Ecole Doctorale SMPC

Thèse présentée pour l'obtention du diplôme de
Docteur de Télécom & Management SudParis

Doctorat conjoint Télécom & Management SudParis -
Université Pierre et Marie Curie

Spécialité : Mathématiques appliquées

par

Selwa RAFI

Chaînes de Markov cachées et séparation non supervisée de sources

soutenue le 11 Juin 2012

devant le jury composé de :

M. MOHAMMAD-DJAFARI Ali , *DR CNRS* (Rapporteur)	Supelec
M. JUTTEN Christian , *Prof.* (Rapporteur)	Université Grenoble I
M. COMON Pierre , *DR CNRS* (Examinateur)	I3S, Sophia-Antipolis
M. DEVILLE Yannick , *Prof.* (Président)	Université Toulouse3
M. BRONIATOWSI Michel , *Prof.* (Examinateur)	Université Paris6
M. CASTELLA Marc, *MdC.* (Co-encadrant)	Télécom SudParis
M. PIECZYNSKI Wojciech, *Prof.* (Directeur de thèse)	Télécom SudParis

Thèse n^o : 2012TELE0020

à toute ma famille.

« Cherchez le Savoir , du berceau jusqu'au tombeau »,
le Prophète Mahomet

Remerciements

Je tiens d'abord à exprimer ma gratitude à M. Wojciech PIECZYNSKI, mon directeur de thèse, pour son encadrement, ses nombreux conseils qui ont été une aide très précieuse tout au long de ces années.

J'exprime également mon énorme reconnaissance a M. Marc CASTELLA pour son encadrement, sa sympathie, sa très grande disponibilité, ses aides et ses conseils tout au long de ces années de thèse.

J'adresse ensuite mes remerciements aux membres du jury pour l'intérêt qu'ils ont porté à mes travaux. Je remercie tout particulièrement M. Christian JUTTEN et M. Ali MOHAMMAD-DJAFARI pour avoir accepté d'être les rapporteurs de ce travail et pour les remarques avisées. Je suis très honorée aussi de la participation de M. Yannick DEVILLE, M. Pierre COMON et M. Michel BRONIATOWSI.

Je remercie toutes les personnes que j'ai pu côtoyer à Télécom SudParis, pour leur sympathie. En particulier, Ihssan mon précieux cadeau du destin.

Quoi que je dise, Quoi que je fasse, je ne saurai jamais remercier les très chères personnes à mon cœurs. A ceux qui ni les mots , ni les gestes, ni rien au monde pourra exprimer mes sentiments envers eux, mes parents ainsi que mes frères et sœurs (Idriss, Latifa, Hanane, Youssef et Morad qui n'a pas hésité à venir de loin). Tous m'ont soutenu , et n'ont jamais douté de moi, Merci.

Table des matières

Symboles et Notations

T	Transposée
$\sim$	Distribution suivie par une variable aléatoire
i	Indice d'états
t	Indice temporel
T	Taille de l'échantillon
K	Nombre d'états du processus caché
M	Nombre d'états du processus auxiliaire
Q	Taille du vecteur des sources
N	Taille du vecteur des observations
$\mathbf{s}_t$	Vecteur source à l'instant t de taille $Q \times 1$
$\mathbf{x}_t$	Vecteur observations à l'instant t taille $N \times 1$
$\mathbf{r}_t$	Processus auxiliaire à l'instant t
$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_T]$	Vecteur des sources ou processus caché
$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]$	Processus observé
$\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_T]$	Processus auxiliaire
$\mathbf{A}$	Matrice de mélange

Ω	Espace des états cachés
Λ	Espace des états auxiliaires
$\mathbf{Z}$	Processus joint de $\mathbf{S}$ et $\mathbf{X}$
$\mathbf{V}$	Processus joint de $\mathbf{S}$ et $\mathbf{R}$
$\mathbf{Y}$	Processus joint de $\mathbf{S}$, $\mathbf{R}$ et $\mathbf{X}$
$\mathcal{N}(\mu, \sigma)$	Loi gaussienne de moyenne μ et de variance σ^2
$\mathcal{L}(\lambda)$	Distribution scalaire de Laplace de paramètre λ
EM	Espérance Maximisation
SEM	Stochastic Expectation Maximization
MPM	Marginal *posterior* Mode
ICE	Iterative Conditional Estimation
CMC	Chaîne de Markov Cachée
CMCouple	Chaîne de Markov Couple
CMT	Chaîne de Markov triplet
ACI	Analyse en Composantes Indépendantes
EQM	Erreur Quadratique Moyenne
$1_{x \neq y}$	Fonction indicatrice $1_{x \neq y} = \begin{cases} 1 & \text{si} \quad x \neq y \\ 0 & \text{sinon} \end{cases}$

