ [8] maps are also given in Figure 1.12.

1.4 Main difficulties in practical cases

In this section, we discuss the main indeterminacies inherent in the model and the main difficulties that we encountered in practice.

1.4.1 Model indeterminacies

Indetermination in a given mixture arises if different sets of parameters yield the same outputs. Without additional information on the sources or mixtures, the model (1.10) admits at least three indeterminacies, namely, scale, permutation and delay indeterminacies (the parameters can only be estimated up to these indeterminacies) [Mou05, Chapter 2][MBI05; RGJ05].

- The scale indetermination in a given mixture i and for a given source j is the result of a factor switch between the sources and amplitudes:

$$a_{ij}s(\lambda; w_j) = a'_{ij}s'(\lambda; w_j), \tag{1.14}$$

  where $a'_{ij} = \frac{a_{ij}}{\eta}$ and $s'(\lambda; w_j) = \eta s(\lambda; w_j)$ with η a constant.

[6] The ℓ_p norm ($p > 0$) of a vector $\boldsymbol{v} = \begin{bmatrix} v_1 & \dots & v_N \end{bmatrix}$ is defined as $\|v\|_p \triangleq \left(\sum_{n=1}^{N} |v_n|^p \right)^{1/p}$.

[7] The ground truth parameters are always followed with the superscript *.

[8] In case of 2D sets of mixtures, the notations $\boldsymbol{A}_{:j}$, $\boldsymbol{C}_{:j}$, $\boldsymbol{W}_{:j}$ represent vectorized 2D maps so they are displayed as an image of dimension $U \times V$.

In order to overcome this indetermination, one can impose the norm of the source signals or their amplitudes. For unimodal sources, a strategy would be to set their maximum to one. In Chapters 3 and 4 we impose the sources to have a unit ℓ_2 norm and in Chapter 5 the maximum of the source is set to 1;

- the delay indetermination in a given mixture i and for a given source j happens in the following case:
$$s(\lambda - c_{ij}; w_j) = s'(\lambda - c'_{ij}; w_j), \tag{1.15}$$
where $s'(\lambda; w_j) = s(\lambda + \eta; w_j)$ and $c'_{ij} = c_{ij} + \eta$ with η a constant.
This indetermination can be alleviated by setting an initial position of the sources. In this thesis, we suppose that the source maximum is centered at its corresponding delay c_{ij}.

- the order (or permutation) indetermination affects the order j in which the sources are estimated. This indetermination cannot be alleviated but does not affect the interpretation.

It is worth mentioning that in some works [PD05; Nio+], the scale and delay indeterminacies are alleviated by picking a "reference" mixture in which the sources are assumed to have unit amplitudes and zero delays. Then the source amplitudes and delays in the other mixtures are relatively estimated to the sources of the reference mixture.

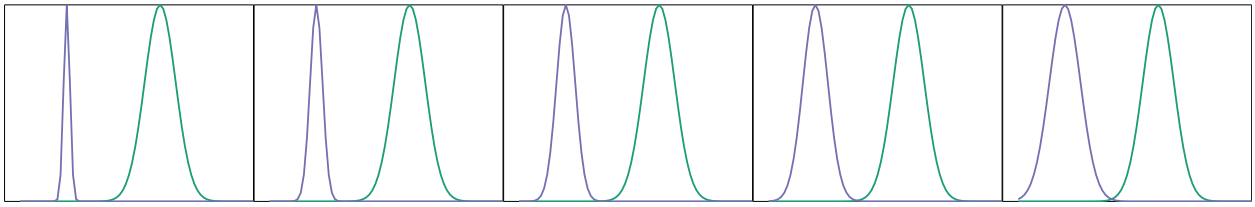
Even though the indeterminacies are alleviated, the estimated parameters do not necessarily correspond to the true parameters. However, in most of the practical cases, this is not an issue as long as the obtained solution is unique.

1.4.2 Correlated sources

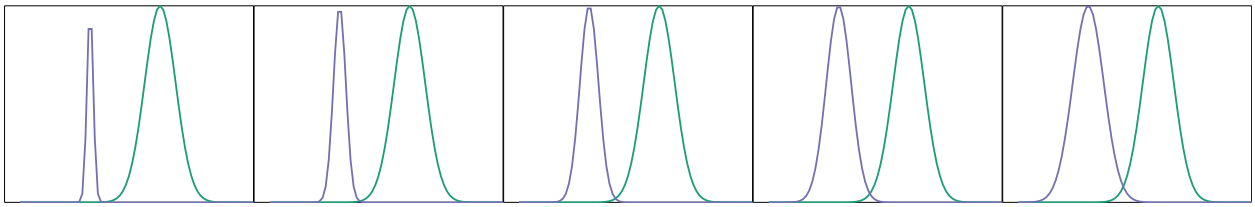
Another difficulty that is encountered when dealing with spectroscopic sources is the fact that they can be highly correlated. The source correlation can be observed when the sources appear in the same mixture and is mainly governed by two factors: the source shape similarities and the source spectral overlap. Figure 1.13 shows the two factors in an example of two sources in a single mixture.

The shape similarity between two sources, in a given mixture i, can be measured by the distance between their shapes given by $|w_{i1} - w_{i2}|$. Two sources with similar shapes may be confused in the mixtures if the sources are associated together using only their shape difference (as it is the case in Chapter 3). This ambiguity motivated us to include the slow evolution of the source parameters in Chapters 4 and 5 as an additional discriminating factor between the sources.

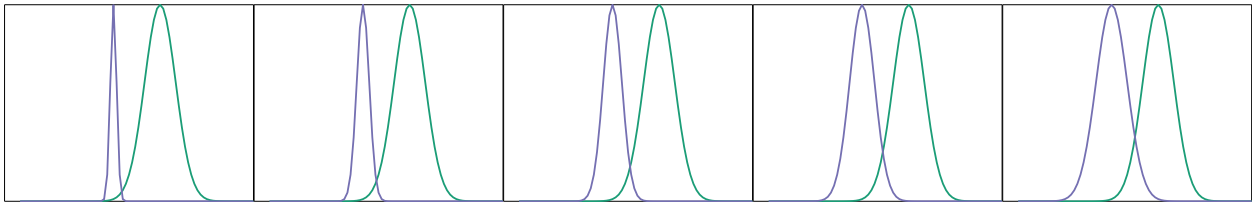
The spectral overlap between two sources, in a given mixture i, can be measured by the distance between their delays given by $|c_{i1} - c_{i2}|$. For some parametrized sources such as Gaussian and Lorentzian functions, the source support practically occupies all the samples of a mixture since they are not limited-support functions. However, the energy of these sources is highly concentrated around their delay (99.7% of a Gaussian function energy is bounded between $-3w$ and $+3w$ around the Gaussian maximum). Therefore, such sources can be assumed to be non-overlapping if the distance between their delays is sufficiently large. The estimation of overlapping source parameters becomes of higher complexity and classical strategies would fail to give an accurate estimation. For instance, when two sources are extremely overlapping, their shape merges and becomes as



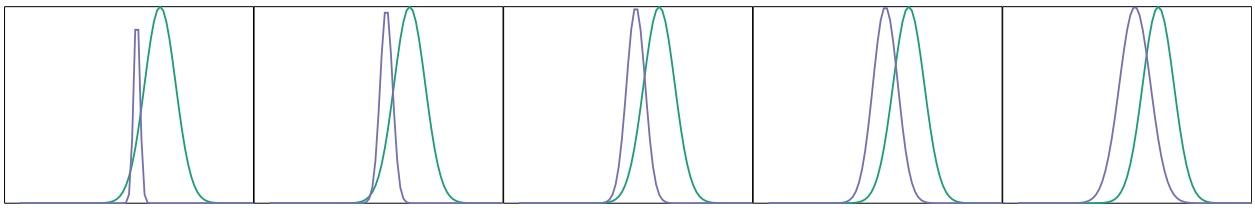
(a) $\nu = 1.8 \cdot 10^{-8}$ (b) $\nu = 1.5 \cdot 10^{-7}$ (c) $\nu = 1.6 \cdot 10^{-6}$ (d) $\nu = 1.7 \cdot 10^{-5}$ (e) $\nu = 1 \cdot 10^{-4}$



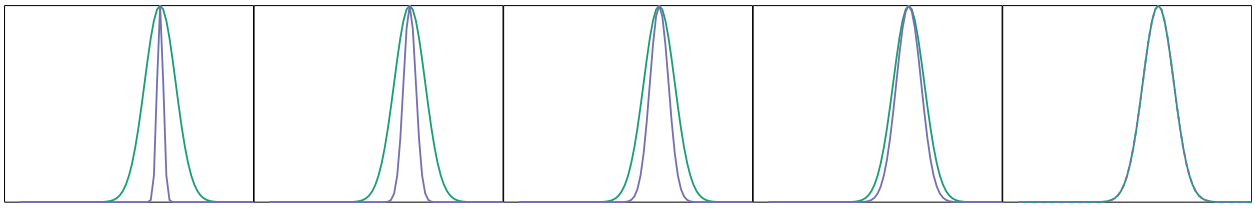
(f) $\nu = 3.6 \cdot 10^{-5}$ (g) $\nu = 1.3 \cdot 10^{-4}$ (h) $\nu = 5.4 \cdot 10^{-4}$ (i) $\nu = 2 \cdot 10^{-3}$ (j) $\nu = 6 \cdot 10^{-3}$



(k) $\nu = 8 \cdot 10^{-3}$ (l) $\nu = 1.7 \cdot 10^{-2}$ (m) $\nu = 3.4 \cdot 10^{-2}$ (n) $\nu = 6.3 \cdot 10^{-2}$ (o) $\nu = 1 \cdot 10^{-1}$

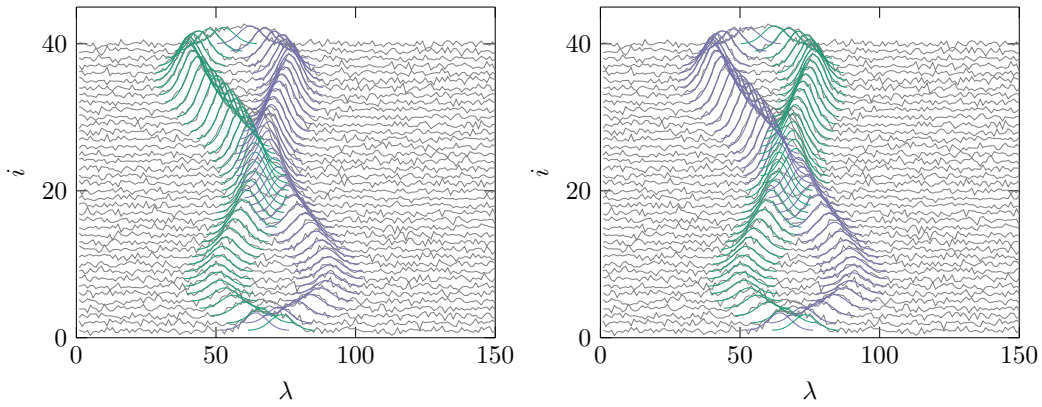


(p) $\nu = 2 \cdot 10^{-1}$ (q) $\nu = 3 \cdot 10^{-1}$ (r) $\nu = 4 \cdot 10^{-1}$ (s) $\nu = 5 \cdot 10^{-1}$ (t) $\nu = 5.7 \cdot 10^{-1}$



(u) $\nu = 6.2 \cdot 10^{-1}$ (v) $\nu = 8.3 \cdot 10^{-1}$ (w) $\nu = 9.3 \cdot 10^{-1}$ (x) $\nu = 9.8 \cdot 10^{-1}$ (y) $\nu = 1$

Figure 1.13: The shape similarity between two sources increases horizontally while their spectral overlap increases vertically. For each case we calculate the coherence level (variates between 0 for non-correlated sources to 1 for extremely correlated sources) defined as $\nu \triangleq \frac{\boldsymbol{s}\left[c_{i1};w_{i1}\right]\boldsymbol{s}\left[c_{i2};w_{i2}\right]^T}{\left\|\boldsymbol{s}\left[c_{i1};w_{i1}\right]\right\|_2\left\|\boldsymbol{s}\left[c_{i2};w_{i2}\right]\right\|_2}$.



(a) scenario 1 (b) scenario 2

Figure 1.14: Two scenarios of separation when $J = 2$ source spectrally overlap in the mixtures around $i = 20$.

if it was a single peak. In case of multiple mixtures, multiple "trajectories" of the source can be observed when two or more sources overlap in some mixtures. To make this idea clearer, we show in Figure 1.14, two possible scenarios of source separation (sources overlapping around the mixture $i = 20$). These two scenarios are visually acceptable since the evolution of the delays is slow in both of them. In Chapter 4 we set a regularization on the source delay evolution in order to favor solutions with the slowest delay evolution. For the same purpose, in Chapter 5 we model the parameter evolutions using B-splines functions.

1.4.3 Varying number of sources

In most applications, the number of sources J (or the model order) is not known. This includes an additional unknown to the problem that is very challenging to estimate. The simplest solution is to fix the source number to a predefined value. An alternative would be to choose a large number of sources then to remove the sources with a small contribution in a post-processing. In addition, we encounter an additional difficulty: the number of sources can vary within the mixtures (see Figure 1.7).

In Chapters 3 and 4 we use sparse approximation methods that aim to correctly approximate the mixtures with the smallest possible number of sources while in Chapter 5 we suppose that the number of sources is known.

1.5 Conclusion

This thesis provides numerical methods to process a set of ordered spectra with slowly evolving peaks. The goal is to get the peak characteristics (intensity, spectral position and shape) and also their temporal or spatial evolution. These data can be found in many applications such as in time-resolved photoelectron spectroscopy or galaxy kinematics. In this thesis, we use a source separation framework to model the problem; the spectra and the peaks are respectively referred to the mixtures and sources. The mixtures are modeled as the noisy sum of parameterized and

delayed sources. However, multiple difficulties can be encountered in practice when dealing with such data and model. Indeed, the model presents indeterminacies, the sources are highly correlated and their number is unknown and vary within the mixtures.

2 State of the art

2.1 Introduction

This chapter presents a survey on methods that, given a 1D or 2D sets of mixtures, aim at achieving two tasks. The first one is to estimate the source parameters (amplitudes, delays and shapes) and the second one is to associate these parameters as they slowly evolve (temporally or spatially) within the mixtures. Two strategies are possible:

- a two-step strategy, that is to first estimate the source parameters independently in each mixture and then apply a post-processing to associate the estimated parameters;
- a one-step strategy, that is to estimate and associate the source parameters jointly.

The first part of this chapter considers the two-step strategy: in Section 2.2, we present different approaches that allow estimating the source parameters (or a subset of them) in a single mixture. A special attention is given to sparse approximation methods which are used in Chapters 3 and 4. In Section 2.3 we present methods that can be used as a post-processing to associate the parameters.

The second part of this chapter considers the one-step strategy. In Section 2.4, Bayesian and joint sparse approximation methods for parameterized sources are presented. In Section 2.5, we present the problem from a source separation point of view.

2.2 Parameter estimation in a single mixture

When considering a unique mixture ($I = 1$), the index i can be alleviated and thus the model writes:

$$\boldsymbol{x} = \sum_{j=1}^{J} a_j \boldsymbol{s}\left[c_j; w_j\right] + \boldsymbol{n}. \tag{2.1}$$

Each source j in this case corresponds to a unique parameterized function whose parameters are the amplitude a_j, delay c_j and shape w_j. Moreover, we consider that all the sources are parameterized by the same function, *e.g.*, Gaussian function. The estimation of source parameters in a single mixture is an inverse problem that has been intensively studied in the past.

2.2.1 Least-squares based methods

The use of a deterministic optimization algorithm is a common approach to estimate the source parameters, in particular for spectroscopic signals. Assuming a white Gaussian noise, the maximum likelihood estimator minimizes the sum of squared differences between the mixture $\boldsymbol{x}$ and its model, *i.e.*, the squared sum of the residual vector $\boldsymbol{r}(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{w}) \in \mathbb{R}^N$:

$$(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{c}}, \widehat{\boldsymbol{w}}) = \underset{\boldsymbol{a},\boldsymbol{c},\boldsymbol{w}}{\operatorname{argmin}} \|\boldsymbol{r}(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{w})\|_2^2 = \underset{\boldsymbol{a},\boldsymbol{c},\boldsymbol{w}}{\operatorname{argmin}} \left\| \boldsymbol{x} - \sum_{j=1}^{J} a_j \boldsymbol{s}\left[c_j; w_j\right] \right\|_2^2, \tag{2.2}$$

where $\boldsymbol{a}$, $\boldsymbol{c}$ and $\boldsymbol{w}$ are vectors of dimension $J \times 1$ that respectively gather the amplitudes a_j, delays c_j and shapes w_j.

The problem (2.2) is a non-linear least squares problem. It is often referred as peak or curve fitting [Bud+96; Sad+05]. Indeed, the criterion may be non-convex because of the non-linearity induced by the delays c_j and shapes w_j. A linearization strategy of the delays and shapes is proposed in [Car+86] for Gaussian sources. This method exploits the natural logarithm of the Gaussian function, that writes:

$$\ln\left(a_j \exp\left(\frac{-(\lambda - c_j)^2}{2w_j^2}\right)\right) = \ln(a_j) - \frac{c_j^2}{2w_j^2} + \frac{2c_j\lambda}{2w_j^2} - \frac{\lambda^2}{2w_j^2}. \tag{2.3}$$

The new expression can be fitted by a second degree polynomial of the form $v_1 + v_2\lambda + v_3\lambda^2$. This method has been shown to be sensitive to noise in [Guo11] where an improvement is proposed by using a weighted least-squares method, the weight being the Gaussian expression. However, the drawback of these methods is that they are only adapted to a unique Gaussian function in the mixture.

In many works such as in [BM68; Na+00] and also in commercial peak fitting softwares [HCS03; Woj10] the problem is solved by using an iterative non-linear optimization algorithm, *e.g.*, the Gauss-Newton and the Levenberg-Marquardt algorithms [NW99]. Moreover, for some non-linear least squares algorithms (*e.g.*, trust-region and sequential quadratic programming algorithms [NW99]) it is possible to have constraints on the parameters like the non-negativity of the amplitudes. This would force the solution to be in a feasible region or would include an additional information on the parameters [Paa97].

These algorithms start from an initial solution of the parameters which is then iteratively updated till it converges towards a local minimum. Setting a good initial solution is important in order to find the best solution but this is not an easy task. Yet, it is possible to use available theoretical or laboratory results [Clé+11] (such as the theoretical spectrum of the gas). Another way is to use the inflection points in the mixture [MPB09]; they are detected from the derivative so as to obtain a first initialization of the source delays. However, this approach is very sensitive

to noise. Hereafter, we give more details on the Gauss-Newton and the Levenberg-Marquardt algorithms [NW99] for unconstrained non-linear least squares.

Let all the source parameters a_j, c_j and w_j be gathered in a vector $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \dots & \theta_{3\cdot J}\end{bmatrix} \in \mathbb{R}^{3\cdot J}$, thus the optimization problem (2.2) reads:

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\boldsymbol{r}(\boldsymbol{\theta})\|_2^2. \tag{2.4}$$

The algorithms find an update step $\boldsymbol{\delta} \in \mathbb{R}^{3\cdot J}$ at each iteration l such that:

$$\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^{(l-1)} + \boldsymbol{\delta}, \tag{2.5}$$

where $\boldsymbol{\theta}^{(l)}$ is the solution at iteration l and $\boldsymbol{\theta}^{(l-1)}$ is the solution found at the previous iteration, while $\boldsymbol{\theta}^{(0)}$ is the initialization.

Gauss-Newton algorithm

By applying the first-order Taylor approximation to $\boldsymbol{r}(\boldsymbol{\theta}^{(l-1)} + \boldsymbol{\delta})$, we get:

$$\boldsymbol{r}(\boldsymbol{\theta}^{(l-1)} + \boldsymbol{\delta}) = \boldsymbol{r}(\boldsymbol{\theta}^{(l-1)}) + \boldsymbol{J}(\boldsymbol{\theta}^{(l-1)})\boldsymbol{\delta}. \tag{2.6}$$

$\boldsymbol{J}(\boldsymbol{\theta}^{(l-1)}) \in \mathbb{R}^{N\times 3\cdot J}$ is the Jacobian matrix of $\boldsymbol{r}$ evaluated at $\boldsymbol{\theta}^{(l-1)}$:

$$\boldsymbol{J}(\boldsymbol{\theta}^{(l-1)}) = \begin{bmatrix} \frac{\partial \boldsymbol{r}(\boldsymbol{\theta}^{(l-1)})}{\partial \theta_1^{(l-1)}} & \dots & \frac{\partial \boldsymbol{r}(\boldsymbol{\theta}^{(l-1)})}{\partial \theta_{3\cdot J}^{(l-1)}} \end{bmatrix}. \tag{2.7}$$

At each iteration, the Gauss-Newton algorithm determines the step $\boldsymbol{\delta}$ by solving the following minimization problem:

$$\min_{\boldsymbol{\delta}} \left\| \boldsymbol{r}(\boldsymbol{\theta}^{(l-1)}) + \boldsymbol{J}(\boldsymbol{\theta}^{(l-1)})\boldsymbol{\delta} \right\|_2^2. \tag{2.8}$$

This is a linear least-squares problem that can be explicitly resolved at each iteration by:

$$\boldsymbol{J}^T(\boldsymbol{\theta}^{(l-1)})\boldsymbol{J}(\boldsymbol{\theta}^{(l-1)})\boldsymbol{\delta} = -\boldsymbol{J}^T(\boldsymbol{\theta}^{(l-1)})\boldsymbol{r}(\boldsymbol{\theta}^{(l-1)}). \tag{2.9}$$

Levenberg-Marquardt algorithm

This method for optimizing non-linear least squares was proposed by Marquardt [Mar63] and is, in fact, a slight variation of the initial method proposed by Levenberg [Lev44]. It is more robust to the initialization than the Gauss-Newton because when the solution is far from the minimum; the algorithm behaves more like a gradient descent algorithm, whereas the Gauss-Newton algorithm assumes the problem to be locally quadratic around each solution. Following the Levenberg-Marquardt algorithm, the step $\boldsymbol{\delta}$ is obtained from the following equation:

$$\left[\boldsymbol{J}^T(\boldsymbol{\theta}^{(l-1)})\boldsymbol{J}(\boldsymbol{\theta}^{(l-1)}) + \eta \operatorname{diag}\left(\boldsymbol{J}^T(\boldsymbol{\theta}^{(l-1)})\boldsymbol{J}(\boldsymbol{\theta}^{(l-1)})\right)\right]\boldsymbol{\delta} = -\boldsymbol{J}^T(\boldsymbol{\theta}^{(l-1)})\boldsymbol{r}(\boldsymbol{\theta}^{(l-1)}), \tag{2.10}$$

where η is a positive parameter. When η is large, the Levenberg-Marquardt is almost a gradient descent algorithm; this is useful when the current solution is far from the minimum. On the contrary,

when η is small, the Levenberg-Marquardt algorithm is almost identical to the Gauss-Newton. The value of η is automatically updated at each iteration according to the following strategy: if the criterion decreases, then η is divided by a constant (typically two), otherwise the update is discarded and η is multiplied by a constant (typically two).

2.2.2 Sparsity-aware methods

Often, spectroscopic mixtures can be expressed as linear combinations of a few sources. That is the unknown number of sources J is usually small. This leads us to sparse approximation (or sparse representation) that aims at approximating a mixture as the linear combination of a small number of elementary signals drawn from a large set of vectors gathered in a matrix denoted by the *dictionary*.

The sparse approximation of a mixture $\boldsymbol{x}$ writes as:

$$\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{\gamma}. \tag{2.11}$$

$\boldsymbol{D} \in \mathbb{R}^{N \times M}$ is the dictionary whose columns $\boldsymbol{d}_m$ are called *atoms*:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{d}_1 & \boldsymbol{d}_2 & \dots & \boldsymbol{d}_M \end{bmatrix}. \tag{2.12}$$

We suppose in the sequel that the atoms are normalized such that $\forall\ m,\ \|\boldsymbol{d}_m\|_2 = 1$. This is a very common assumption in sparse approximation [Ela10]. Moreover, the dictionary in sparse approximation is overdetermined; the number of atoms (number of columns) is much larger than the mixture length (number of rows), *i.e.*, $M \gg N$. $\boldsymbol{\gamma} = \begin{bmatrix} \alpha_1 & \dots & \alpha_M \end{bmatrix}^T \in \mathbb{R}^M$ is a sparse vector (most of its elements are zeros). We denote by $\mathcal{S}$ the support of $\boldsymbol{\gamma}$ gathering the indices of non-zero elements in $\boldsymbol{\gamma}$:

$$\mathcal{S} = \{m : \gamma_m \neq 0\}. \tag{2.13}$$

The cardinal of the support indicates the sparsity level of the vector. The indices in $\mathcal{S}$ indicate which atoms are used to approximate $\boldsymbol{x}$ and the values of the non-zeros elements in $\boldsymbol{\gamma}$ correspond to the atom weights, *i.e.*,

$$\boldsymbol{x} \approx \sum_{j \in \mathcal{S}} \gamma_j \boldsymbol{d}_j. \tag{2.14}$$

Parametric dictionary

The construction of the dictionary is a crucial element to get a good approximation of the mixture. For instance, in image processing, wavelet dictionaries are highly used [RBE10] since most natural images are approximately sparse when projected onto such dictionaries. In compressed sensing [Don06a], the atoms are required to be uncorrelated in order to have an accurate recovery of the signal when sampled at a lower rate than Nyquist's frequency. Here, randomly generated dictionaries are a simple and good choice since it has proven that they respect the restricted isometry propriety [EK12] which ensures that the atoms extracted from the dictionary are almost orthonormal [CRT06; Don06a]. However, verifying this property is an NP-hard problem [TP14], a simpler metric to measure the non-correlation of a dictionary is the mutual coherence defined as

the maximum inner product between two distinct normalized atoms, *i.e.*, :

$$\mu_D \triangleq \max_{m \neq m'} |d_m^T d_{m'}|. \tag{2.15}$$

μ_D varies between 0 and 1, and the dictionary is more correlated as this values increases.

For parameter estimation, the dictionary is constructed by sampling the parameters of interest (mostly the non-linear ones) over a predefined grid characterized by a sampling step [Aus+10; YDD09]. In other words, each atom corresponds to the parametric model evaluated at a certain value on the grid. The main drawback of such an approach is that the parameters are discretely estimated.

An example of a parametric dictionary can be seen in spectral line estimation [BCI07; DB13] where the goal is to estimate the unknown frequencies of multiple mixed sinusoidal signals. Here, the atoms are sinusoidal signals, each with a certain grid frequency. Another example of using a parametric dictionary is sparse spike deconvolution [Bou+11; DP15; DM05; DP17]. Here all the sources in (2.1) have the same known shape parameter w and the goal is to retrieve their delays and amplitudes and thus the dictionary gathers the delayed source versions. In [BMS11] a parametric dictionary is constructed for estimating the main features of astrophysical mixtures (MUSE-like spectra). A mixture is modeled as the superposition of three features: emission and absorption peaks with unknown width and position, step spectrum with an unknown center to model the mixtures discontinuity and a continuum modeled by sinusoids with unknown frequencies. To find the unknown parameters for each feature, the dictionary is composed of three sub-dictionaries that respectively sample the parameters of the three features.

Sparse approximation

The sparse solution γ is the one that allows to fit at best a mixture with the smallest number of atoms. Mathematically speaking, this corresponds to solve the following optimization problem:

$$\min_{\gamma} \|\gamma\|_0 \quad \text{such that} \quad \|\boldsymbol{x} - \boldsymbol{D}\gamma\|_2^2 \leq \epsilon, \tag{2.16}$$

where $\|\cdot\|_0$ is the ℓ_0 pseudo-norm that counts the number of non-zero elements in a vector and ϵ is a small positive constant. In this formulation, the desired sparsity level is unknown, rather an upper bound is set on the residual norm. An equivalent formulation sets a maximal sparsity level constraint K on the sparse vector:

$$\min_{\gamma} \|\boldsymbol{x} - \boldsymbol{D}\gamma\|_2^2 \quad \text{such that} \quad \|\gamma\|_0 \leq K. \tag{2.17}$$

A third formulation, which is not equivalent to both previous formulations (from an optimization point of view), uses a regularization to trade-off between the data-fit term and the sparsity level:

$$\min_{\gamma} \|\boldsymbol{x} - \boldsymbol{D}\gamma\|_2^2 + \eta \|\gamma\|_0 . \tag{2.18}$$

Indeed, as the value of the regularization term η gets larger, the sparsity level decreases since it would cost more to add an atom to the support.

The ℓ_0 optimization problem is known to be NP-hard [AK98]. The optimal solution can be obtained by an exhaustive search over all the possibilities with the desired sparsity level, which is only tractable for very small scale problems. Alternative solutions have been proposed over the past years, the two main method categories are presented in the sequel.

Greedy algorithms

These methods provide an approximated solution to the ℓ_0 optimization problem (mostly to the problem (2.17)) and construct the solution iteratively. At each iteration, one or more atoms that yield the best improvement are added to the solution support: this is a *forward* strategy. A *forward-backward* strategy may add or remove atoms at each iteration. Most of the greedy algorithms share the following structure: selection or withdraw of an atom, estimation of the non-zero elements of the sparse vector, and residual update. The main difference between the methods lies in the atom selection step and the weight estimation strategies.

Matching Pursuit (MP) [MZ93] is a simple iterative algorithm: at each iteration, the atom that is the most correlated to the residual $\boldsymbol{r}$ is selected (the residual is defined as the mixture $\boldsymbol{x}$ from which the previously selected atoms have been removed). This is done by maximizing the inner-product between the residual and the normalized atoms $\boldsymbol{d}_m$:

$$\widehat{m} = \underset{m}{\operatorname{argmax}} \, |\boldsymbol{d}_m^T \boldsymbol{r}|. \tag{2.19}$$

Then, the weight of the newly selected atom is estimated by $\widehat{\gamma}_{\widehat{m}} = \boldsymbol{d}_{\widehat{m}}^T \boldsymbol{r}$ and the residual is updated. MP has a low computational time but is known to propagate erroneous atom for highly correlated dictionaries [Tro04] and to re-select the same atoms several times.

Orthogonal Matching Pursuit (OMP) [PRK93; Tro04] tries to overcome the MP limitations by projecting, at each iteration, the mixture onto the space defined by the currently selected atoms:

$$\widehat{\boldsymbol{\gamma}}_{\mathcal{S}} = \underset{\boldsymbol{\gamma}_{\mathcal{S}}}{\operatorname{argmin}} \, \|\boldsymbol{x} - \boldsymbol{D}_{:\mathcal{S}} \boldsymbol{\gamma}_{\mathcal{S}}\|_2^2, \tag{2.20}$$

where $\boldsymbol{\gamma}_{\mathcal{S}} = \{\gamma_m; m \in \mathcal{S}\}$. This step is computationally expensive, relatively to MP, since it includes a matrix inversion. Assuming that $\boldsymbol{D}_{:\mathcal{S}}$ is full column rank, the selected atom weights estimation writes:

$$\widehat{\boldsymbol{\gamma}}_{\mathcal{S}} = (\boldsymbol{D}_{:\mathcal{S}}^T \boldsymbol{D}_{:\mathcal{S}})^{-1} \boldsymbol{D}_{:\mathcal{S}}^T \boldsymbol{x}. \tag{2.21}$$

Nevertheless, strategies like QR or Cholesky matrix factorizations can be used to efficiently implement the iterative inversions [SC12]. In addition, non-negative weight constraint can be considered [BEZ08; YWD15; Ngu+17].

Orthogonal Least Squares (OLS) [CBL89] selects the atom that minimizes the residual error. This implies the resolution of a linear least squares problem for each tested atom to estimate its weight in addition to the weights of the atoms which are in the support. Obviously, this requires high computation time. The difference between OMP and OLS is addressed in [BD07].

The Single Best Replacement algorithm (SBR) [Sou+11] aims at minimizing the criterion in (2.18). This algorithm can be interpreted as a forward–backward extension of OLS where each iteration consists in choosing between adding or retiring an atom.

All the greedy algorithms presented above start with an empty support, but other algorithms start from an initial solution. Orthogonal Matching Pursuit with Replacement (OMPR) [JTD11] adds and retires one atom from the support at each iteration. First, the best atom in a gradient descent sense is added and then a threshold step retires the atom with the lowest contribution (in the gradient decent sense) from the support.

SWAP [VB13] computes the residual error for each switch between an atom in the support and all the other atoms and the one yielding the lowest criterion is considered. This operation is repeated until no further decrease of the criterion is possible. The authors present this algorithm as a solution when the dictionary is highly correlated, however, this algorithm is very time-consuming.

Another popular algorithm that can be classified as a greedy algorithm (in the wide sense) is the Iterative Hard Thresholding (IHT) [BD09] that optimizes the problem (2.18). The main IHT iteration consists of a gradient descent step followed by hard thresholding. However, this method performs poorly in the presence of noise [BD09, Section 8].

More details on greedy pursuit algorithms can be found in [TGS06], and a comparative study is given in [RS09].

ℓ_0 relaxation algorithms

The idea of this family of methods is to replace the non-convex ℓ_0 pseudo-norm by another norm or function which are easier to solve and favor sparse solutions. We distinguish between the ℓ_1 convex relaxation and non-convex relaxations.

The ℓ_1 norm is the closest convex norm to the ℓ_0 pseudo-norm. It calculates the absolute sum of the vector entries, *i.e.*, $\|\boldsymbol{\gamma}\|_1 = \sum_{m=1}^{M} |\gamma_m|$ and it has been proven that its minimization favors sparse solutions [DE03]. Especially, relaxing the ℓ_0 optimization problem in (2.18), leads to the Lasso problem:

$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\gamma}\|_2^2 + \eta\|\boldsymbol{\gamma}\|_1. \tag{2.22}$$

The solution of the ℓ_1 relaxed optimization problem is identical to the initial ℓ_0 optimization problem under some conditions [Don06b]. Nevertheless, the ℓ_1 relaxation algorithms can be used to yield an approximated solution to the original ℓ_0 optimization problem.

Despite, the convexity of the ℓ_1 norm, the optimization remains challenging because the ℓ_1-norm is not smooth. A first family of optimization methods reformulate the problem (2.22) as a smooth differentiable constrained optimization problem. Some methods in this family are based on gradient projection [FNW07], interior-points [Kim+07] or alternating direction [YZ11] algorithms. A second family of methods exploits the proximal operator [PB14] to iteratively solve computationally easier sub-problems than the original ℓ_1 optimization problem. A popular algorithm in this family is the Iterative Shrinkage Thresholding Algorithm (ISTA) [DDDM04]. At iteration l, ISTA calculates the sparse solution $\boldsymbol{\gamma}^{(l)}$ as:

$$\boldsymbol{\gamma}^{(l)} = \text{shrink}\left(\boldsymbol{\gamma}^{(l-1)} - \eta\boldsymbol{D}^T(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\gamma}^{(l-1)}), \gamma\eta\right), \tag{2.23}$$

where η is the step size. The operator shrink is the element-wise soft thresholding operator defined as $\text{shrink}(a, b) = \text{sign}(a)\max\{|a| - b, 0\}$ where a and b are constants. This operator is the proximal

operator for the ℓ_1-norm. Fast ISTA (FISTA) [BT09] is an improvement of ISTA that accelerates its convergence. A survey on the different ℓ_1 relaxation methods can be found in [Zha+15].

A number of non-convex relaxation approaches, such as the log penalty [Fri12], the smoothly clipped absolute deviation penalty [FL01] and the minimax concave penalty [Zha10] have also been proposed.

Continuous sparse approximation

One of the limitations of parameter estimation using a sparse approximation method is that the parameters are estimated among a set of discretized values. Obviously, a small sampling step yields a better estimation but yields a large dictionary. Therefore, the performance of sparse approximation algorithms is negatively affected since the dictionary mutual coherence may increases. Also, the computation time increases and additional memory resources may be required as the dictionary size increases. In the recent years, multiple methods have been proposed to get a continuous estimation using sparse approximation techniques.

In [ETS11], the Continuous Basis Pursuit (C-BP) algorithm is proposed to estimate the delays. The idea is to generate a dictionary with auxiliary interpolation functions to express local translations of the parameterized function. The first studied interpolation is based on the Taylor approximation which is motivated by the fact that delays can be linearized using the source derivative. Specifically, let c_{ij} be the continuous delay and $\widetilde{c}_{ij}$ is its discrete approximation, then $c_{ij} = \widetilde{c}_{ij} + \psi$ with ψ is the discrete estimation error. Following the first order Taylor approximation, the error ψ can be linearized such that:

$$s(\lambda - c_{ij}; w_{ij}) = s(\lambda - \widetilde{c}_{ij} - \psi; w_j) \approx s(\lambda - \widetilde{c}_{ij}; w_j) - \psi \dot{s}(\lambda - \widetilde{c}_{ij}; w_j), \tag{2.24}$$

where $\dot{s}(\lambda; w_j)$ is the derivative of $s(\lambda; w_j)$. Indeed by finding the discrete estimation $\widetilde{c}_{ij}$ and by using $s(\lambda - c_{ij}; w_{ij})$ as the data, one can find ψ to get a continuous estimation. Therefore, the mixture is approximated as $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{\gamma} - \dot{\boldsymbol{D}}\dot{\boldsymbol{\gamma}}$ where $\boldsymbol{D}$ and $\dot{\boldsymbol{D}}$ respectively gather the delayed sources and their derivatives. This resulting optimization problem [ETS11, equation (13)] is solved using a a standard ℓ_1 optimization method (*e.g.*, interior points methods). The C-BP algorithm is discussed and evaluated in [DP17]. In [CPF14] the Taylor interpolation is used to estimate parameters of dimension greater than 1 and the resulting criterion is minimized using a proximal iterative algorithm.

Another interpolation strategy is proposed in [ETS11]: the so-called "polar" interpolation. This strategy empirically outperforms the Taylor interpolation as shown in [ETS11]. The idea comes from the observation that the manifold of shifted signals lies on the surface of a hypersphere since the ℓ_2 norm of the atoms is preserved under translation. Thus, any segment on the manifold can be approximated by an arc, in particular, the segment induced by three adjacent atoms. In [FDJ15; Knu+14] the "polar" interpolation is extended to a greedy algorithm inspired from OMP. In [PPZ17] a polynomial interpolation method, and in [ZY17] a least squares gradient-based method is used as post-processing to estimate the continuous parameters.

Although the above methods yield a continuous estimation, they are implicitly based on a finite dictionary and thus still dependent on the grid sampling step so that the parameters may not be recovered perfectly. Methods based on the minimization of the atomic norm [Tan+12; BTR13] have

been proposed without predefining a grid. However, this approach is mainly adapted to spectral line estimation problem.

2.2.3 Time-delay estimation

Time delay estimation methods estimate the delays when the shape of the sources is known. Applications can be found in radar, sonar and audio signal processing [Car87]. A classical method is to determine the delay that maximizes the cross-correlation between the delayed mixture and a reference non-delayed mixture [KC76]. This method yields discrete estimations governed by the resolution of the cross-correlation function. In the presence of white noise, this method is equivalent to the optimal matched filter. However, in the case of multiple sources, this method fails since multiple peaks appear in the cross-correlation.

Another approach is to convert the delay estimation problem to phase estimation problem by using the Fourier transform leading to a line spectral estimation problem [BTR13]. Traditional approaches to recover signals are the Multiple Signal Classification (MUSIC) algorithm [Sch86] or the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) algorithm [RK89]. However, these methods usually suppose the signals to be spatially uncorrelated and their performance would decrease when the sources are highly correlated [YX16]. Another strategy is the use of sparse approximation techniques using a dictionary of pure frequencies as shown in the previous section.

2.2.4 Stochastic methods

Stochastic optimization is a powerful tool for optimizing criteria which are very costly or slow to evaluate [BCDF10]. The maximum entropy [UB75], the maximum-likelihood [Cap79] and the Bayesian approaches [Ben00; Bre13; Maz05; HG02; RFA03; FD05] are widespread stochastic methods.

In [Fli+05; VC14] the galaxy kinematics is considered. The mixtures from each spatial position are processed independently and a single Gaussian parameterized source is considered. The source parameters are estimated using an expectation-maximization algorithm in [Fli+05]. A Monte Carlo Markov Chain (MCMC) algorithm is used in [VC14] in addition to an image deconvolution task to remove the blur effect of the PSF induced by the acquisition instrument. Also, in [Maz05, Chapter 4] the estimation of multiple source parameters in a single mixture is done with an MCMC algorithm. A Reversible Jump MCMC (RJMCMC) algorithm [Gre95] is used in [HG02; RFA03; FD05]. This algorithm permits the simulation of the posterior distribution of varying dimension. The difference between these methods lies in the choice of the priors. Even though the statistical methods are effective and have attractive convergence guarantees, they suffer from high computational time [HKW16].

2.3 Parameter association

At this stage, we suppose that we have a set of $I \times J$ value for each of the three parameters (amplitudes a_{ij}, delays c_{ij} and shapes w_{ij}), which are independently estimated for each mixture.