Introduction

Cette thèse a été financée par une bourse de l'Institut Télécom/Télécom SudParis et s'est déroulée au sein du département Communication, Image et Traitement de l'Information (CITI). Ce projet de recherche s'inscrit également dans le cadre du laboratoire SAMOVAR, UMR-CNRS 5157. Le travail qui a été effectué concerne la résolution du problème de restauration qui consiste à estimer un ensemble de données originales à partir d'un ensemble de données dont nous disposons. Ce problème est rencontré dans différents domaines, notamment en traitement du signal et de l'image, en biologie ou encore en finance. En effet, des données multidimensionnelles subissent une transformation (un mélange) inconnue, et donnent naissance à des données observées. La récupération des données originales à partir des données observées peut être vue comme un problème de séparation aveugle de sources, dans lequel on s'intéresserait à estimer l'opérateur de transformation (de mélange). Une solution connue dans ce cadre est l'Analyse en Composantes Indépendantes (ACI). Celle-ci suppose que les sources originales sont indépendantes et cherche à les retrouver de telle façon que les sources estimées soient indépendantes elles aussi. Une extension du modèle classique d'analyse en composantes indépendantes sera proposé dans ce travail. Il permettra d'élargir le champ d'applicabilité du modèle classique au cas où les données pourraient aussi être dépendantes.

D'autre part, en présence de bruit la restauration se complique de manière notable. Les méthodes classiques fondées sur l'analyse en composantes indépendantes ne permettent pas la modélisation explicite du bruit. Une solution, en particulier dans le cas de données à valeurs discrètes, consiste en la segmentation bayésienne. Celle-ci permet la prise en compte explicite du

bruit, et la relaxation de toute hypothèse sur la nature du mélange et sur les dépendances mutuelles entre données dans ce cadre bien précis. Dans ce document, nous allons utiliser les modèles développés récemment, Chaînes de Markov Couples (CMCouples) et Chaîne de Markov Triplets (CMT) qui ont prouvé leur grande efficacité pour les signaux monodimensionnels. Nous allons mettre en relief leur efficacité dans le problème de séparation de données multidimensionnelles à valeurs discrètes.

Organisation du document

Le document est organisé de la manière suivante : le premier chapitre est consacré aux généralités. Nous y introduisons les problématiques traitées dans nos études, et nous y rappellerons les principaux modèles et méthodes qui seront utilisés par la suite. Ce chapitre est divisé en trois parties distinctes : dans la première partie nous présentons la problématique de la séparation de sources, nous présentons brièvement les différents modèles de mélange. Nous nous arrêterons ensuite sur le modèle d'analyse en composantes indépendantes. Dans la deuxième partie de ce chapitre, nous définissons la segmentation bayésienne. Et dans la troisième partie nous ferons une liaison entre le problème de segmentation et celui de la séparation de sources. Dans ce cadre, nous mettons en relief les méthodes fondées sur les modèles de Markov. En particulier, nous y détaillerons la modélisation par les chaînes de Markov cachées et les algorithmes qui y interviennent.

Dans le deuxième chapitre, nous passerons aux généralisations du modèle des chaînes de Markov cachées. En effet, des modèles dits chaînes de Markov Couples et chaînes de Markov Triplets ont été récemment proposés. Grâce à leur richesse, ces modèles permettent une modélisation plus adéquate des données dont les dépendances mutuelles, temporelles et celles du bruit sont complexes. Nous détaillerons ces deux modèles ainsi que toutes les méthodes nécessaires à la segmentation fondée sur ces modèles.

Dans le troisième chapitre, les modèles présentés dans le chapitre précédent seront expérimentés dans le cadre de la séparation de sources. En effet, nous montrerons que la segmentation peut être utilisée efficacement pour la séparation des données discrètes. Des exemples concrets avec des images

synthétiques et réelles seront présentés.

Dans le quatrième chapitre, nous reviendrons sur le modèle d'analyse en composantes indépendantes. Nous présenterons une nouvelle méthode qui permet de surmonter la restriction de l'hypothèse d'indépendance supposée par ce modèle. Un processus stochastique est introduit pour gérer la dépendance/indépendance mutuelles des sources. Enfin le chapitre 5 présente les conclusions de ce travail.