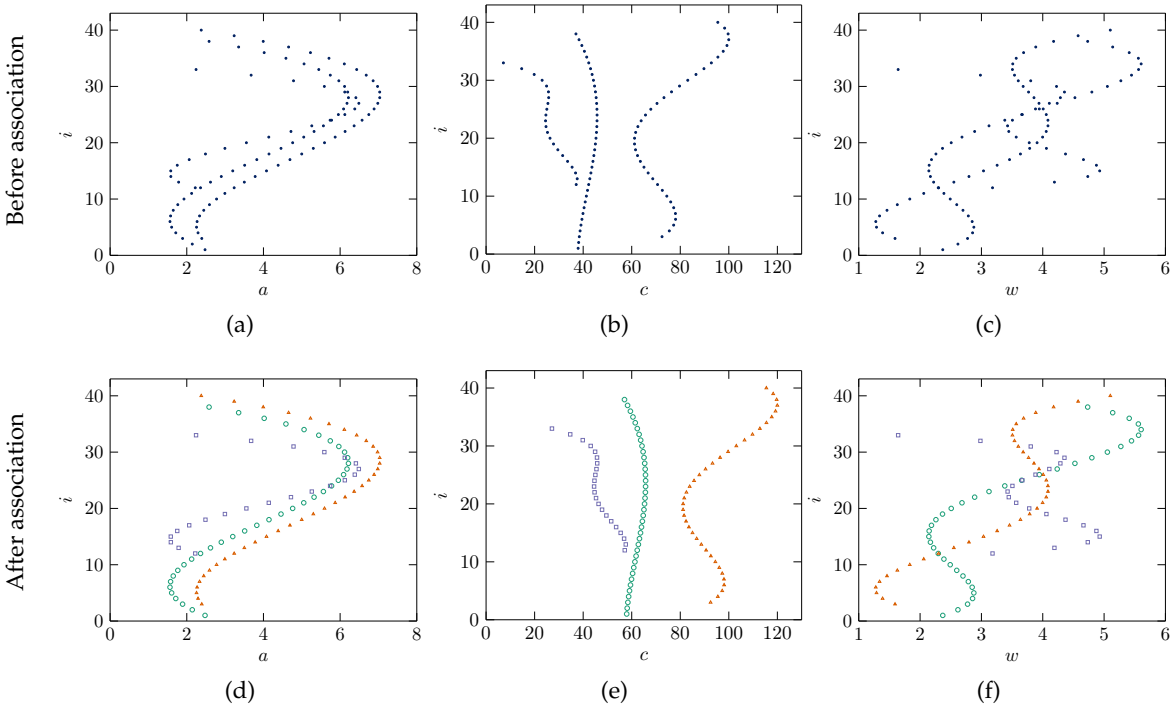


Figure 2.1: First row: a possible output of estimating the parameters independently in each mixture of Figure 1.11d. Second row: the expected association where each color and marker corresponds to a source.

Indeed, nothing imposes that the parameters with the same index j belong to the same physical source due to the independent processing of the mixtures. Therefore, a post-processing is required to associate the independently estimated parameters in such way to give a joint meaning for the index j; all the parameters sharing the same index j belong to a slow evolving source within the mixture. Let us consider again the mixtures in Figure 1.11. A possible estimation of the source parameters is given in the first row of Figure 2.1 and we wish to associate the parameters in order to have the results of the second row in Figure 2.1.

This problem can be assimilated to multiple object tracking where the objects are data points (scatters). They usually evolve with respect to time instants. Tracking can be found in radar to detect targets [Bla86] or in computer vision to follow the trajectories of vehicles [Coi+98]. The works in this literature consider two main steps: object detection and tracking. In this section, we concentrate on the tracking step that yields "trajectories" of the objects which are so often smooth. Hereafter, we suppose that we deal with a temporal sequence of data point objects representing the estimated parameters, whilst the time instants in our case are equivalent to the discrete mixture index i. The methods used for this purpose can be grouped into deterministic and statistical methods.

The deterministic methods define a cost of associating a source in the mixture i to one of the other sources in the mixture $i-1$. The cost usually measures the distance between the estimates (*e.g.* their Euclidean distance). However, the assignment problem is a combinatorial problem. Yet, good solutions can be found using adapted assignment algorithms such as the Hungarian [Kuh55],

greedy [AOT00] or metaheuristic algorithms [BR03].

In [MSD13], similar synthetic 1D sets of mixtures to the ones used in this thesis are considered. After estimating the source parameters using the SBR algorithm, a matrix $\boldsymbol{P}^{J\times J}$ is computed, whose elements $p_{jj'}$ represent the cost of associating the source $j \in \{1,\dots,J\}$ in the mixture i to the source $j' \in \{1,\dots,J\}$ in the mixture $i-1$:

$$p_{jj'} = \frac{(a_{ij} - a_{(i-1)j'})^2}{h_a} + \frac{(c_{ij} - c_{(i-1)j'})^2}{h_c} + \frac{(w_{ij} - w_{(i-1)j'})^2}{h_w}, \tag{2.25}$$

where h_a, h_c and h_w are constants that are manually tuned. Then, the Hungarian algorithm finds the best match and the same process is repeated for the other mixtures sequentially. The assignment methods may suffer from high computation time especially when the number of sources is larger. Also, an erroneous association for a mixture leads to a propagation error due to the sequential assignment.

Another deterministic strategy, often used for spectroscopic signals, is Dynamic Time Warping method (DTW) [Mül07], that is a strategy to cancel the delays by aligning the mixtures. A cost matrix of dimension $N \times N$ (N is the length of a mixture) must be calculated to measure the distance between mixture samples. Then, a discrete optimization algorithm is used to solve the assignment problem, that is to find the best path with the least global distance, which is the sum of costs along the path. An example of a cost matrix between two synthetic and noisy mixtures with 2 sources is represented in Figure 2.2. In Figure 2.3 the alignment result of DTW is displayed. However, DTW can only be applied to two mixtures. Also, this method only cancels the delays and does not yield an explicit delay estimation.

Stochastic methods, especially, the Kalman [Hay04] and the particle filters [Aru+02] are very popular methods to achieve multiple object tracking since they allow to model the uncertainties in the model. These algorithms are based on two steps; prediction and correction. The first step predicts the parameters to track (*e.g.*, the delays in a mixture) in a mixture i by using a predefined parameter evolution model and the estimated parameters at the previous mixture $i-1$. Let $\boldsymbol{\theta}_i$ be the vector of parameters that we want to track, the prediction step writes:

$$\boldsymbol{\theta}_i = f(\boldsymbol{\theta}_{i-1}) + \boldsymbol{v}_i, \tag{2.26}$$

where f is the parameter evolution model and $\boldsymbol{v}_i$ is the process noise vector. The correction step, corrects the predicted parameters $\boldsymbol{\theta}_i$ by exploiting the actual measurements at index i given in $\boldsymbol{y}_i$:

$$\boldsymbol{y}_i = g(\boldsymbol{\theta}_i) + \boldsymbol{b}_i, \tag{2.27}$$

where g is the model mapping the observation to the parameters and $\boldsymbol{b}_i$ is the measurement noise vector. The Kalman filter considers that the functions f and g are linear and that the process and measurement noise are Gaussian random variables. It starts from an initial guess and the parameters are then sequentially estimated. If the models representing the parameter evolution and the model mapping are non-linear, the Taylor series expansion is used to linearize them, yielding the extended Kalman filter [Hay04]. However, the drawback of this method lies in the need to predefine the dynamic models analytically, which is not always possible.

Contrary to the Kalman filter, the particle filter [VDM+01] can deal with non-linear function

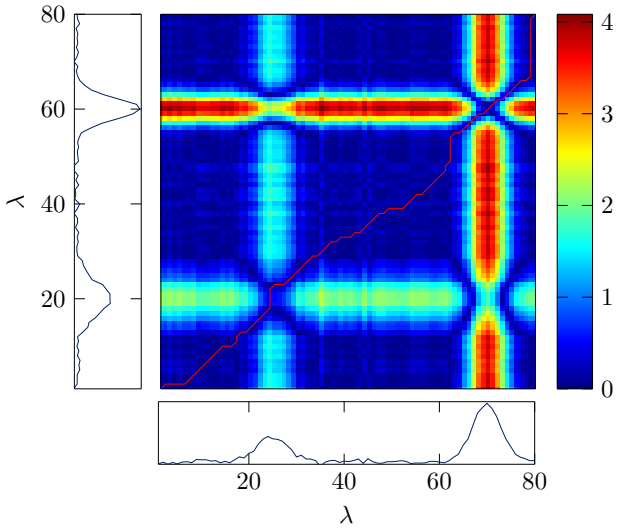
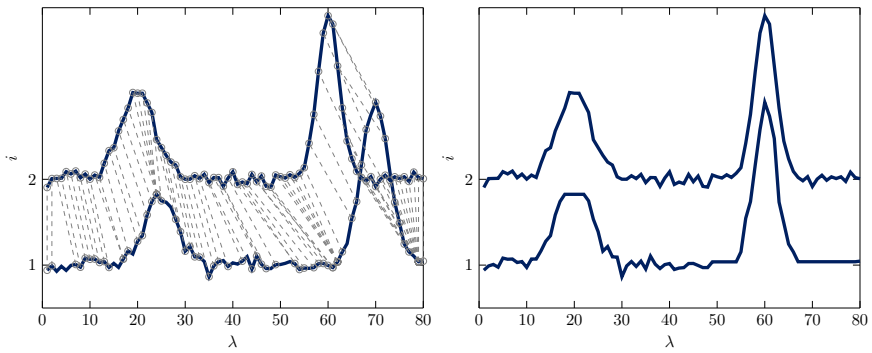


Figure 2.2: An example of a dynamic time warping cost matrix calculated using the Manhattan distance (sum of absolute values of the difference between the mixtures samples). The red line shows the best assignment path.



(a) (b)

Figure 2.3: (a) the generated two mixtures and the DTW assignments of each sample of the first mixture to a sample of the second mixture. (b) The result of the DWT where the first mixture $i = 1$ is aligned with the second mixture $i = 2$.

and with non-Gaussian noise distributions. The particle filter uses simulation methods, *e.g.*, Monte-Carlo sampling strategy to generate estimates of the parameters. The parameters are estimated in the maximum likelihood sense and are propagated to the next mixture. The quality of the estimation is related to the number of generated particles and thus this method can be computationally expensive if an accurate estimation is required.

2.4 Joint parameterized source decomposition

Addressing independently the problems of parameter estimation and association is not ideal because such a strategy does not exploit the high similarity between two adjacent mixtures. Indeed,

both problems are dependent since the parameters evolve slowly from one mixture to another one. In addition to this, a two-step approach may lead to error propagation. In this section, we present methods that address the parameter estimation and association jointly (one-step strategy). In particular, we present the Bayesian and joint sparse approximation frameworks.

2.4.1 Bayesian Framework

In [Maz11] the problem of decomposing a sequence of a 1D set of mixtures is considered. The mixtures are modeled as the noisy sum of Gaussian functions (the number of sources is supposed to be known), and the peak parameters are supposed to evolve slowly from a mixture to its adjacent ones. To estimate the source parameters, a Bayesian generative model is used. Priors are set on each parameter to impose some physical information and to impose their slow evolution from one mixture to another one by using smoothness derivate operator. An MCMC algorithm is proposed, where the source parameters are updated using a Gibbs sampler. In particular, the delays c and shapes w are sampled with a random walk Metropolis-Hastings algorithm and the amplitudes a are directly sampled from a Gaussian posterior distribution.

This previous work is extended in [Maz+15] to estimate the number of sources in each mixture. For this purpose, a RJMCMC [Gre95] algorithm is used. In addition, Markovian priors are set on the parameters to ensure their slow evolution. However, the methods in [Maz11; Maz+15] would suffer from high computation time when the data dimension becomes large and are only adapted to a 1D set of mixtures (their extension to the case of a 2D sets of mixtures is not straightforward).

2.4.2 Joint sparse approximation

In Section 2.2.2, we discussed the problem of sparse approximation of a single mixture. Given I mixtures, we can rewrite the I sparse approximations problems (2.17) together as:

$$\forall\, i, \qquad \min_{\boldsymbol{\gamma}_i} \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\gamma}_i\|_2^2 \;\text{ such that }\; \|\boldsymbol{\gamma}_i\|_0 \leq K. \tag{2.28}$$

This is an independent mixture approximation strategy that does not exploit the relationship between the mixtures.

When multiple mixtures can be sparse approximated by the same dictionary $\boldsymbol{D}$, the sparse approximation problem can be written as:

$$\boldsymbol{X} = \boldsymbol{D}\boldsymbol{\Gamma}, \tag{2.29}$$

where $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \dots & \boldsymbol{x}_I\end{bmatrix} \in \mathbb{R}^{N\times I}$ is the set of observed mixtures, and $\boldsymbol{\Gamma} = \begin{bmatrix}\boldsymbol{\gamma}_1 & \dots & \boldsymbol{\gamma}_I\end{bmatrix} \in \mathbb{R}^{M\times I}$ is the set of the sparse approximation vectors corresponding to each mixture after being decomposed following the dictionary $\boldsymbol{D} \in \mathbb{R}^{N\times M}$.

Most of the works that consider joint sparse approximation are based on the assumption that all the sparse vectors share a common support, this is known as the Multiple Measurement Vector (MMV) problem [TGS06]. Several methods developed for the case of a single mixture are extended to the MMV problem. Simultaneous OMP (S-OMP) [TGS06] is an extension of OMP. In [BDB14] SBR is also extended. Five other greedy algorithms are extended in [Bla+14] and a survey on greedy

extensions can be found in [Rak11]. However, the assumption that all the observed signals must share the same support is too restrictive in many applications, especially when the parameters vary from one mixture to another such as the delays in the considered thesis applications.

Dynamic approaches where the support varies from one sparse vector to another can be found in the literature. Recursive sparse approximation [VZ16] compute the sparse approximation of a mixture from the knowledge of the sparse approximation of the previous spectrum. More precisely, the new support $\mathcal{S}_i$ resulting from the sparse approximation of mixture $\boldsymbol{x}_i$ is calculated from the support of the previously approximated mixture such that $\mathcal{S}_i = \mathcal{S}_{i-1} \cup T \setminus P$ where T and P respectively gather the set of indices to add and retire from the support. A review on recursive sparse approximation methods can be found in [VZ16].

The concept of social sparsity was introduced in [KSD13] by promoting a structure between the sparse approximation of consecutive signals with less restrictive support assumptions. However, this approach requires a precise definition of the neighborhood which is not always possible.

2.5 Source separation

In its general definition, source separation aims at recovering source signals from their noisy mixtures by exploiting all the available information on the mixing process and the sources [CJ10]. The sources are usually non-parameterized, therefore, in this section, the parameterized source notation $s(\lambda; w_j)$ is replaced by $s_j(\lambda)$ which is a non-parameterized signal. This problem can be found in many fields: in audio signal processing [YR04; Vin+14], the goal would be to retrieve the voice of multiple persons having a conversation from their recordings (the so-called *cocktail party problem*). In digital communications [TVP96], the goal would be to separate signals containing information from different emitters. In astrophysics, the goal would be to retrieve several astrophysical components from their hyperspectral image [Bob+15].

2.5.1 Mixing model

The mixing model describes the relationship between the mixtures and the sources. The most general mixing model is the convolutive model, where each mixture is the sum of J convolution product between the sources and filters with unknown impulsion response $h_{ij}(\lambda)$:

$$\forall\, i, \qquad x_i(\lambda) = \sum_{j=1}^{J} (h_{ij} * s_j)(\lambda) + n_i(\lambda). \tag{2.30}$$

This model is not in the scope of this thesis. However, a review of works considering this model can be found in [CJ10].

Several special cases of the convolutive model are studied in the literature. Especially, the instantaneous mixing model is the most used one. Another special case is the delayed (or anechoic) mixing model that interests us the most since it allows to model the delays. These two special cases are discussed in the sequel.

Instantaneous source separation

The simplest model of source separation is the instantaneous model: the filter $h_{ij}(\lambda)$ is a unshifted Dirac of amplitude a_{ij}, thus each mixture $x_i(\lambda)$ is the linear combination of the source $s_j(\lambda)$:

$$\forall\, i, \qquad x_i(\lambda) = \sum_{j=1}^{J} a_{ij} s_j(\lambda) + n_i(\lambda). \tag{2.31}$$

To resolve this ill-posed inverse problem, additional information must be made on the sources. The Independent Component Analysis (ICA) separation framework assumes that the sources are statistically independent and that their distributions are non-Gaussian. These family of methods tries to find the mixing matrix $\boldsymbol{A}$ in such a way that the sources are as independent as possible. Among the many possible measurements used to measure the statistical independence *e.g.*, one find the Kullback–Leibler divergence, the mutual information or higher-order statistics [LAC97]. Most of the methods that have been proposed for the ICA problem exploit second or higher-order statistics to discriminate the sources [CJ10]. However, the assumption of statistical independence is very restrictive in many real-world applications. Some methods reduce the importance of this assumption by introducing the non-negativity constraint such as in spectral decomposition [MBI05] or image processing [Lee+01].

The Alternating Least Squares (ALS) is a strategy often used when the amplitudes and source are non-negative: the amplitude and sources are estimated in two distinct steps. This approach guarantees the convergence to a stationary point [GS00], but it requires to solve a non-negative linear least squares optimizations which can be computationally expensive.

The Non-negative Matrix Factorization (NMF) [ZC07] also alternates between the amplitude and source estimation but presents a computationally cheaper solution by exploiting the multiplicative update rules [AHK12]. This strategy updates the amplitudes and sources in a gradient descent fashion but the solution obtained is not unique.

Delayed source separation

In between the general convolutive and the simplest instantaneous mixtures, lies the delayed source separation mixing model, also known as the anechoic mixing model. The mixture model is the linear combination of delayed and scaled versions of sources, without permitting multiple occurrences of the same source in the mixtures. For this model, the filter impulse responses are given by $h_{ij}(\lambda) = a_{ij}\delta_{c_{ij}}(\lambda)$ where $\delta_{c_{ij}}(\lambda)$ is the Dirac delta function positioned on c_{ij}. Figure 2.4 displays a scheme for the model of a mixture for spectroscopic bell-shaped sources. The corresponding mixing model can be written as:

$$\forall\, i, \qquad x_i(\lambda) = \sum_{j=1}^{J} a_{ij} s_j(\lambda - c_{ij}) + n_i(\lambda). \tag{2.32}$$

Note that the source shapes do not vary within the mixtures ($s_j(\lambda)$ is independent from the index i).

Usually, delayed source separation is proposed for under-determined applications (when the number of mixtures is lower than the number of sources). Additional assumptions are needed in

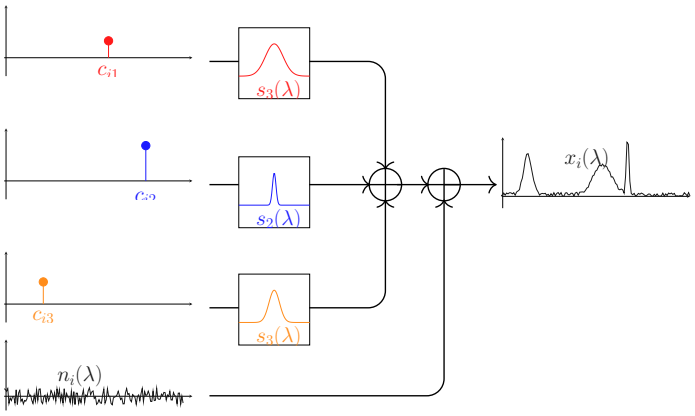


Figure 2.4: Delayed source separation mixing mode for the ith mixture of $J = 3$ sources.

this case due to the infinite number of solutions. In [JRY00] the W-Disjoint Orthogonality (WDO) assumption first introduced to discriminate the sources; they are assumed to be orthogonal in the frequency domain. The DUET algorithm proposed in [YR04] also make the WDO assumption and retrieves the delays and amplitudes in the case of two mixtures by calculating the ratio of the time-frequency representation of the mixtures. The DEMIX anechoic method proposed in [AGB10] relies on the assumption that in the neighborhood of some time-frequency points, only one source contributes to the mixture. However, the under-determined setting does not correspond to the considered spectroscopic data and assumption such as the WDO are not valid since the sources may spectrally overlap and thus cannot be orthogonal in the frequency domain.

Many methods have been also proposed for the over-determined applications (when the number of mixtures is larger than the number of sources). In [CB06; JF11] the Taylor expansion is applied to the mixtures in the temporal domain to linearize the delays and hence use instantaneous source separation methods. Another linearization strategy is to analyze the mixtures in the frequency [MMH07a; Oue+14] or time-frequency [Nio+; OG11] domains, so that delays become phases which allows using instantaneous source separation approaches coupled with phase estimation methods.

The above methods are based on the independence and non-correlation assumptions and may be valid in many real-world applications [Dua10]. Efforts have been done though to relax these assumptions. In [PD05], the AD-TiFROM algorithm simplifies the strong W-disjoint orthogonality assumption needed in the DUET algorithm [YR04] and the assumption that the sources must be orthogonal in the time-frequency domain is relaxed: each source must be dominant at least in one time-frequency window, AD-TiFROM main task is to detect these windows. However, this assumption becomes invalid when the evolution of a source delay is fast from one mixture to another, o or when two delayed source signals significantly overlap in some mixtures. For spectroscopic signals, in particular, some delayed source separation methods have been proposed. In [HHL03; HH03], a time warping strategy was used to cancel the effect of the delays on each mixture after finding the delays over a predefined discrete grid using an exhaustive search strategy. The sources and amplitudes were then estimated in the linear least squares sense. However, the exhaustive search step is not feasible when the delays take too many values. In [MMH07b], NMF was extended to consider delays: non-negative amplitudes and sources were found by using

multiplicative updates, and delays were estimated with a gradient descent algorithm coupled with a maximizing cross-correlation procedure to reduce the effect of local minima.

2.6 Conclusion

In this chapter, we presented state-of-the art methods that aim at estimating and associating the source parameters (amplitudes, delays and shapes) in a set of mixtures. Methods from two strategies are presented. In the first one, the parameters are estimated independently for each mixture, then the parameters are associated together in order to have a slowly evolving parameter set for each source. However, a two-step strategy does not exploit the relation between the mixtures during the estimation step. In the second one, the parameter estimation and association are done jointly which present a more optimal solution than a two-step strategy.

In this thesis, we adopt a one-step source separation framework and we explicitly exploit the knowledge of the parametrized source, thus the source estimation is recast to the estimation of their shape parameters. This knowledge would allow us to significantly relax the strong assumptions often made in the source separation framework such as the source statistical independence and non-correlation. Three one-step methods are proposed in the following three chapters, the first two (Chapters 3 and 4) are based on sparse approximation strategies whereas the third one (Chapters 5) is based on B-spline parameter modeling.

3

Separation of parameterized and delayed sources

3.1 Introduction

This chapter presents a first approach to address the separation of delayed and parameterized sources. The source parameters (amplitudes, delays and shapes) are jointly estimated and associated in a one-step strategy. Specifically, the sources are only discriminated by the difference of their shape parameters which are assumed to be constant for each source. So, in this chapter, we do not impose the slow evolution of any parameter within the mixtures.

We address the problem as a constrained optimization problem of a data-fit criterion (Section 3.2). The latter is minimized using an Alternating Least Squares (ALS) scheme of two steps (Section 3.3). First, the shape parameters are estimated with the Levenberg-Marquardt algorithm. Second, an algorithm inspired from Orthogonal Matching Pursuit (OMP) estimates the amplitudes the delays and it is designed in such a way to associate the sources from different mixtures following their shape similarities (Section 3.4). Results on synthetic 1D and 2D sets of mixtures, in addition to a discussion about some of the method settings, are presented in Section 3.5.

3.2 Model and criterion

3.2.1 Model and assumptions

Let us recall the mixing model of delayed and parameterized sources:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[c_{ij}; w_j\right] + \boldsymbol{n}_i. \tag{3.1}$$

Here, i is the index of a mixture (corresponding to either a 1D or 2D position) and all the parameters sharing the same index j belong to the same source. Besides, this model does not permit multiple occurrences of the same source in the mixtures and each source has a unique shape parameters (w_j is independent from i).

We consider that the amplitudes are non-negative ($a_{ij} \geq 0$). Without loss of generality, we suppose that the sources are all parameterized with the same function and that the corresponding shape parameter is a scalar. Specifically, in this chapter, we make the following assumptions which are motivated by the sparse approximation method used in the sequel:

- the delays c_{ij} are discretized over a grid with a sampling step Δ:

$$c_{ij} = \ell_{ij}\Delta, \tag{3.2}$$

  where $\ell_{ij} \in \mathbb{N}$ is the delay index. Note that the sampling step Δ can be smaller than the mixture sampling step *i.e.*, $\Delta < 1$, so that a higher allocation resolution can be obtained for the delays.

- the sources are energy-normalized and invariant to translation (their unit ℓ_2-norm is preserved when delayed):

$$\forall\, i,\ j, \qquad \left\|\boldsymbol{s}\left[c_{ij}; w_j\right]\right\|_2 = 1. \tag{3.3}$$

  As stated in Section 1.4.1, the source normalization allows us to overcome the scale indetermination. Recall that the sources are centered around their delays c_{ij} allowing us to overcome the delay indetermination.

Finally, the model considering the above assumptions now reads:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; w_j\right] + \boldsymbol{n}_i, \tag{3.4}$$

where the normalization factors $\frac{1}{\left\|\boldsymbol{s}\left[c_{ij};w_j\right]\right\|_2}$ are merged with the amplitudes a_{ij}.

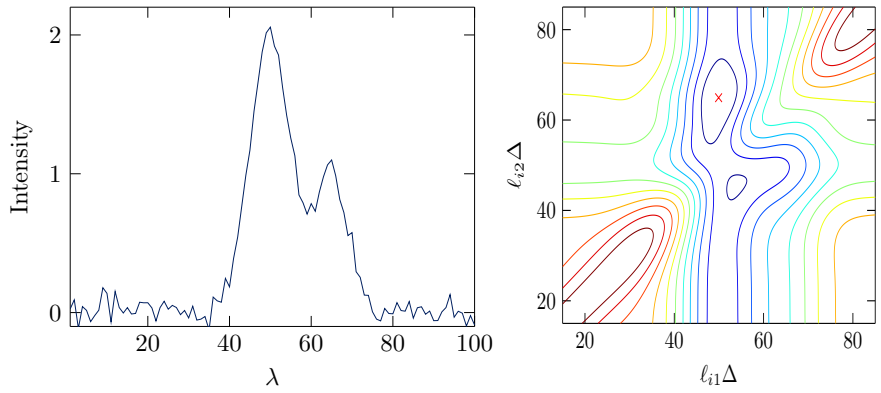
3.2.2 Criterion

The delayed and parameterized source separation problem is recast to a parameter estimation problem, where the parameters to estimate are the amplitudes a_{ij}, delays c_{ij} and shape parameters w_j. Supposing the noise to be white and Gaussian, the maximum likelihood estimator is obtained by minimizing the criterion:

$$E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) = \sum_i \varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \boldsymbol{w}), \tag{3.5}$$

where $\varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \boldsymbol{w})$ is the quadratic error related to mixture i:

$$\varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \boldsymbol{w}) = \left\|\boldsymbol{x}_i - \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; w_j\right]\right\|_2^2, \tag{3.6}$$



(a) Simulated mixture (b) $\boldsymbol{L}_{1:} \to \varepsilon(\boldsymbol{A}^*_{1:}, \boldsymbol{L}_{1:}, \boldsymbol{w}^*)$

Figure 3.1: (a) A noisy mixture of two Gaussian sources ($I = 1$, $J = 2$) with $a^*_{11} = 5.95, a^*_{12} = 2.66$, $\ell^*_{11}\Delta = 50$, $\ell^*_{12}\Delta = 65$, $w^*_1 = 5$, $w^*_2 = 4$ and $N = 100$. (b) The criterion $\boldsymbol{L}_{1:} \to \varepsilon(\boldsymbol{A}^*_{1:}, \boldsymbol{L}_{1:}, \boldsymbol{w}^*)$ admits local minimizers and flat surfaces. The red $\times$ indicates the global minimum.

and $\boldsymbol{A} \in \mathbb{R}_+^{I\times J}$, $\boldsymbol{L} \in \mathbb{N}$, $\boldsymbol{w} \in \mathbb{R}^J$ respectively gather the amplitudes a_{ij}, delays ℓ_{ij} and shape parameters w_j for mixtures i and sources j (see equation (1.9)).

Therefore, the parameterized source separation problem is formulated as the following constrained minimization problem:

$$\min_{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{L},\boldsymbol{w}} E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}). \tag{3.7}$$

3.3 Sparse-based alternating least squares

The optimization problem (3.7) is challenging because of the non-convexity of the criterion E, induced by the nonlinearity of model (3.4) with respect to $\boldsymbol{L}$ and $\boldsymbol{w}$. As an example, Figure 3.1 displays the variations of criterion $E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w})$ with respect to $\boldsymbol{L}$ in the case of $I = 1$ mixture and $J = 2$ sources: one can see that it admits multiple local minimizers as well as flat regions, making its optimization difficult even for this simple example.

Therefore, we choose an ALS scheme to optimize the criterion. This is an iterative descent strategy consisting of minimizing the criterion E with respect to a block of variables while fixing the others, and vice-versa. The algorithm stops if the criterion decrease at one iteration becomes lower than a constant ρ. It is not guaranteed that the ALS scheme converges towards the global minimizer of (3.7) since it is a block minimization of a non-convex criterion. Yet, this scheme is often used in delayed source separation, where, *e.g.*, the sources, delays and amplitudes are alternately estimated [Nio+; MMH07a; HH03]. However, the methods proposed by these authors, suffer from several limitations, as stated in Section 2.5.1 (source independence assumption [Nio+; MMH07a] or unfeasible exhaustive search [HH03]) making them unsuitable for the considered problem.

The proposed ALS scheme is given in Algorithm 1 and optimizes the criterion alternately with respect to the source shape parameters $\boldsymbol{w}$ on the one hand and with the delays $\boldsymbol{L}$ and amplitudes $\boldsymbol{A}$ on the other hand. More precisely:

Algorithm 1: ALS scheme for $\min_{\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}} E(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w})$.

Initialization: $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{L}} = \boldsymbol{0}_{I\times J}$, $\widehat{\boldsymbol{w}}$ randomly
do
  $(\boldsymbol{A}^0, \boldsymbol{L}^0, \boldsymbol{w}^0) \leftarrow (\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{w}})$
  for $i = 1 : I$ **do**
    $(\widehat{\boldsymbol{A}}_{i:}, \widehat{\boldsymbol{L}}_{i:}) \leftarrow \underset{\boldsymbol{A}_{i:},\boldsymbol{L}_{i:}}{\operatorname{argmin}}\, \varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \widehat{\boldsymbol{w}})$
  end
  $\widehat{\boldsymbol{w}} \leftarrow \underset{\boldsymbol{w}}{\operatorname{argmin}}\, E\big(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \boldsymbol{w}\big)$
while $\frac{E(\boldsymbol{A}^0,\boldsymbol{L}^0,\boldsymbol{w}^0) - E(\widehat{\boldsymbol{A}},\widehat{\boldsymbol{L}},\widehat{\boldsymbol{w}})}{E(\boldsymbol{A}^0,\boldsymbol{L}^0,\boldsymbol{w}^0)} \geq \rho$;

- the shape estimation is a continuous non-linear least–squares problem:

$$\widehat{\boldsymbol{w}} \leftarrow \underset{\boldsymbol{w}}{\operatorname{argmin}}\, E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}), \tag{3.8}$$

  and is solved using the Levenberg-Marquardt algorithm (see Section 2.2.1). In case some constraints (*e.g.*, non-negativity or lower and upper bounds) are required for the shape parameters, methods like trust-region-reflective or sequential quadratic programming can be used instead;

- the amplitude and delay estimation step writes:

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow \underset{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{L}}{\operatorname{argmin}}\, E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}). \tag{3.9}$$

  This is the main challenging problem and is detailed hereafter.

3.4 Amplitude and delay estimation

It follows from (3.5) that problem (3.9) is separable to I independent sub-problems:

$$\min_{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{L}} E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) \Leftrightarrow \forall i, \qquad \min_{\boldsymbol{A}_{i:}\geq\boldsymbol{0},\boldsymbol{L}_{i:}} \varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \boldsymbol{w}). \tag{3.10}$$

We propose to minimize each $\varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \boldsymbol{w})$ using a sparse approximation algorithm. The principle of sparse approximation is to represent each mixture i with a few atoms of an overcomplete dictionary (see Section 2.2.2). This choice is justified in the case of spectroscopic signals which contain very few peaks.

3.4.1 Dictionary formulation

Let us introduce the dictionary $\boldsymbol{S}$ as the matrix composed by J blocks denoted as $\boldsymbol{S}[j] \in \mathbb{R}^{N\times M}$:

$$\boldsymbol{S} = \begin{bmatrix}\boldsymbol{S}[1] & \dots & \boldsymbol{S}[j] & \dots & \boldsymbol{S}[J]\end{bmatrix}. \tag{3.11}$$

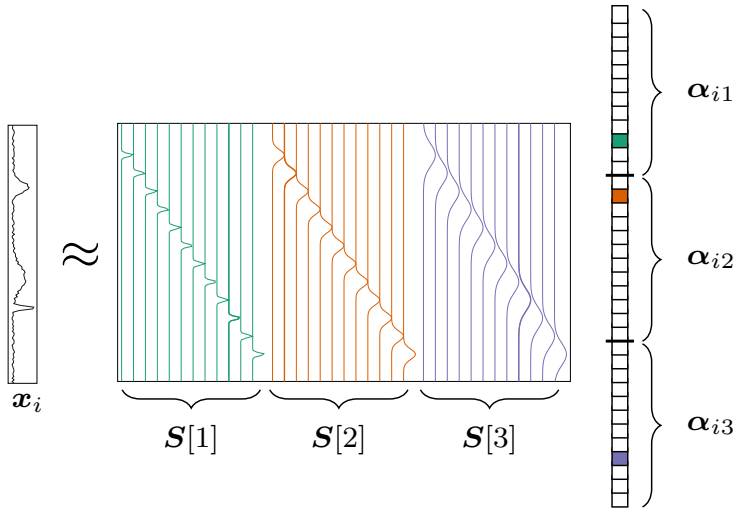


Figure 3.2: Sparse representation model of a mixture $\boldsymbol{x}_i$ with $J = 3$ sources. Each block $\boldsymbol{S}[j]$ of the dictionary gathers the delayed versions of $\boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]$. The sparse representation is structured so that each block vector $\boldsymbol{\alpha}_{ij}$ is 1-sparse.

Each block contains M possible version of source j sampled over the delay grid:

$$\boldsymbol{S}[j] \triangleq \left[\boldsymbol{s}\left[0; w_j\right] \quad \boldsymbol{s}\left[\Delta; w_j\right] \quad \ldots \quad \boldsymbol{s}\left[(M-1)\Delta; w_j\right]\right]. \tag{3.12}$$

Each mixture $\boldsymbol{x}_i$ is approximated as:

$$\boldsymbol{x}_i \approx \sum_{j=1}^{J} \boldsymbol{S}[j]\boldsymbol{\alpha}_{ij}, \tag{3.13}$$

where $\boldsymbol{\alpha}_{ij} = \begin{bmatrix} 0 & \ldots & 0 & a_{ij} & 0 & \ldots & 0 \end{bmatrix}^T \in \mathbb{R}_+^M$ is a 1-sparse vector, so that each source appears at most once in each mixture (see Fig. 3.2). The value and index of the non-zero element in $\boldsymbol{\alpha}_{ij}$ respectively indicate the amplitude a_{ij} and delay ℓ_{ij} of source j in mixture i. Moreover, once can notice that all the delayed versions selected from the same block dictionary $\boldsymbol{S}[j]$ from different mixtures will be associated together.

In consequence, the optimization problem (3.10) can be rewritten as:

$$\min_{\forall j, \boldsymbol{\alpha}_{ij} \geq \boldsymbol{0}} \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} \boldsymbol{S}[j]\boldsymbol{\alpha}_{ij} \right\|_2^2 \text{ s.t. } \forall\, j, \quad \|\boldsymbol{\alpha}_{ij}\|_0 \leq 1. \tag{3.14}$$

3.4.2 OMP-like implementation for delayed source separation

Greedy algorithms are effective and efficient when the sparsity level J is small and overcome the performance of ℓ_1 relaxation methods for highly correlated dictionaries [Tro04]. Also, they can easily be adapted to obtain a structured sparse solution such as a 1-sparse per block solution. In particular, we choose to adapt the Non-Negative OMP (NN-OMP) algorithm [BEZ08].

Let us first recall the main steps in each iteration of NN-OMP:

Algorithm 2: NN-OMP-like implementation for $(\widehat{\boldsymbol{A}}_{i:}, \widehat{\boldsymbol{L}}_{i:}) \leftarrow \underset{\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}}{\operatorname{argmin}}\, \varepsilon(\boldsymbol{A}_{i:}, \boldsymbol{L}_{i:}, \widehat{\boldsymbol{w}})$

Initialization: $\widehat{\boldsymbol{A}}_{i:} = \widehat{\boldsymbol{L}}_{i:} = \boldsymbol{0}_{1\times J}$, $\mathcal{J} = \emptyset$, $\boldsymbol{r}_i = \boldsymbol{x}_i$

1 **while** $k \leq J$ **and** $\|\boldsymbol{r}_i\|_2^2 \geq \varsigma$ **do**
2 $\quad k \leftarrow k + 1$
3 $\quad$ **for** $j \in \{1, \ldots, J\} \backslash \mathcal{J}$ **do**
4 $\quad\quad \tilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}}\ \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+$
5 $\quad$ **end**
6 $\quad \widehat{j} \leftarrow \underset{j\notin\mathcal{J}}{\operatorname{argmax}}\ \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\tilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+$
7 $\quad \widehat{\ell}_{i\widehat{j}} \leftarrow \tilde{\ell}_{i\widehat{j}}$
8 $\quad \mathcal{J} \leftarrow \mathcal{J} \cup \{\widehat{j}\}$
9 $\quad$ Update amplitudes $\widehat{\boldsymbol{A}}_{i\mathcal{J}}$ according to (3.15)
10 $\quad \boldsymbol{r}_i \leftarrow \boldsymbol{x}_i - \sum_{j\in\mathcal{J}} \widehat{a}_{ij} \boldsymbol{s}\left[\widehat{\ell}_{ij}\Delta; \widehat{w}_j\right]$
11 **end**
12 **if** $\varepsilon(\boldsymbol{A}^0_{i:}, \boldsymbol{L}^0_{i:}, \widehat{\boldsymbol{w}}) < \varepsilon(\widehat{\boldsymbol{A}}_{i:}, \widehat{\boldsymbol{L}}_{i:}, \widehat{\boldsymbol{w}})$ **then**
13 $\quad (\widehat{\boldsymbol{A}}_{i:}, \widehat{\boldsymbol{L}}_{i:}) \leftarrow (\boldsymbol{A}^0_{i:}, \boldsymbol{L}^0_{i:})$
14 **end**

1. first, the so-called forward selection step consists in choosing the column of the dictionary that is the most positively correlated with the residual;
2. second, the amplitudes corresponding to the chosen columns are updated by solving a non-negative least–squares estimation problem [LH95];
3. finally, the residual is updated by removing the contributions of the chosen columns.

NN-OMP considers the constraint in (3.14) to be $\sum_{j=1}^{J} \|\boldsymbol{\alpha}_{ij}\|_0 \leq J$. However, this constraint does not enforce the sources to appear at most once in each mixture. Therefore, the proposed implementation consists in forcing the sparse vector to be structured in blocks, each $\boldsymbol{\alpha}_{ij}$ being 1-sparse (this allows to jointly associate the source with their amplitude and delay estimation).

The proposed algorithm is given in Algorithm 2 and reproduces the three steps of an iteration of NN-OMP:

- the column in the dictionary that is the most positively correlated with the residual is selected, yielding the corresponding source $\widehat{j}$ and its delay $\widehat{\ell}_{i\widehat{j}}$ (lines 3–7). The selected source $\widehat{j}$ is then added to the list $\mathcal{J}$ of selected sources (line 8) so that the sub-block dictionary $\boldsymbol{S}[j]$ for $j \in \mathcal{J}$ will not be tested in the next iterations;

- the amplitudes of the selected sources gathered in $\mathcal{J}$ are estimated by solving the non-negative linear least–squares problem (line 9):

$$\widehat{\boldsymbol{A}}_{i:} \leftarrow \underset{\boldsymbol{A}_{i:}}{\operatorname{argmin}}\, \varepsilon(\boldsymbol{A}_{i:}, \widehat{\boldsymbol{L}}_{i:}, \widehat{\boldsymbol{w}}) \;\text{ s.t. }\; \begin{cases} \boldsymbol{A}_{i\mathcal{J}} \geq \boldsymbol{0} \\ \boldsymbol{A}_{i\overline{\mathcal{J}}} = \boldsymbol{0} \end{cases} \tag{3.15}$$

  where $\overline{\mathcal{J}}$ denotes the complementary subset of $\mathcal{J}$, *i.e.*, $\overline{\mathcal{J}} = \{1, \ldots, J\} \setminus \mathcal{J}$;

- lastly, the residual vector $\boldsymbol{r}_i$ (defined as the difference between mixture i and the sum of the estimated sources) is updated (line 10).

The main difference between NN-OMP and our implementation lies in the first step. Also, lines 11–13 have been added to ensure a decrease of the criterion by invalidating the estimates if they produce a criterion value that is greater than the value obtained at the previous iteration of Algorithm 1.

Furthermore, Algorithm 2 offers the possibility to obtain a variable number of sources per mixture. To do so, an additional stopping criterion is added in line 1 of Algorithm 2 such that the loop breaks if the squared residual norm $\|\boldsymbol{r}_i\|_2^2$ becomes lower than a threshold ς, *e.g.*, related to the noise variance. So, to enforce J sources to appear in each mixture, ς must be set to zero.

3.5 Results

3.5.1 Synthetic 1D set of mixtures

The performance of the proposed algorithm is evaluated on a synthetic 1D set of mixtures. The following settings are set. All the sources are modeled by a Gaussian function: $s(\lambda; w_j) = \exp(-\lambda^2/2w_j^2)$ where the shape parameters w_j correspond to the Gaussian widths. The number of mixtures is equal to $I = 40$, each with $N = 150$ samples and the number of sources is set to $J = 3$ where two of them are spectrally overlapping. The delay sampling step is set to $\Delta = 0.2$. Moreover, all the sources do not appear in all the mixtures, to deal with this situation we set the threshold ς in Algorithm 2 to $1.5\sigma_n^2$ where σ_n^2 is the noise variance and the ALS stopping parameter ρ is set to 10^{-6}. The amplitudes and delays are generated using polynomials with random coefficients, and the ground-truth shape parameters are $\boldsymbol{w}^* = \begin{bmatrix} 1.3 & 3.6 & 5.2 \end{bmatrix}$.

The ground truth and the estimated parameters are superimposed in Figure 3.4 for the amplitudes, delays and shapes. The shape parameter estimation is adequate as we have $\widehat{\boldsymbol{w}} = \begin{bmatrix} 1.22 & 1.78 & 5.2 \end{bmatrix}$. The amplitude estimation is correct with a little perturbation induced by the noise. The delay estimation largely corresponds to the ground truth except around the mixture $i = 20$ where the estimates are slightly affected because of the spectral source overlap. Moreover, this example shows that the algorithm is able to successfully deal with a varying number of sources within the mixtures. Therefore, when the source shape parameters are sufficiently different the separation (association) is satisfactory without the need to include any additional discriminating factors.

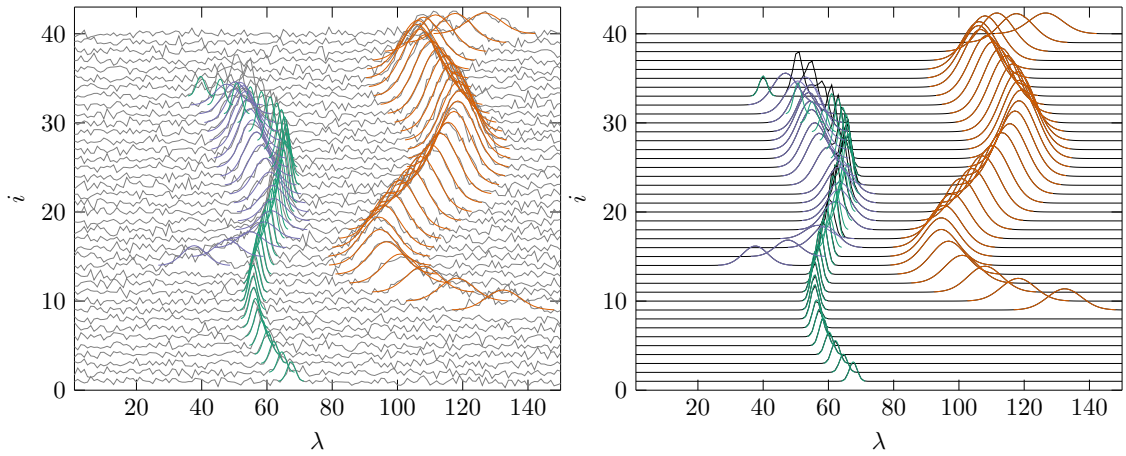


Figure 3.3: (a) A synthetic noisy set of 1D mixtures superimposed by the ground-truth sources. (b) the mixture reconstruction superimposed by the estimated sources.

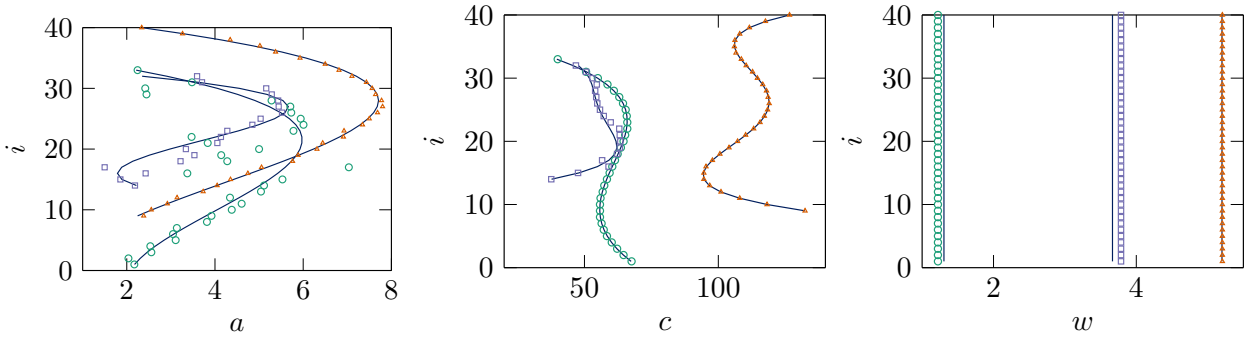


Figure 3.4: The estimated parameters (amplitudes $\widehat{a}_{ij}$, delays $c_{ij} = \widehat{\ell}_{ij}\Delta$ and shapes $\widehat{w}_j$) of the estimated sources in Figure 3.3. The ground truth parameters are plotted in plain lines.

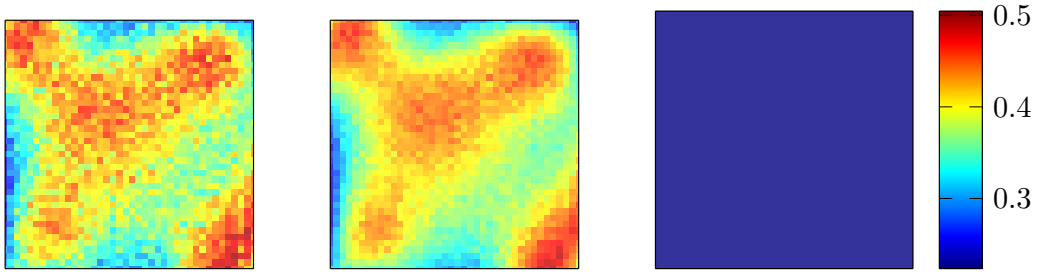
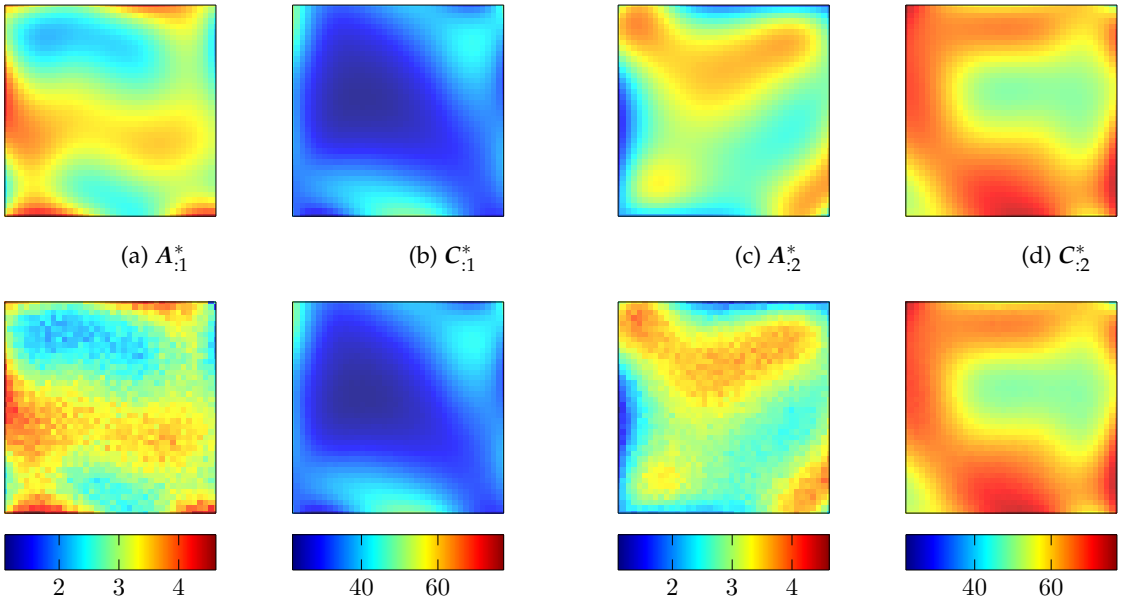


Figure 3.5: The white images of the generated multispectral image, the reconstruction and the absolute residual.



(a) $A^*_{:1}$ (b) $C^*_{:1}$ (c) $A^*_{:2}$ (d) $C^*_{:2}$

(e) $\widehat{A}_{:1}$ (f) $\widehat{C}_{:1}$ (g) $\widehat{A}_{:2}$ (h) $\widehat{C}_{:2}$

Figure 3.6: First row: the ground truth parameters (amplitudes $\boldsymbol{A}$ and delays $\boldsymbol{C} = \boldsymbol{L}\Delta$) of the two sources. Second row: the corresponding estimated parameters.

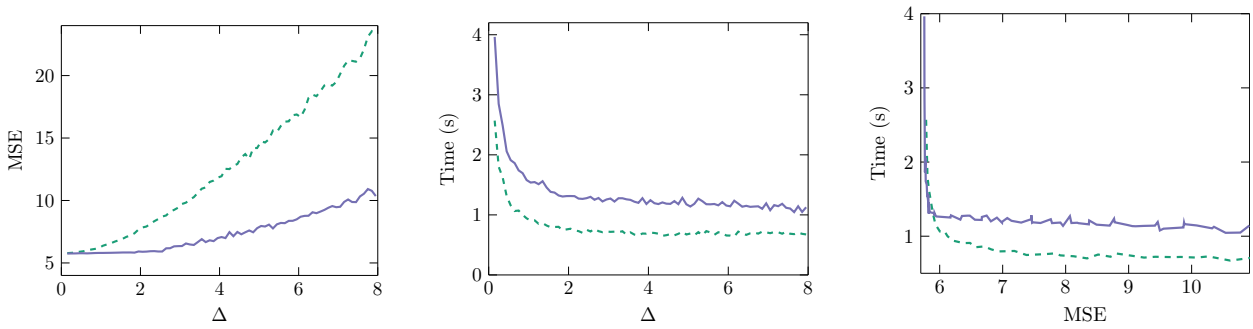
3.5.2 Synthetic 2D sets of mixtures

The proposed algorithm is adapted for both 1D and 2D sets of mixtures. We evaluate the performance of the algorithm in case of a 2D sets of mixtures (a synthetic multispectral image) with dimension $40 \times 40 \times 100$ and $J = 2$ Gaussian parameterized sources. The SNR is set to 15 dB. The parameters are generated using cubic B-spline surfaces with random control points except for the shape parameters which are $\boldsymbol{w}^* = \begin{bmatrix}1.37 & 3.84\end{bmatrix}$. The sampling step Δ is set to 0.2 and the ALS stopping parameter ρ is set to 10^{-6}. The generated multispectral image, its reconstruction and the residual image are displayed in Figure 3.5. This figure shows that the proposed algorithm correctly reconstructs and denoises the mixtures. The computation time is equal to 15 s.

The estimated shape parameter are very close to the ground truth values: $\widehat{\boldsymbol{w}} = \begin{bmatrix}1.37 & 3.85.\end{bmatrix}$. The ground truth and the estimated maps of the amplitudes and delays are shown in Figure 3.6, one can observe that they are well estimated.

3.5.3 The impact of the delay discretization

Delay estimation is constrained to be on a grid. The easiest way to refine its estimation is to decrease the value of Δ, but this results in a bigger dictionary, and in turn, an increase of the computational burden. Recently, Fyhn *et al.* [FDJ15, Section IV] proposed a continuous extension of OMP, named IBOMP, for sparse deconvolution. Their approach makes use of an interpolation strategy introduced in [ETS11]. Specifically, this approach first requires to estimate the delays on a sampling grid and then to refine the delay estimation by using an interpolation between two consecutive delays. The authors in [ETS11; FDJ15] advocate the use of polar interpolation which



(a) (b) (c)

Figure 3.7: Comparison between both versions of the non-regularized method, without (dashed) and with (solid) delay interpolation.

empirically outperforms other interpolation techniques in terms of estimation quality. This can be done in our algorithm by refining the estimation of $\widehat{\ell}_{i\widehat{j}}$ with $\widehat{\ell}_{i\widehat{j}} + \eta$ where $\eta \in [-\Delta/2, \Delta/2]$ is estimated as in [FDJ15].

To compare the proposed algorithm with and without delay interpolation, 30 datasets with random parameters are generated, each with $I = 40$ mixtures, $J = 3$ Gaussian parameterized sources, $N = 200$ samples and an average SNR equals to 15 dB. Besides, the stopping threshold is set to $\rho = 10^{-6}$. The data are processed with a delay sampling step Δ varying from 0.1 to 8. Recall that the sampling step of the mixtures is set to 1 by default. Figure 3.7 shows the results in terms of estimation quality and the computation time. Results show that introducing the interpolation scheme improves the estimation quality, especially for large values of Δ, at the price of an increasing computation time. More important, this simulation shows that for a given estimation quality, the implementation without using interpolation remains faster than the one using the interpolation as shown in Figure 3.7c.

3.5.4 Shape discriminating limit

Algorithm 1 is able to assign the estimated peaks to the right source because the sources can be discriminated by their shape parameter (see Figure 3.2). We are interested in finding the resolution limit, that is the least difference between the shape parameters of two sources beyond which the sources can be discriminated. For this purpose, we consider $I = 40$ mixtures each with $N = 200$ samples, and $J = 2$ Gaussians sources of widths w_1 and w_2 and with constant delays through the mixtures. We gradually vary the ratio w_2/w_1 from 0.5 to 1.5. For each ratio, we measure the switch percentage defined as the percentage of wrongly assigned peaks over the total number of peaks (a peak is wrongly assigned if it belongs to source 1 while it is assigned to the source 2 and *vice versa*). The experiment is repeated for three SNRs (Figure 3.8).

The results show that for high SNR, the proposed method can separate the sources whatever the ratio w_2/w_1 except of course when both parameters are equal. In Figure 3.9 we show an example of $I = 15$ mixtures and $J = 2$ sources with the same shape parameters $w_1^* = w_2^* = 4$ and SNR equals to 15 dB. The resulting source separation and delays estimation show clearly that the method fails to separate the sources in this case even though the sources do not spectrally overlap. This

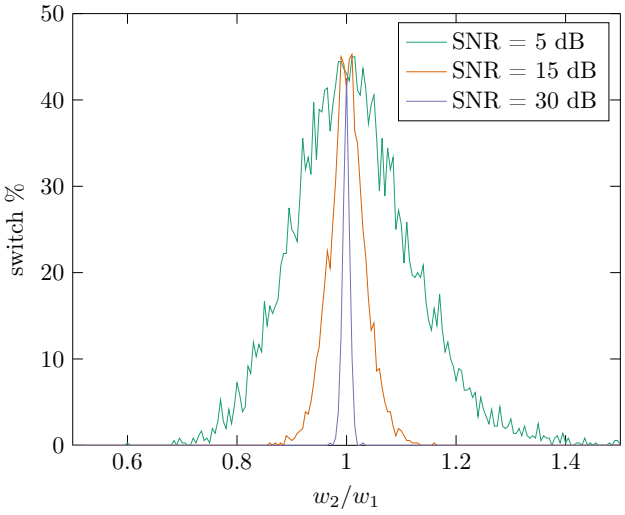
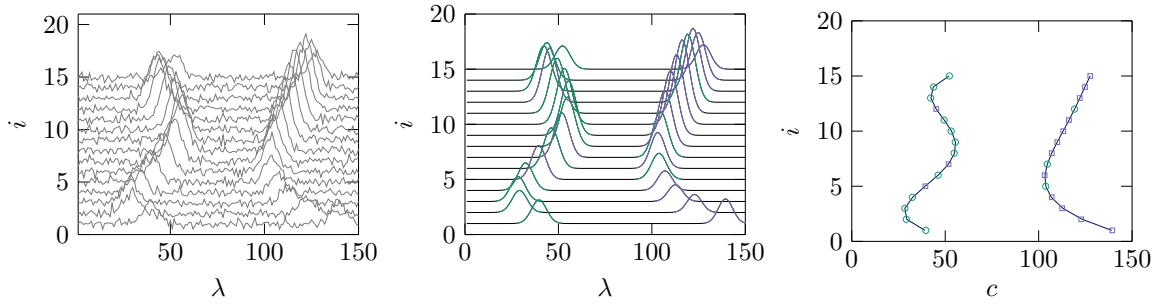


Figure 3.8: Switch percentage with respect to the ratio of two Gaussian widths, for three different SNR.



(a) (b) (c)

Figure 3.9: (a) a synthetic noisy set of 1D mixtures with two sources sharing the same shape. (b) the mixture reconstruction obtained by the proposed method. (c) the delays $\boldsymbol{C}$ estimation.

motivates us to introduce a regularization term to overcome this limitation in Chapter 4 or to model the parameter evolution using B-splines in Chapter 5.

3.6 Conclusion

We propose a method to separate delayed and parameterized sources from their mixtures where the sources are discriminated by the difference of their shape parameters. The problem is addressed as an optimization problem of a data-fit criterion which is resolved using an ALS scheme. First, the sources are estimated by estimating their shape parameters using the Levenberg-Marquardt. Second, the amplitude and delays are jointly estimated using an adapted version of the OMP algorithm. The latter finds a structured sparse solution (1-sparse per block) after projecting each mixture on an overcomplete block-dictionary: each block gathers the delayed version of a source

where the delays are sampled over a grid. This method is validated on both 1D and 2D sets of mixtures. However, when two sources share very similar shapes, the separation becomes ambiguous. This motivates us to introduce a new discriminating factor, the slow parameter evolution, in Chapters 4 and 5.

4

Separation of parameterized sources with slow delay evolution

4.1 Introduction

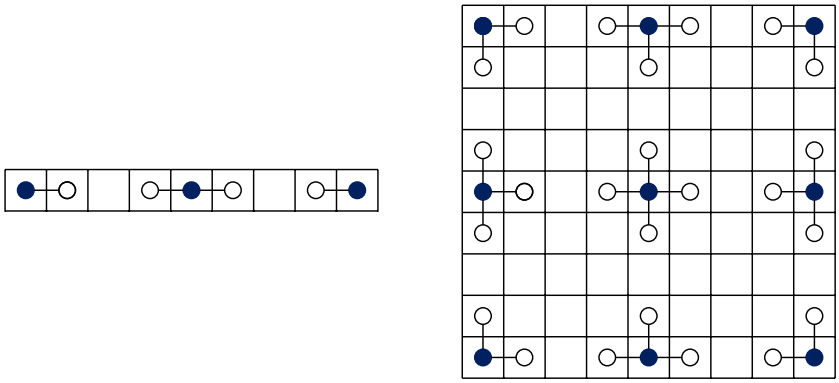
In this chapter, we give a first solution to take into account the temporal or spatial delay slow evolution of the sources within the mixtures. The motivation of considering the slow delay evolution is twofold. First, as stated in Chapter 1, this is physically justified in practice as it can be the consequence of a short acquisition time between the measurement of the mixture signals (such as in photoelectron spectroscopy [Glo+05]) or neighboring spatial sensors (such as in galaxy kinematics [GWK89] or audio recorded mixtures [CB06]). Second, it helps to better discriminate sources that have similar shapes and overlap in the spectral domain (highly correlated sources).

Similarly to Chapter 3 the delayed and parameterized source separation problem is considered. To favor the slow delay evolution, we introduce a regularization term to the data-fit criterion (Section 4.2). An ALS scheme is used to optimize the criterion, where the shape parameters are estimated using the Levenberg-Marquardt. The amplitudes and delays are estimated in all the mixtures using a joint greedy sparse approximation method inspired from OMP and coupled with the Iterated Conditional Modes (ICM) algorithm (Section 4.3). Results on synthetic 1D and 2D sets of mixtures are given in Section 4.4.

4.2 Delay regularization

4.2.1 Criterion

To get a slow delay evolution for each source, we propose to add a regularization term to the data-fit term (defined in Section 3.2.1) and minimize the augmented criterion with respect to the source



(a) 1D set of mixtures (b) 2D sets of mixtures

Figure 4.1: Neighbor mixtures (white dots) for several mixtures (blue dots) in case of 1D and 2D sets of mixtures.

amplitudes, delays and shape parameters. The aim of the regularization term is to favor slow temporal (respectively spatial) delay evolution of each source within a 1D set (respectively 2D sets) of mixtures by measuring the delay variation of each source. The latter is measured, for a given delay $\ell_{ij}\Delta$, by calculating the difference between this delay and the delays of neighboring mixtures for the same source j. Therefore, we need to set a generalized definition of mixture neighboring that can be adapted to both 1D and 2D sets of mixtures.

In case of 1D set of mixtures, the neighbors of a mixture i are the mixtures $i-1$ and $i+1$ (see Figure 4.1a) and in case of 2D sets of mixtures, the neighbors are the 4 connected spatial positions (see Figure 4.1b). Of course, the neighbors are less on the edges (*e.g.*, for $i=1$ and $i=I$ in case of 1D set of mixtures). Moreover, we use the notion of *clique* as defined for Markov random fields [PL97]. In a simple definition, a subset is called a clique of order two if any indices i and i' in this subset are neighbors. Let $\mathcal{G}$ be the set of all the cliques of order two.

Now, we can define the regularization term $R(\boldsymbol{L})$:

$$R(\boldsymbol{L}) = \sum_{j=1}^{J} \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij}\Delta - \ell_{i'j}\Delta\right)^2 = \Delta^2 \sum_{j=1}^{J} \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}\right)^2. \tag{4.1}$$

with $\ell_{ij} \in \mathbb{N}$ is the delay index and Δ is the delay sampling step. Finally, the regularized criterion $F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w})$ composed of the data-fit term $E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) = \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} a_{ij} \boldsymbol{s}\left[\ell_{ij}\Delta; w_j\right] \right\|_2^2$ and regularization term $R(\boldsymbol{L})$ writes:

$$F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) = E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) + \tau R(\boldsymbol{L}), \tag{4.2}$$

with τ the regularization parameter that allows to trade-off between the two terms. Note that we do not consider the problem of determining the best value of τ, rather it is manually tuned by the user.

Algorithm 3: ALS scheme for $\min_{\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}} F(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w})$.

Initialization: $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{L}} = \boldsymbol{0}_{I\times J}$, $\widehat{\boldsymbol{w}}$ randomly
1 **do**
2 $\quad(\boldsymbol{A}^0, \boldsymbol{L}^0, \boldsymbol{w}^0) \leftarrow (\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{w}})$
3 $\quad(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow \underset{\boldsymbol{A},\boldsymbol{L}}{\operatorname{argmin}}\, F(\boldsymbol{A},\boldsymbol{L},\widehat{\boldsymbol{w}})$
4 $\quad\widehat{\boldsymbol{w}} \leftarrow \underset{\boldsymbol{w}}{\operatorname{argmin}}\, F\big(\widehat{\boldsymbol{A}},\widehat{\boldsymbol{L}},\boldsymbol{w}\big)$
5 **while** $\frac{F(\boldsymbol{A}^0,\boldsymbol{L}^0,\boldsymbol{w}^0)-F(\widehat{\boldsymbol{A}},\widehat{\boldsymbol{L}},\widehat{\boldsymbol{w}})}{F(\boldsymbol{A}^0,\boldsymbol{L}^0,\boldsymbol{w}^0)} \geq \rho$;

4.2.2 ALS scheme

Criterion $F(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w})$ is optimized using Algorithm 3 which is based on an the ALS scheme:

- the shape estimation is a continuous non-linear least–squares problem:

$$\widehat{\boldsymbol{w}} \leftarrow \underset{\boldsymbol{w}}{\operatorname{argmin}}\, F(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}), \tag{4.3}$$

  Since $R(\boldsymbol{L})$ does not depend on $\boldsymbol{w}$, the shape estimation (line 4) is identical to the optimization of $E(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w})$ with respect to $\boldsymbol{w}$ given in equation (4.3) and is solved using the Levenberg-Marquardt algorithm (see Section 2.2.1);

- the amplitudes and delays are jointly estimated:

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow \underset{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{L}}{\operatorname{argmin}}\, F(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w}). \tag{4.4}$$

  This problem differs from the one seen in Section 3.4 because $F(\boldsymbol{A},\boldsymbol{L},\boldsymbol{w})$ does not read as a separable sum with respect to $\boldsymbol{L}_{i:}$. Indeed, the terms $(\ell_{ij}-\ell_{i'j})^2$ appearing in the regularization term $R(\boldsymbol{L})$ not only depend on ℓ_{ij} but also on delays in neighboring mixtures. Therefore, a new optimization strategy must be proposed to address (4.4).

In the sequel we will detail the estimation strategy for $\boldsymbol{A}$ and $\boldsymbol{L}$.

4.3 Amplitude and slow delay estimation

A sparse approximation strategy is used to estimate the amplitudes and delays. Recall that when using the dictionary formulation (see Section 3.4.1), each mixture is approximated such that: $\boldsymbol{x}_i = \sum_{j=1}^{J} \boldsymbol{D}\boldsymbol{\alpha}_{ij}$. The delay ℓ_{ij} can be obtained from the support of the sparse vectors $\boldsymbol{\alpha}_{ij}$. Therefore, since the delays are imposed to evolve slowly, the support of the sparse vectors will move slowly as well. Figure 4.2 visualizes the sparse solutions $\boldsymbol{\alpha}_{ij}$ for neighboring mixtures when the regularization term τ is very small, moderate and very high. One can remark that the regularization goal is to get slowly moving support for each source and thus allows to get an intermediate solution between the case where the mixtures are approximated independently and the MMV case where the sparse vectors share a common vector.

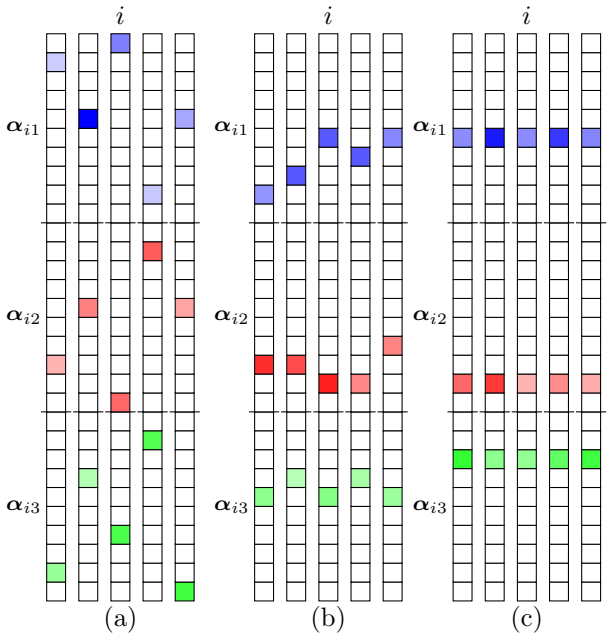


Figure 4.2: Examples of a sparse model for $J = 3$ sources and $I = 6$ mixtures where the white and the colored squares respectively represent the zero and non-zero amplitudes. Each column corresponds to a mixture i, and is divided into J blocks representing the sources (each source is indicated with a unique color). (a) The case of no delay regularization. (b) The case of a slow-moving regularization. (c) The case of a very strong regularization, resulting in a simultaneous sparse approximation with constant support for all mixtures.

Algorithm 4 is proposed to optimize $F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w})$ with respect to $\boldsymbol{A}$ and $\boldsymbol{L}$. It is a greedy algorithm in the spirit of Algorithm 2. However, Algorithm 4 takes all mixtures as inputs while Algorithm 2 considers a single mixture.

At each iteration, the source $\widehat{j}$ inducing the largest decrease of criterion F is selected (lines 2–5) and added to the list of selected sources $\mathcal{J}$ (line 6). The corresponding delays $\boldsymbol{L}_{:\widehat{j}}^{\text{temp}}$ are computed. Then, the amplitudes of the selected sources in $\mathcal{J}$ in all the mixtures are estimated using a non-negative least squares solver (line 8). Finally, the residual vectors of all the mixtures are updated (line 11). Note that adding a new source results in a decrease of the data-fit term $E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w})$ but makes the regularization term $R(\boldsymbol{L})$ increasing. In other words, the criterion $F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) = E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) + \tau R(\boldsymbol{L})$ can either increase or decrease. Therefore, the condition in line 9 is set to break the loop in case of an increase of F, so that Algorithm 4 is indeed a descent algorithm. Similarly, the condition in line 13 ensures a decrease of the criterion by invalidating the estimates if they produce a criterion value that is larger than the value obtained at the previous iteration of Algorithm 3.

Let us now specify the rule for selecting the source $\widehat{j}$ among all candidate sources $j \notin \mathcal{J}$. Since $\widehat{j}$ is defined as the source yielding the largest decrease of criterion F, lines 4–6 in Algorithm 4 aim at estimating, for each source j that has not already been selected, the value of the corresponding delays $\boldsymbol{L}_{:j}$ (denoted by $\tilde{\boldsymbol{L}}_{:j}$) as well as the value of F obtained while considering the set of sources $\mathcal{J} \cup \{j\}$. To do so, OMP selection rule philosophy is used: the amplitudes $a_{ij'}$ and delays $c_{ij'}$ of

Algorithm 4: Implementation of $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow \underset{\boldsymbol{A},\boldsymbol{L}}{\operatorname{argmin}} F(\boldsymbol{A}, \boldsymbol{L}, \widehat{\boldsymbol{w}})$

Initialization: $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{L}} = \boldsymbol{L}^{\text{temp}} = \boldsymbol{0}_{I\times J}$, $\mathcal{J} = \emptyset$, $\boldsymbol{r}_i = \boldsymbol{x}_i \; \forall i$

1 **for** $k = 1 \to J$ **do**

2 **for** $j \in \{1, \dots, J\} \backslash \mathcal{J}$ **do**

3 Compute $\widetilde{\boldsymbol{L}}_{:j}$ defined in (4.8) using the ICM algorithm

4 **end**

5 $\widehat{j} \leftarrow \underset{j\notin\mathcal{J}}{\operatorname{argmax}} \sum_{i=1}^{I} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\widetilde{\ell}_{ij} - \widetilde{\ell}_{i'j}\right)^2$

6 $\mathcal{J} \leftarrow \mathcal{J} \cup \{\widehat{j}\}$

7 $\boldsymbol{L}^{\text{temp}}_{:\widehat{j}} \leftarrow \widetilde{\boldsymbol{L}}_{:\widehat{j}}$

8 $\boldsymbol{A}^{\text{temp}} \leftarrow \underset{\boldsymbol{A}}{\operatorname{argmin}} F(\boldsymbol{A}, \boldsymbol{L}^{\text{temp}}, \widehat{\boldsymbol{w}})$ s.t. $\{\boldsymbol{A}_{:\mathcal{J}} \geq \boldsymbol{0}, \boldsymbol{A}_{:\overline{\mathcal{J}}} = \boldsymbol{0}\}$

9 **if** $F(\boldsymbol{A}^{\text{temp}}, \boldsymbol{L}^{\text{temp}}, \widehat{\boldsymbol{w}}) > F(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{w}})$ **then** Break **end**

10 $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow (\boldsymbol{A}^{\text{temp}}, \boldsymbol{L}^{\text{temp}})$

11 **for** $i = 1 \to I$ **do** $\boldsymbol{r}_i \leftarrow \boldsymbol{x}_i - \sum_{j\in\mathcal{J}} \widehat{a}_{ij} \boldsymbol{s}\left[\widehat{\ell}_{ij}\Delta; \widehat{w}_j\right]$ **end**

12 **end**

13 **if** $F(A^0, L^0, \widehat{\boldsymbol{w}}) < F(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{w}})$ **then** $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow (A^0, L^0)$ **end**

previously selected sources ($j' \in \mathcal{J}$) are fixed and one source j is considered. Thus, one needs to consider the minimization of $F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w})$ with respect to $\boldsymbol{A}_{:j}$ and $\boldsymbol{L}_{:j}$, while fixing the values of $\boldsymbol{L}_{:j'}$ and $\boldsymbol{A}_{:j'}$ for $j' \in \mathcal{J}$:

$$\widetilde{\boldsymbol{L}}_{:j} \leftarrow \underset{\boldsymbol{L}_{:j}}{\operatorname{argmin}} \min_{\boldsymbol{A}_{:j}\geq\boldsymbol{0}} \sum_{i=1}^{I} \left\| \boldsymbol{r}_i - a_{ij} \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right] \right\|_2^2 + \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}\right)^2, \tag{4.5}$$

where $\boldsymbol{r}_i$ is the current residual of mixture i by removing the previously selected sources, *i.e.*, :

$$\forall\, i, \qquad \boldsymbol{r}_i = \boldsymbol{x}_i - \sum_{j'\in\mathcal{J}} \widehat{a}_{ij'} \boldsymbol{s}\left[\widehat{\ell}_{ij'}\Delta; \widehat{w}_{j'}\right]. \tag{4.6}$$

The minimization of (4.5) with respect to $\boldsymbol{A}_{:j}$ while fixing $\boldsymbol{L}_{:j}$ has a closed form solution:

$$\forall i, \;\; a_{ij} = \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+. \tag{4.7}$$

Plugging back (4.7) into (4.5), (4.5) simplifies to:

$$\widetilde{\boldsymbol{L}}_{:j} \leftarrow \underset{\boldsymbol{L}_{:j}}{\operatorname{argmax}} \sum_{i} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}\right)^2. \tag{4.8}$$

which is the cost function appearing at line 7 in Algorithm 4. We now detail the adapted strategy to resolve this discrete optimization problem.

4.3.1 Delay estimation with an ICM-like algorithm

The optimization problem (4.8) is combinatorial, therefore we propose to use Iterated Conditional Modes (ICM) [Bes86] which is a popular iterative coordinate-wise optimization method in image processing. While it converges to a local optimizer, it generally gives good results. At each ICM iteration, the optimization problem (4.8) is replaced by a series of I optimization problems of dimension one. More precisely, at each ICM iteration, all the mixtures $i \in \{1, \dots, I\}$ are *swept* and whenever a mixture i is visited, the related delay $\widetilde{\ell}_{ij}$ is estimated by maximizing the criterion in (4.8) with respect to ℓ_{ij} whilst fixing the other delays $\widetilde{\ell}_{i'j}$ with $i' \neq i$. Then:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \widetilde{\ell}_{i'j}\right)^2. \tag{4.9}$$

The convergence of ICM is considered to be reached when the relative distance between the estimates of two consecutive iterations is small enough. Mathematically, the stopping condition reads $\|\widetilde{\boldsymbol{L}}_{:j}^{(t)} - \widetilde{\boldsymbol{L}}_{:j}^{(t-1)}\|_2^2 / \|\widetilde{\boldsymbol{L}}_{:j}^{(t-1)}\|_2^2 < \xi$ where t indicates the current ICM iteration.

The order in which the mixtures are swept in an ICM iteration t is important and can have an influence on the estimation. In order to exploit at best the relationship between the mixtures we propose two sequential sweeping strategies that are respectively applied in case of a 1D and a 2D sets of mixtures.

Two-direction sweeping for 1D set of mixtures

The local optimization problem (4.9) can be explicitly detailed in the case of 1D set of mixtures following the visited mixture i, yielding the following updates:

- if $i = 1$:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2(\ell_{ij} - \widetilde{\ell}_{(i+1)j})^2, \tag{4.10}$$

- if $1 < i < I$:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2(\ell_{ij} - \widetilde{\ell}_{(i-1)j})^2 - \tau\Delta^2(\ell_{ij} - \widetilde{\ell}_{(i+1)j})^2, \tag{4.11}$$

- if $i = I$:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2(\ell_{ij} - \widetilde{\ell}_{(i-1)j})^2. \tag{4.12}$$

The sequential sweeping strategy is as follows. First, a starting mixture $i_0 \in \{1, \dots, I\}$ is randomly chosen, the corresponding delay $\widetilde{\ell}_{i_0 j}$ is estimated using one of the above updates. Then, the delays in mixtures $i_0 + 1$ to I are sequentially estimated, and the same sequential procedure is used for mixtures $i_0 - 1$ to 1.

Of course, the delays are not known at first ICM iteration so equations (4.10), (4.11) and (4.12) cannot be computed as they require the delays in the adjacent mixtures of a given mixture i.

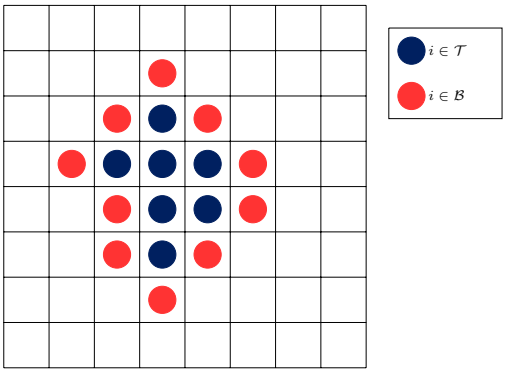


Figure 4.3: An illustration where the mixture spatial positions represented by a blue dot belong to $\mathcal{T}$ and the ones represented by a red dot belong to $\mathcal{B}$. Other spatial positions without dots have not been swept.

Therefore, we propose a recursive initialization strategy (only for the first ICM iteration, $t = 0$) based on (4.5) after discarding the unknown delays, that is detailed now.

The starting mixture i_0 is the one maximizing the data-fit term. The corresponding delay is then easily estimated by using only this data-fit term, leading to:

$$(i_0, \widetilde{\ell}_{i_0 j}) \leftarrow \underset{i,\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2. \tag{4.13}$$

Then, the delays are sequentially estimated from mixtures $i = i_0 + 1$ to I. Two regularization terms depending on ℓ_{ij} may appear in (4.8). The term $\left(\ell_{ij} - \ell_{(i-1)j}\right)^2$ can be computed since $\ell_{(i-1)j} = \tilde{\ell}_{(i-1)j}$ is known whereas $\left(\ell_{ij} - \ell_{(i+1)j}\right)^2$ is discarded since $\ell_{(i+1)j}$ has not been estimated yet. Hence,

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2(\ell_{ij} - \tilde{\ell}_{(i-1)j})^2. \tag{4.14}$$

Finally, the delays ℓ_{ij} in the mixtures going from $i = i_0 - 1$ down to $i = 1$ are sequentially estimated. Only the regularization term $\left(\ell_{ij} - \ell_{(i+1)j}\right)^2$ is taken into account in (4.8). Hence:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2(\ell_{ij} - \tilde{\ell}_{(i+1)j})^2. \tag{4.15}$$

Region growing inspired sweeping for 2D sets of mixtures

Contrary to the case of a 1D set of mixtures, multiple strategies can be used to sequentially sweep a 2D sets of mixtures departing from a starting mixture i_0. For example, the mixtures can be swept by columns or following the Hilbert-Peano sequential trajectory [CT07]. We choose a correlation-driven sequential sweeping strategy (the mixtures with higher correlations with the residual are swept first) that is inspired from the popular region growing image segmentation algorithm [AB94].

The region growing inspired sweeping procedure for an ICM iteration (t) starts with empty sets denoted by $\mathcal{T}$ and $\mathcal{B}$ and iterates as follows:

1. a starting mixture $i_0 \in \{1, \dots, I\}$ is randomly selected, added to $\mathcal{B}$ and its corresponding delay $\widetilde{\ell}_{i_0 j}$ is estimated using (4.9);
2. the mixture with the highest correlation with the residual and belonging to $\mathcal{B}$ is selected. Let i_s be the index of this mixture, then $i_s \leftarrow \underset{i \in \mathcal{B}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+^2$;
3. i_s is retired from $\mathcal{B}$ and added to $\mathcal{T}$;
4. the neighbors of i_s which are neither in $\mathcal{T}$ and $\mathcal{B}$ are added to $\mathcal{B}$ and their delays are estimated using (4.9);
5. repeat from 2 till $\mathcal{B}$ is empty.

Figure 4.3 illustrates an example showing spatial positions i that belong to $\mathcal{T}$ and $\mathcal{B}$.

For the first ICM iteration, the delays are not known so the optimization problem (4.9) cannot be solved as it requires the delays in the neighboring mixtures of a given mixture i. This implies to only modify the steps 1 and 4 of the procedure given below (the optimization problem (4.9) is only needed in these steps). For step 1, the starting mixture i_0 is selected and its corresponding delay is estimated by only using the data-fit term:

$$(i_0, \widetilde{\ell}_{i_0 j}) \leftarrow \underset{i, \ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2. \tag{4.16}$$

For step 4, we redefine the optimization problem (4.9) in such a way to only consider the delays in the mixtures which have been swept *i.e.*, the mixtures that belong to $\mathcal{T}$ or $\mathcal{B}$. The redefined optimization problem writes:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{\substack{(i,i') \in \mathcal{G} \\ i' \in \mathcal{T} \cup \mathcal{B}}} \left(\ell_{ij} - \widetilde{\ell}_{i'j}\right)^2. \tag{4.17}$$

4.3.2 Remarks

If no regularization is considered (*i.e.*, $\tau = 0$), then $F(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w}) = E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{w})$. We recommend to use Algorithm 1 rather that Algorithm 3 since the former exploits the separability of the criterion. Conversely, when τ tends to infinity, the delays related to each source are necessarily constant: $\forall i \neq i', \ \ell_{ij} = \ell_{i'j}$. Therefore, the sparse vectors $\boldsymbol{\alpha}_{ij}$ $\forall i$ share a common support (see Fig. 4.2(c)). The estimation of $\boldsymbol{L}_{:j}$ in (4.8) becomes:

$$\widetilde{\boldsymbol{L}}_{:j} = [\widetilde{\ell}_j, \dots, \widetilde{\ell}_j] \quad \text{with} \quad \widetilde{\ell}_j \leftarrow \underset{\ell}{\operatorname{argmax}} \sum_i \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell\Delta; \widehat{w}_j\right]\right)_+^2; \tag{4.18}$$

and can be obtained using an exhaustive search by testing the M possibilities for ℓ. This is very similar to the S-OMP algorithm [TGS06] (which is a greedy algorithm for sparse recovery of vectors having a common support) with the difference that the vectors $\boldsymbol{\alpha}_{ij}$ are 1-sparse (see equation (3.14)).

In Appendix A we present an extension of Algorithm 4 that allows obtaining a varying number of sources in each mixture, *i.e.*, a source that appears in a contiguous subset of mixtures. In

particular, sweeping strategies for both 1D and 2D sets of mixtures are presented; they consist of modified versions of the sweeping procedures detailed above.

In Appendix B we present a new method to estimate the amplitudes, delays and shapes. These parameters are also supposed to vary slowly within the mixtures in contrary to this chapter where the shape parameters are constant for each source. The resulting algorithm estimates all the parameters jointly and is very similar to Algorithm 4.