Ce travail de thèse a donné lieu aux publications suivantes :

Articles de conférences :

- S. Rafi, M. Castella and W. Pieczynski, **"Pairwise Markov model applied to unsupervised image separation"**, International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA), Innsbruck, Austria, 16-18 February 2011.

- S. Rafi, M. Castella and W. Pieczynski, **"An extension of the ICA model using latent variable"**, *Proc. of ICASSP*, pp.3712-3715, 22-27, Prague, Czech Republic, May 2011.

- S. Rafi, M. Castella and W. Pieczynski, **"Une extension avec variables latentes du modèle d'analyse en composantes indépendantes"**, *Actes XXème Colloque GRETSI*, Bordeaux, France, September 2011.

Articles journaux :

- M. Castella, S. Rafi, P. Comon and W. Pieczynski, **"Separation of instantaneous mixtures of dependent sources using classical ICA methods"**, *EURASIP Journal on Advances in Signal Processing*, submitted.

Chapitre 1

Contexte et généralités

1.1 Séparation de sources

Dans cette partie, nous considérons le problème de séparation de sources [1, 2]. C'est un domaine de recherche actif au cours des dernières décennies. Il suscite un grand intérêt grâce à ses diverses applications telles que les télécommunications, le biomédical [3–5], l'acoustique [6–11] ou encore le traitement de la parole [7, 12, 13] ou de l'image [14, 15]. Son principe consiste à retrouver la configuration initiale de données inconnues appelées *sources* subissant une transformation, et donnant naissance à des observées appelées *observations*.

L'illustration classique de la séparation de sources est le problème dit "cocktail party problem" (figure 1.1). Imaginons que, lors d'une soirée, on dispose N capteurs dans une salle, où N personnes discutent par groupes de tailles diverses. Chaque capteur enregistre la superposition des discours des personnes à ses alentours et le problème consiste à retrouver la voix de chaque personne en éliminant l'effet des autres voix. Les premiers travaux qui se sont intéressés à la résolution de ce problème remontent aux années 80 [16, 17]. Ces travaux visaient à analyser le mouvement chez les vertébrés [18] dans le cadre des réseaux de neurones. Ensuite, la formalisme mathématique du problème de la séparation de sources a donné naissance aux premiers algorithmes permettant d'avoir une solution [19, 20]. La notion d'*Analyse en Composantes Indépendantes* [21, 22] a formalisé mathématiquement le

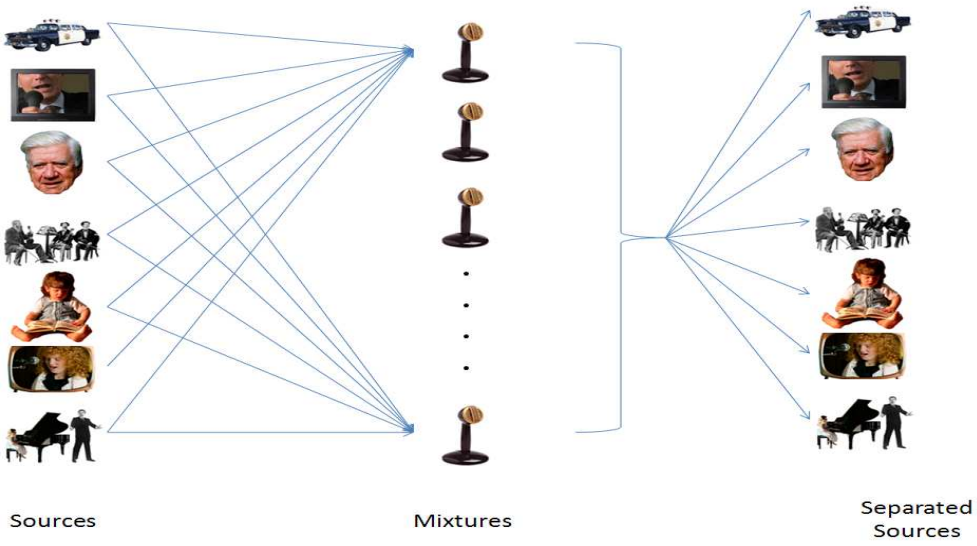


FIGURE 1.1 – Illustration du "cocktail party problem"

problème dans le cadre des mélanges linéaires.

De façon générale le problème de séparation de sources est décrit de la manière suivante (voir figure 1.2) : nous notons $\mathbf{s}_t = [s_t^1, \ldots, s_t^Q]^{\mathrm{T}}$ le vecteur contenant les Q sources à un instant donné t, et de façon similaire $\mathbf{x}_t = [x_t^1, \ldots, x_t^N]^{\mathrm{T}}$ le vecteur des N observations à l'instant t. On suppose disponibles T échantillons des vecteurs d'observations (de taille N) ; ce que l'on note $\mathbf{X} = (\mathbf{x}_{1\}, \ldots, \mathbf{x}_T)}$. Ces données observées sont issues de mélange de T échantillons (*sources*) inconnus de taille Q notés $\mathbf{S} = (\mathbf{s}_t, \ldots, \mathbf{s}_T)$. Retrouver les sources inconnues $\mathbf{s}_t$ à partir des observations $\mathbf{x}_t$ est l'objectif des méthodes de **séparation aveugle des sources**. Sachant que les observations $\mathbf{X}$ sont un mélange (linéaire ou non) des sources $\mathbf{S}$ par un système $\mathcal{A}$, la restauration de $\mathbf{S}$ peut être réalisée si nous arrivons à estimer un opérateur de séparation , que nous notons $\mathcal{B}$. En d'autres termes, soit l'équation du mélange :

$$\mathbf{X} = \mathcal{A}(\mathbf{S}) \tag{1.1}$$

où $\mathcal{A}$ est un opérateur inconnu. Typiquement, la séparation est réalisée dès que l'on peut estimer $\mathcal{B}$.

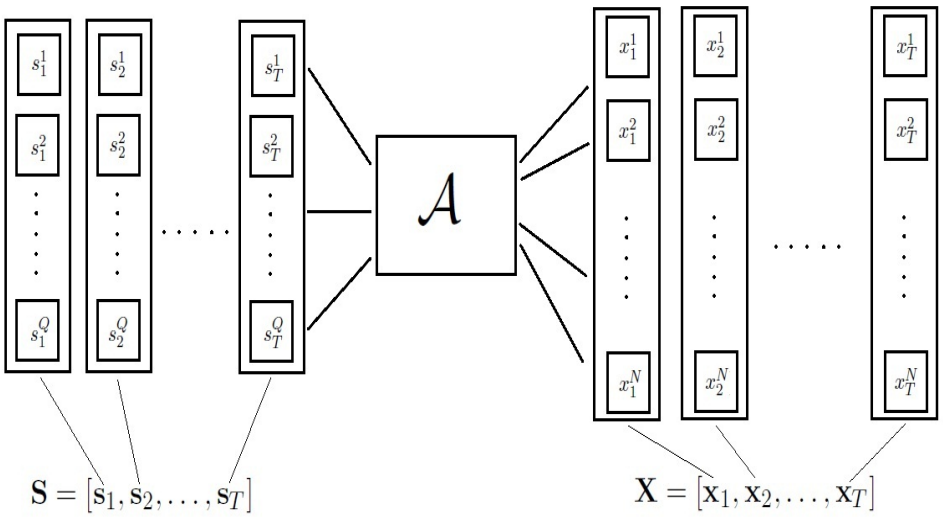


FIGURE 1.2 – Séparation de sources

Nous parlons, par définition, de séparation "aveugle" de sources lorsque seules les données observées sont à disposition. Bien que le terme "aveugle" soit utilisé, des hypothèses sur le mélange sont tout de même considérées, en particulier, sa structure. La prise en compte des connaissances *a priori* et d'hypothèses supplémentaires permet de mieux définir le problème ; celles-ci permettent de mieux décrire le modèle du mélange ainsi que les caractéristiques des sources.

Un grand nombre de travaux a étudié la problématique de la séparation de sources dans différents contextes allant des mélanges linéaires instantanés [23–26] jusqu'aux mélanges linéaires convolutifs [27–33] ou encore les mélanges non linéaires [17, 34, 35]. Ces études ont abouti à des méthodes qui balayent un champ d'applications variées. Des algorithmes ont été développés en fonction du modèle de mélange et de la nature des signaux.