4.4 Results

4.4.1 Synthetic 1D set of mixtures

The performance of the proposed algorithm is evaluated on a synthetic 1D set of mixtures. The following settings are set. All the sources are modeled by a Gaussian function: $s(\lambda; w_j) = \exp(-\lambda^2/2w_j^2)$. The number of mixtures is equal to $I = 40$, each with $N = 150$ samples and the number of sources is set to $J = 4$ where two of them are spectrally overlapping. The SNR defined in equation (1.13) is equal to 15 dB. The amplitudes and delays are generated using polynomials with random coefficients, and the shape parameters are $\boldsymbol{w}^* = \begin{bmatrix}1.17 & 1.45 & 1.49 & 5.49\end{bmatrix}$. One can remark that the first three sources have very similar shape parameters. The delay sampling step is fixed to $\Delta = 0.2$, the regularization parameter is tuned $\tau = 0.02$, the ALS and the ICM stopping parameters ρ and ξ are respectively set to to 10^{-6}.

The generated mixtures are displayed in Figure 4.4, and the mixture denoised reconstruction is displayed in Figure 4.4b superimposed by the estimated sources. These results show that the reconstruction is adequate and that the sources are well recovered. The computation time to obtain this result is 1.5 s.

The ground-truth and the estimated amplitudes, delays and shapes are displayed in Figure 4.5. The shape parameter estimation is adequate as we have $\widehat{\boldsymbol{w}} = \begin{bmatrix}1.07 & 1.45 & 1.47 & 5.6\end{bmatrix}$. The amplitude estimation suffers from a little perturbation induced by the noise. Moreover, despite the fact that some sources share similar shapes and some others are overlapping, the delay estimates are good and evolve slowly due to the regularization.

4.4.2 Synthetic 2D sets of mixtures

The proposed algorithm is tested on a 2D sets of mixtures with dimension $40 \times 40 \times 100$ and $J = 2$ Gaussian parameterized sources. The SNR is set to 15 dB. The parameters are generated using cubic B-spline surfaces with random control points except for the shape parameters which are exactly the same $\boldsymbol{w}^* = \begin{bmatrix}5 & 5\end{bmatrix}$. The delay sampling step is fixed to $\Delta = 0.2$, the regularization parameter is tuned to $\tau = 0.04$, the ALS and the ICM stopping parameters ρ and ξ are respectively set to to 10^{-6}.

The generated multispectral image, its reconstruction and the residual image resulting from the absolute difference between the image and its reconstruction are displayed in Figure 4.6. This figure shows that the proposed algorithm correctly reconstructs the mixtures. The computation time equals 50 s.

The estimated shape parameter are very close to the ground truth values: $\widehat{\boldsymbol{w}} = \begin{bmatrix}5.2 & 5.1\end{bmatrix}$. The ground truth and the estimated maps of the amplitudes and delays are shown in Figure 4.7. One

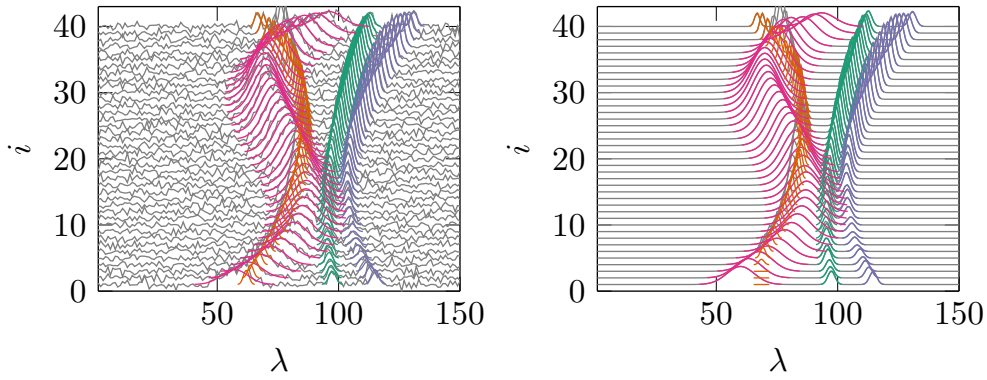


Figure 4.4: (a) a synthetic noisy set of 1D mixtures superimposed by the ground-truth sources. (b) the mixture reconstruction superimposed by the estimated sources.

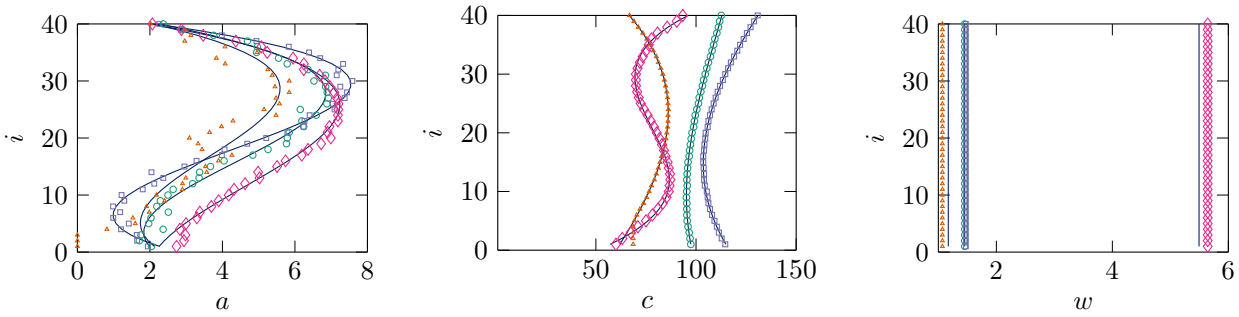
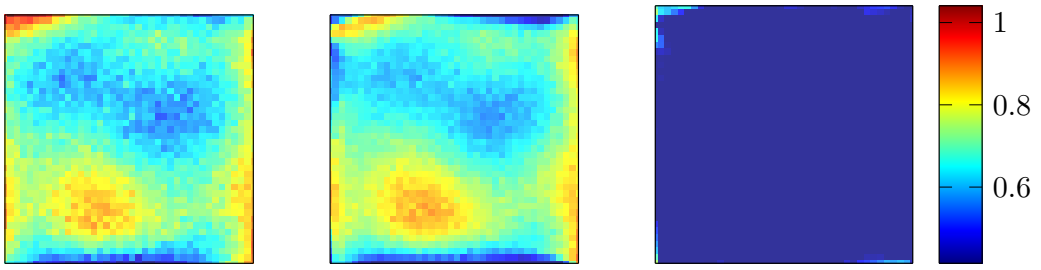


Figure 4.5: The estimated parameters (amplitudes $\widehat{a}_{ij}$, delays $c_{ij} = \widehat{\ell}_{ij}\Delta$ and shapes $\widehat{w}_j$) of the estimated sources in Figure 4.4. The ground truth parameters are plotted in plain lines.

can observe that the amplitude and delay parameters are well estimated even though the sources share the same shape parameter. The delay maps are smooth due to the regularization.

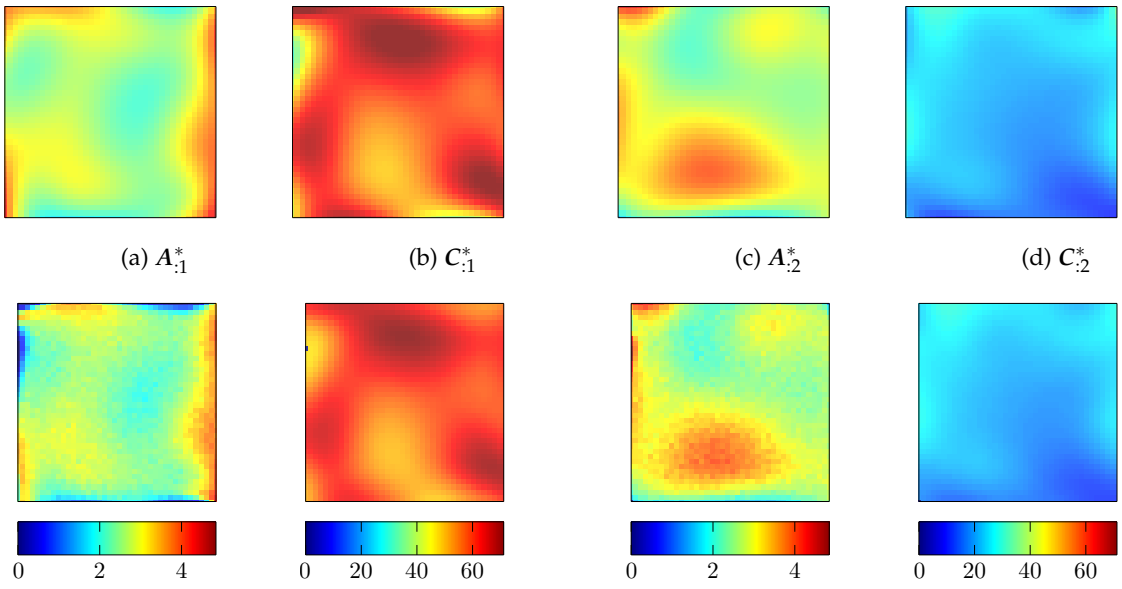
4.4.3 Regularization parameter influence on very noisy mixtures

To show the benefit of the regularization term for mixtures with high noise level (low SNR) we set the following experiment. We generate a 1D set of $I = 40$ mixtures, $N = 70$ and $J = 1$ source with SNR equals to -3 dB (Figure 4.8). The delay sampling step is fixed to $\Delta = 0.2$, the ALS and the ICM stopping parameters ρ and ξ are respectively set to 10^{-6}. Then, the proposed algorithm is run for three different regularization parameters τ. First, τ is set to zero, the delay estimation is mainly affected by the noise and the estimation is not precise this is shown in the mixture reconstruction and the estimated source (Figure 4.9a) and the estimated delays (Figure 4.9d). Second, we set a tuned value of τ equals to 0.01, the obtained results (Figures 4.9b, 4.9e) show that the delay estimation is very close to the ground truth. Finally, we set a very high τ equals to 8, obviously, the estimated delays are constant (Figures 4.9c, 4.9f) since the regularization term is much favored over the data-fit term. This simulation demonstrates that the proposed algorithm can deal with very



(a) Data (b) Reconstruction (c) Absolute residual

Figure 4.6: The white images of the generated multispectral image, the reconstruction and the absolute residual.



(a) $A^*_{:1}$ (b) $C^*_{:1}$ (c) $A^*_{:2}$ (d) $C^*_{:2}$

(e) $\widehat{A}_{:1}$ (f) $\widehat{C}_{:1}$ (g) $\widehat{A}_{:2}$ (h) $\widehat{C}_{:2}$

Figure 4.7: First row: the ground truth parameters (amplitudes $\boldsymbol{A}$ and delays $\boldsymbol{C} = \boldsymbol{L}\Delta$) of the two sources. Second row: the corresponding estimated parameters.

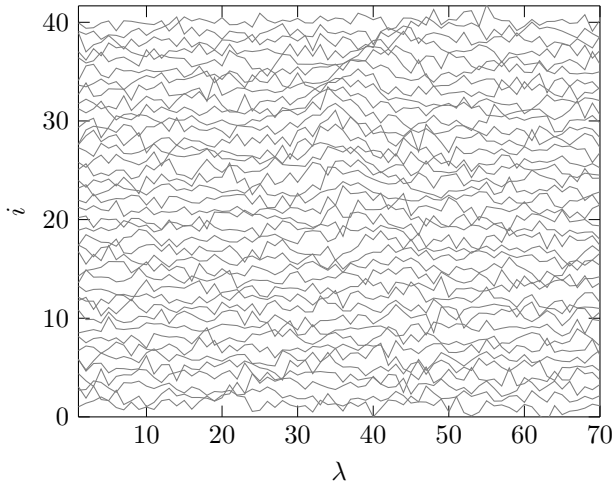
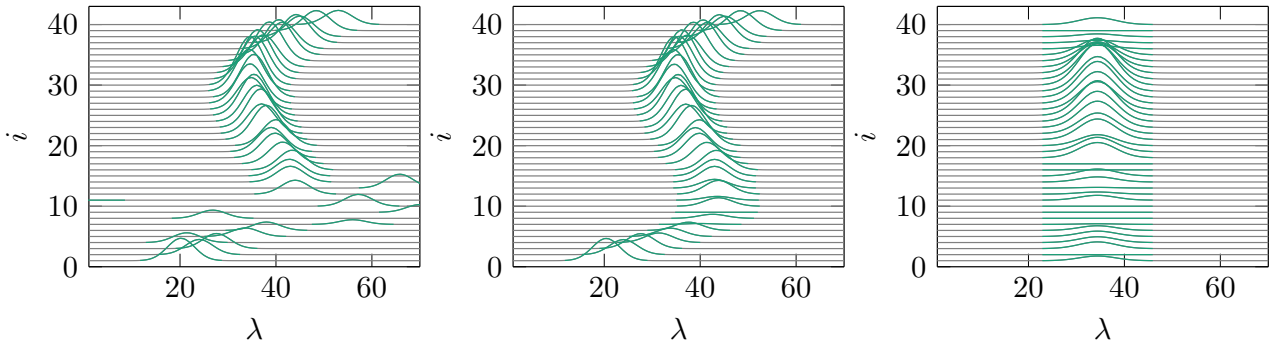
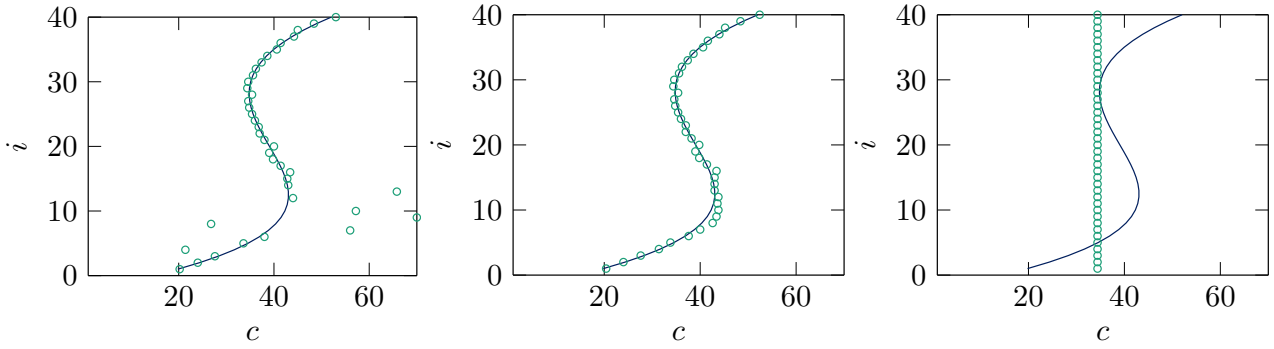


Figure 4.8: A very noisy synthetic set of 1D mixtures (SNR = -3 dB) with $J = 1$ source.



(a) $\tau = 0$ (b) $\tau = 0.01$ (c) $\tau = 8$



(d) $\tau = 0$ (e) $\tau = 0.01$ (f) $\tau = 8$

Figure 4.9: The first and second row respectively display the estimated sources superimposed on the mixture reconstruction and the estimated delays superimposed on the ground-truth delays. Each column shows the results obtained for a different regularization parameter τ value.

noisy mixtures and that the regularization allows the algorithm to be more robust to high noise levels.

4.5 Conclusion

Taking into account the slow parameter evolution is important since it allows to consider a physical observation that is valid in many applications and since it allows to better discriminate the sources. In this chapter, we propose to favor the slow evolution of the parameters by adding a regularization term to the data-fit term. The resulting criterion is optimized using an ALS scheme. The amplitudes and delays are estimated jointly within this scheme by using a joint greedy algorithm inspired by OMP. Results on synthetic 1D and 2D sets of mixtures validate the proposed methods and prove that the consideration of the slow delay evolution is useful, especially for highly correlated sources. In the next chapter, the slow evolution of the parameter is considered in an alternative way; the parameter evolutions are modeled using B-splines.

5 Source parameter modeling using B-splines

5.1 Introduction

In this chapter we exploit an alternative strategy, of the two previous chapters, to estimate the parameters (amplitudes, delays and shape parameters) and to consider their slow evolution. We exploit the B-splines to model the temporal or spatial evolution of the parameters. This modeling is embedded in the delayed and parameterized source mixing model, with the possibility to have varying shape parameters. The motivations for using B-splines are numerous. First, it allows ensuring the slow evolution of the parameter, due to the polynomial aspect of B-splines, which is a description of the considered physical phenomena. Second, the number of the parameters can be highly reduced because only a few B-spline weights (control points) describes a parameter evolution.

The road-map of this chapter is as follows. Some basic elements on B-spline basis functions and curves are presented in Section 5.2. In Section 5.3 we present the mixing model combined with the parameter B-spline modeling (1D set of mixtures) where the problem is addressed as a constrained non-linear least squares problem. The adopted optimization strategy is detailed in Section 5.4. We then extend the model and the optimization strategy for 2D sets of mixtures by modeling the spatial evolution of the source parameters as B-spline surfaces in Section 5.5. Finally, numerical results to quantify the performance of the proposed method are presented in Section 5.6.

5.2 B-splines

B-splines (*Basis* splines) were first introduced in 1946 by Schoenberg [Sch46] and are a set of piece-wise polynomial functions, the latter are referred as B-spline basis functions with very interesting properties (some of them are given later). The linear combination of these functions allows one to express a continuous function with a countable number of parameters, called control points.

B-splines have been used in many signal processing topics and have proved their effectiveness. Especially, a common application of B-splines is interpolation [Rei67; Uns99]. In image resizing and re-sampling, B-spline interpolation is used to find the pixel values over continuous positions after some geometric transformations. Moreover, B-Splines were used in image registration [Rue+06], contour detection [MZ92], signal compression [MY86] and 3D modeling [MGB94]. For spectroscopic signals, B-splines were used as a fit of absolute reflectance measurements in [Thu+03].

5.2.1 B-spline basis function

Two main elements are required to define the set of B-spline basis functions:

- a degree d (or order $d+1$), that specifies the maximal degree of the polynomial functions;
- a knot vector $\boldsymbol{k}$ that is a sequence of increasing real numbers, *i.e.*, $\boldsymbol{k} = \begin{bmatrix} k_0 & k_1 & \dots & k_K \end{bmatrix}$ with $k_0 \leq k_1 \leq \dots k_K$.

The number of knots in $\boldsymbol{k}$ is $K+1$ where each couple of two consecutive knots yields a knot interval in which a polynomial function of degree d is defined. A B-spline basis function of degree d is defined over $d+2$ consecutive knots from the support knot vector $\boldsymbol{k}$. Therefore, the number of B-spline basis functions of degree d that can be generated over a knot vector $\boldsymbol{k}$ with $K+1$ knots is: $M = K - d$.

We denote by $b_m^d(x)$ the mth B-spline basis function ($m \in \{1, \dots, M\}$) of degree d and defined over $k_m, \dots, k_{m+d+1}$. x denotes here an arbitrary variable and must be distinguished from a mixture sample $x(\lambda)$. Following the *Cox de Boor recursion formula* [DB72], a B-spline basis function can be generated using the following recurrence relations:

$$b_m^d(x) = \frac{x - k_m}{k_{m+d} - k_m} b_m^{d-1}(x) + \frac{k_{m+d+1} - x}{k_{m+d+1} - k_{m+1}} b_{m+1}^{d-1}(x), \tag{5.1}$$

$$b_m^0(x) = \begin{cases} 1, & \text{if} \quad k_m \leq x < k_{m+1} \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Another way to generate a B-spline basis function is to apply a d repetitive convolution on $b_m^0(x)$, such that:

$$b_m^d(x) = \underbrace{(b_m^0 * b_m^0 * \dots * b_m^0)}_{\text{d times}}(x). \tag{5.3}$$

Figure 5.1 shows the B-spline basis functions b_0^0, b_0^1, b_0^2, and b_0^3 and their corresponding knots.

B-spline basis function properties

In the following list, we recall the main properties of B-spline basis functions (for detailed study of these and other properties, the reader is referred to [DB72]):

- Local support: the support of a B-spline basis function is limited:

$$b_m^d(x) = 0 \quad \text{if } x \notin [k_m, k_{m+d+1}]; \tag{5.4}$$

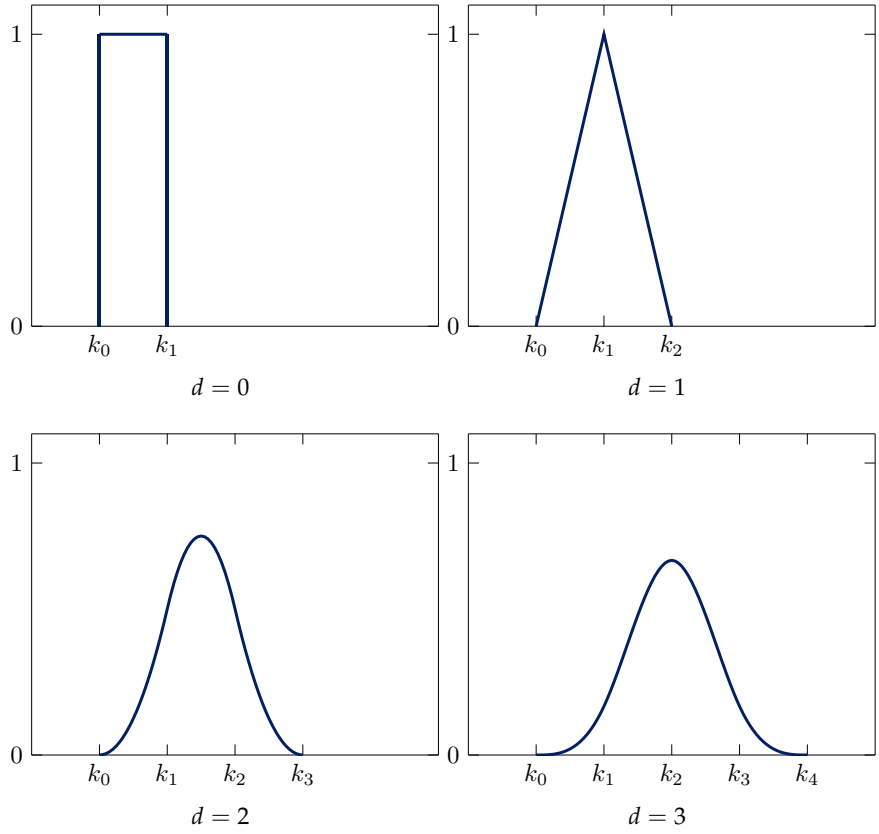


Figure 5.1: B-spline basis functions of degree 0, 1, 2 and 3 and their corresponding uniform (equidistant) knots.

- the B-spline basis functions are null on the extreme knots defining their interval:

$$\forall\, d \geq 1, \qquad b_m^d(k_m) = b_m^d(k_{m+d+1}) = 0; \tag{5.5}$$

- non-negativity: a B-spline basis function is non-negative:

$$\forall x \in \mathbb{R}, \qquad b_m^d(x) \geq 0; \tag{5.6}$$

- sum-to-one: the sum of all the B-splines basis on the knot interval $[k_m, k_{m+1})$ is 1, if and only if, $d+1$ basis functions are non-null on the knot interval:

$$\sum_{m=1}^{M} b_m^d(x) = 1; \tag{5.7}$$

- continuity: B-spline basis functions belong to the highest possible class of continuity, that is all the derivatives of $b_m^d(x)$ exist in the interior of a knot interval:

$$b_m^d(x) \in \mathcal{C}^{d-1}([k_m, k_{m+d+1}]). \tag{5.8}$$

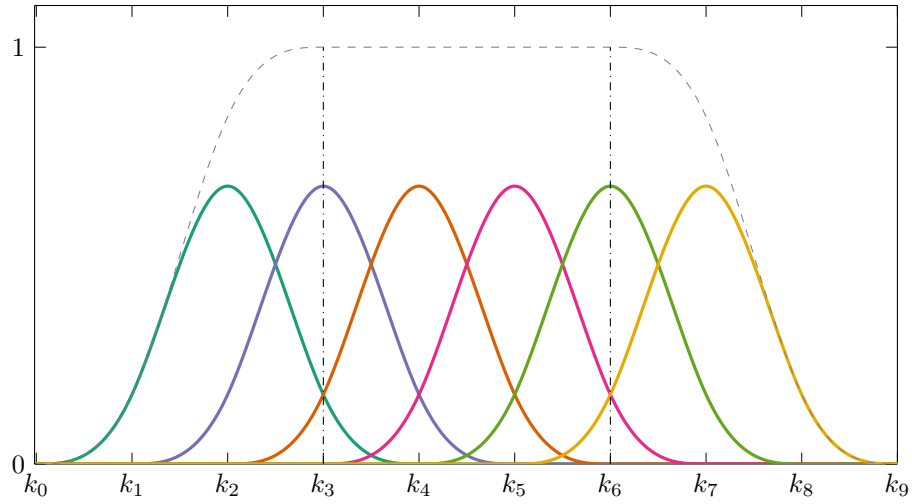


Figure 5.2: 6 cubic B-spline basis functions ($d = 3$) with 10 uniform knots. The sum of the basis functions is plotted in dashed gray line. The interval inside of the two vertical lines is the interval where the sum-to-one property is valid.

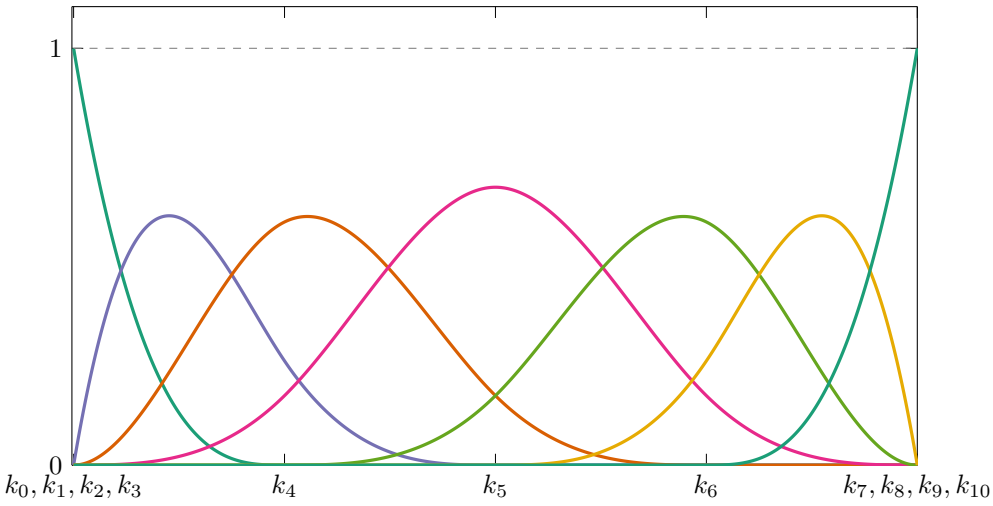


Figure 5.3: 6 cubic B-spline basis functions ($d = 3$) with 11 knots at 5 distinct positions (the first and last $d + 1$ knots are coincident). The sum of the basis functions is plotted in dashed gray line. The sum-to-one property is valid everywhere.

B-spline settings

The definition of B-splines needs to choose its degree and the associated knot vector. Common choices are:

- the degree d: cubic B-splines ($d = 3$) are a good compromise between effectiveness and computational complexity and have the minimal curvature property [Uns99]. Cubic B-splines are considered in this thesis;

- the knot vector: the most common setting is to consider a uniform knot vector, that is all the knots are equally spaced. This has some computational advantage because in this case, the B-spline basis functions are expressed as shifted versions of each others, *i.e.*,

$$\forall\, m \in \{1, \dots M\}, \qquad b_m^d(x) = b_0^d(x - k_m). \tag{5.9}$$

  Figure 5.2 shows an example of cubic B-spline basis functions ($d = 3$) with uniform knot vector. A non-uniform knot vector gives the ability to have a variable resolution of the data representation; closer knots increase the resolution.

- sum-to-one enforcement by coincident knots: if a knot vector $\boldsymbol{k} \in \mathbb{R}^{K+1}$ is defined over $K+1$ different positions, the sum-to-one property cannot be satisfied on the first and last d knot intervals as it is shown in Figure 5.2, and the B-spline is always null at the boundary knots k_0 and k_K (when $d \geq 1$). A common setting to overcome this limitation, is to use $d+1$ multiplicity on the boundary knots k_0 and k_K, that is $k_0 = k_1 = \ldots = k_d$ and $k_K = k_{K-1} = \ldots = k_{K-d}$. Thus, the $K+1$ knot vector elements are:

$$k_0 = \ldots = k_d \leq \ldots \leq k_{K-d} = \ldots = k_K.$$

  Figure 5.3 displays an example of cubic B-spline basis functions ($d = 3$) with coincident boundary knots. In the sequel, this setting is adopted for constructing the knot vectors.

5.2.2 Linear combination of B-splines

We stated in section 5.2.1 that one can construct $M = K - d$ B-spline basis functions of degree d defined over a knot vector containing $K+1$ knots. The linear combination of these M basis functions yields a continuous B-spline curve $f(x)$ of degree d, such that:

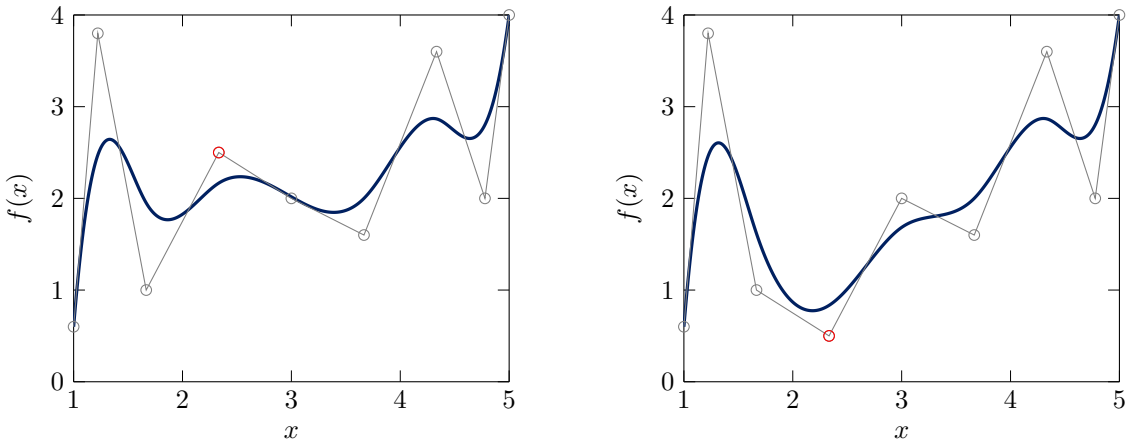
$$f(x) = \sum_{m=1}^{M} u_m b_m^d(x), \tag{5.10}$$

where $u_m \in \mathbb{R}$ is the control point of the mth B-spline basis function $b_m^d(x)$. The control points are sometimes denoted as coefficients or weights. Equation (5.10) can be rewritten in vector notation as:

$$f(x) = \boldsymbol{b}[x]^T \boldsymbol{u}, \tag{5.11}$$

where $\boldsymbol{b}[x] \triangleq \begin{bmatrix} b_1^d(x) & b_2^d(x) & \ldots & b_M^d(x) \end{bmatrix}^T$ and $\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 & \ldots & u_M \end{bmatrix}^T$. We omit the degree d from the notation $\boldsymbol{b}[x]$ for brevity. Note that, since the B-spline functions $b_m^d(x)$ have local supports, the control points u_m only have influence on the curve for $x \in [k_m\ k_{m+d+1}]$. Figure 5.4 plots an example of a B-spline curve and shows the local influence of the control points.

Two applications arise as a direct application of B-spline curve modeling: interpolation and approximation problems. Given a set of data points, the B-spline interpolation aims to build a knot vector and estimate the B-spline weights in such way that the B-spline curve passes exactly through the data points. However, in the presence of noise, interpolation becomes useless since it models the noise in addition to the data. In practice, B-spline approximation allows one to smoothly approximate the data by tolerating a small error. Namely, the data are approximated by minimizing



(a) (b)

Figure 5.4: (a) a B-spline curve with knot vector $\boldsymbol{k} = [1\ 1\ 1\ 1\ 1.6\ 2.3\ 3\ 3.6\ 4.3\ 5\ 5\ 5\ 5]$ and control points $\boldsymbol{u} = [0.6\ 3.8\ 1\ 2.5\ 2\ 1.6\ 3.6\ 2\ 4]$. Each circle is associated with a B-spline basis function with the ordinate of the corresponding control point. (b) The same B-spline curve as in (a) is considered except for the 4th element of $\boldsymbol{u}$ that is set to 0.5 instead of 2.5; this shows the local influence of a control point.

the error that is often the squared difference between the data ordinates and the B-spline curve at the position of these data.

5.3 Source parameter B-spline modeling

In this section, we present how to model the source parameters (amplitudes, delays, and shapes) using B-splines so as to favor their slow evolution. This parameter modeling is then embedded in the mixing model.

Let us first consider an example of $I = 40$ noisy 1D set of mixtures with a unique source $J = 1$ parameterized by a Gaussian function (Figure 5.5). A B-spline curve with random control points is used to generate each parameter. The parameters are then obtained by evaluating the curves at the discrete mixture positions i. This is the direct problem, but the inverse problem is to estimate the B-spline curve control points from the observed mixtures. Further, in this example, one can remark the high reduction in the unknowns number to estimate: only 21 control points must be estimated (7 for each parameter since cubic B-splines and 10 knots are used) instead of $40 \times J \times 3 = 120$ parameters.

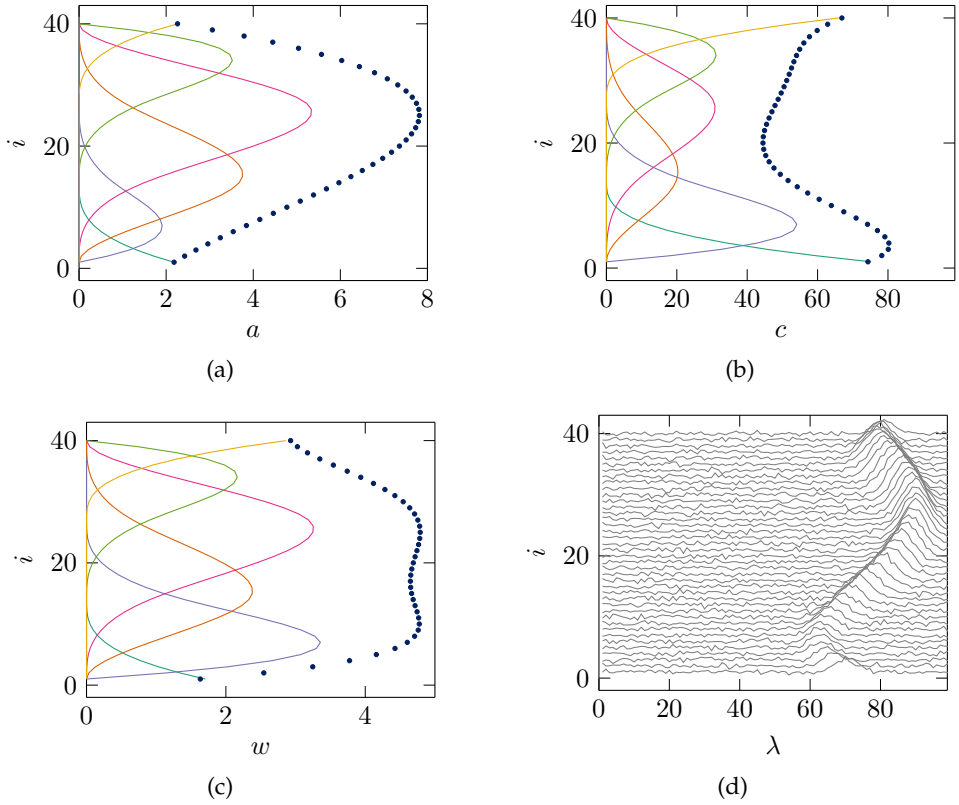


Figure 5.5: A generative example. The amplitudes, delays, and shapes (dots) of a source generated as B-spline curves and the B-spline basis functions weighted by their randomly generated control points (a)–(c). The same knot vector is used for all the parameters and is given by $\boldsymbol{k} = [1\ 1\ 1\ 1\ 14\ 27\ 40\ 40\ 40\ 40]$. (d) The generated mixtures with a Gaussian parameterized source using the parameters in (a)–(c).

5.3.1 Parameter modeling using B-spline curves

In this chapter we consider the mixing model where the shape parameters vary within the mixtures. Let us recall the corresponding model:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[c_{ij}; w_{ij}\right] + \boldsymbol{n}_i. \tag{5.12}$$

The evolution through the mixtures of the amplitudes a_{ij}, delays c_{ij} and shapes w_{ij} for each source is modeled by cubic B-spline curves with unknown control points. Without loss of generality, we suppose that all the parameters are modeled with B-splines curves defined over the same knot vector $\boldsymbol{k}$. The knot positions are bounded by $i = 1$ and $i = I$, and they can either be defined on discrete or real positions. Moreover, the first $d + 1$ knots are coincident for the reasons stated in Section 5.2.1. Similarly, the last $d + 1$ knots are coincident.

The amplitude evolution B-spline curve modeling for any source j can be seen as a continuous function of i, *i.e.*, $a_j(i)$. But, only the discrete values of $i \in \{1, \dots, I\}$ interest us:

$$\forall\, j, \qquad a_{ij} \triangleq a_j(i) = \sum_{m=1}^{M} \phi_m^j b_m^d(i) = \boldsymbol{b}[i]^T \boldsymbol{\phi}_j, \tag{5.13}$$

where $\boldsymbol{\phi}_j = \begin{bmatrix} \phi_1^j & \phi_2^j & \dots & \phi_M^j \end{bmatrix}^T$ is a vector gathering the amplitude control points of source j and $\boldsymbol{b}[i] = \begin{bmatrix} b_1^d(i) & b_2^d(i) & \dots & b_M^d(i) \end{bmatrix}^T$ is is a vector gathering the B-spline basis functions evaluated at mixture index i. Similarly, the delay and shape modeling writes as:

$$\forall\, j, \qquad c_{ij} = \sum_{m=1}^{M} \sigma_m^j b_m^d(i) = \boldsymbol{b}[i]^T \boldsymbol{\sigma}_j, \tag{5.14}$$

$$\forall\, j, \qquad w_{ij} = \sum_{m=1}^{M} \omega_m^j b_m^d(i) = \boldsymbol{b}[i]^T \boldsymbol{\omega}_j, \tag{5.15}$$

where the delay and shape control points of source j are respectively gathered in $\boldsymbol{\sigma}_j = \begin{bmatrix} \sigma_1^j & \sigma_2^j & \dots & \sigma_M^j \end{bmatrix}^T$ and $\boldsymbol{\omega}_j = \begin{bmatrix} \omega_1^j & \omega_2^j & \dots & \omega_M^j \end{bmatrix}^T$.

It is worth noting that the B-spline modeling reduces the number of the unknown for each parameter from $(I \times J)$ to $(M \times J)$ with M much smaller than I.

Note that, the proposed parameter modeling can easily consider the particular case where the shape parameters of sources are invariant within the mixtures (the model used in Chapters 3 and 4). To consider this model, the shapes must be modeled as a B-spline curve with degree $d = 0$ and two knots where first is positioned on $i = 1$ and the second on $i = I$. Indeed, the B-splines of degree 0 are constant piecewise functions.

5.3.2 Source separation meets B-spline curve modeling

By combining the mixing model (5.12) and the parameter B-spline modeling in (5.13), (5.14) and (5.15) we get:

$$\boldsymbol{x}_i = \sum_{j=1}^{J} (\boldsymbol{b}[i]^T \boldsymbol{\phi}_j) \boldsymbol{s}\left[\boldsymbol{b}[i]^T \boldsymbol{\sigma}_j; \boldsymbol{b}[i]^T \boldsymbol{\omega}_j\right] + \boldsymbol{n}_i \quad i = 1, \dots, I. \tag{5.16}$$

The problem of estimating the parameter with the assumption of slow evolution is recast to the estimation of the control points. Supposing the noise to be Gaussian and i.i.d, the maximum likelihood estimation is obtained by minimizing the following criterion:

$$\mathcal{L}(\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} (\boldsymbol{b}[i]^T \boldsymbol{\phi}_j) \boldsymbol{s}\left[\boldsymbol{b}[i]^T \boldsymbol{\sigma}_j; \boldsymbol{b}[i]^T \boldsymbol{\omega}_j\right] \right\|_2^2, \tag{5.17}$$

where:

$$\begin{aligned}\boldsymbol{\Phi} &= \begin{bmatrix}\boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \dots & \boldsymbol{\phi}_J\end{bmatrix} \in \mathbb{R}^{M\times J},\\ \boldsymbol{\Sigma} &= \begin{bmatrix}\boldsymbol{\sigma}_1 & \boldsymbol{\sigma}_2 & \dots & \boldsymbol{\sigma}_J\end{bmatrix} \in \mathbb{R}^{M\times J},\\ \boldsymbol{\Omega} &= \begin{bmatrix}\boldsymbol{\omega}_1 & \boldsymbol{\omega}_2 & \dots & \boldsymbol{\omega}_J\end{bmatrix} \in \mathbb{R}^{M\times J}.\end{aligned} \tag{5.18}$$

In contrary to the methods proposed in Chapter 4, we do not need a regularization term to favor the parameter slow evolution; the data-fit term is enough since the slow evolution is directly embedded in the model. Therefore, the optimization problem reads:

$$\min_{\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}} \mathcal{L}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}). \tag{5.19}$$

Constraints

The specificities of the spectroscopic application are considered by adding constraints to the optimization problem (5.19). The amplitudes must be non-negative, the delays are bounded by the two extreme values of the wavelength vector. Besides, we impose the shapes to be bounded by $w_{\min}$ and $w_{\max}$. We apply the same constraints on the control points modeling the parameters. Thus we have the following constrained optimization problem:

$$\min_{\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}} \mathcal{L}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}) \qquad \text{such that } \mathcal{C} = \begin{cases}\boldsymbol{\Phi} \in \mathbb{R}_+^{M\times J}\\ \boldsymbol{\Sigma} \in [1,N]^{M\times J}\\ \boldsymbol{\Omega} \in [w_{\min}, w_{\max}]^{M\times J}\end{cases}. \tag{5.20}$$

These constraints on the control points allow one to respect the constraints on the physical parameters a_{ij}, c_{ij} and w_{ij}. However, some possible solutions are excluded since the parameters are the linear combination of the control points and the B-spline basis functions. For instance, the non-negativity constraint on the physical amplitudes, yields:

$$a_{ij} \geq 0, \tag{5.21}$$

therefore:

$$\forall\, i,j, \qquad \sum_{m=1}^{M} \phi_m^j b_m^d(i) \geq 0. \tag{5.22}$$

Knowing that $b_m^d(i) \geq 0\ \forall\ m,i$ and that $\sum_m b_m^d(i) = 1$, for some ϕ_m^j that are non-positive the equation (5.22) might still holds. Even though some solutions are excluded by considering the constraints in $\mathcal{C}$. We choose to tolerate this limitation for implementation reasons and since this only have negligible influence on the result.

5.4 Proposed optimization strategy

The joint optimization problem (5.20) can be addressed as a constrained non-linear least squares problem. Indeed, the non-linearity of model (5.16) is due to the presence of the delay σ_m^j and shape

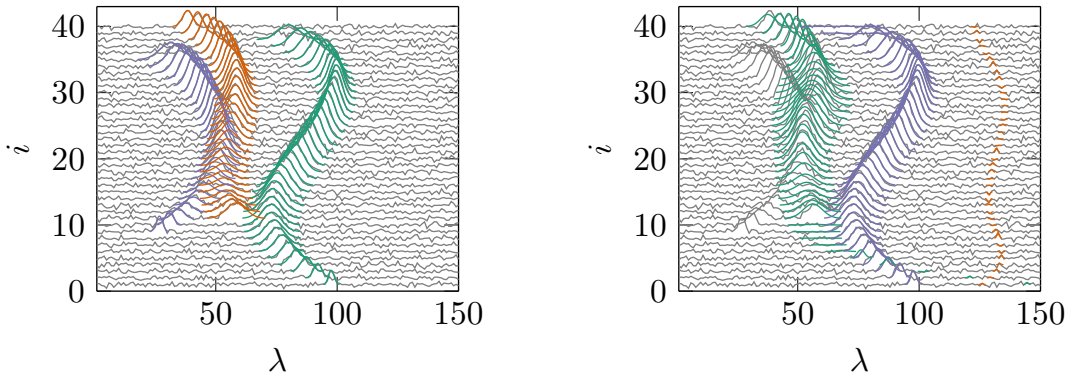


Figure 5.6: (a) $I = 40$ noisy mixtures and the ground truth sources. (b) the noisy mixtures superimposed by the estimated sources obtained by minimizing equation (5.20) initialized by random control points. All the parameters are modeled with cubic B-spline curves with knot vector $\boldsymbol{k} = [1\ \ 1\ \ 1\ \ 1\ \ 10.7\ \ 20.5\ \ 30.2\ \ 40\ \ 40\ \ 40\ \ 40]$.

ω_m^j control points. We propose to use a Sequential Quadratic Programming (SQP) algorithm [NW99] to solve this optimization problem; it gives the possibility to solve constrained non-linear least squares problem. SQP iteratively updates the solution, starting with an initial solution. We respectively denote by $\boldsymbol{\Phi}^0$, $\boldsymbol{\Sigma}^0$ and $\boldsymbol{\Omega}^0$ the initial solutions for the amplitude, delay and shape control points.

The criterion (5.17) is non-convex and the choice of the initial solutions for optimization solver is crucial, especially for the delay control points (see Figure 3.1). The choice of the initial solution is not an easy task, and it has a significant influence on the output results. Figure 5.6 shows the output result of a $I = 40$ mixtures with $J = 3$ sources. The result is obtained by solving the optimization problem (5.20) using SQP algorithm where the control points are randomly initialized. The result is not satisfactory since one source models the noise. The reason is that the (random) initialization of the delay control points is far from the actual source delays. Therefore we propose to exploit information from the observed mixtures to initialize the delay control points.

We adopt an iterative strategy where the number of estimated sources is incremented by one at each iteration. In addition to the estimation of a new source, the parameters of the previous ones are also updated at each iteration. We propose to initialize the delay control points as the wavelength of the highest energy in the average mixture (computed by summing up all the mixtures) after removing the contributions of the previously estimated sources; this is a simple and efficient choice for finding a potential delay, around which a source takes place.

The proposed algorithm is given in Algorithm 5. At each iteration, the number of sources to estimate is incremented from $k = 1$ to $k = J$. In other words, at each iteration, the optimization problem (5.20) is solved with respect to a subset of the parameters, namely the first k columns of $\boldsymbol{\Phi}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ by using an SQP algorithm. At iteration $k+1$, the solution previously found at iteration k is used to initialize the first k sources (corresponds to the k first columns of $\boldsymbol{\Phi}^0$, $\boldsymbol{\Sigma}^0$ and $\boldsymbol{\Omega}^0$) . The $k+1$ column of $\boldsymbol{\Phi}^0$ and $\boldsymbol{\Omega}^0$ is set to a random value generated following a uniform distribution $\mathcal{U}$ over the intervals in $\mathcal{C}$. All the elements in the $k+1$ column of $\boldsymbol{\Sigma}^0$ are set, as said above, to the wavelength with the highest intensity in the average residual spectrum $\lambda_{\max}$.

Algorithm 5: Minimization of $\mathcal{L}(\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{\Omega})$ s.t. $(\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{\Omega}) \in \mathcal{C}$

Outputs: $\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Omega}}$
Initialization: $\widehat{\mathbf{\Phi}} = \mathbf{0}_{M \times J}$, $\widehat{\mathbf{\Sigma}} = \mathbf{0}_{M \times J}$, $\widehat{\mathbf{\Omega}} = \mathbf{0}_{M \times J}$
for $k = 1 : J$ **do**

Solve

$$(\widehat{\mathbf{\Phi}}_{:,1:k}, \widehat{\mathbf{\Sigma}}_{:,1:k}, \widehat{\mathbf{\Omega}}_{:,1:k}) \leftarrow \underset{\mathbf{\Phi}_{:,1:k}, \mathbf{\Sigma}_{:,1:k}, \mathbf{\Omega}_{:,1:k}}{\operatorname{argmin}} \mathcal{L}(\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{\Omega}) \text{ s.t. } (\mathbf{\Phi}_{:,1:k}, \mathbf{\Sigma}_{:,1:k}, \mathbf{\Omega}_{:,1:k}) \in \mathcal{C}$$

using SQP algorithm, initialized as follows:

$$\mathbf{\Phi}^0_{:,1:k} = \begin{bmatrix} \widehat{\mathbf{\Phi}}_{:,1:k-1} & \boldsymbol{\phi}^0_k \end{bmatrix} \text{ with } \boldsymbol{\phi}^0_k \sim \mathcal{U}(0, +\infty)$$

$$\mathbf{\Omega}^0_{:,1:k} = \begin{bmatrix} \widehat{\mathbf{\Omega}}_{:,1:k-1} & \boldsymbol{\omega}^0_k \end{bmatrix} \text{ with } \boldsymbol{\omega}^0_k \sim \mathcal{U}(w_{\min}, w_{\max})$$

$$\mathbf{\Sigma}^0_{:,1:k} = \begin{bmatrix} \widehat{\mathbf{\Sigma}}_{:,1:k-1} & \boldsymbol{\sigma}^0_k \end{bmatrix} \text{ with } \boldsymbol{\sigma}^0_k = \lambda_{\max} \mathbf{1}_M$$

$\lambda_{\max}$ is the wavelength with the highest energy in the residual pixel sum:

$$\lambda_{\max} = \underset{\lambda}{\operatorname{argmax}} \sum_i \left(x_i(\lambda) - \sum_{j=1}^{k-1} \boldsymbol{b}[i]^T \widehat{\boldsymbol{\phi}}_j s(\lambda - \boldsymbol{b}[i]^T \widehat{\boldsymbol{\sigma}}_j; \boldsymbol{b}[i]^T \widehat{\boldsymbol{\omega}}_j) \right)^2$$

end

5.5 Extension to two dimensional mixtures

The extension of B-splines to the case of two-dimensional functions is straightforward. In this section, we introduce the bivariate B-spline basis functions, before detailing their linear combinations to construct B-spline surfaces. Finally, we combine the source separation model in the case of multispectral images and the parameter modeling as B-spline surfaces with unknown control points.

5.5.1 Bivariate B-spline basis functions

Let us suppose a two-dimensional space defined by the variables u and v. Let $b_p^{d_u}(u)$ be a one-dimensional B-spline basis function of degree d_u and knot vector $\boldsymbol{k}_u \in \mathbb{R}^{K_u+1}$ defined over the first spatial dimension u (see Section 5.2.1). Similarly, $b_q^{d_v}(v)$ is a B-spline basis function of degree d_v and knot vector $\boldsymbol{k}_v \in \mathbb{R}^{K_v+1}$ defined over the second spatial dimension v.

The two knot vectors $\boldsymbol{k}_u$ and $\boldsymbol{k}_v$ generate a knot grid containing $(K_u+1) \times (K_v+1)$ knots. Again, the knot vectors can be either uniform, non-uniform and with coincident boundary knots (see Section 5.2.1). For the sake of brevity, and without loss of generality, we suppose that $d_u = d_v = d$.

We denote by $i = \begin{bmatrix} u & v \end{bmatrix} \in \mathbb{N}^2$ the spatial index of a pixel in an image where u and v are respectively the column and the row of the pixel. The bivariate B-spline basis function with degree d at the spatial position i is denoted as:

$$\beta^d_{p,q}(i) = b^d_p(u) b^d_q(v). \tag{5.23}$$

Figure 5.7 shows a cubic ($d = 3$) bivariate B-spline basis function $\beta^3_{0,0}(i)$.

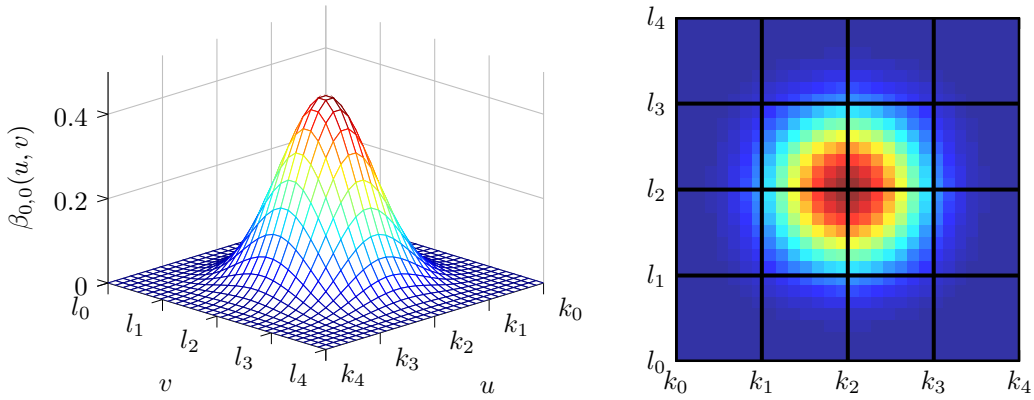


Figure 5.7: (a) a perspective view of a bivariate B-spline basis function of degree $d = 3$. (b) a top view of the bivariate B-spline basis function and the knot grid in black lines. The knot grid contains $(d+2)^2 = 25$ knots.

5.5.2 B-spline surfaces

A B-spline surface function $f(i) \triangleq f(u,v)$ is the linear combination of $M = P \times Q$ bivariate B-spline basis functions such that:

$$f(i) = \sum_{p=1}^{P} \sum_{q=1}^{Q} \mu_{p,q} \beta_{p,q}^{d}(i), \tag{5.24}$$

where $\mu_{p,q}$ is the control point of the B-spline bivariate basis function $\beta_{p,q}(i)$. Using vector notation, equation (5.24) writes:

$$f(i) = \boldsymbol{\beta}[i]^T \boldsymbol{\mu}, \tag{5.25}$$

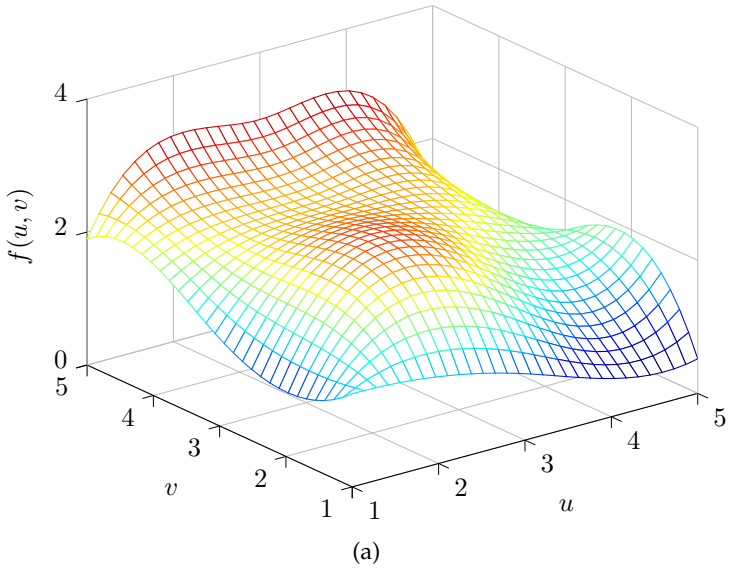
where $\boldsymbol{\beta}[i] = \begin{bmatrix} \beta_{1,1}^{d}(i) & \beta_{1,2}^{d}(i) & \dots & \beta_{1,P}^{d}(i) & \beta_{2,1}^{d}(i) & \dots & \dots & \beta_{P,Q}^{d}(i) \end{bmatrix}^T \in \mathbb{R}^M$.
and $\boldsymbol{\mu} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \dots & \mu_{1,P} & \mu_{2,1} & \dots & \dots & \mu_{P,Q} \end{bmatrix}^T \in \mathbb{R}^M$. An example of a cubic B-spline surface is given in Figure 5.8.

5.5.3 Parameter B-spline surface modeling

The amplitudes, delays, and shapes are modeled as B-spline surfaces. Therefore, the amplitude, delay and shape parameters modeling at mixture i for a given source j respectively write:

$$\forall\, j, \qquad a_{ij} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \phi_{p,q}^{j} \beta_{p,q}^{d}(i) = \boldsymbol{\beta}[i]^T \boldsymbol{\phi}_j, \tag{5.26}$$

$$\forall\, j, \qquad c_{ij} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sigma_{p,q}^{j} \beta_{p,q}^{d}(i) = \boldsymbol{\beta}[i]^T \boldsymbol{\sigma}_j, \tag{5.27}$$



(a)

Figure 5.8: A B-spline surface with degree $d_x = d_y = d = 3$ and knot vectors $\boldsymbol{k} = \boldsymbol{l} = [1\ 1\ 1\ 1\ 2.3\ 3.6\ 5\ 5\ 5\ 5]$. The corresponding control point vector $\boldsymbol{\mu}$ is not displayed.

$$\forall\, j, \qquad w_{ij} = \sum_{p=1}^{P}\sum_{q=1}^{Q} \omega_{p,q}^{j}\beta_{p,q}^{d}(i) = \boldsymbol{\beta}[i]^T\boldsymbol{\omega}_j, \tag{5.28}$$

where the amplitude, delay and shape control points of source j are respectively gathered in:

$$\begin{aligned} \boldsymbol{\phi}_j &= \begin{bmatrix} \phi_{1,1}^j & \phi_{1,2}^j & \dots & \phi_{1,P}^j & \phi_{2,1}^j & \dots & \dots & \phi_{P,Q}^j \end{bmatrix}^T, \\ \boldsymbol{\sigma}_j &= \begin{bmatrix} \sigma_{1,1}^j & \sigma_{1,2}^j & \dots & \sigma_{1,P}^j & \sigma_{2,1}^j & \dots & \dots & \sigma_{P,Q}^j \end{bmatrix}^T, \\ \text{and } \boldsymbol{\omega}_j &= \begin{bmatrix} \omega_{1,1}^j & \omega_{1,2}^j & \dots & \omega_{1,P}^j & \omega_{2,1}^j & \dots & \dots & \omega_{P,Q}^j \end{bmatrix}^T. \end{aligned} \tag{5.29}$$

Source separation meets B-spline surface modeling

The new model is obtained by combining the source separation model equation (5.12) and the B-spline parameters in (5.26), (5.27) and (5.28):

$$\boldsymbol{x}_i = \sum_{j=1}^{J} (\boldsymbol{\beta}[i]^T\boldsymbol{\phi}_j)\boldsymbol{s}\left[\boldsymbol{\beta}[i]^T\boldsymbol{\sigma}_j; \boldsymbol{\beta}[i]^T\boldsymbol{\omega}_j\right] + \boldsymbol{n}_i \quad \forall i. \tag{5.30}$$

The source separation problem is recast to the estimation of the control points gathered in $\boldsymbol{\phi}_j$, $\boldsymbol{\sigma}_j$ and $\boldsymbol{\omega}_j$, therefore we define the following criterion:

$$\mathcal{O}(\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} (\boldsymbol{\beta}[i]^T\boldsymbol{\phi}_j)\boldsymbol{s}\left[\boldsymbol{\beta}[i]^T\boldsymbol{\sigma}_j; \boldsymbol{\beta}[i]^T\boldsymbol{\omega}_j\right] \right\|_2^2. \tag{5.31}$$

where,

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \dots & \boldsymbol{\phi}_J \end{bmatrix}, \tag{5.32}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & \sigma_2 & \dots & \sigma_J \end{bmatrix}, \tag{5.33}$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\omega}_1 & \boldsymbol{\omega}_2 & \dots & \boldsymbol{\omega}_J \end{bmatrix}. \tag{5.34}$$

The corresponding constrained optimization problem writes:

$$\min_{\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}} \mathcal{O}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}) \quad \text{such that} \quad (\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega}) \in \mathcal{C}. \tag{5.35}$$

The same strategy that is used to optimize the criterion $\mathcal{L}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega})$ is used to solve the optimization problem equation (5.35). The Algorithm to optimize $\mathcal{O}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega})$ is similar to Algorithm 5 up to minor modifications:

- $\mathcal{O}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega})$ replaces $\mathcal{L}(\boldsymbol{\Phi},\boldsymbol{\Sigma},\boldsymbol{\Omega})$;
- $\lambda_{\max} \leftarrow \underset{\lambda}{\operatorname{argmax}} \sum_i \left(x_i(\lambda) - \sum_{j=1}^{k} \boldsymbol{\beta}[i]^T \widehat{\boldsymbol{\phi}}_j s\left(\lambda - \boldsymbol{\beta}[i]^T \widehat{\boldsymbol{\sigma}}_j; \boldsymbol{\beta}[i]^T \widehat{\boldsymbol{\omega}}_j\right) \right)^2$.

5.6 Results

5.6.1 Synthetic 1D set of mixtures

The performance of the proposed algorithm is evaluated on a synthetic 1D set of mixtures. The following settings are set. All the sources are modeled by a Gaussian function: $s(\lambda; w_j) = \exp(-\lambda^2/2w_j^2)$ where the shape parameters correspond the Gaussian width. The number of mixtures is equal to $I = 40$, each with $N = 150$ samples and the number of sources is set to $J = 3$ where two of them are spectrally overlapping. The SNR defined in equation (1.13) is equal to 15 dB. The amplitudes and delays and shapes are generated using polynomials with random coefficients. The amplitudes, delays and shapes are modeled using cubic B-splines. The amplitude knot vector is $\boldsymbol{k} = \begin{bmatrix} 1 & 1 & 1 & 1 & 20 & 40 & 40 & 40 & 40 \end{bmatrix}$, whereas the delay and shape are modeled with more knots using this knot vector $\boldsymbol{k} = \begin{bmatrix} 1 & 1 & 1 & 1 & 10 & 20 & 30 & 40 & 40 & 40 & 40 \end{bmatrix}$. The lower and upper bounds of the shape parameters control points are respectively $w_{\min} = 0.5$ and $w_{\max} = 8$. The generated mixtures and the mixture denoised reconstruction are displayed superimposed by the estimated sources are displayed in Figure 5.9. One can remark that the reconstruction and the source recovery are very satisfactory. The computation time to obtain this result is 6.1 s.

The ground truth and the estimated amplitudes, delays and shapes are displayed in Figure 5.10. These results show that the parameters are well estimated and their evolution is smooth due to their B-spline modeling.

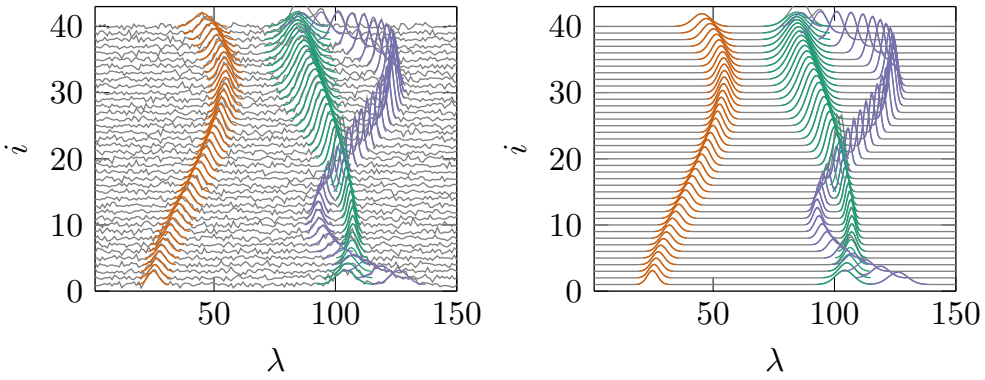


Figure 5.9: (a) a synthetic noisy set of 1D mixtures superimposed by the ground-truth sources. (b) the mixture reconstruction superimposed by the estimated sources.

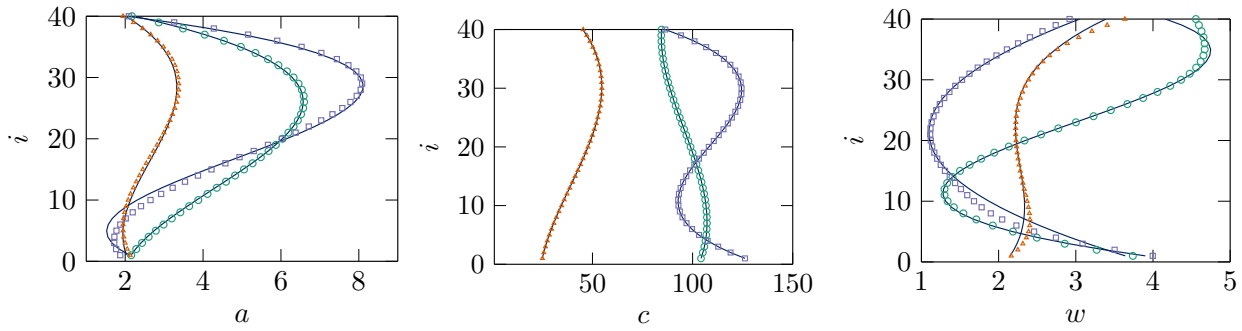


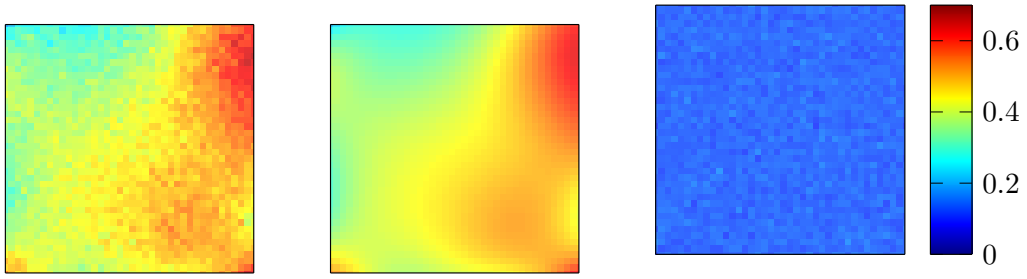
Figure 5.10: The estimated parameters (amplitudes $\widehat{a}_{ij}$, delays c_{ij} and shapes $\widehat{w}_j$) of the estimated sources in Figure 5.9. The ground truth parameters are plotted in plain lines.

5.6.2 Synthetic 2D sets of mixtures

The proposed algorithm is tested on a 2D sets of mixtures with dimension $40 \times 40 \times 100$ and $J = 2$ Gaussian parameterized sources. The SNR is set to 15 dB. The ground-truth parameters are generated using cubic B-spline surfaces with random control points. The parameters are modeled using cubic surface B-splines with the following knot vectors $\boldsymbol{k}_u = \boldsymbol{k}_v \left[1\ 1\ 1\ 1\ 20\ 40\ 40\ 40\ 40\right]$. The lower and upper bounds of the shape parameters control points are respectively $w_{\min} = 0.5$ and $w_{\max} = 8$.

The generated multispectral image, its reconstruction and the residual image resulting from the absolute difference between the image and its reconstruction are displayed in Figure 5.11. This figure shows that the proposed algorithm correctly reconstructs the mixtures. The computation time is equal to 50 s.

The ground-truth and the estimated maps of the amplitudes, delays and shapes are shown in Figure 5.12. One can observe that the amplitude, delay and shape parameters are very well estimated and spatially slow evolving as expected.



(a) data (b) reconstruction (c) absolute residual

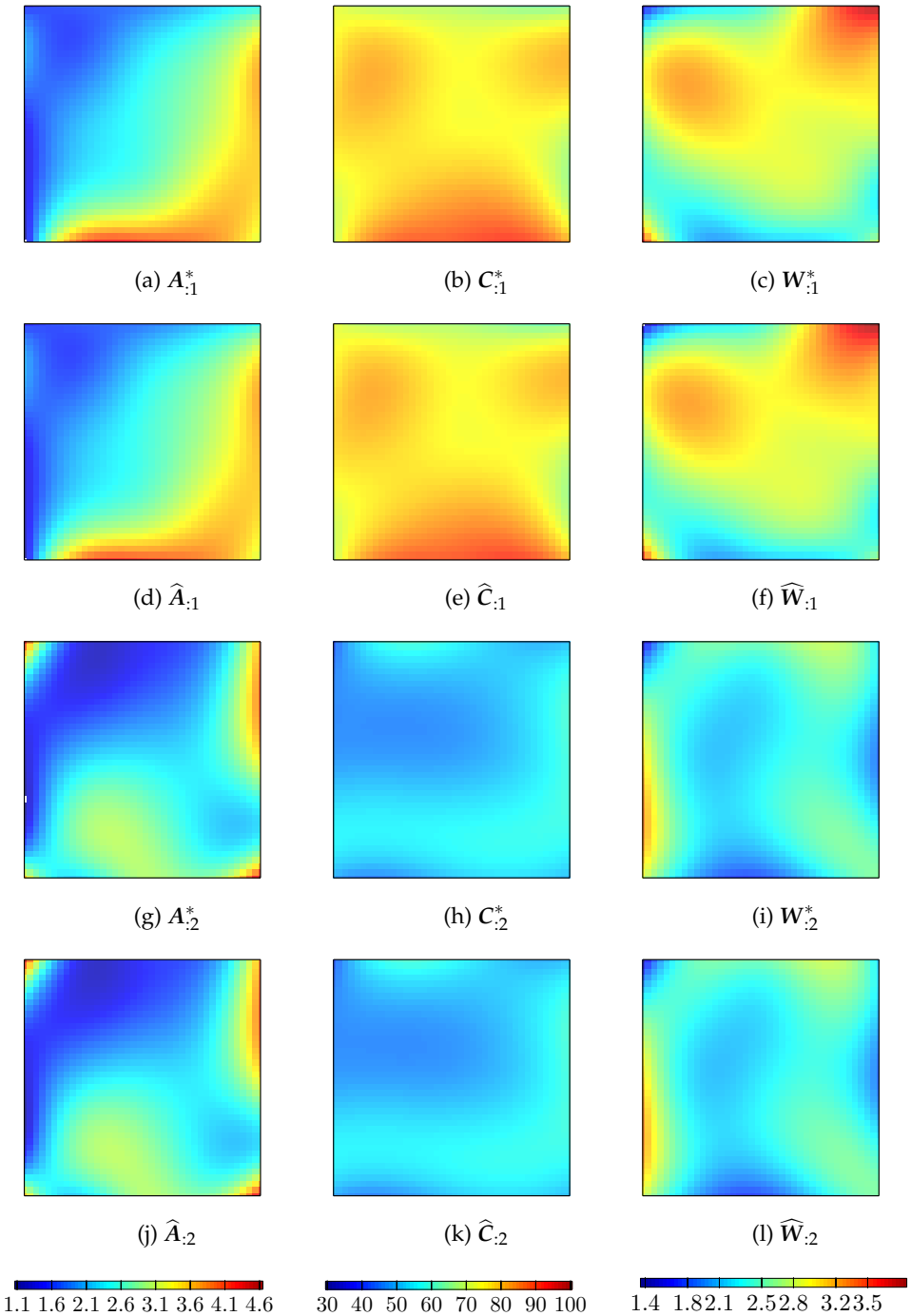
Figure 5.11: The white images of the synthetic multispectral image, its reconstruction and the residual error.

5.6.3 SNR and knot number influence

To study the impact of the knot number, we consider modeling the parameters (a_{ij}, c_{ij} and w_{ij}) with cubic B-spline curves ($d = 3$) defined over a knot vector containing different numbers of knots. Namely, we variate the knot number from 10 to 14 knots. The knot positions are equally-spaced. The five calls of the algorithm with different knot numbers are compared following different criteria with respect to the SNR which varies between 0 and 30 dB with a step of 2 dB. We generate 200 sequences of $I = 40$ noisy mixtures with SNR equal to 15 dB, each with $J = 3$ sources and $N = 150$ samples. The parameters are generated using polynomial functions with degrees varying from 3 to 5. The average of the 200 results are displayed in Figure 5.13 where we compare the effect of knot number in means of the MSE[1] and computation time.

Results show that the MSE and the parameter estimation qualities decrease as the SNR increases. On the other hand, as the number of knots increases the MSE is lower at the price of higher computation time which is expected since a larger number of knots yields more estimation flexibility (less regularized solutions) but the number of control points to estimate increases (Figure 5.13a). On the other hand, the computation time increases as the SNR increases for all the versions (Figure 5.13b). This is due to the non-linear least solvers that require more iterations for high SNR mixtures to find an accurate local minimizer where such an accurate solution cannot be found when the SNR is low.

[1] $\text{MSE} = \frac{1}{N \cdot I} \sum_i \| \boldsymbol{x}_i - \sum_{j=1}^{J} \widehat{a}_{ij} \boldsymbol{s}[\widehat{c}_{ij}; \widehat{w}_{ij}] \|_2^2$.



(a) $\boldsymbol{A}^*_{:1}$ (b) $\boldsymbol{C}^*_{:1}$ (c) $\boldsymbol{W}^*_{:1}$

(d) $\widehat{\boldsymbol{A}}_{:1}$ (e) $\widehat{\boldsymbol{C}}_{:1}$ (f) $\widehat{\boldsymbol{W}}_{:1}$

(g) $\boldsymbol{A}^*_{:2}$ (h) $\boldsymbol{C}^*_{:2}$ (i) $\boldsymbol{W}^*_{:2}$

(j) $\widehat{\boldsymbol{A}}_{:2}$ (k) $\widehat{\boldsymbol{C}}_{:2}$ (l) $\widehat{\boldsymbol{W}}_{:2}$

Figure 5.12: The estimated and the ground-truth parameters.

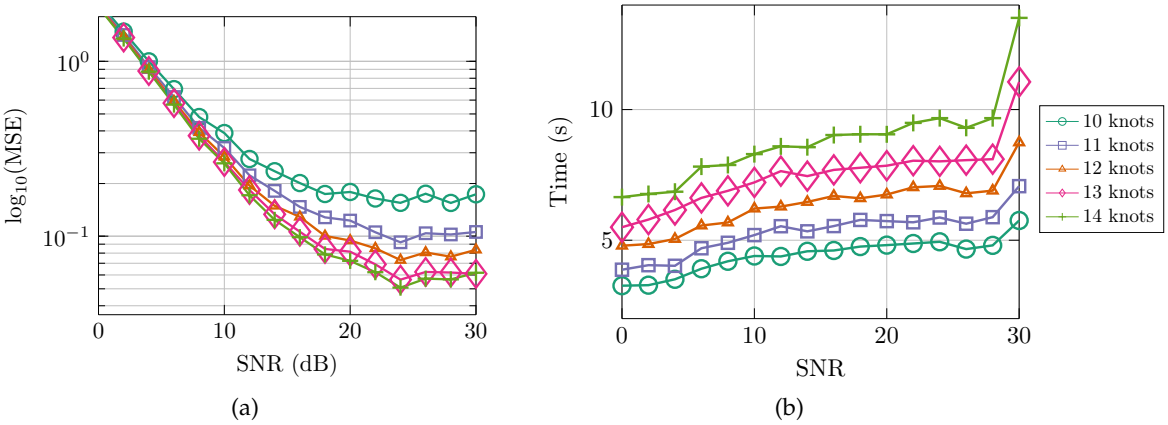


Figure 5.13: MSE and Computation time comparison with respect to the SNR for different knot numbers. The blue, green, red, cyan and purple plots respectively correspond the number of knots varying from 10 to 14.

5.7 Conclusion

In this chapter, we propose to model the parameter evolution within the mixture using B-splines. In cases of 1D set of mixtures, B-spline curves are used whereas in the case of 2D sets of mixtures B-spline surfaces are used. This modeling has numerous advantages. First, it naturally yields a slow evolution of the parameters. Second, the B-spline modeling results in a significant reduction of the number of unknowns, which is particularly interesting in the case of multispectral images. So, the parameter estimation is replaced by the estimation of their corresponding control points. The latter are found using a constrained non-linear least squares solver (SQP). Results on synthetic examples show that the proposed method can successfully achieve the goal in very acceptable computation time and that the method is robust to the noise level.

6 Simulations and application to real data

6.1 Introduction

This chapter is dedicated to illustrate the performance of the proposed methods and to compare them with state-of-the-art methods. The proposed methods are the methods of Chapter 3, 4 and 5 and in the sequel they will respectively be referred to the non-regularized, regularized and B-spline methods. The non-regularized and regularized methods both deal with delayed and parameterized source separation problem where the shape parameters of each source are constant within the mixtures, with the difference that the regularized method considers the slow delay evolution by using a regularization. The B-spline method considers a more general model by allowing for varying shape parameters for each source. The amplitudes, delays and shapes are modeled using B-splines to ensure their slow evolution.

The proposed and state-of-the-art methods will be compared in terms of the following criteria:

- computation time;
- Mean Squared Error (MSE) defined as:

$$\text{MSE} \triangleq \frac{1}{N \cdot I} \sum_{i}^{I} \|\boldsymbol{r}_i\|_2^2, \tag{6.1}$$

  where $\boldsymbol{r}_i$ is the residual of mixture i, *i.e.*, $\boldsymbol{r}_i = \boldsymbol{x}_i - \sum_{j=1}^{J} \widehat{a}_{ij} \boldsymbol{s}\left[\widehat{c}_{ij}; \widehat{w}_{ij}\right]$.

- amplitude, delay and shape errors, respectively defined as:

$$\frac{\|\boldsymbol{C}^* - \widehat{\boldsymbol{C}}\|_F^2}{\|\boldsymbol{C}^*\|_F^2}, \quad \frac{\|\boldsymbol{A}^* - \widehat{\boldsymbol{A}}\|_F^2}{\|\boldsymbol{A}^*\|_F^2} \quad \text{and} \quad \frac{\|\boldsymbol{W}^* - \widehat{\boldsymbol{W}}\|_F^2}{\|\boldsymbol{W}^*\|_F^2}, \tag{6.2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and the superscript * refers to the ground-truth parameters. Note that in Chapters 3 and 4, the shape error was computed as $\frac{\|\boldsymbol{w}^*-\widehat{\boldsymbol{w}}\|_2^2}{\|\boldsymbol{w}^*\|_2^2}$ (constant source shapes) and $\widehat{\boldsymbol{C}} = \Delta\widehat{\boldsymbol{L}}$ is the estimated delay matrix.

This chapter is organized as follows. First, a synthetic set of 1D mixtures is used for comparison between the proposed methods, two other delayed source separation methods with non-parameterized sources and a Bayesian method (Section 6.2). Second, we establish statistical comparisons with respect to different noise levels (SNR's) (Section 6.3). Finally, in Section 6.4, we show results obtained with the proposed methods on real time-resolved photoelectron spectra (1D set of mixtures) and a multispectral image of a galaxy (2D sets of mixtures).

6.2 Comparison on a synthetic 1D set of mixtures

The three proposed methods are compared with the following state-of-the-art methods:

- the method of [MMH07b] considers the delayed source separation problem where the mixtures are analyzed in the frequency domain. The number of sources J is supposed to be known and to be constant in all the mixtures. No slow evolution prior on the parameters is supposed. This method extends NMF to deal with delays: the sources and amplitudes are estimated using multiplicative updates and the delays are estimated with a gradient descent algorithm coupled with a maximizing cross-correlation procedure;
- the method of [Nio+] also considers the delayed source separation problem and the mixtures are analyzed in the time-frequency domain. The number of sources J is supposed to be known and to be constant in all the mixtures with no slow evolution prior on the parameters. An ALS scheme of two steps is used where the source signals and the amplitudes and delays are estimated in the least-squares sense;
- the method of [Maz+15] is a Bayesian method that addresses the separation of parameterized sources where the amplitudes, delays and shape parameters are supposed to slowly evolve within a 1D set of mixtures (this method is not applicable to 2D sets of mixtures) by setting Markovian priors. Moreover, this method can deal with varying number of sources in the mixtures by using the RJMCMC algorithm [Gre95].

The methods are evaluated on $I = 40$ synthetic mixtures of $N = 200$ samples, $J = 4$ sources, each with a constant shape parameter. The SNR defined as:

$$\text{SNR} = 10\log_{10}\frac{\sum_i \left\| \sum_{j=1}^{J} a_{ij}^* \boldsymbol{s}\left[c_{ij}^*; w_j^*\right] \right\|_2^2}{N \cdot I \cdot \sigma_n^2}, \tag{6.3}$$

is set to 15 dB (noise variance σ_n^2 equals to 0.12). The sources are Gaussian, *i.e.*, $s(\lambda; w_j) = \exp(-\lambda^2/2w_j^2)$ where the shape parameters w_j refers to the Gaussian widths. The amplitudes and delays of each Gaussian source are continuously generated from polynomials of degree 2, 3 or 4. Besides, the shape parameters are set to $\boldsymbol{w}^* = \begin{bmatrix}4 & 4 & 2 & 6\end{bmatrix}^T$. The mixtures are displayed

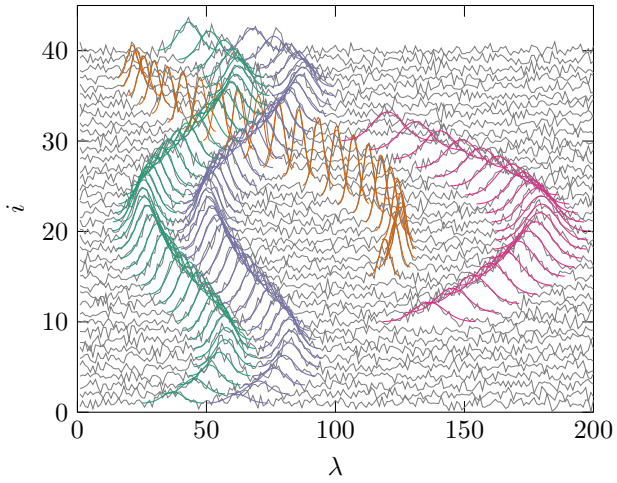


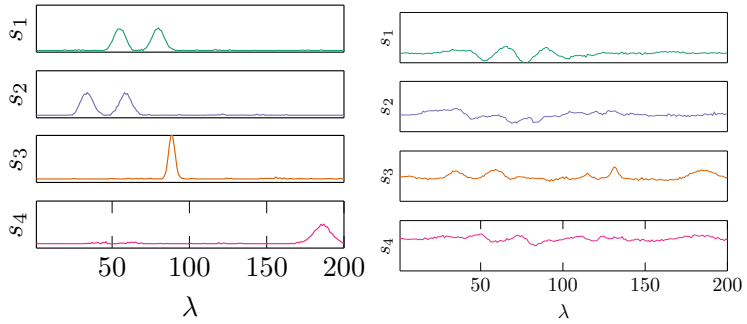
Figure 6.1: Synthetic data used for comparison: 40 mixtures with 4 sources and SNR = 15 dB. Each source is represented with a different color.

in Figure 6.1 and show noticeable behavior: two sources (around $\lambda = 50$) are very close and highly correlated (Gaussians with same shape parameter, equals to 4); there are several overlaps of the sources; and the number of sources per mixture is varying (*e.g.*, two sources for mixtures $i < 10$ and four sources for mixtures ≥ 10).

6.2.1 Evaluation of non-parameterized methods

We first discuss the results obtained with the non-parameterized methods [MMH07b; Nio+]. Figure 6.2 displays the estimated source signals $\boldsymbol{s}_j \in \mathbb{R}^N$ and Figure 6.3 shows the mixture reconstruction and the estimated amplitudes and delays.