Le traitement du problème de séparation de sources est différent selon la nature du mélange. La méthode de séparation doit être adaptée à la nature du mélange déterminée par la structure de l'opérateur $\mathcal{A}$. En effet, nous distinguons les mélanges *linéaires* dont l'opérateur $\mathcal{A}$ est une matrice, des mélange *non-linéaires* pour lesquels la forme de $\mathcal{A}$ est quelconque. D'autre

part, dans le cadre linéaire, nous distinguons les mélanges instantanés des mélanges convolutifs [27–33].

1.1.1 Analyse en Composantes Indépendantes

Au début des années 90, le problème de séparation aveugle de sources a été formalisé dans le cas des mélanges linéaires instantanés ; c'est-à-dire dans le cas où le système $\mathcal{A}$ est un opérateur linéaire correspondant à la multiplication par une matrice $\mathbf{A}$. Les premières contributions ont donné naissance au modèle d'Analyse en Composantes Indépendantes. Ce modèle est, en fait, une méthode d'analyse de données permettant d'exprimer un vecteur multidimensionnel en terme de composantes statistiquement indépendantes [36]. Dans le contexte de la séparation aveugle de sources, l'analyse en composantes indépendantes peut se définir comme l'ensemble des méthodes de séparation de sources basées sur l'hypothèse d'indépendance [21, 22]. Celle-ci permet l'estimation de la matrice de séparation $\mathbf{B} = \mathbf{A}^{-1}$ qui maximise un critère d'indépendance [23–26]. L'équation de mélange spécifique aux mélanges linéaires est la suivante :

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t \tag{1.2}$$

telle que,

- $\mathbf{A}$ est une matrice fixée
- pour tout t, $\mathbf{s}_t$ est un vecteur aléatoire à composantes statistiquement indépendantes
- toutes les composantes du vecteur sources $\mathbf{s}_t$ sont non gaussiennes sauf éventuellement une [37].

Dès lors que ces hypothèses sont vérifiées, la séparation est théoriquement possible [21, 37]. Dans le cadre aveugle, la matrice $\mathbf{A}$ est inconnue. D'autre part, lorsque $\mathbf{A}$ est de taille $N \times Q$ nous pouvons alors distinguer trois types de mélanges selon les valeurs de N et Q :

– ***Mélanges déterminés*** $(N = Q)$: un mélange pour lequel il y a autant de sources que de capteurs ; pour la séparation d'un tel mélange, identifier la matrice de mélange $\mathbf{A}$ ou son inverse $\mathbf{B}$ fournit directement

les sources à condition que la matrice **A** soit inversible (régulière).

- ***Mélanges sous-déterminés*** $(N < Q)$: Ici, le nombre de capteurs N est inférieur au nombre de sources Q. Dans ce cas, la matrice de mélange **A** est non inversible à gauche, et même en ayant **A** la déduction des sources **s** n'est pas simple. Nous avons alors un problème à double complexité : d'abord il faut identifier la matrice de mélange **A** et ensuite il faut restaurer les sources **s**. Sans *a priori* suffisamment informatif, la solution est impossible. Dans le cas de mélange linéaire, lorsque la matrice **A** est connue, le problème se résout à un vecteur aléatoire près [38]. Si la matrice **A** est inconnue et les sources sont discrètes ou parcimonieuses, une solution est possible.

- ***Mélanges sur-déterminés*** $(N > Q)$: dans un mélange sur-déterminé le nombre d'observations N est supérieur aux nombres de sources Q. Dans ce cas, la séparation peut se faire en se ramenant au cas déterminé. En effet, à condition que la matrice de mélange soit de rang plein, on peut réduire la dimension des observations à l'aide d'un blanchiment ou d'une analyse en composantes principales.

Dans ce travail, nous allons nous limiter au cas déterminé $N = Q$.

Dans le cadre aveugle, des indéterminations de permutations et de facteurs d'échelles sont indissociables de la séparation par les méthodes d'analyse en composantes indépendantes. à cause de ces indéterminations, l'estimation de la puissance et de l'ordre des sources est impossible. Toutefois, des contraintes imposées permettent de remédier à ces ambiguïtés : on peut imposer l'unicité des éléments diagonaux de la matrice du mélange **A** ou fixer arbitrairement la puissance des sources estimées par normalisation ou pré-blanchiment.