First, the method of [MMH07b] yields two sources ($\boldsymbol{s}_3$ and $\boldsymbol{s}_4$) with one Gaussian shaped peak and two sources ($\boldsymbol{s}_1$ and $\boldsymbol{s}_2$) with two Gaussian-shaped peaks (Figure 6.2a). The sources $\boldsymbol{s}_3$ and $\boldsymbol{s}_4$ are more adequate with the ground-truth source which are Gaussian functions, while the sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ do not correspond to any ground truth source. This result can be expected since this method does not impose the source shapes or any constraints such as the unimodality of the sources. Yet, the mixture reconstruction corresponds to a denoised version of the mixtures. For the estimated amplitudes and delays shown in Figure 6.3, multiple observations can be made. On the one hand, the amplitudes and delays of $\boldsymbol{s}_3$ and $\boldsymbol{s}_4$ are close to the ground-truth in the mixtures where the sources are present and have low amplitudes and arbitrary delays in the other mixtures. On the other hand, the delays and amplitudes of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are affected by the non-accurate source signal estimates. Their amplitudes alternate in values; when the amplitude of $\boldsymbol{s}_1$ is close to the ground-truth, the amplitude of $\boldsymbol{s}_2$ is low and vice-versa. This indicates that these two sources are redundant, which is validated when observing their delay estimates which are almost identical and are situated in between the ground truth delays of the two sources with identical shapes. The MSE, computation time and amplitude and delay errors of this method are presented in Table 6.1 which numerically compares the methods based on the criteria defined in Section 6.1. These results show that the method of [MMH07b] can yield satisfactory parameter estimates only for unimodal



(a) Method of [MMH07b] (b) Method of [Nio+]

Figure 6.2: The estimated source s_j with non-parameterized delayed source separation methods applied to the mixtures in Figure 6.1.

sources. However, unimodal source estimates cannot be guaranteed and, in practice, multimodal sources are often obtained with this method.

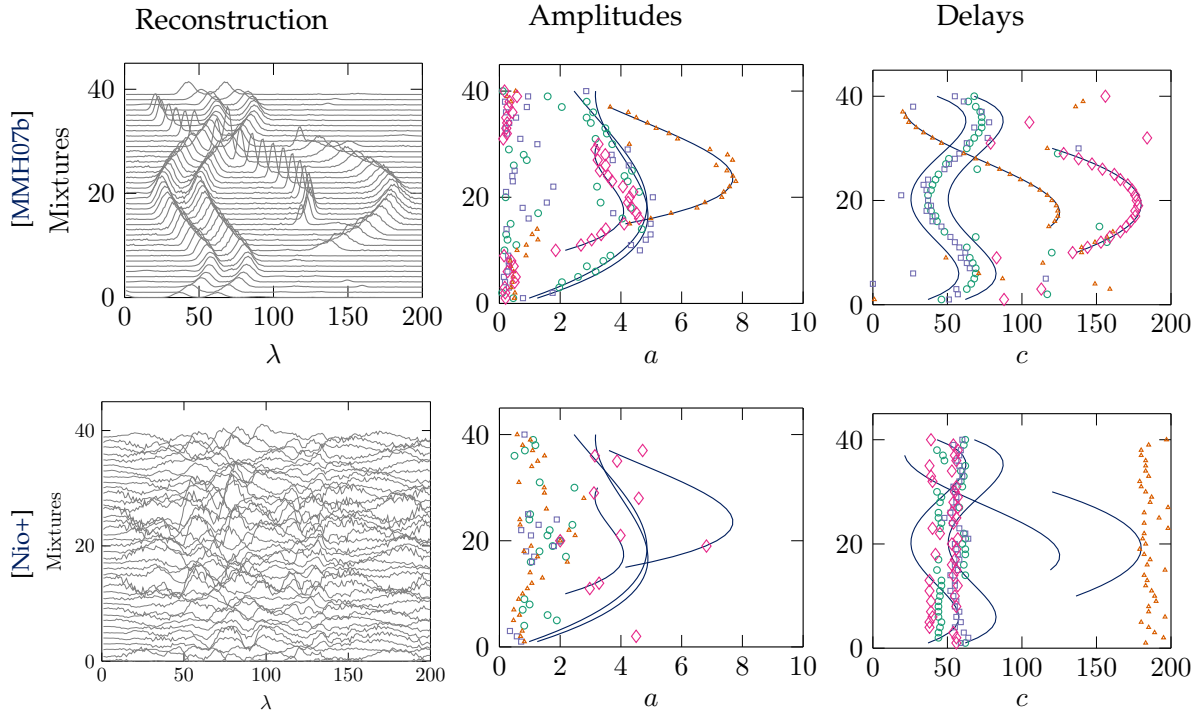


Figure 6.3: The mixture reconstruction and the estimated amplitudes and delays with the method of [MMH07b] (first row) and the method of [Nio+] (second row) applied to the mixtures in Figure 6.1. The circle, square, triangle and diamond markers respectively represent s_1, s_2, s_3 and s_4.

Second, from the source estimates of the method of [Nio+] displayed in Figure 6.2b, it is obvious that this method was not able to produce peak-shaped source signals, but rather oscillating signals with numerous peaks. This may be due to the source statistical independence assumption made in [Nio+] which is not valid for highly correlated sources such as in this example. In

	non-regularized	regularized	B-splines	[Maz+15]	[MMH07b]
Time (s)	3.9	**1.0**	7.5	41.5	9.8
MSE	**0.123**	0.126	0.25	0.124	0.15
Shape error	$3.9 \cdot 10^{-4}$	$\mathbf{3.5 \cdot 10^{-4}}$	$4 \cdot 10^{-3}$	0.2	n/a
Delay error	$3.2 \cdot 10^{-2}$	$5.6 \cdot 10^{-5}$	$\mathbf{2.7 \cdot 10^{-5}}$	$3.6 \cdot 10^{-5}$	$1.3 \cdot 10^{-1}$
Amplitude error	$3.9 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$4.1 \cdot 10^{-2}$	$\mathbf{6.8 \cdot 10^{-3}}$	$2.1 \cdot 10^{-1}$

Table 6.1: Numerical performance of the compared methods on the synthetic mixtures plotted in Figure 6.4. The amplitude, delay and shape errors are defined as in (6.2) and are only calculated in the mixtures where ground truth sources are present.

consequence, the mixture reconstruction is not satisfactory and the amplitude and delay estimates do not correspond to the ground-truth as shown in Figure 6.3. The method of [Nio+] clearly does not yield good results for the family of mixtures that we consider in this thesis.

6.2.2 Comparison with a parameterized method

We now compare the non-regularized, regularized and B-spline methods respectively presented in Chapters 3, 4 and 5 with the method of [Maz+15] that explicitly exploits the knowledge of parametrized source shapes. Again, the methods are tested with the mixtures displayed in Figure 6.1. The four methods model the sources by Gaussian signals. The ALS stopping constant for the non-regularized and regularized methods is $\rho = 10^{-6}$ and the delay sampling step is $\Delta = 0.2$. The non-regularized and regularized methods are able to deal with varying number of sources per mixture, to do so we set the threshold ς appearing in Algorithm 2 to $1.2\sigma_n^2$ and the threshold κ relative to the regularized method (see Appendix A) to $1.5\sigma_n$. The regularization parameter and the ICM stopping constant of the regularized method are respectively set to $\tau = 8 \cdot 10^{-2}$ and $\xi = 10^{-6}$. For the B-spline method, the amplitudes and delays are modeled using cubic B-splines with a knot vector $\boldsymbol{k} = \begin{bmatrix} 1 & 1 & 1 & 1 & 14 & 27 & 40 & 40 & 40 \end{bmatrix}$, while the shape parameters are modeled using B-splines with degree 0 and two knots $\boldsymbol{k} = \begin{bmatrix} 1 & 40 \end{bmatrix}$ in order to force the shape parameters to be constant for each source. The optimization of shape control points is constrained with a lower bound $w_{\min} = 0.5$ and an upper bound $w_{\max} = 8$. Note that the B-spline modeling reduces the number of unknowns from 480 (amplitudes, delays and shapes) to 72 control points.

The mixture reconstruction and the estimated parameters are displayed in Figure 6.4. Besides, the methods are numerically compared in Table 6.1. One can remark that the reconstruction is equally good for the non-regularized, regularized and B-spline methods. The B-spline method yields lower reconstruction quality since it is not designed to deal with a varying number of sources.

The amplitude estimates of the non-regularized and regularized methods and the method of [Maz+15] are adequate. However, the B-spline method is affected by the varying number of sources; the amplitudes gradually attenuate when a source becomes absent and thus the amplitudes are shifted towards estimates lower than the ground-truth values.

The delay estimates of the non-regularized method are not satisfactory for the two sources with the similar shape as the switch percentage between these sources is 45 % (see Section 3.5.4), this is because when sources have the same shape parameter, the identification becomes ambiguous. On the contrary, the regularized and B-spline methods and the method of [Maz+15], yield improved delay estimates due to the slow delay evolution consideration. Note that the delay evolution keeps

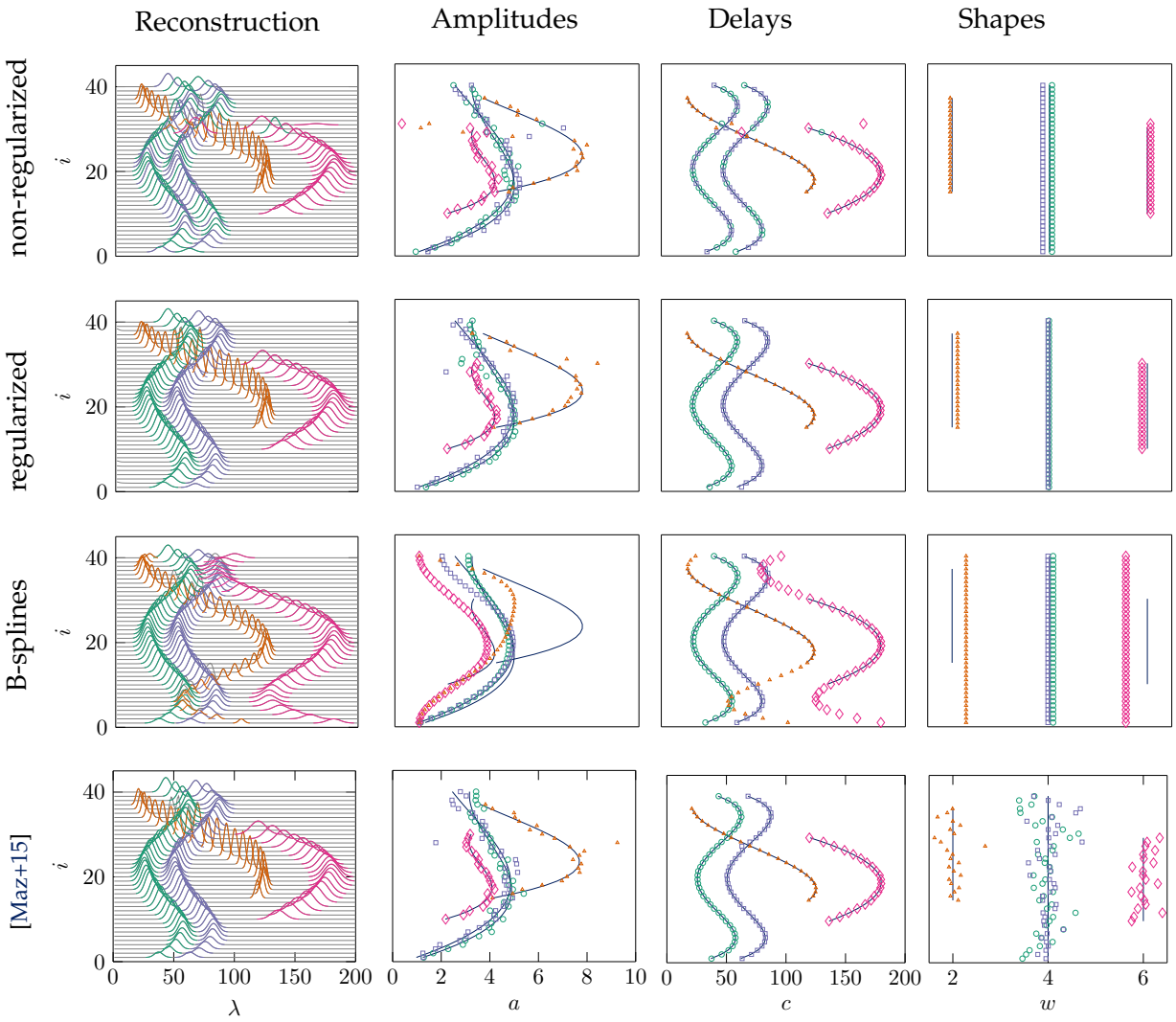


Figure 6.4: Results on the synthetic data of Figure 6.1. The first column displays the reconstructed mixtures as well as the estimated sources. The second, third and fourth columns display the estimated amplitudes, delays and shapes. Each source is represented with unique color and marker. The ground truth delays and amplitudes are plotted in dark blue lines.

being smooth in the mixtures where sources are absent for the B-spline method. These results are numerically validated in Table 6.1.

The estimated shape parameters of the non-regularized and regularized methods are very close to the ground-truth, they are respectively given by: $\widehat{\boldsymbol{w}} = \begin{bmatrix}4 & 3.8 & 1.9 & 6\end{bmatrix}^T$ and $\widehat{\boldsymbol{w}} = \begin{bmatrix}3.9 & 3.9 & 2.1 & 5.9\end{bmatrix}^T$. The B-spline method yields less accurate shape estimation for the two sources that do not appear in all the mixtures $\widehat{\boldsymbol{w}} = \begin{bmatrix}4 & 3.9 & 2.28 & 5.5\end{bmatrix}^T$. For the method of [Maz+15], the shape parameters can vary within a source, but the variations remain small around the ground-truth.

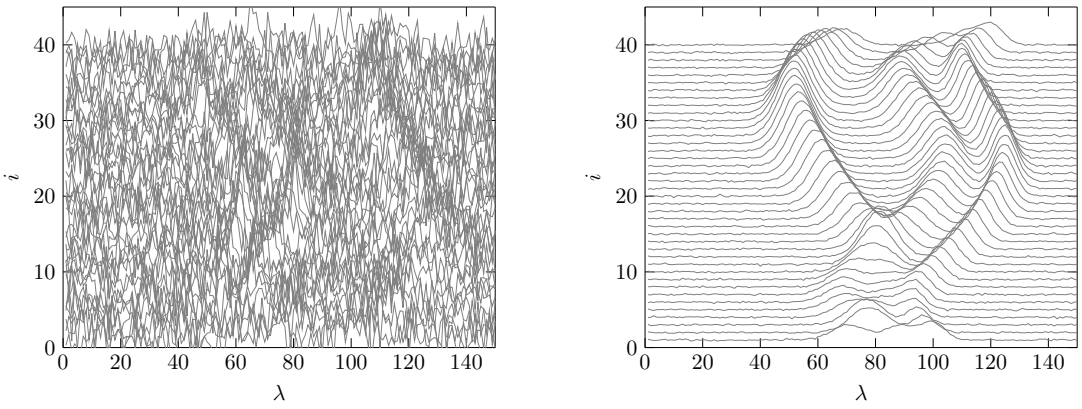
These simulations show that including the parameterized source knowledge is of interest when dealing with highly correlated sources and allow to overcome difficulties encountered when using non-parameterized methods as [MMH07b; Nio+]. The proposed regularization of Chapter 4 and the B-spline parameter modeling of Chapter 5 ensure to separate correctly the sources with very similar shape parameters. The proposed methods give accurate results within a very low computation time, especially the regularized method is the fastest. On the contrary, the method of [Maz+15], while very effective, is 12 to 40 times slower than the proposed non-regularized and regularized methods and 5 times slower than the B-spline method. Note, however, that the latter model is more versatile since the shape parameter are allowed to vary within a source and additional priors on the shapes and amplitudes are considered.

6.3 Influence of the SNR

We perform statistical simulations to compare the three proposed methods and the method of [Maz+15] with respect to the SNR. The methods of [Nio+] is not tested since it does not yield accurate outputs. Also, the method of [MMH07b] is not tested since it is not straightforward to obtain the delays and amplitudes from the algorithm output especially when the sources are not unimodal.

The simulations are set for SNR values varying between 0 and 30 dB with a step of 2 dB. For each SNR, 1000 1D sets of $I = 30$ mixtures with Gaussian sources are generated, each $J = 3$ sources that appear in all the mixtures and $N = 200$ samples. The shape parameters are chosen randomly between 0.5 and 5; the delays and amplitudes are generated by polynomial functions with random coefficients and degrees varying between 2 and 4. Figure 6.5 displays two examples of mixtures with the lowest and the highest considered SNR's (0 and 30 dB). For the non-regularized and regularized methods, the delay sampling step is equal to $\Delta = 0.2$ and the ALS stopping constant is set to $\rho = 10^{-6}$. The thresholds ς and κ are set to zero so that the sources are present in all the mixtures. The ICM stopping constant and the regularization parameter of the regularized method are respectively set to $\xi = 10^{-4}$ and $\tau = 0.1$. For the B-spline method, cubic B-splines with knot vector $\boldsymbol{k} = \begin{bmatrix}1 & 1 & 1 & 1 & 14 & 27 & 40 & 40 & 40 & 40\end{bmatrix}$ are used to model the amplitudes, delays, and B-splines with zero degree and knot vector $\boldsymbol{k} = \begin{bmatrix}1 & 40\end{bmatrix}$ are used to model the constant shapes of each source. The methods are compared in terms of the MSE, the parameter errors and the computation time. The results are shown in Figure 6.6 where for each SNR, the average of the results obtained for the 1000 generated datasets is plotted.

As expected, the MSE decreases as the SNR increases for all the methods, and it should be



(a) SNR = 0 dB (b) SNR = 30 dB

Figure 6.5: Two examples of mixtures with the two extreme SNR values used in the simulations of Section 6.3.

noticed that the variations between the methods are very similar. Indeed, the non-regularized method yields the best MSE since it minimizes only the data-fit criterion (equivalent to the MSE) without any additional regularizations.

For the amplitude error, the B-spline method slightly over-perform the competitors which behave almost similarly. The delay error for the non-regularized method decreases as the SNR increases but remains higher than its competitors. On the contrary, the regularized and B-spline methods and the method [Maz+15] that impose the slow delay evolution give substantially improved estimation of the delays and the estimates are more robust to the noise variation. The shape error of the method of [Maz+15] is higher than its competitors; this is because this method estimates varying shape parameters while the ground truth ones are constant for each source. The three proposed methods give equally good shape estimation quality which is also robust to the SNR. The amplitude, delay and shape errors suggests that the B-spline method is more effective than its competitors. However, this method is favored, in these simulations, since the ground-truth parameters are generated using polynomial functions that can be perfectly fitted by B-splines.

Again, it is shown that the method of [Maz+15] is much slower than the proposed methods (approximately 14 times slower than the non-regularized and regularized methods and 6 times slower than the B-spline method). Therefore, the regularized and B-spline methods are competitive with the method of [Maz+15] with much lower computation time.

6.4 Real data

In this section, we present results obtained with the proposed methods on real data. Two sets of real data are considered: the first is a 1D set of time-resolved photoelectron spectra and the second consists of 2D sets of spectra representing a galaxy multispectral image.

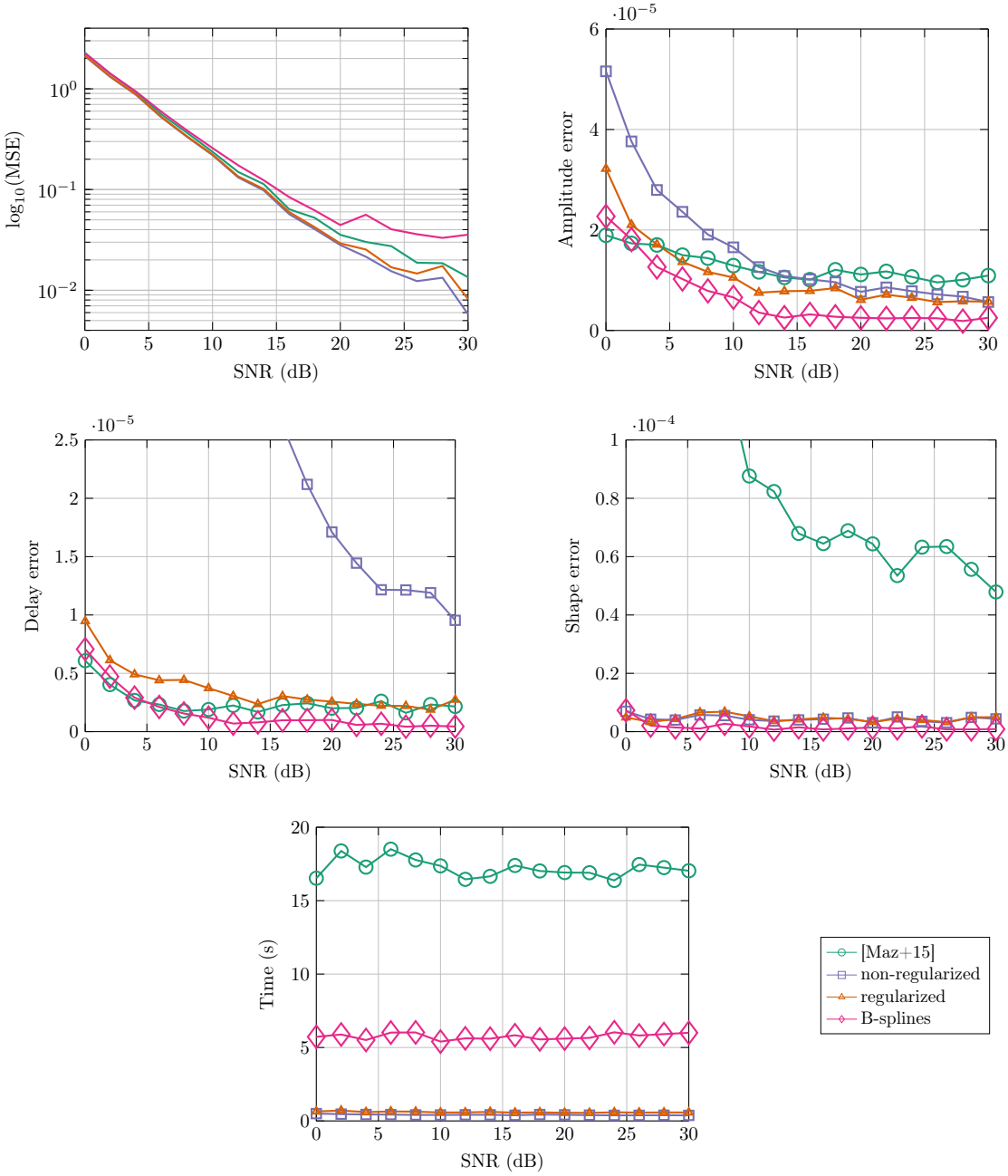


Figure 6.6: Comparison of the three proposed methods and the method of [Maz+15] with respect to the SNR. A zoom is made on the delay and shape error results, so some points with high values are not shown for the non-regularized method (delay error) and the method of [Maz+15] (shape error).

6.4.1 Time-resolved photoelectron spectra

The non-regularized, regularized and B-spline methods and the method of [Maz+15] were applied to the sequence of photoelectron mixtures (the experiment studies the relaxation of an atom of barium [Mas+10; Maz+15]) shown in Figure 1.4. The sequence gathers $I = 44$ mixtures (covering a duration of 3.47 ps), each of $N = 181$ samples (from 0.02 eV to 2.52 eV). The sampling in both energy (λ) and time (i) is irregular, but we suppose the sampling to be regular for simplicity and because it is of less influence.

For the non-regularized and regularized methods, the delay sampling step is set to $\Delta = 5 \cdot 10^{-4}$ and the ALS stopping constant is set to $\rho = 10^{-6}$. All the sources are supposed to appear in all the mixtures, therefore the thresholds ς and κ are set to zero. For the regularized method, we set the regularization term to $\tau = 100$ and the ICM stopping constant is set to $\xi = 10^{-6}$. For the B-spline method, cubic B-splines with knot vector $\boldsymbol{k} = \begin{bmatrix} 1 & 1 & 1 & 1 & 15 & 29 & 44 & 44 & 44 & 44 \end{bmatrix}$ is used to model the amplitudes, delays and shapes. The optimization of the shape control points is constrained with a lower bound $w_{\min} = 0.01$ and an upper bound $w_{\max} = 0.1$. The use of B-splines allows to reduce the number of parameters from 729 to 108.

Besides, the photoelectron mixtures include a background that must be estimated and removed. The background in each mixture i can be modeled as an exponential of the form $h_i \exp(-\lambda/q)$. Therefore, the additional parameters q and h_i (gathered in a vector $\boldsymbol{h} \in \mathbb{R}^I$) must be estimated in addition to the amplitudes, delays and shapes. To do so, an additional step is added to the ALS scheme in Algorithms 1 and 3 (respectively representing the algorithms of Chapters 3 and 4) to estimate the new unknowns by using the Levenberg-Marquardt algorithm. Let $\boldsymbol{e}(q) \triangleq \begin{bmatrix} \exp(-1/q) & \exp(-2/q) & \dots & \exp(-N/q) \end{bmatrix}^T$ be a vector gathering the exponential function samples for $\lambda \in \{1, \dots, N\}$, then the new optimization problem writes as:

$$\left(\widehat{q}, \widehat{\boldsymbol{h}}\right) \leftarrow \underset{q,\boldsymbol{h}}{\operatorname{argmin}} \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} \widehat{a}_{ij} \boldsymbol{s}\left[\widehat{\ell}_{ij}\Delta; \widehat{w}_j\right] - h_i \boldsymbol{e}(q) \right\|_2^2 . \tag{6.4}$$

This optimization problem takes place in between the lines 6 and 7 of Algorithm 1 and in between the lines 4 and 5 of Algorithm 3. Indeed, the amplitudes, delays and shapes are estimated in the ALS scheme by considering $\widehat{q}$ and $\widehat{\boldsymbol{h}}$ to subtract the background form the mixtures. The estimated backgrounds using the non-regularized and regularized methods and the method of [Maz+15] are presented in Figure 6.8. These results show that these methods estimate correctly the background and that they can be easily extended to consider this additional application requirement. Moreover, one can note that the results obtained by the proposed methods are smoother than the one obtained with the method of [Maz+15]. The B-spline method is applied to data from which the background estimated by the regularized method is removed in a pre-processing.

The reconstruction and the estimated parameters of the compared methods are presented in Figure 6.9. The reconstruction quality of the four compared methods is equivalently good, which can be visually confirmed. Let us now compare the parameter estimates. The delay seems to be the most significant parameter in this comparison:

- the non-regularized method yields non-satisfactory delay estimation. Only one source (around 0.5 eV) presents a slow evolution, whereas the other five sources present multiple discontinuities in their evolutions;

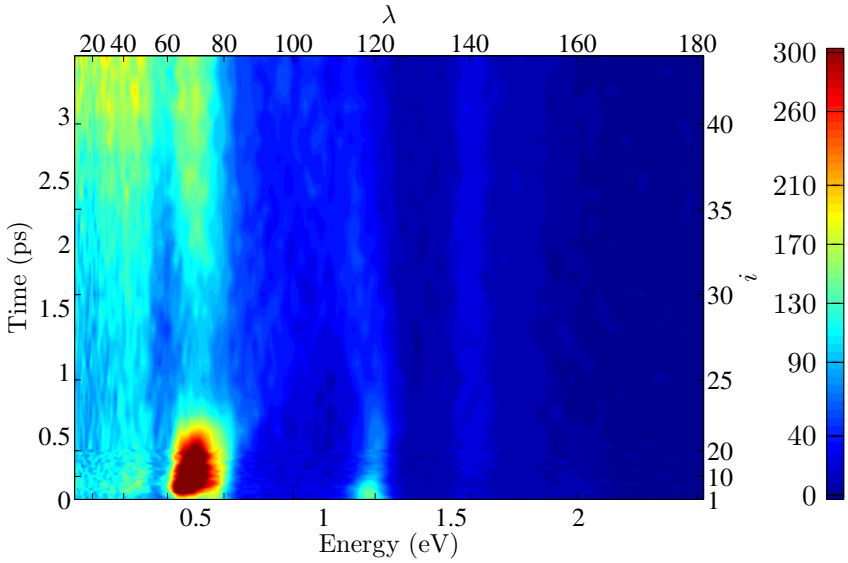


Figure 6.7: Time-resolved photoelectron spectra ($I = 44$ mixtures) represented as an intensity map of dimension $I \times N$.

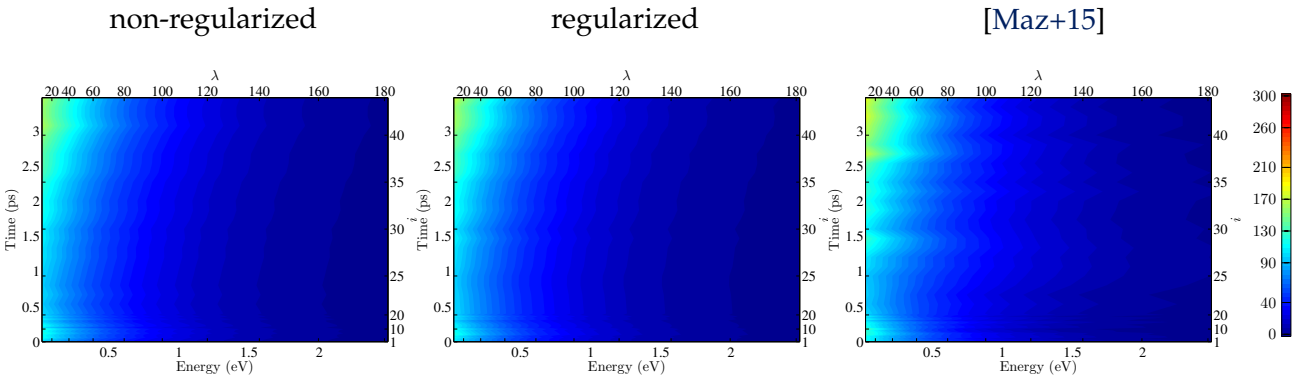


Figure 6.8: The estimated background using the methods of Chapters 3 and 4 and the method of [Maz+15].

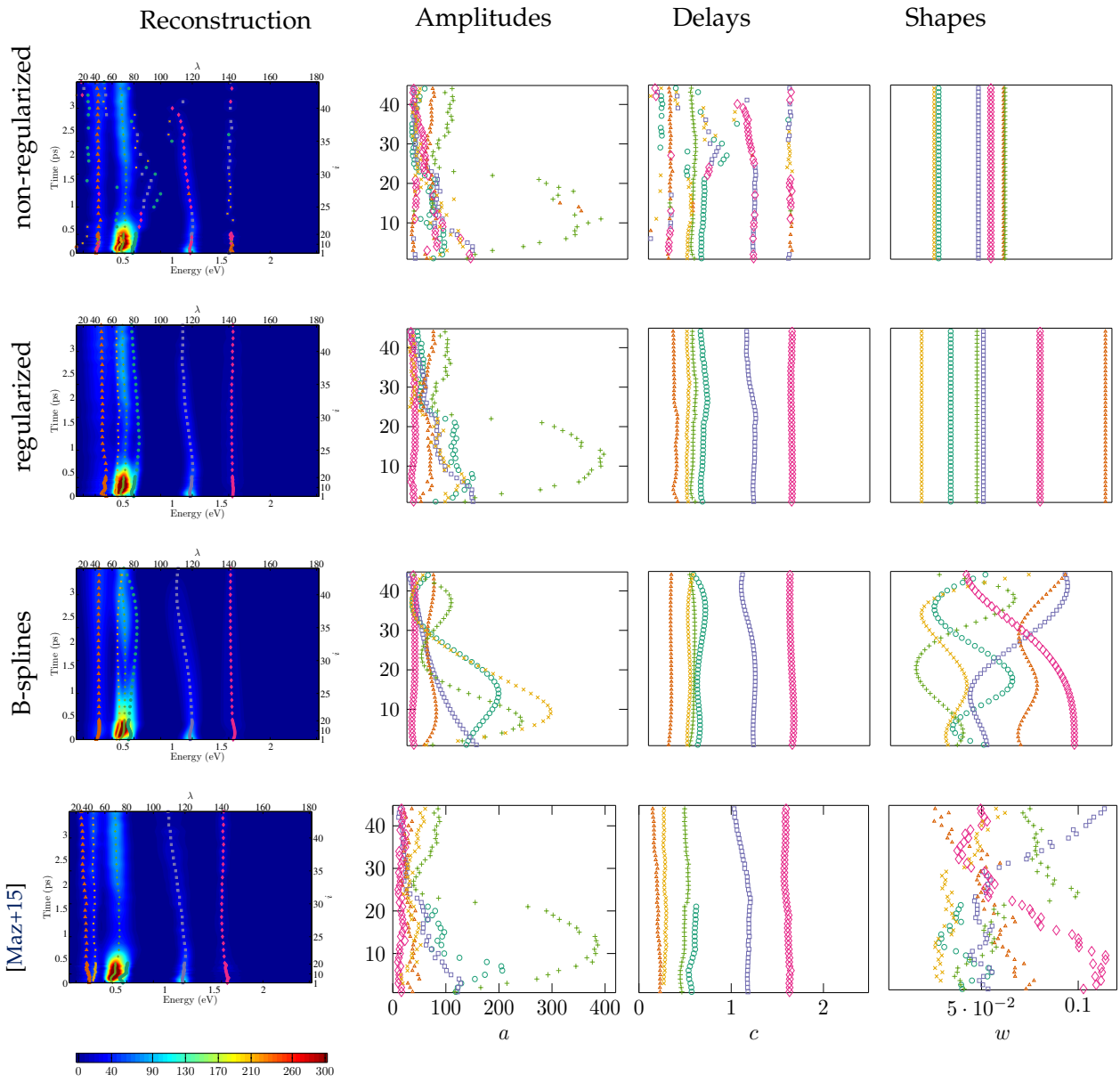


Figure 6.9: First column: mixture reconstruction superimposed by the estimated delays using the regularized, non-regularized and B-spline methods and the method of [Maz+15]. The second, third and fourth columns respectively represent the estimated amplitudes, delays and shapes.

- the regularized method yields a smoother and better delay evolution than the non-regularized method due to the delay regularization. Thus, the practitioner is capable to track the temporal evolution of the peaks issuing from the same molecule energy levels;
- the B-spline method gives similar results as the regularized method, except for the three sources in the main band which are tighter;
- the method of [Maz+15] yields similar delay evolution for five sources out of six when compared to the regularized and B-spline methods (two sources represent the band around 0.2 eV instead of one in the proposed methods). The regularized and B-spline methods rather improve the shape of the main band (around 0.5 eV) by addition of nearby new sources than focusing on the low-intensity energy ones as seen for method [Maz+15]. Such behavior looks reasonable according to the shape of the decay simulated [Mas+14].

Next, regarding the amplitude estimates, one can remark that the behavior of the non-regularized and regularized methods and the method of [Maz+15] is roughly similar. Yet, the B-spline method yields remarkably smoother evolution and different solutions than the competitors for the three sources representing the main band around 0.5 eV. This may suggest that the amplitude B-spline modeling is not required in this application. For the shape parameter estimation, the regularized and non-regularized methods yield constant shapes for each source. The B-spline method and the method of [Maz+15] allow for shape varying and give comparable solutions. Again the B-spline method gives a smoother solution.

Note that the method of [Maz+15] is able to deal with a varying number of sources through the mixtures. This was not possible to consider in the proposed methods since setting a threshold would eliminate some sources with very small amplitudes (like the source around 1.6 eV). A solution would be to consider different thresholds for each source.

The regularized and B-spline methods are respectively about 127 and 27 times faster than the method of [Maz+15] (respectively 33 seconds, 156 seconds and 70 minutes). Note that for the B-spline method, the time for estimating the background has not been considered.

In conclusion, the slow parameter evolution, especially for delays, seems to be crucial to obtain good solutions and the proposed methods appear to be both effective and efficient.

6.4.2 Galaxy NGC-4254

The proposed method is applied to a real observation in the radio band (around the HI peak at 21 cm) of the galaxy NGC-4254 [PVM93]. The multispectral image is of dimension $140 \times 140 \times 42$ ($I = 140^2$ mixtures) and its white image (the average multispectral images following the wavelength dimension) is displayed in Figure 6.10. It is worth mentioning that a synthetic kinematic structure has been manually added to the image by an astronomer. This structure occupies a circular small part in the top-left of the galaxy. Its amplitude is spatially Gaussian, while the delays and shapes are constant within this structure. Adding this synthetic structure would allow us to test the performance of the proposed methods.

The method of [Maz+15] is not applied to this galaxy multispectral image since it is not adapted to deal with 2D sets of mixtures. For the non-regularized and regularized methods, the number of sources is set to $J = 3$, the delay sampling step is set to $\Delta = 0.2$ and the ALS stopping constant

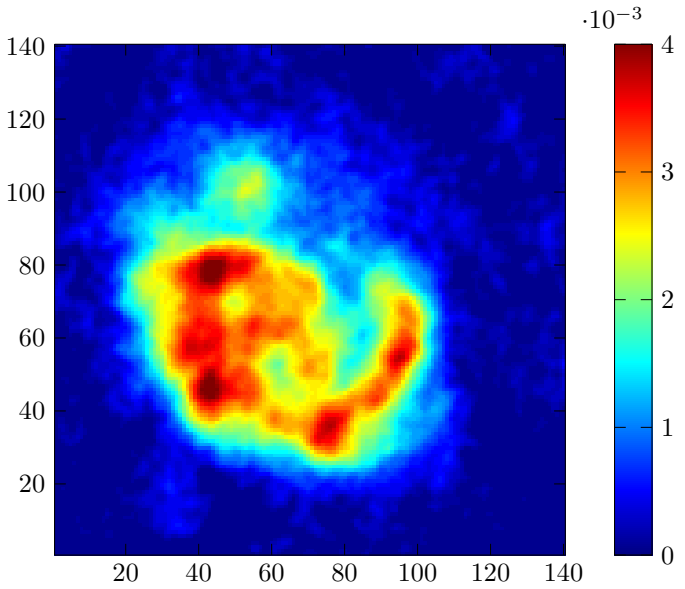


Figure 6.10: The white image of the galaxy NGC-4254. The center synthetic structure appears on the spatial position (50,100).

is set to $\rho = 10^{-6}$. Moreover, we set the threshold $\varsigma = \sigma_n^2$ for the non-regularized method, while we use the algorithm extension presented in Appendix A to the regularized method by setting the threshold $\kappa = 2\sigma_n$ (the NGC-4254 noise variance has been estimated in Section 1.3.4). For the regularized method, we set the regularization term to $\tau = 2 \cdot 10^{-4}$ and the ICM stopping constant is set to $\xi = 10^{-6}$. For the B-spline method, the number of sources is set to $J = 2$ and cubic surface B-splines with the following knots are used to model the delays and shapes: $\boldsymbol{k}_u = \boldsymbol{k}_v = \begin{bmatrix}1 & 1 & 1 & 1 & 24 & 47 & 70 & 93 & 116 & 140 & 140 & 140 & 140\end{bmatrix}$. The optimization of shape control points is constrained with a lower bound $w_{\min} = 0.5$ and an upper bound $w_{\max} = 6$. The amplitudes are modeled with more knots so as to fit the complex behavior of the intensities in the multispectral image: $\boldsymbol{k}_u = \boldsymbol{k}_v = \begin{bmatrix}1 & 1 & 1 & 1 & 12 & 33 & 44 & 54 & 65 & 76 & 87 & 97 & 108 & 119 & 129 & 140 & 140 & 140 & 140\end{bmatrix}$. Overall, the number of unknowns is thus reduced from 39200 to only 706.

Figure 6.11 displays the white images of the reconstruction and the residual error obtained by the proposed methods. The reconstruction of the non-regularized method is better than the other two methods. This is confirmed when observing the residual maps, also this is expected since this method minimizes the residual error without considering the slow evolution of parameters. In addition, the reconstruction of the B-spline appears to be "blurred"; this is due to the smooth amplitude, delay and shape estimates.

First, the good reconstruction quality of the non-regularized method does not imply good source separation quality as confirmed in Figure 6.12 that displays the estimated parameters. In fact, the three estimated sources do not have any physical meaning as they do not represent the kinematic structures of the galaxy. The amplitudes of the first source present some regions with high intensities while the amplitudes of the third source are low everywhere. Moreover, the delay estimate of the three sources does not show a smooth evolution. The computation time of this method is 150 s.

Second, despite a lower reconstruction quality of the regularized method compared to the

non-regularized method, the estimated source parameters of the method of Chapter 4 (Figure 6.13) represent well the galaxy kinematics of NGC-4254. Specifically, the first estimated source represents the main galaxy kinematic structure: two spiral arms rotating around the center as shown in the amplitude map. The delay maps of this source clearly show that the gas is rotating within the galaxy as the lower part is mostly delayed towards red regions and the upper part is mostly delayed towards blue regions of the electromagnetic spectrum. Thus the upper part is moving slower away from the Earth than the lower part. The second estimated source represents a very attenuated source as shown in the amplitude map. Also, it has a constant delay in the red region; this structure moves away from the earth in a constant and relatively high speed. Note that, to detect this very attenuated source, its starting mixture in the ICM sweeping i_0 (see Section 4.3.1) has been manually selected. The third estimated source represents the added synthetic kinematic structure. The amplitude map has a Gaussian form, the delays and shapes are constant and the estimates correspond to the ground-truth of the synthetic kinematic structure. These results show that the regularized method is capable to estimate the galaxy kinematics of real data from a high number of mixtures in a very fast computation time (271 s). Besides, the threshold κ introduced in Appendix A allows us to obtain sources that only appear in a connected subset of mixtures (pixels). However, this method estimates constant shapes for each source and thus the velocity dispersion field information that can be extracted from the shape variation cannot be obtained.

Third, we evaluate the source separation of the B-spline method. In Figure 6.14 we represent the estimated parameters after applying a threshold on the amplitudes (the same threshold is then applied to the delay and shape maps) so that we only display the sources in the mixtures with relatively high amplitudes. Indeed, the first estimated source represents the main kinematic structure of the galaxy and can be compared to the first source of the regularized method. Again, the delay map allows us to determine the kinematic structure rotation velocity but also the shape parameter variation that yields additional information for the astronomer which is the velocity dispersion field. The estimated parameters of the second source highlight the presence of three sources which can be visually distinguished. The source on the top is the added synthetic structure. The source on the left is very faint and difficult to find but the proposed method was able to find it automatically. This source can be compared with the second source obtained with the regularized method. Finally, the source on the right is mostly a part of the main kinematic source but it was estimated in a second source. This method allows retrieving the three main kinematic structures of the galaxy. However, a post-processing is required in order to apply a threshold and to distinguish between multiple sources which are estimated as a unique source. Moreover, this method is much slower than the non-regularized and regularized methods with almost 6 hours of computation time.

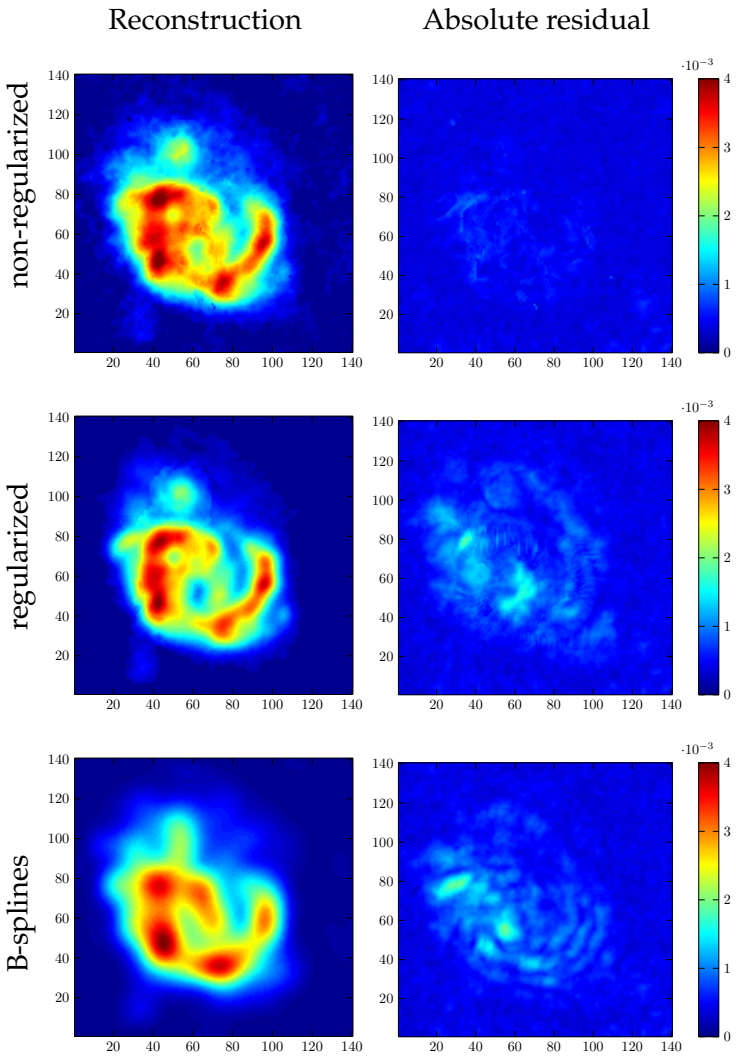


Figure 6.11: The white images of the reconstruction (first column) and the absolute residual (second column) obtained with the three proposed methods.

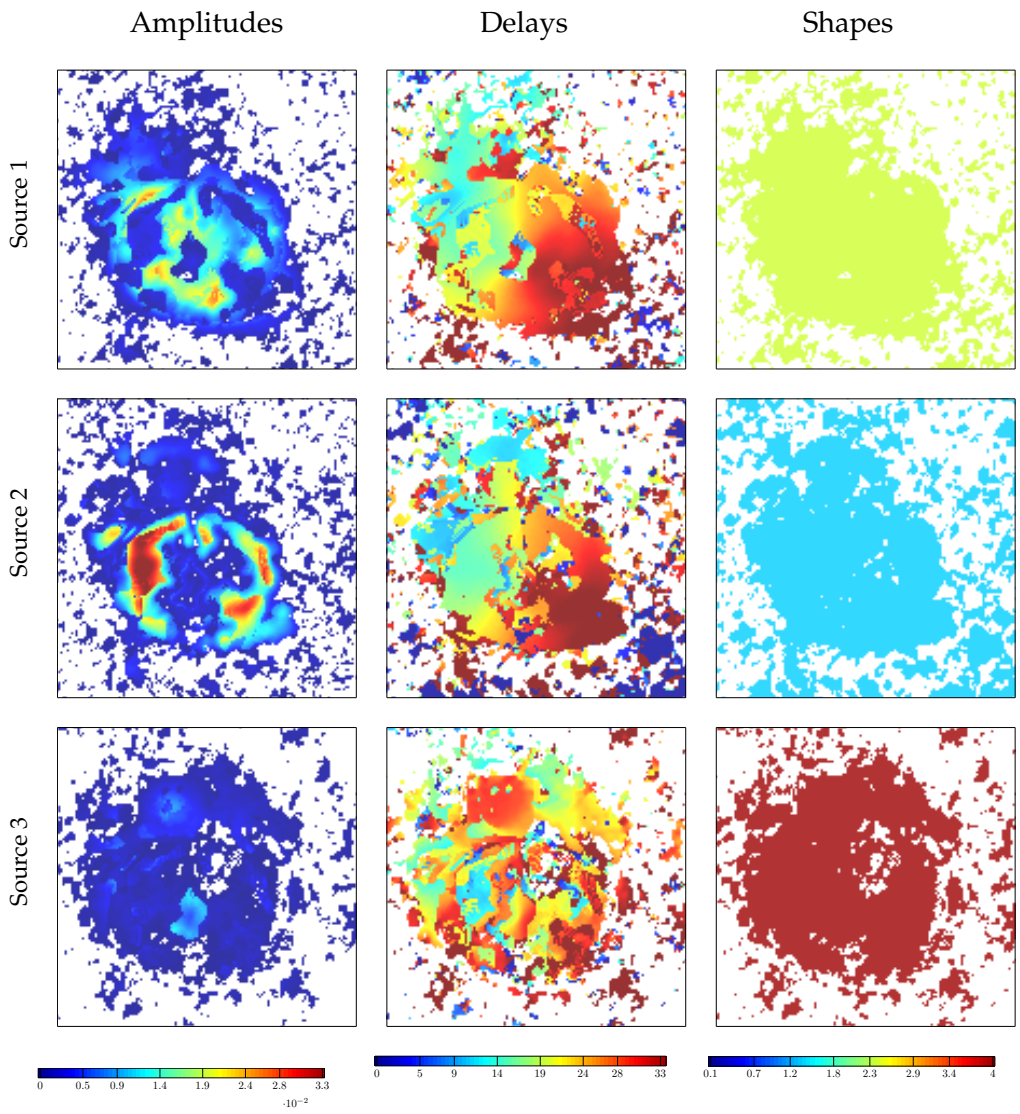


Figure 6.12: The estimated parameter maps of $J = 3$ sources using the non-regularized method of Chapter 3.

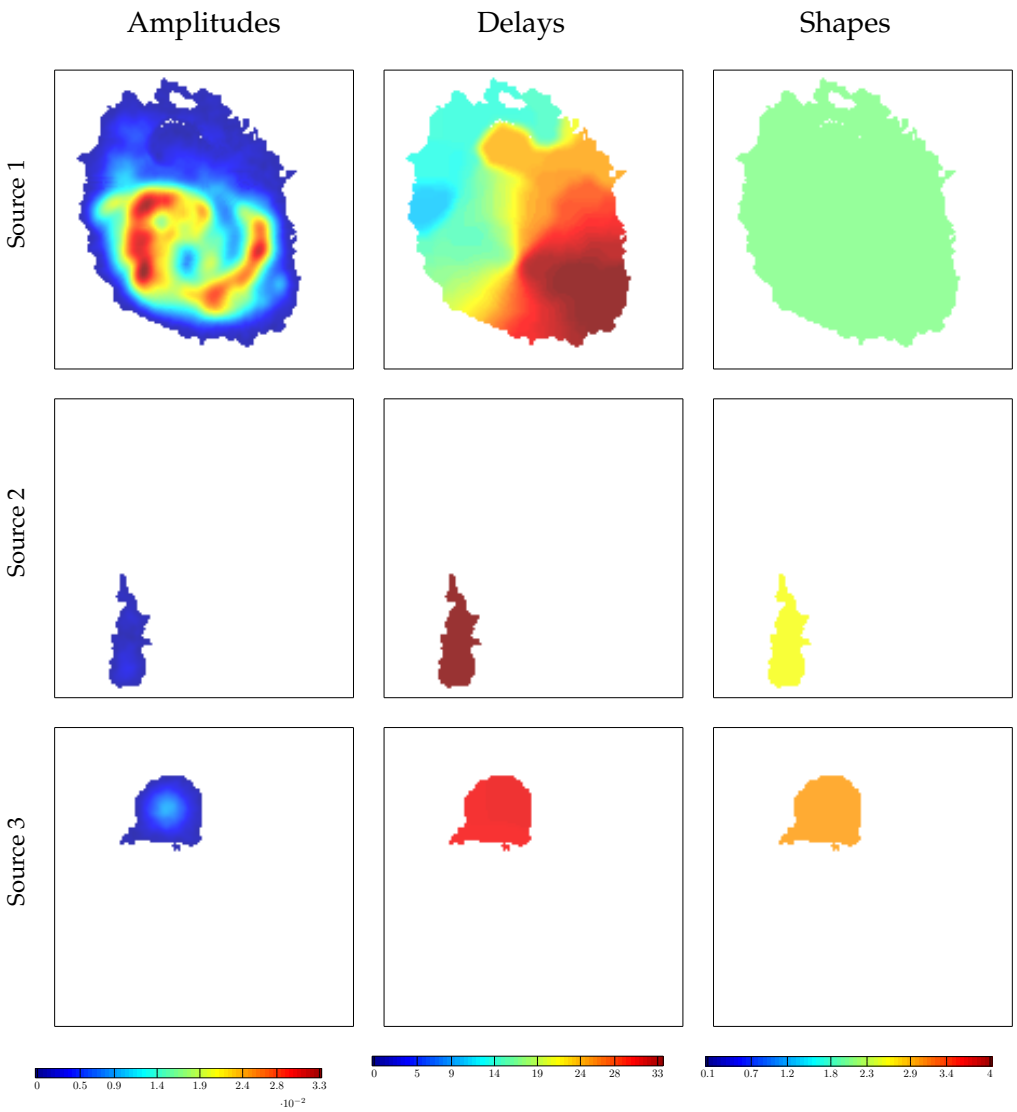


Figure 6.13: The estimated parameter maps of $J = 3$ sources using the regularized method of Chapter 4.

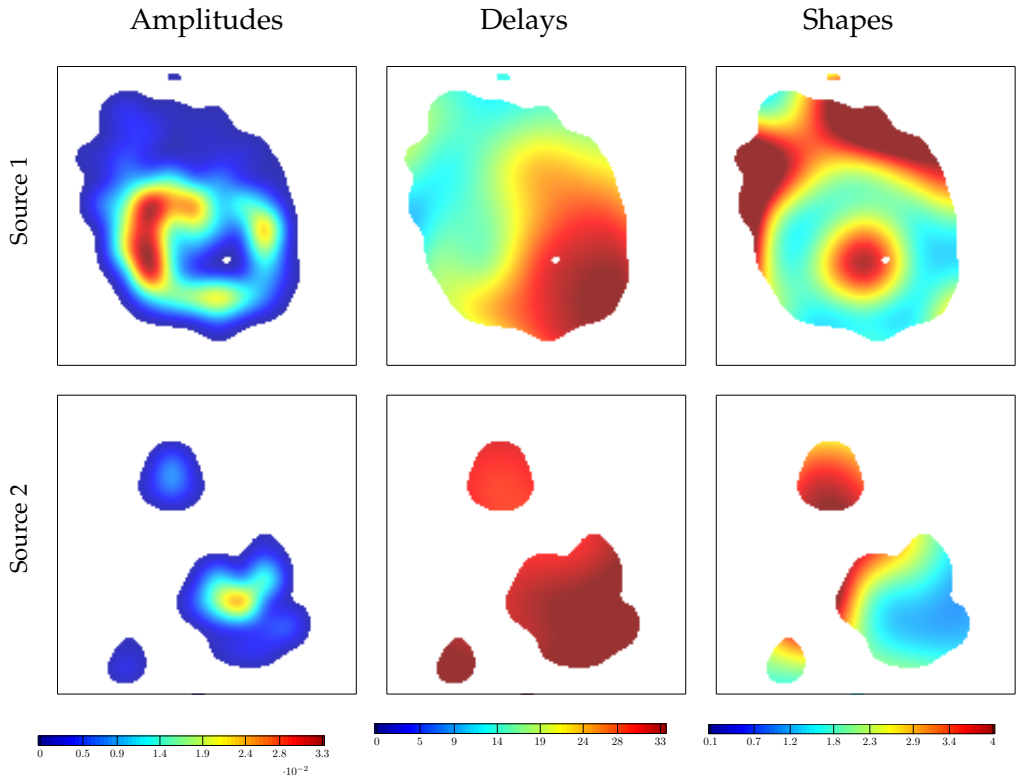


Figure 6.14: The estimated parameter maps of $J = 2$ sources using the B-spline method of Chapter 5.

6.5 Conclusion

This chapter presents numerical simulations on synthetic and real data. The three proposed method of Chapter 3 (non-regularized method), Chapter 4 (regularized method) and Chapter 5 (B-spline method) are compared together and with state-of-the-art methods. Namely, the state-of-the-art methods are two delayed source separation methods that deal with non-parameterized sources and a Bayesian method adapted for the decomposition of spectra and peak association. The results show that:

- The two non-parameterized methods that we test yield non-satisfactory results when compared to the parameterized methods. Thus, the use of parameterized sources, when it is possible, is beneficiary;
- the consideration of the slow parameter evolution, especially the delays, is crucial to obtain good results especially for highly correlated sources;
- the proposed methods are as effective as the Bayesian method but are much faster;
- the proposed methods can be applied to real-data (time-resolved photoelectron spectra and galaxy multispectral image) and yield physically justified results.

Conclusion

The work in this thesis aims at achieving two goals. First, estimate the peak characteristics (intensity, spectral position and width) in a temporal sequence of spectra or multispectral image. Second, associate the peaks of different spectra as they evolve slowly from one spectrum to its neighboring ones. Two applications motivate this work: time-resolved photoelectron spectroscopy and galaxy kinematics. We propose to address the two goals jointly in a delayed source separation framework where the mixtures and the delayed sources respectively refer to spectra and peaks with varying spectral positions. Although the sources may be highly correlated, the three proposed approaches are effective due to the source parameterization.

The first method (Chapter 3) addresses the delayed and parameterized source separation problem [Mor+17b; Mor+17a]. An ALS scheme is proposed to minimize a data-fit criterion. On one hand, the shape parameters are estimated using a non-linear least squares solver, namely the Levenberg-Marquardt algorithm. On second hand, the amplitudes and delays are estimated by a greedy sparse approximation algorithm. In addition, we show that choosing a sufficiently small delay sampling step yields equally good but faster results than using a continuous extension to the OMP-like algorithm. The contributions are:

- the use of a sparse strategy to jointly approximate multiple spectroscopic measurements;
- the sparse vectors are enforced to be structured so that they are 1-sparse per block. This ensures a single occurrence of the same source in the mixtures and thus associate the sources jointly with their parameter estimation;
- the varying number of sources within the mixtures is taken into account by the OMP-like algorithm.

The second method (Chapter 4) considers the slow evolution of the delays of each source within the mixtures by adding a regularization term to the data-fit criterion defined in Chapter 3 [Mor+19]. This implies a new sparse approximation strategy coupled with ICM to estimate the amplitudes and delays jointly in all the mixtures. The contributions are:

- a new regularization to favor delay slow evolution in spectroscopic data;
- the sparse approximation algorithm produces solutions with a slow-moving support that can be seen as an intermediate solution between the solutions given by independent simultaneous sparse approximation;
- the use of classical image processing algorithms (namely, ICM and region growing) within the sparse approximation algorithm.

The third method (Chapter 5) uses B-splines to model the slow evolution of the amplitudes, delays and shapes [Mor+18]. The number of unknowns is significantly reduced since the parameters are replaced by the B-spline control points. The problem is then addressed as the optimization of a data-fit term using an SQP algorithm. The contribution is the use of B-splines in the context of spectroscopic data to model the parameter evolution.

The proposed methods are validated on both synthetic and real data. They overcome the state-of-the-art method limitations since they can deal with highly correlated sources and are effective and efficient.

Perspectives

The OMP-like implementation algorithms in Chapters 3 and 4 are forward greedy algorithms. This means that each iteration of these algorithms is dedicated to make a local optimization so as to pick a new source while keeping the previously selected sources. This yields sub-optimal solutions which cannot be improved, even with more iterations. A first possible solution to overcome this limitation is to use forward-backward sparse approximation algorithms such as as [Sou+11; VB13; JTD11] where a selected source can be retired in favor of a better one. Another solution can be found in [CDP17] where a continuous optimization step is added at each iteration to update the selected sources.

In this thesis, the mixtures are analyzed in their natural domain. An alternative solution would be to analyze the mixtures in the frequency or time-frequency domains, as it is often proposed in delayed source separation [MMH07b; PD05]. This has the advantage to linearize the delays and thus yielding a simpler optimization problem. The resulting problem shares similarities with spectral line estimation problem [BCI07; DB13].

The knot vector in the B-spline modeling (Chapter 5) is a set of equidistant knots covering all the mixtures. This strategy is the simplest but does not consider sources that only appear in a subset of the mixtures. Also, additional knots are required between the mixtures in case a source has a complex evolution behavior. Therefore, a strategy where the knots positions are optimized with respect to the data would be interesting. In the literature, this is known as free or adaptive knots optimization [Jup78; DGK01].

Appendices

A

Varying source number extension for the method of Chapter 4

In the considered application, a source does not appear necessary in all mixtures. However, the amplitudes and delays in Chapter 4 are estimated in all the mixtures $i \in \{1, \dots, I\}$ with Algorithm 4 (Section 4.3). In this appendix, we present an extension so that this algorithm only yields estimates in a subset of mixtures. Moreover, we constraint this subset to be contiguous; all the mixtures are connected.

Let $\mathcal{I}_j$ be the set of mixture indices in which the source j appears. This implies that the amplitudes of source j in the mixtures that do not belong to $\mathcal{I}_j$ are null, *i.e.*,

$$\forall\, j, i \notin \mathcal{I}_j \qquad a_{ij} = 0. \tag{A.1}$$

A.1 Amplitudes and slow delay evolution in $\mathcal{I}_j$

The modified version of Algorithm 4 that considers the subsets $\mathcal{I}_j$ is given in Algorithm 6. $\mathcal{I}_j$ is computed simultaneously with the estimation of the delays of a source j in line 5. More precisely, the computation of $\mathcal{I}_j$ takes place in the ICM sweeping strategies presented for both 1D and 2D sets of mixtures in Section 4.3.1.

It is straightforward to show that, when considering a subset of mixtures $\mathcal{I}_j$, that the optimization problem to estimate the delays of source j ($\widetilde{\boldsymbol{L}}_{\mathcal{I}_j j}$) given in (4.8) implies the sum of the correlation products to be only over the mixtures in $\mathcal{I}_j$. In addition, the two mixtures i and i' appearing in the regularization term must both belong to $\mathcal{I}_j$:

$$\widetilde{\boldsymbol{L}}_{\mathcal{I}_j j} \leftarrow \underset{\boldsymbol{L}_{\mathcal{I}_j j}}{\operatorname{argmax}} \sum_{i \in \mathcal{I}_j} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{\substack{(i,i') \in \mathcal{G} \\ i,i' \in \mathcal{I}_j}} \left(\ell_{ij} - \ell_{i'j}\right)^2. \tag{A.2}$$

Algorithm 6: Implementation of $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow \underset{\boldsymbol{A},\boldsymbol{L}}{\operatorname{argmin}} F(\boldsymbol{A}, \boldsymbol{L}, \widehat{\boldsymbol{w}})$, considering the subsets $\mathcal{I}_j$

Initialization: $\boldsymbol{A}^{\text{new}} = \boldsymbol{L}^{\text{new}} = \widehat{\boldsymbol{A}} = \widehat{\boldsymbol{L}} = \boldsymbol{0}_{I\times J}$, $\mathcal{J} = \emptyset$, $\boldsymbol{r}_i = \boldsymbol{x}_i \ \forall i$, $k = 0$

do
- $k \leftarrow k + 1$
- $(\boldsymbol{A}^{\text{old}}, \boldsymbol{L}^{\text{old}}) \leftarrow (\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}})$
- **for** $j \in \{1, \ldots, J\} \backslash \mathcal{J}$ **do**
  - Compute $\mathcal{I}_j$ and $\widetilde{\boldsymbol{L}}_{\mathcal{I}_j j}$ defined in (A.2) using the ICM algorithm
- **end**
- $\widehat{j} \leftarrow \underset{j\notin\mathcal{J}}{\operatorname{argmax}} \sum_{i\in\mathcal{I}_j} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{\substack{(i,i')\in\mathcal{G} \\ i\in\mathcal{I}_j}} \left(\widetilde{\ell}_{ij} - \widetilde{\ell}_{i'j}\right)^2$
- $\mathcal{J} \leftarrow \mathcal{J} \cup \{\widehat{j}\}$
- $\boldsymbol{L}^{\text{new}}_{\mathcal{I}_{\widehat{j}}\widehat{j}} \leftarrow \widetilde{\boldsymbol{L}}_{\mathcal{I}_{\widehat{j}}\widehat{j}}$
- **for** $i \in \mathcal{I}_{\widehat{j}}$ **do**
  - $\boldsymbol{A}^{\text{new}}_{i\mathcal{J}} \leftarrow \underset{\boldsymbol{A}_{i\mathcal{J}}\geq\boldsymbol{0}}{\operatorname{argmin}} F(\boldsymbol{A}, \boldsymbol{L}^{\text{new}}, \widehat{\boldsymbol{w}})$
- **end**
- **if** $F(\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}}, \widehat{\boldsymbol{w}}) < F(\boldsymbol{A}^{\text{old}}, \boldsymbol{L}^{\text{old}}, \widehat{\boldsymbol{w}})$ **then**
  - $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow (\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}})$
  - **for** $i \in \mathcal{I}_{\widehat{j}}$ **do** $\boldsymbol{r}_i \leftarrow \boldsymbol{x}_i - \sum_{j\in\mathcal{J}} \widehat{a}_{ij} \boldsymbol{s}\left[\widehat{\ell}_{ij}\Delta; \widehat{w}_j\right]$ **end**
- **end**

while $k \leq J$ **and** $F(\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}}, \widehat{\boldsymbol{w}}) < F(\boldsymbol{A}^{\text{old}}, \boldsymbol{L}^{\text{old}}, \widehat{\boldsymbol{w}})$;

if $F(\boldsymbol{A}^0, \boldsymbol{L}^0, \widehat{\boldsymbol{w}}) < F(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{w}})$ **then**
- $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) \leftarrow (\boldsymbol{A}^0, \boldsymbol{L}^0)$

end

Consequently, the optimization of (A.2) with respect to one delay while fixing all the others (ICM strategy) reads:

$$\widetilde{\ell}_{ij} \leftarrow \underset{\ell_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; \widehat{w}_j\right]\right)_+^2 - \tau\Delta^2 \sum_{\substack{(i,i')\in\mathcal{G} \\ i,i'\in\mathcal{I}_j}} \left(\ell_{ij} - \widetilde{\ell}_{i'j}\right)^2. \tag{A.3}$$

Furthermore, we introduce a threshold κ in the ICM sweeping procedure. It is compared to the correlation product $\left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+$ for a delay estimate $\widetilde{\ell}_{ij}$: the corresponding mixture i is added to $\mathcal{I}_j$ only if the correlation product is larger than κ. This comparison is justified since the correlation product corresponds to the amplitude source j as demonstrated inequation (4.7). Therefore, κ can be related to the noise standard variation (*e.g.*, $3\sigma_n$). For instance, in astronomy, it is common to set such a threshold to $3\sigma_n$ with σ_n is the noise standard-deviation.

In the sequel, we detail how the threshold κ is included in the 1D and 2D ICM sweeping procedures.

A.2 Determining $\mathcal{I}_j$ and $\widetilde{L}_{\mathcal{I}_j j}$ in 1D ICM sweeping

Each iteration of the ICM is as follows when dealing with a 1D set of mixtures. First, a starting mixture i_0 is randomly selected and its corresponding delay $\widetilde{\ell}_{i_0 j}$ is estimated using (A.3). i_0 is added to $\mathcal{I}_j$ only if the correlation product $\left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{i_0 j}\Delta; \widehat{w}_j\right]\right)_+$ is larger than κ. Otherwise, $\widetilde{\ell}_{i_0 j}$ is discarded and a new starting mixture is randomly selected; this is repeated till a correlation higher than κ is obtained. Next, the mixtures with index going from $i_0 + 1$ to I are sequentially visited and their corresponding delays $\widetilde{\ell}_{ij}$ are estimated using (A.3). But, if a delay estimate for a mixture $i \in \{i_0 + 1, \ldots, I\}$ results in a correlation product $\left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+$ smaller than κ, then the mixtures from i to I are not added to $\mathcal{I}_j$ and their delays are discarded. A similar procedure is applied when sweeping the mixtures from $i_0 - 1$ to 1.

A.3 Determining $\mathcal{I}_j$ and $\widetilde{L}_{\mathcal{I}_j j}$ in 2D ICM sweeping

In Section 4.3.1 we detailed the sequential sweeping procedure when dealing with a 2D sets of mixtures which is inspired from the region-growing algorithm. Introducing a threshold to this sweeping strategy is straightforward since the region-growing algorithm is designed to give "regions" of connected subsets of spatial positions. The sweeping procedure given in Section 4.3.1 is modified as follows to consider the threshold κ:

1. a starting mixture $i_0 \in \{1, \ldots, I\}$ is randomly selected, and its corresponding delay $\widetilde{\ell}_{i_0 j}$ is estimated using (A.3). i_0 is added to $\mathcal{B}$ only if the corresponding correlation product $\left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{i_0 j}\Delta; \widehat{w}_j\right]\right)_+$ is larger than κ. Otherwise, a new starting mixture is randomly selected;
2. the mixture with the highest correlation and belonging to $\mathcal{B}$ is selected. Let i_s be the index of this mixture, then $i_s \leftarrow \text{argmax}_{i \in \mathcal{B}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+^2$;
3. i_s is retired from $\mathcal{B}$ and added to $\mathcal{I}_j$;
4. the neighbors of i_s which are neither in $\mathcal{I}_j$ and $\mathcal{B}$ are considered. Their delays are estimated using (A.3) and only the mixture indices of the ones yielding correlation product $\left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\widetilde{\ell}_{ij}\Delta; \widehat{w}_j\right]\right)_+$ larger than κ are added $\mathcal{B}$;
5. repeat from 2 till $\mathcal{B}$ is empty.

B

Joint sparse approximation: amplitudes, slow evolving delays and shapes

B.1 Delay and shape sampling and regularization

In this appendix, we consider a new method inspired from Algorithm 4; a joint greedy algorithm inspired from OMP to estimate the amplitudes and the slow evolving delays within the mixtures. In this appendix, we extend Algorithm 4 to also estimate the shape parameters which are supposed to evolve slowly, in contrary to Chapter 4 where the shapes are supposed to be constant for each source within the mixtures. Another difference with Chapter 4 is that all the parameters are estimated in one step, hence avoiding an ALS scheme.

B.1.1 Delay and shape regularized criterion

Similar to the delay sampling in Chapter 4 ($c_{ij} = \ell_{ij}\Delta$ with Δ is the delay sampling step), the shape parameters are also sampled over a grid. This is due to the use of a sparse approximation strategy to estimate the delays and shapes. The shape sampling step is denoted by Θ, such that $w_{ij} = p_{ij}\Theta$ with $p_{ij} \in \mathbb{N}$. Besides, p_{ij} is sampled in such a way that $w_{\min} \leq p_{ij}\Theta \leq w_{\max}$ where $w_{\min}$ and $w_{\max}$ are respectively predefined lower and upper bounds of the shape parameters. Therefore, the new mixing model considering the delay and shape sampling writes as:

$$\forall\, i, \qquad \boldsymbol{x}_i = \sum_j a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; p_{ij}\Theta\right] + \boldsymbol{n}_i. \tag{B.1}$$

The shape slow evolution is favored by a regularization term $H(\boldsymbol{P})$ which is added to the data-fit term $E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{P}) = \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{J} a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; p_{ij}\Theta\right] \right\|_2^2$ and the delay regularization term

$R(\boldsymbol{L}) = \Delta^2 \sum_{j=1}^{J} \sum_{(i,i')\in\mathcal{G}} (\ell_{ij} - \ell_{i'j})^2$. The shape regularization term is defined similarly to the delay regularization term $R(\boldsymbol{L})$, *i.e.*, :

$$H(\boldsymbol{P}) = \Theta^2 \sum_{j=1}^{J} \sum_{(i,i')\in\mathcal{G}} (p_{ij} - p_{i'j})^2. \tag{B.2}$$

where $\boldsymbol{P} \in \mathbb{R}^{I\times J}$ is a matrix gathering the shape indices p_{ij} and $\mathcal{G}$ is the set of all the cliques of order two.

Hence, the global criterion reads:

$$G(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{P}) = E(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{P}) + \tau\Delta^2 R(\boldsymbol{L}) + \mu\Theta^2 H(\boldsymbol{P}), \tag{B.3}$$

with τ and μ are respectively the delay and shape regularization terms.

B.1.2 Amplitudes and slow delays and shapes estimation

The considered optimization problem is:

$$\left(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{P}}\right) \leftarrow \underset{\boldsymbol{A}\geq\boldsymbol{0}, \boldsymbol{L}, \boldsymbol{P}}{\operatorname{argmin}}\, G(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{P}). \tag{B.4}$$

We propose to optimize the criterion G jointly (without using an ALS scheme as it was done for the optimization of E and F in Chapters 3 and 4).

The proposed algorithm is given in Algorithm 7 which is a greedy algorithm that shares similarities with Algorithm 4 (therefore some aspects are not detailed as they were developed in Section 4.3). At each iteration the algorithm follows the OMP philosophy: the delays and shapes of a source j are estimated while fixing the other sources (the next section details how this is done). j is then added to the list of estimated source $\mathcal{J}$. Next, the amplitudes of the sources in $\mathcal{J}$ are estimated using a non-negative least squares solver (line 6). Finally, the residual vectors of all the mixtures are updated (line 9). The condition in line 7 is set to break the loop in case of an increase of G induced by the regularization terms.

B.1.3 Delay and shape estimation with an ICM-like algorithm

Let us now detail the estimation of the delays and shapes of source j at each iteration of Algorithm 7 (line 5). First, one needs to consider the minimization of $G(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{P})$ with respect to $\boldsymbol{A}_{:j}$,$\boldsymbol{L}_{:j}$ and $\boldsymbol{P}_{:j}$, while fixing the values of previously estimated sources:

$$\left(\boldsymbol{L}_{:j}^{\text{new}}, \boldsymbol{P}_{:j}^{\text{new}}\right) \leftarrow \underset{\boldsymbol{L}_{:j}, \boldsymbol{P}_{:j}}{\operatorname{argmin}} \min_{\boldsymbol{A}_{:j}\geq\boldsymbol{0}} \sum_{i=1}^{I} \left\| \boldsymbol{r}_i - a_{ij}\boldsymbol{s}\left[\ell_{ij}\Delta; p_{ij}\Theta\right] \right\|_2^2 + \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}\right)^2 + \mu\Theta^2 \sum_{(i,i')\in\mathcal{G}} \left(p_{ij} - p_{i'j}\right)^2. \tag{B.5}$$

Algorithm 7: Implementation of $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{P}}) \leftarrow \underset{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{L},\boldsymbol{P}}{\operatorname{argmin}} G(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{P})$

Initialization: $\boldsymbol{A}^{\text{new}} = \boldsymbol{L}^{\text{new}} = \boldsymbol{P}^{\text{new}} = \widehat{\boldsymbol{A}} = \widehat{\boldsymbol{L}} = \widehat{\boldsymbol{P}} = \boldsymbol{0}_{I\times J}$, $\mathcal{J} = \emptyset$, $\boldsymbol{r}_i = \boldsymbol{x}_i \ \forall i$, $j = 0$

do
- $j \leftarrow j + 1$
- $\mathcal{J} \leftarrow \mathcal{J} \cup \{j\}$
- $(\boldsymbol{A}^{\text{old}}, \boldsymbol{L}^{\text{old}}, \boldsymbol{P}^{\text{old}}) \leftarrow (\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}}, \boldsymbol{P}^{\text{new}})$
- Compute $\boldsymbol{L}_{:j}^{\text{new}}$ and $\boldsymbol{P}_{:j}^{\text{new}}$ defined in (B.6) using the ICM algorithm
- $\boldsymbol{A}^{\text{new}} \leftarrow \underset{\boldsymbol{A}}{\operatorname{argmin}}\, G(\boldsymbol{A}, \boldsymbol{L}^{\text{new}}, \boldsymbol{P}^{\text{new}})$ s.t. $\{\boldsymbol{A}_{:\mathcal{J}} \geq \boldsymbol{0}, \boldsymbol{A}_{:\overline{\mathcal{J}}} = \boldsymbol{0}\}$
- **if** $G(\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}}, \boldsymbol{P}^{\text{new}}) < G(\boldsymbol{A}^{\text{old}}, \boldsymbol{L}^{\text{old}}, \boldsymbol{P}^{\text{old}})$ **then**
  - $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{P}}) \leftarrow (\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}}, \boldsymbol{P}^{\text{new}})$
  - **for** $i = 1 : I$ **do** $\boldsymbol{r}_i \leftarrow \boldsymbol{x}_i - \sum_{j\in\mathcal{J}} \widehat{a}_{ij}\boldsymbol{s}\left[\widehat{\ell}_{ij}\Delta; \widehat{p}_{ij}\Theta\right]$ **end**
- **end**

while $j \leq J$ **and** $G(\boldsymbol{A}^{\text{new}}, \boldsymbol{L}^{\text{new}}, \boldsymbol{P}^{\text{new}}) < F(\boldsymbol{A}^{\text{old}}, \boldsymbol{L}^{\text{old}}, \boldsymbol{P}^{\text{old}})$;

After using the amplitude closed-form estimation in (4.7), (B.5) simplifies to:

$$\left(\boldsymbol{L}_{:j}^{\text{new}}, \boldsymbol{P}_{:j}^{\text{new}}\right) \leftarrow \underset{\boldsymbol{L}_{:j},\boldsymbol{P}_{:j}}{\operatorname{argmax}} \sum_{i=1}^{I} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; p_{ij}\Theta\right]\right)_+^2 - \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}\right)^2 - \mu\Theta^2 \sum_{(i,i')\in\mathcal{G}} \left(p_{ij} - p_{i'j}\right)^2. \tag{B.6}$$

Again, we propose to use the ICM algorithm to solve this combinatorial problem. At each ICM iteration all the mixtures are swept. Whenever a mixture is visited, its delay ℓ_{ij} and shape p_{ij} are estimated while fixing all the other couples $(\ell_{i'j}, p_{i'j})$ with $i' \neq i$. This leads to the following estimation problem with two unknowns:

$$\left(\ell_{ij}^{\text{new}}, p_{ij}^{\text{new}}\right) \leftarrow \underset{\ell_{ij},p_{ij}}{\operatorname{argmax}} \left(\boldsymbol{r}_i^T \boldsymbol{s}\left[\ell_{ij}\Delta; p_{ij}\Theta\right]\right)_+^2 - \tau\Delta^2 \sum_{(i,i')\in\mathcal{G}} \left(\ell_{ij} - \ell_{i'j}^{\text{new}}\right)^2 - \mu\Theta^2 \sum_{(i,i')\in\mathcal{G}} \left(p_{ij} - p_{i'j}^{\text{new}}\right)^2. \tag{B.7}$$

This optimization problem can be easily solved by trying all the combinations of the couples ℓ_{ij} and p_{ij} (an optimized implementation allows to resolve this problem very efficiently).

At each ICM iteration, the mixtures are swept sequentially so that the neighboring information is effectively exploited. The same sequential sweeping strategies as in Section 4.3.1 are used to find ℓ_{ij} and p_{ij} for a source j. The only difference would be to resolve the local optimization problem (B.7) instead of (4.9).

B.2 Results

Algorithm 7 is tested on a synthetic 1D set of mixtures and real data, namely, time-resolved photoelectron spectra and the galaxy NGC-4254.

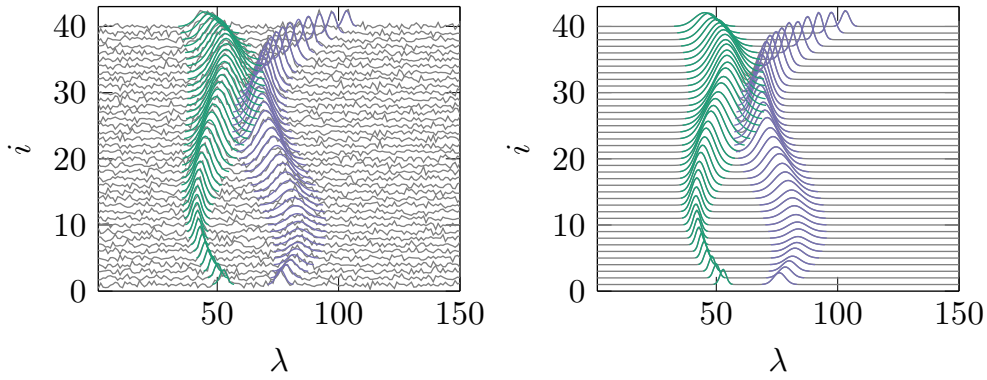


Figure B.1: (a) a synthetic noisy set of 1D mixtures. (b) the mixture reconstruction obtained by the proposed method.

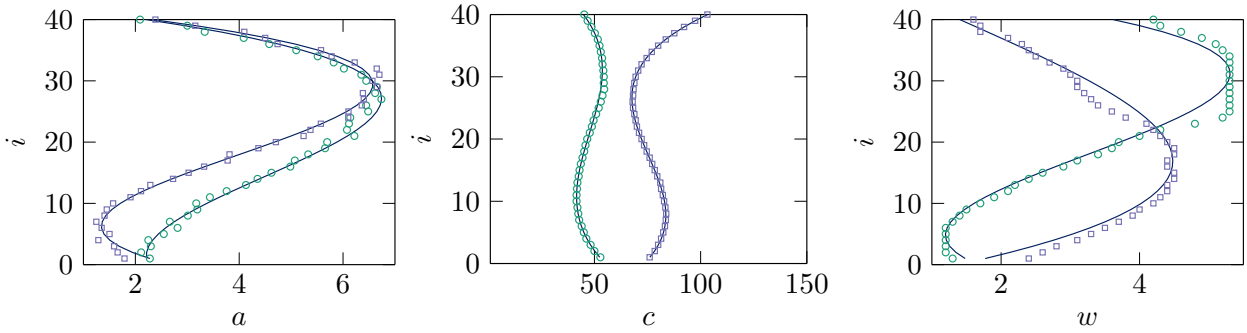
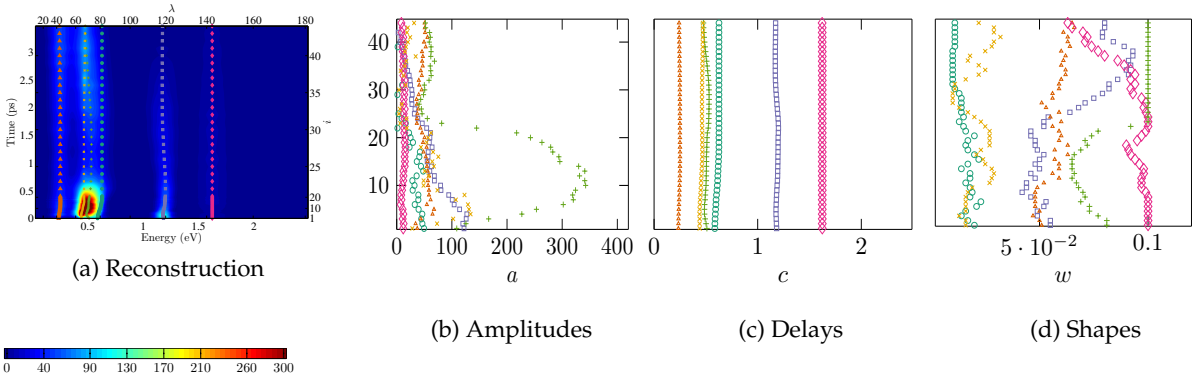


Figure B.2: The estimated parameters (amplitudes $\widehat{a}_{ij}$, delays $c_{ij} = \widehat{\ell}_{ij}\Delta$ and shapes $\widehat{w}_{ij} = \widehat{p}_{ij}\Theta$) of the estimated sources. The ground truth parameters are plotted in plain lines.

B.2.1 Synthetic 1D set of mixtures

The performance of the proposed algorithm is evaluated on synthetic 1D set of mixtures. The following settings are set. All the sources are modeled by a Gaussian function: $s(\lambda; w_j) = \exp(-\lambda^2/2w_j^2)$ where the shape parameters w_j represent the Gaussian widths. The number of mixtures is equal to $I = 40$, each with $N = 150$ samples and the number of sources is set to $J = 2$. The delay sampling step is set to $\Delta = 0.2$. The shape sampling step is set to $\Theta = 0.1$ with $w_{\min} = 1$ and $w_{\max} = 5.5$. The delay and shape regularization parameters are respectively tuned to $\tau = 0.1$ and $\mu = 5$.

The generated mixtures, their reconstruction and the estimated sources are displayed in Figure B.1 and the estimated parameters are shown in Figure B.2. The reconstruction and the source recovery are satisfactory. In addition, the delays and shapes evolve slowly as expected. However, around the mixture $i = 30$ the estimation is slightly affected when the two sources are spectrally close. These results are obtained in only 0.3 s.



(a) Reconstruction (b) Amplitudes (c) Delays (d) Shapes

Figure B.3: First column: mixture reconstruction superimposed by the estimated delays. The second, third and fourth columns respectively represent the estimated amplitudes, delays and shapes.

B.2.2 Time-resolved photoelectron spectroscopy

Figure B.3 presents the results obtained on a real 1D set of mixtures, namely real time-resolved spectra (the real data are presented in Figure 6.7). Note that the background has been removed from the mixtures before the processing using the results of Chapter 4. The delay sampling step is set to $\Delta = 5 \cdot 10^{-4}$. The shape sampling step is set to $\Theta = 1 \cdot 10^{-3}$ with $w_{\min} = 0.01$ and $w_{\max} = 0.1$. The delay and shape regularization parameters are respectively tuned to $\tau = 3000$ and $\mu = 100$. The obtained results are similar to the method of Chapter 4 (displayed in Section 6.4.1), in terms of mixture reconstruction and amplitude and delay estimates. The shape parameter evolution is slow and smoother that the one obtained by the method of [Maz+15] (see Figure 6.9). The results are obtained in 16 s.

B.2.3 Galaxy NGC-4254

Finally, the algorithm is applied to the galaxy NGC-4254 multispectral image presented in Figure 6.10. The delay sampling step is set to $\Delta = 0.2$. The shape sampling step is set to $\Theta = 0.1$ with $w_{\min} = 1$ and $w_{\max} = 5$. The delay and shape regularization parameters are respectively tuned to $\tau = 3 \cdot 10^{-4}$ and $\mu = 5 \cdot 10^{-4}$. The sources appear only in a subset of the mixtures, therefore the extension used in Appendix A is applied to Algorithm 7 and we set the threshold $\kappa = 2\sigma_n$.

The amplitude, delay and shape estimates of the three sources are given in Figure B.4. Concerning, the parameter estimates, the amplitude and delay estimates are similar to the ones obtained with the method of Chapter 4 (regularized method in Figure 6.12). The shape parameters between the two methods are different, namely for the first source as we get a map of varying shapes indicating the velocity dispersion of the main kinematic structure. The results are obtained in 43 s.

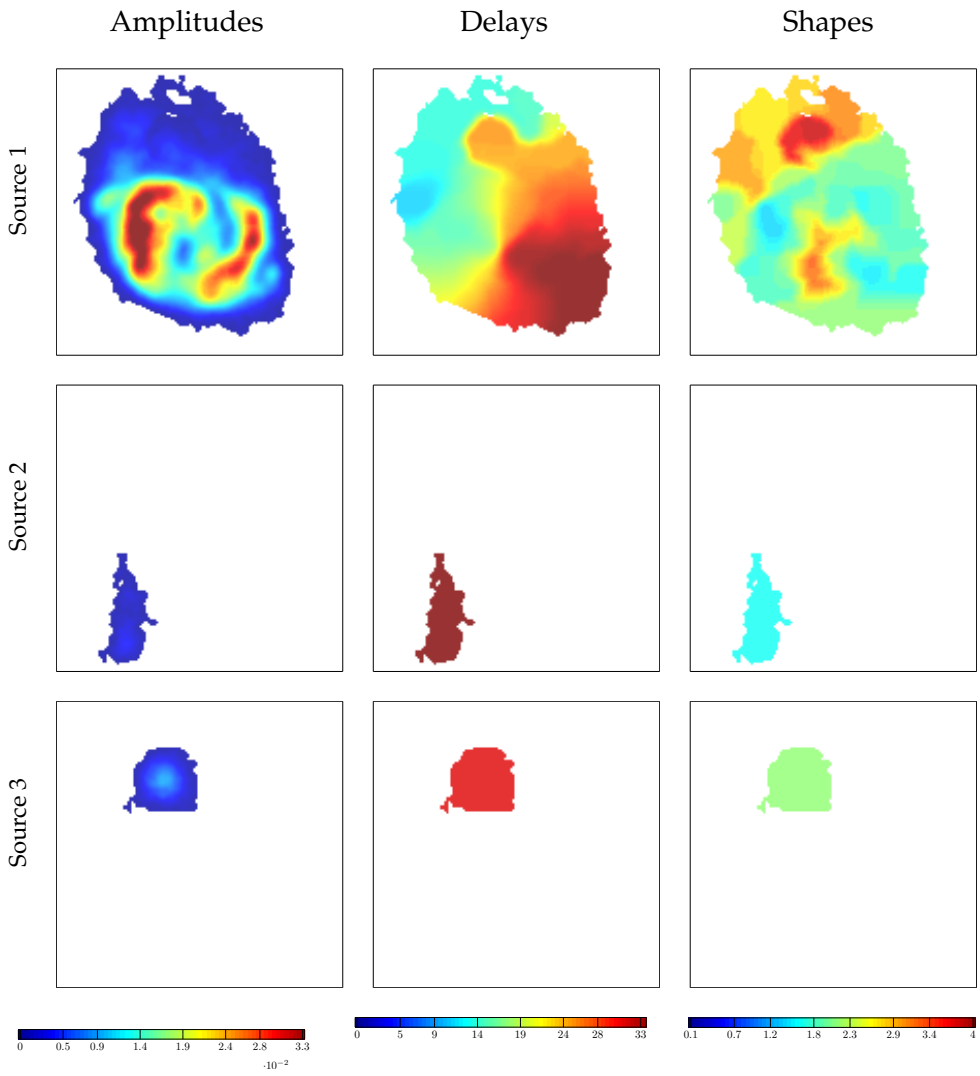


Figure B.4: The estimated parameter maps of $J = 3$ sources with the method of Appendix B.

Bibliography

[AB94] R. Adams and L. Bischof. "Seeded region growing". In: *IEEE Transactions on pattern analysis and machine intelligence* 16.6 (1994), pp. 641–647.

[AGB10] S. Arberet, R. Gribonval, and F. Bimbot. "A robust method to count and locate audio sources in a multichannel underdetermined mixture". In: *IEEE Transactions on Signal Processing* 58.1 (2010), pp. 121 –133.

[AHK12] S. Arora, E. Hazan, and S. Kale. "The Multiplicative Weights Update Method: a Meta-Algorithm and Applications." In: *Theory of Computing* 8.1 (2012), pp. 121–164.

[AK98] E. Amaldi and V. Kann. "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems". In: *Theoretical Computer Science* 209.1-2 (1998), pp. 237–260.

[Ang01] D. Anglos. "Laser-induced breakdown spectroscopy in art and archaeology". In: *Applied Spectroscopy* 55.6 (2001), 186A–205A.

[AOT00] R. K. Ahuja, J. B. Orlin, and A. Tiwari. "A greedy genetic algorithm for the quadratic assignment problem". In: *Computers & Operations Research* 27.10 (2000), pp. 917–934.

[Aru+02] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on Signal Processing* 50.2 (2002), pp. 174–188.

[Aus+10] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin. "On the relation between sparse reconstruction and parameter estimation With Model Order Selection". In: *IEEE Journal of Selected Topics in Signal Processing* 4.3 (2010), pp. 560–570.

[BCDF10] E. Brochu, V. M. Cora, and N. De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning". In: *arXiv preprint:1012.2599* (2010).

[BCI07] S. Bourguignon, H. Carfantan, and J. Idier. "A sparsity-based method for the estimation of spectral lines from irregularly sampled data". In: *IEEE Journal of Selected Topics in Signal Processing* 1.4 (2007), pp. 575–585.

[BD07] T. Blumensath and M. E. Davies. *On the difference between orthogonal matching pursuit and orthogonal least squares*. Tech. rep. 2007.

[BD09] T. Blumensath and M. E. Davies. "Iterative hard thresholding for compressed sensing". In: *Applied and computational harmonic analysis* 27.3 (2009), pp. 265–274.

[BDB14] L. Belmerhnia, E.-H. Djermoune, and D. Brie. "Greedy methods for simultaneous sparse approximation". In: *Proceedings of the European Signal Processing Conference, EUSIPCO*. IEEE. 2014, pp. 1851–1855.

[Ben00] N. Benitez. "Bayesian photometric redshift estimation". In: *The Astrophysical Journal* 536.2 (2000), p. 571.

[Bes86] J. Besag. "On the statistical analysis of dirty pictures". In: *Journal of the Royal Statistical Society B* 48.3 (1986), pp. 48–259.

[BEZ08] A. M. Bruckstein, M. Elad, and M. Zibulevsky. "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations". In: *IEEE Transactions on Information Theory* 54.11 (2008), pp. 4813–4820.

[BH09] A. Barth and P. I. Haris. *Biological and biomedical infrared spectroscopy*. Vol. 2. IOS press, 2009.

[Bla+14] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jing. "Greedy algorithms for joint sparse recovery." In: *IEEE Transactions on Signal Processing* 62.7 (2014), pp. 1694–1704.

[Bla86] S. S. Blackman. *Multiple-Target Tracking with Radar Applications*. Norwood, MA: Artech House, 1986.

[BM68] A. Bauder and R. J. Myers. "Least squares curve fitting of EPR spectra". In: *Journal of Molecular Spectroscopy* 27.1-4 (1968), pp. 110–116.

[BMS11] S. Bourguignon, D. Mary, and É. Slezak. "Restoration of Astrophysical Spectra With Sparsity Constraints: Models and Algorithms". In: *IEEE Journal of Selected Topics in Signal Processing* 5.5 (2011), pp. 1002–1013.

[Bob+15] J. Bobin, J. Rapin, A. Larue, and J.-L. Starck. "Sparsity and adaptivity for the blind separation of partially correlated sources". In: *IEEE Transactions on Signal Processing* 63.5 (2015), pp. 1199–1213.

[Bos81] A. Bosma. "21-cm line studies of spiral galaxies. II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types". In: *The Astronomical Journal* 86.12 (1981), pp. 1825–1846.

[Bou+11] S. Bourguignon, C. Soussen, H. Carfantan, and J. Idier. "Sparse deconvolution: Comparison of statistical and deterministic approaches". In: *2011 IEEE Statistical Signal Processing Workshop, SSP*. IEEE. 2011, pp. 317–320.

[BR03] C. Blum and A. Roli. "Metaheuristics in combinatorial optimization: Overview and conceptual comparison". In: *ACM Computing Surveys* 35.3 (2003), pp. 268–308.

[Bre13] G. L. Bretthorst. *Bayesian spectrum analysis and parameter estimation*. Vol. 48. Springer Science & Business Media, 2013.

[BT09] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM journal on imaging sciences* 2.1 (2009), pp. 183–202.

[BTR13] B. N. Bhaskar, G. Tang, and B. Recht. "Atomic norm denoising with applications to line spectral estimation". In: *IEEE Transactions on Signal Processing* 61.23 (2013), pp. 5987–5999.

[Bud+96] D. E. Budil, S. Lee, S. Saxena, and J. H. Freed. "Nonlinear-least-squares analysis of slow-motion EPR spectra in one and two dimensions using a modified Levenberg–Marquardt algorithm". In: *Journal of Magnetic Resonance, Series A* 120.2 (1996), pp. 155–189.

[Cap79] J. Capon. “Maximum-likelihood spectral estimation”. In: *Nonlinear methods of spectral analysis*. Springer, 1979, pp. 155–179.

[Car+86] R. A Caruana, R. B. Searle, T. Heller, and S. I. Shupack. “Fast algorithm for the resolution of spectra”. In: *Analytical chemistry* 58.6 (1986), pp. 1162–1167.

[Car87] G. C. Carter. “Coherence and time delay estimation”. In: *Proceedings of the IEEE* 75.2 (1987), pp. 236–255.

[CB06] G. Chabriel and J. Barrère. “An instantaneous formulation of mixtures for blind separation of propagating waves”. In: *IEEE Transactions on Signal Processing* 54.1 (2006), pp. 49–58.

[CBL89] S. Chen, S. A. Billings, and W. Luo. “Orthogonal least squares methods and their application to non-linear system identification”. In: *International Journal of control* 50.5 (1989), pp. 1873–1896.

[CDP17] P. Catala, V. Duval, and G. Peyré. “A low-rank approach to off-the-grid sparse deconvolution”. In: *Journal of Physics: Conference Series*. Vol. 904. 1. IOP Publishing. 2017, p. 012015.

[CJ10] P. Comon and C. Jutten. *Handbook of blind source separation: independent component analysis and applications*. Academic Press, 2010.

[Clé+11] H. Clénet, P. Pinet, Y. Daydou, F. Heuripeau, C. Rosemberg, D. Baratoux, and S. Chevrel. “A new systematic approach using the modified Gaussian model: insight for the characterization of chemical composition of olivines, pyroxenes and olivine–pyroxene mixtures”. In: *Icarus* 213.1 (2011), pp. 404–422.

[Coi+98] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik. “A real-time computer vision system for vehicle tracking and traffic surveillance”. In: *Transportation Research Part C: Emerging Technologies* 6.4 (1998), pp. 271–288.

[CPF14] E. Chouzenoux, J. C. Pesquet, and A. Florescu. “A multi-parameter optimization approach for complex continuous sparse modeling”. In: *In Proceedings of the 19th International Conference on Digital Signal Processing, DSP*. 2014, pp. 817–820.

[CRT06] E. J. Candes, J. K. Romberg, and T. Tao. “Stable signal recovery from incomplete and inaccurate measurements”. In: *Communications on pure and applied mathematics* 59.8 (2006), pp. 1207–1223.

[CT07] J. W. Cannon and W. P. Thurston. “Group invariant Peano curves”. In: *Geometry & Topology* 11.3 (2007), pp. 1315–1355.

[DB13] M. F. Duarte and R. G. Baraniuk. “Spectral compressive sensing”. In: *Applied and Computational Harmonic Analysis* 35.1 (2013), pp. 111–129.

[DB72] C. De Boor. “On calculating with B-splines”. In: *Journal of Approximation theory* 6.1 (1972), pp. 50–62.

[DDDM04] I. Daubechies, M. Defrise, and C. De Mol. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. In: *Communications on Pure and Applied Mathematics* 57.11 (2004), pp. 1413–1457.

[DE03] D. L. Donoho and M.l Elad. "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization". In: *Proceedings of the National Academy of Sciences* 100.5 (2003), pp. 2197–2202.

[DGK01] I. DiMatteo, C. R. Genovese, and R. E. Kass. "Bayesian curve-fitting with free-knot splines". In: *Biometrika* 88.4 (2001), pp. 1055–1071.

[DM05] C. Dossal and S. Mallat. "Sparse spike deconvolution with minimum scale". In: *Proceedings of Signal Processing with Adaptive Sparse Structured Representations, SPARS*. Citeseer. 2005, pp. 1–4.

[Don06a] D. L. Donoho. "Compressed sensing". In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.

[Don06b] D. L. Donoho. "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 59.6 (2006), pp. 797–829.

[DP15] V. Duval and G. Peyré. "Exact support recovery for sparse spikes deconvolution". In: *Foundations of Computational Mathematics* 15.5 (2015), pp. 1315–1355.

[DP17] V. Duval and G. Peyré. "Sparse spikes super-resolution on thin grids II: the continuous basis pursuit". In: *Inverse Problems* 33.9 (2017), p. 095008.

[Dua10] J. Duan. "Restoration and separation of piecewise polynomial signals. Application to Atomic Force Microscopy". PhD thesis. Université Henri Poincaré, Nancy, 2010.

[EK12] Y. C. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[Ela10] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer New York, 2010.

[ETS11] C. Ekanadham, D. Tranchina, and E. P. Simoncelli. "Recovery of sparse translation-invariant signals with continuous basis pursuit". In: *IEEE Transactions on Signal Processing* 59 (2011), pp. 4735–4744.

[FD05] R Fischer and V Dose. "Analysis of mixtures in physical spectra". In: *Monographs of official statistics: Bayesian methods with applications to science, policy, and official statistics*. 2005, pp. 145–154.

[FDJ15] K. Fyhn, M. F. Duarte, and S. H. Jensen. "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation". In: *IEEE Transactions on Signal Processing* 63.4 (2015), pp. 870–881.

[FL01] J. Fan and R. Li. "Variable selection via nonconcave penalized likelihood and its oracle properties". In: *Journal of the American statistical Association* 96.456 (2001), pp. 1348–1360.

[Fli+05] F. Flitti, C. Collet, B. Vollmer, and F. Bonnarel. "Multiband segmentation of a spectroscopic line data cube: application to the HI data cube of the spiral galaxy NGC-4254". In: *EURASIP Journal on Advances in Signal Processing* 2005.15 (2005), p. 921039.

[FNW07] M. Figueiredo, R. D. Nowak, and S. J. Wright. "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems". In: *IEEE Journal of selected topics in signal processing* 1.4 (2007), pp. 586–597.

[Fri12] J. H. Friedman. "Fast sparse regression and classification". In: *International Journal of Forecasting* 28.3 (2012), pp. 722–738.

[Glo+05] E. Gloaguen, J.-M. Mestdagh, L. Poisson, F. Lepetit, J.-P. Visticot, B. Soep, M. Coroiu, A. Eppink, and D. H. Parker. "Experimental evidence for ultrafast electronic relaxation in molecules, mediated by diffuse states". In: *Journal of the American Chemical Society* 127.47 (2005), pp. 16529–16534.

[Goe+85] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. "Imaging spectrometry for earth remote sensing". In: *Science* 228.4704 (1985), pp. 1147–1153.

[Gre95] P. J. Green. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination". In: *Biometrika* 82.4 (1995), pp. 711–732.

[GS00] L. Grippo and M. Sciandrone. "On the convergence of the block nonlinear Gauss–Seidel method under convex constraints". In: *Operations research letters* 26.3 (2000), pp. 127–136.

[Guo11] H. Guo. "A simple algorithm for fitting a Gaussian function". In: *IEEE Signal Processing Magazine* 28.5 (2011), pp. 134–137.

[GWK89] G. Gilmore, R. FG Wyse, and K. Kuijken. "Kinematics, chemistry, and structure of the Galaxy". In: *Annual review of Astronomy and Astrophysics* 27.1 (1989), pp. 555–627.

[Had23] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Yale University Press, 1923.

[Hay04] S. Haykin. *Kalman filtering and neural networks*. Vol. 47. John Wiley & Sons, 2004.

[HCS03] R. Hesse, T. Chassé, and R. Szargan. "Unifit 2002—universal analysis software for photoelectron spectra". In: *Analytical and bioanalytical chemistry* 375.7 (2003), pp. 856–863.

[HG02] N. M. Haan and S. J. Godsill. "Bayesian models for DNA sequencing". In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. Vol. 4. IEEE. 2002, pp. IV–4020.

[HH03] S. Hong and R. A. Harshman. "Shifted factor analysis—part II: algorithms". In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 17.7 (2003), pp. 379–388.

[HHL03] Richard A Harshman, Sungjin Hong, and Margaret E Lundy. "Shifted factor analysis—Part I: Models and properties". In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 17.7 (2003), pp. 363–378.

[HKW16] G. A Hanasusanto, D. Kuhn, and W. Wiesemann. "A comment on "computational complexity of stochastic programming problems"". In: *Mathematical Programming* 159.1-2 (2016), pp. 557–569.

[Hol04] J. M. Hollas. *Modern spectroscopy*. John Wiley & Sons, 2004.

[JF11] N. Jiang and D. Farina. "Covariance and time-scale methods for blind separation of delayed sources". In: *IEEE Transactions on Biomedical Engineering* 58.3 (2011), pp. 550–556.

[JRY00] A. Jourjine, S. Rickard, and O. Yilmaz. "Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures". In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. Vol. 5. 2000, pp. 2985–2988.

[JTD11] P. Jain, A. Tewari, and I. S. Dhillon. "Orthogonal matching pursuit with replacement". In: *In proceedings of the 25th Annual Conference on Neural Information Processing Systems, NIPS*. 2011, pp. 1215–1223.

[Jup78] D. L. Jupp. "Approximation to data by splines with free knots". In: *SIAM Journal on Numerical Analysis* 15.2 (1978), pp. 328–343.

[KC76] C. Knapp and G. Carter. "The generalized correlation method for estimation of time delay". In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. Vol. 24. IEEE. 1976, pp. 320–327.

[Kim+07] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. "An interior-point method for large-scale ℓ_1-regularized least squares". In: *IEEE journal of selected topics in signal processing* 1.4 (2007), pp. 606–617.

[KJ98] Robert C. Kennicutt J. "Star formation in galaxies along the Hubble sequence". In: *Annual Review of Astronomy and Astrophysics* 36.1 (1998), pp. 189–231.

[Knu+14] K. C. Knudson, J. Yates, A. Huk, and J. W. Pillow. "Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit". In: *Proceedings of Advances in neural information processing systems, NIPS*. 2014, pp. 1215–1223.

[KSD13] M. Kowalski, K. Siedenburg, and M. Dörfler. "Social sparsity! neighborhood systems enrich structured shrinkage operators". In: *IEEE transactions on signal processing* 61.10 (2013), pp. 2498–2511.

[Kuh55] Harold W Kuhn. "The Hungarian method for the assignment problem". In: *Naval Research Logistics* 2.1-2 (1955), pp. 83–97.

[LAC97] J.-L. Lacoume, P.-O. Amblard, and P. Comon. *Statistiques d'ordre supérieur pour le traitement du signal*. MASSON, 1997.

[Lee+01] J. S. Lee, D. D. Lee, S. Choi, and D. S. Lee. "Application of nonnegative matrix factorization to dynamic positron emission tomography". In: *Proceedings of International Conference on Independent Component Analysis and Signal Separation, ICA*. 2001, pp. 629–632.

[Lev44] K. Levenberg. "A method for the solution of certain non-linear problems in least squares". In: *Quarterly of applied mathematics* 2.2 (1944), pp. 164–168.

[LH95] C. L. Lawson and R. J. Hanson. *Solving least squares problems*. SIAM, 1995.

[Mal02] E. R Malinowski. *Factor analysis in chemistry*. Wiley, 2002.

[Mar63] D. W Marquardt. "An algorithm for least-squares estimation of nonlinear parameters". In: *Journal of the society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441.

[Mas+10] A. Masson, L. Poisson, M.A. Gaveau, B. Soep, J.M. Mestdagh, V. Mazet, and F. Spiegelman. "Dynamics of highly excited barium atoms deposited on large argon clusters. I. General trends". In: *Journal of Chemical Physics* 133.5 (2010), p. 054307.

[Mas+14] A Masson, M.C. Heitz, J.M. Mestdagh, M.A. Gaveau, L. Poisson, and F. Spiegelman. "Coupled electronic and structural relaxation pathways in the postexcitation dynamics of Rydberg states of BaAr N Clusters". In: *Physical Review Letters* 113.12 (2014), p. 123005.

[Maz+15] V. Mazet, S. Faisan, S. Awali, M.-A. Gaveau, and L. Poisson. "Unsupervised joint decomposition of a spectroscopic signal sequence". In: *Signal Processing* 109 (2015), pp. 193–205.

[Maz05] V. Mazet. "Développement de méthodes de traitement de signaux spectroscopiques: estimation de la ligne de base et du spectre de raies". PhD thesis. Université Henri Poincaré-Nancy I, 2005.

[Maz11] V. Mazet. "Joint Bayesian decomposition of a spectroscopic signal sequence". In: *IEEE Signal Processing Letters* 18.3 (2011), pp. 181–184.

[MB81] D. Mihalas and J. Binney. *Galactic astronomy: Structure and kinematics*. WH Freeman and Co., 1981.

[MBI05] S. Moussaoui, D. Brie, and J. Idier. "Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution". In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. Vol. 5. 2005, pp. 289–292.

[MCV] V. Mazet, C. Collet, and B. Vollmer. "Decomposition and Classification of Spectral Lines in Astronomical Radio Data Cubes". In: *SCIA 2009*.

[MGB94] H. Michael, F. Georg, and G. Bernd. "Modeling and animation of facial expressions based on B-splines". In: *The Visual Computer* 11.2 (1994), pp. 87–95.

[MMH07a] M. Mørup, K. H. Madsen, and L. K. Hansen. "Shifted independent component analysis". In: *Proceedings of International Conference on Independent Component Analysis and Signal Separation, ICA*. Springer. 2007, pp. 89–96.

[MMH07b] M. Mörup, K. H. Madsen, and L. K. Hansen. "Shifted non-negative matrix factorization". In: *IEEE Workshop on Machine Learning for Signal Processing*. 2007.

[Mor+17a] H. Mortada, V. Mazet, C. Soussen, and C. Collet. "Séparation de sources retardées et paramétriques". In: *GRETSI*. 2017.

[Mor+17b] H. Mortada, V. Mazet, C. Soussen, and C. Collet. "Separation of delayed parameterized sources". In: *Proceedings of the European Signal Processing Conference, EUSIPCO*. 2017.

[Mor+18] H. Mortada, V. Mazet, C. Soussen, and C. Collet. "Spectroscopic decomposition of astronomical multispectral images using B-splines". In: *Workshop on Hyperspectral Images and Signal Processing, WHISPERS*. 2018.

[Mor+19] H. Mortada, V. Mazet, C. Soussen, C. Collet, and L. Poisson. "Parameterized source separation for delayed spectroscopic signals". In: *Signal Processing* 158 (2019), pp. 48–60.

[Mou05] S. Moussaoui. “Séparation de sources non-négatives : Application au traitement des signaux de spectroscopie”. PhD thesis. Université Henri Poincaré, Nancy, 2005.

[MPB09] H. D. Makarewicz, M. Parente, and J. L. Bishop. “Deconvolution of VNIR spectra using modified Gaussian modeling (MGM) with automatic parameter initialization (API) applied to CRISM”. In: *WHISPERS*. IEEE. 2009, pp. 1–5.

[MSD13] V. Mazet, C. Soussen, and E.-H. Djermoune. “Décomposition de spectres en motifs paramétriques par approximation parcimonieuse”. In: *GRETSI*. 2013, pp. 1–4.

[MY86] G. Medioni and Y. Yasumoto. “Corner detection and curve representation using cubic B-splines”. In: *IEEE International Conference on Robotics and Automation*. Vol. 3. 1986, pp. 764–769.

[MZ92] S. Mallat and S. Zhong. “Characterization of signals from multiscale edges”. In: *IEEE Transactions on pattern analysis and machine intelligence* 14.7 (1992), pp. 710–732.

[MZ93] S. G. Mallat and Z. Zhang. “Matching pursuits with time-frequency dictionaries”. In: *IEEE Transactions on signal processing* 41.12 (1993), pp. 3397–3415.

[Mül07] M. Müller. “Dynamic time warping”. In: *Information retrieval for music and motion* (2007), pp. 69–84.

[Na+00] R. Na, I.-M. Stender, L. Ma, and H. C. Wulf. “Autofluorescence spectrum of skin: component bands and body site variations”. In: *Skin Research and Technology* 6.3 (2000), pp. 112–117.

[Ngu+17] T. T. Nguyen, C. Soussen, J. Idier, and E.-H. Djermoune. “An optimized version of non-negative OMP”. In: *GRETSI*. 2017.

[Nio+] D. Nion, B. Vandewoestyne, S. Vanaverbeke, K. Abeele, H. Gersem, and L. De Lathauwer. “A Time-Frequency Technique for Blind Separation and Localization of Pure Delayed Sources”. In: *LVA/ICA 2010.*

[NW99] J. Nocedal and S. Wright. “Numerical optimization”. In: *Springer Science* 35.67-68 (1999), p. 7.

[OG11] Lars Omlor and M. A. Giese. “Anechoic Blind Source Separation Using Wigner Marginals”. In: *J. Mach. Learn. Res.* 12 (2011), pp. 1111–1148.

[OL77] John J Olivero and RL Longbothum. “Empirical fits to the Voigt line width: A brief review”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 17.2 (1977), pp. 233–236.

[Oue+14] W. Ouedraogo, B. Nicolas, B. Oudompheng, J. I. Mars, and C. Jutten. “A frequency method for blind separation of an anechoic mixture”. In: *Proceedings of the European Signal Processing Conference, EUSIPCO*. 2014.

[Paa97] P. Paatero. “Least squares formulation of robust non-negative factor analysis”. In: *Chemometrics and intelligent laboratory systems* 37.1 (1997), pp. 23–35.

[Pav+08] D. L. Pavia, G. M. Lampman, G. S. Kriz, and J. A. Vyvyan. *Introduction to spectroscopy*. Cengage Learning, 2008.

[PB14] N. Parikh and S. Boyd. “Proximal algorithms”. In: *Foundations and Trends in Optimization* 1.3 (2014), pp. 127–239.

[PD05] M. Puigt and Y. Deville. “Time–frequency ratio-based blind separation methods for attenuated and time-delayed sources”. In: *Mechanical Systems and Signal Processing* 19.6 (2005), pp. 1348–1379.

[Pea81] G. Peach. “Theory of the pressure broadening and shift of spectral lines”. In: *Advances in Physics* 30.3 (1981), pp. 367–474.

[Pet95] V. V. Petrov. *Limit theorems of probability theory: sequences of independent random variables*. Tech. rep. Oxford, New York, 1995.

[Phi+86] M. Phillips, C. Jenkins, M. Dopita, E. Sadler, and L. Binette. “Ionized gas in elliptical and S0 galaxies. I-A survey for H-alpha and forbidden N II emission”. In: *The Astronomical Journal* 91 (1986), pp. 1062–1085.

[PL97] R. Paget and D. Longstaff. “Extracting the cliques from a neighbourhood system”. In: *IEE Proceedings-Vision, Image and Signal Processing* 144.3 (1997), pp. 168–170.

[PPZ17] T. A. R. Passarin, D. R. Pipa, and M. V. W. Zibetti. “A minimax dictionary expansion for sparse continuous reconstruction”. In: *Proceedings of the European Signal Processing Conference, EUSIPCO*. 2017, pp. 2136–2140.

[PRJ59] R. V. Pound and Glen A. Rebka J. “Gravitational red-shift in nuclear resonance”. In: *Physical Review Letters* 3.9 (1959), p. 439.

[PRK93] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition”. In: *Proceedings of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*. IEEE. 1993, pp. 40–44.

[PVM93] B. Phookun, S. N. Vogel, and L. G. Mundy. “NGC 4254: a spiral galaxy with an m= 1 mode and infalling gas”. In: *The Astrophysical Journal* 418 (1993), pp. 113–122.

[RBE10] R. Rubinstein, A. M. Bruckstein, and M. Elad. “Dictionaries for sparse representation modeling”. In: *Proceedings of the IEEE* 98.6 (2010), pp. 1045–1057.

[Rei67] C. H. Reinsch. “Smoothing by spline functions”. In: *Numerische mathematik* 10.3 (1967), pp. 177–183.

[RFA03] S. G. Razul, W. Fitzgerald, and C. Andrieu. “Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 497.2-3 (2003), pp. 492–510.

[RGJ05] B. Rivet, L. Girin, and C. Jutten. “Solving the indeterminations of blind source separation of convolutive speech mixtures”. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. Vol. 5. IEEE. 2005, pp. 533–536.

[RK89] R. Roy and T. Kailath. “ESPRIT-estimation of signal parameters via rotational invariance techniques”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37.7 (1989), pp. 984–995.

[RPL13] I. Ramírez, C. Prieto, and D. Lambert. “Oxygen abundances in nearby FGK stars and the galactic chemical evolution of the local disk and halo”. In: *The Astrophysical Journal* 764.1 (2013), p. 78.

[RS09] G. Rath and A. Sahoo. “A comparative study of some greedy pursuit algorithms for sparse approximation”. In: *Proceedings of the European Signal Processing Conference, EUSIPCO*. IEEE. 2009, pp. 398–402.

[Rue+06] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers. “Diffeomorphic registration using B-splines”. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2006, pp. 702–709.

[Sad+05] A. Sadezky, H. Muckenhuber, H. Grothe, R. Niessner, and U. Pöschl. “Raman microspectroscopy of soot and related carbonaceous materials: spectral analysis and structural information”. In: *Carbon* 43.8 (2005), pp. 1731–1742.

[SBN04] A. Stolow, A. E. Bragg, and D. M. Neumark. “Femtosecond time-resolved photoelectron spectroscopy”. In: *Chemical reviews* 104.4 (2004), pp. 1719–1758.

[SC12] B. L. Sturm and M. G. Christensen. “Comparison of orthogonal matching pursuit implementations”. In: *Proceedings of the European Signal Processing Conference, EUSIPCO*. IEEE. 2012, pp. 220–224.

[Sch46] I. J. Schoenberg. “Contributions to the problem of approximation of equidistant data by analytic functions. Part B. On the problem of osculatory interpolation. A second class of analytic approximation formulae”. In: *Quarterly of Applied Mathematics* 4.2 (1946), pp. 112–141.

[Sch86] R. Schmidt. “Multiple emitter location and signal parameter estimation”. In: *IEEE transactions on antennas and propagation* 34.3 (1986), pp. 276–280.

[Sou+11] C. Soussen, J. Idier, D. Brie, and J. Duan. “From Bernoulli–Gaussian deconvolution to sparse signal restoration”. In: *IEEE Transactions on Signal Processing* 59.10 (2011), pp. 4572–4584.

[Tan+12] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. “Compressive sensing off the grid”. In: *Proceedings of Allerton Conference on Communication, Control, and Computing*. 2012, pp. 778–785.

[Ten11] J. Tennyson. *Astronomical Spectroscopy: An Introduction to the Atomic and Molecular Physics of Astronomical Spectra Second Edition*. World Scientific Publishing Company, 2011.

[TGS06] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit”. In: *Signal Processing* 86.3 (2006), pp. 572–588.

[Thu+03] P. Thueler, I. Charvet, F. P. Bevilacqua, M. Saint Ghislain, G. Ory, P. Marquet, P. Meda, B. Vermeulen, and C. D. Depeursinge. “In vivo endoscopic tissue diagnostics based on spectroscopic absorption, scattering, and phase function properties”. In: *Journal of biomedical optics* 8.3 (2003), pp. 495–504.

[TP14] A. M. Tillmann and M. E. Pfetsch. “The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing”. In: *IEEE Transactions on Information Theory* 60.2 (2014), pp. 1248–1259.

[Tro04] J. A. Tropp. “Greed is good: Algorithmic results for sparse approximation”. In: *IEEE Transactions on Information theory* 50.10 (2004), pp. 2231–2242.

[TVP96] S. Talwar, M. Viberg, and A. Paulraj. "Blind separation of synchronous co-channel digital signals using an antenna array. I. Algorithms". In: *IEEE Transactions on Signal Processing* 44.5 (1996), pp. 1184–1197.

[UB75] T. J. Ulrych and T. N. Bishop. "Maximum entropy spectral analysis and autoregressive decomposition". In: *Reviews of Geophysics* 13.1 (1975), pp. 183–200.

[Uns99] M. Unser. "Splines: A perfect fit for signal and image processing". In: *IEEE Signal processing magazine* 16.6 (1999), pp. 22–38.

[VB13] D. Vats and R. Baraniuk. "When in doubt, SWAP: High-dimensional sparse recovery from correlated measurements". In: *Proceedings of the conference on Neural Information Processing Systems, NIPS*. 2013, pp. 989–997.

[VC14] E. Villeneuve and H. Carfantan. "Nonlinear deconvolution of hyperspectral data with MCMC for studying the kinematics of galaxies". In: *IEEE Transactions on Image Processing* 23.10 (2014), pp. 4322–4335.

[VDM+01] R. Van D. M., A. Doucet, N. De Freitas, and E. A. Wan. "The unscented particle filter". In: *Proceedings of the conference on Advances in Neural Information Processing Systems, NIPS*. 2001, pp. 584–590.

[Vin+14] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot. "From blind to guided audio source separation: how models and side information can improve the separation of sound". In: *IEEE Signal Processing Magazine* 31.3 (2014), pp. 107–115. ISSN: 1053-5888.

[VZ16] N. Vaswani and J. Zhan. "Recursive recovery of sparse signal sequences from compressive measurements: A review". In: *IEEE Transactions on Signal Processing* 64.13 (2016), pp. 3523–3549.

[Woj10] M. Wojdyr. "Fityk: a general-purpose peak fitting program". In: *Journal of Applied Crystallography* 43.5-1 (2010), pp. 1126–1128.

[YDD09] M. Yaghoobi, L. Daudet, and M. E. Davies. "Parametric Dictionary Design for Sparse Coding". In: *IEEE Transactions on Signal Processing* 57.12 (2009), pp. 4800–4810.

[YR04] Ö. Yilmaz and S. Rickard. "Blind separation of speech mixtures via time-frequency masking". In: *IEEE Transactions on Signal Processing* 52.7 (2004), pp. 1830–1847.

[YWD15] M. Yaghoobi, D. Wu, and M. E. Davies. "Fast non-negative orthogonal matching pursuit". In: *IEEE Signal Processing Letters* 22.9 (2015), pp. 1229–1233.

[YX16] Z. Yang and L. Xie. "Exact joint sparse frequency recovery via optimization methods". In: *IEEE Transactions on Signal Processing* 64.19 (2016), pp. 5145–5157.

[YZ11] J. Yang and Yin Zhang. "Alternating direction algorithms for ℓ_1-problems in compressive sensing". In: *SIAM journal on scientific computing* 33.1 (2011), pp. 250–278.

[ZC07] R. Zdunek and A. Cichocki. "Nonnegative matrix factorization with constrained second-order optimization". In: *Signal Processing* 87.8 (2007), pp. 1904–1916.

[Zew88] A. H. Zewail. "Laser femtochemistry". In: *Science* 242.4886 (1988), pp. 1645–1653.

[Zha+15] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang. "A survey of sparse representation: Algorithms and applications". In: *IEEE Access* 3 (2015), pp. 490–530.

[Zha10] C.-H. Zhang. “Nearly unbiased variable selection under minimax concave penalty”. In: *The Annals of statistics* 38.2 (2010), pp. 894–942.

[ZY17] W. Zhang and F. Yu. “Off-the-grid compressive time delay estimation via manifold-based optimization”. In: *IEEE Communications Letters* 21.5 (2017), pp. 983–986.

[Rak11] A. Rakotomamonjy. “Surveying and comparing simultaneous sparse approximation (or group-Lasso) algorithms”. In: *Signal processing* 91.7 (2011), pp. 1505–1526.

Hassan MORTADA

Separation of parameterized and delayed sources: application to spectroscopic and multispectral data

Résumé – Ce travail est motivé par la spectroscopie de photoélectrons et l'étude de la cinématique des galaxies où les données correspondent respectivement à une sequence temporelle de spectres et à une image multispectrale. L'objectif est d'estimer les caractéristiques (amplitude, position spectrale et paramètre de forme) des raies présentes dans les spectres, ainsi que leur évolution au sein des données. Dans les applications considérées, cette évolution est lente puisque deux spectres voisins sont souvent très similaires : c'est une connaissance *a priori* qui sera prise en compte dans les méthodes developpées. Ce problème inverse est abordé sous l'angle de la séparation de sources retardées, où les spectres et les raies sont attribués respectivement aux mélanges et aux sources. Les méthodes de l'état de l'art sont inadéquates car elles supposent la décorrélation ou l'indépendance des sources, ce qui n'est pas le cas. Nous tirons partie de la connaissance des sources pour les modéliser par une fonction paramétrique. Nous proposons une première méthode de moindres carrés alternés : les paramètres de formes sont estimés avec l'algorithme de Levenberg-Marquardt, tandis que les amplitudes et les positions sont estimées avec un algorithme inspiré d'Orthogonal Matching Pursuit. Une deuxième méthode introduit un terme de régularisation pour prendre en compte l'évolution lente des positions; un nouvel algorithme d'approximation parcimonieuse conjointe est alors proposée. Enfin, une troisième méthode contraint l'évolution des amplitudes, positions et paramètres de forme par des fonctions B-splines afin de garantir une évolution lente conforme à la physiques des phénomène observés. Les points de contrôle des B-splines sont estimés par un algorithme de moindre carrés non-linéaires. Les résultats sur des données synthétiques et réelles montrent que les méthodes proposées sont plus efficaces que les méthodes de l'état de l'art et aussi efficaces qu'une méthode bayésienne adaptée au problème mais avec un temps de calcul sensiblement réduit.

Mots-clés : séparation de source retardées, mélange anéchoïque, approximation parcimonieuse, décompostion de spectres, images multispectrales, B-splines

Abstract – This work is motivated by photoelectron spectroscopy and the study of galaxy kinematics where data respectively correspond to a temporal sequence of spectra and a multispectral image. The objective is to estimate the characteristics (amplitude, spectral position and shape) of peaks embedded in the spectra, but also their evolution within the data. In the considered applications, this evolution is slow since two neighbor spectra are often very similar: this *a priori* knowledge that will be taken into account in the developed methods. This inverse problem is approached as a delayed source separation problem where spectra and peaks are respectively associated with mixtures and sources. The state-of-the-art methods are inadequate because they suppose the source decorrelation and independence, which is not the case. We take advantage of the source knowledge in order to model them by a parameterized function. We first propose an alternating least squares method: the shape parameters are estimated with the Levenberg-Marquardt algorithm, whilst the amplitudes and positions are estimated with an algorithm inspired from Orthogonal Matching Pursuit. A second method introduces a regularization term to consider the delay slow evolution; a new joint sparse approximation algorithm is thus proposed. Finally, a third method constrains the evolution of the amplitudes, positions and shape parameters by B-spline functions to guarantee their slow evolution. The B-spline control points are estimated with a non-linear least squares algorithm. The results on synthetic and real data show that the proposed methods are more effective than state-of-the-art methods and as effective as a Bayesian method which is adapted to the problem. Moreover, the proposed methods are significantly faster.

Keywords : delayed source separation, anechoic mixing, sparse approximation, spectra decomposition, multispectral images, B-splines