En se basant sur le modèle d'analyse en composantes indépendantes, de nombreux algorithmes ont été proposés [33]. Nous citons, l'algorithme JADE [39] dont le nom vient de l'acronyme anglais de "Joint Approximation Diagonalization of Eigenmatrices". Son principe est basé sur la diagonalisation d'un tenseur de cumulants d'ordre quatre. En effet, la séparation est atteinte lorsque le critère d'indépendance est maximisé c'est-à-dire lorsque

les éléments non diagonaux du tenseur tendent vers zéro. Nous citons aussi l'algorithme CoM2 qui maximise une fonction de contraste basée sur les cumulants d'ordre quatre [21] : le contraste dans cet algorithme correspond à la somme des modules carrés des cumulants marginaux des sorties. Aussi, l'algorithme Fast-ICA a été ensuite proposé [40].

1.1.2 Mélanges bruités

Plus réalistes que les modèles sans bruit, les mélanges bruités sont les plus souvent rencontrés dans les applications réelles. Soit, $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_T)$ l'ensemble des échantillons du bruit, l'équation du mélange bruité s'écrit de la manière suivante :

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{b}_t \tag{1.3}$$

La présence du bruit complique notablement le traitement des données. Les modèles classiques d'analyse en composantes indépendantes ne prennent pas en compte la présence du bruit. Cependant, l'approche Bayésienne a été proposée comme une alternative aux méthodes classiques d'analyse en composantes indépendantes [41–52]. Elle permet de prendre en compte du bruit additif explicitement dans le modèle, en exploitant les informations *a priori* sur les paramètres. En effet, une distribution est associée à chaque paramètre, et les probabilités *a posteriori* sont alors déduites grâce à la règle de Bayes. Les paramètres des distributions de chaque élément à estimer sont obtenus par maximisation des lois *a posteriori* [42, 43, 53].

Dans le paragraphe suivant, nous allons nous intéresser particulièrement au problème de séparation de sources discrètes c'est-à-dire des sources dont les composantes prennent leurs valeurs dans d'un ensemble fini d'état. Ce problème sera considéré dans un contexte général : sans tenir compte de la nature du mélange ni de la structure de dépendance mutuelle entre sources. En effet, nous allons voir dans ce qui suit que la séparation des sources discrète peut être vue comme un problème de segmentation.

1.2 La segmentation

La segmentation d'un signal mono-dimensionnel ou d'une image bidimensionnelle est la technique permettant de diviser ce signal ou cette image en un nombre fini de zones homogènes, souvent appelées classes. La segmentation statistique consiste à rechercher ces zones à partir du signal (ou de l'image) observé(e), ce qui revient à estimer les composantes d'un processus aléatoire *caché* $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_T)$, à partir d'un processus *observé* $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$. On suppose que chaque composante de $\mathbf{S}$ prend ses valeurs dans un ensemble d'états $\Omega = \{\omega_1, \ldots, \omega_K\}$. Les liens entre les processus $\mathbf{S}$ et $\mathbf{X}$ sont modélisés par la probabilité jointe $p(\mathbf{S}, \mathbf{X})$. Dans les approches classiques cette loi peut s'écrire

$$p(\mathbf{S}, \mathbf{X}) = p(\mathbf{S})p(\mathbf{X}|\mathbf{S})$$

$p(\mathbf{S})$ est dite loi *a priori*, elle modélise les liens entre les données cachées, tandis que $p(\mathbf{X}|\mathbf{S})$, appelée parfois loi d'attache aux données, modélise le bruit. L'étape clé dans une segmentation statistique est celle du calcul des diverses quantités liées à la loi *a posteriori* $p(\mathbf{S}|\mathbf{X})$. La segmentation est alors effectuée dès que ce calcul est possible avec une complexité raisonnable.

En réalité, le problème de la segmentation est complexe. Le calcul de la loi *a posteriori* n'est pas toujours possible ; en particulier, il peut présenter une complexité trop grande dans les modèles trop généraux. La difficulté du calcul survient lorsque la taille de l'échantillon augmente. Dans les cas des échantillons de grande taille une solution, qui s'est souvent montrée efficace, consiste à utiliser les modèles de chaînes de Markov cachées qui d'une part, tiennent compte des dépendances entre composantes, et d'autre part permettent le calcul exact des lois *a posteriori*.

Le modèle le plus simple des chaînes de Markov est le modèle de chaîne de Markov cachée à bruit indépendant, que l'on notera CMC-BI, dans lequel on considère la markovianité du processus $\mathbf{S}$. Ce modèle a été ensuite généralisé aux chaînes de Markov couples, notées CMCouples, dans lesquelles on considère la markovianité du couple $(\mathbf{S}, \mathbf{X})$ [54, 55]. Ces modèles permettent de modéliser des processus à dépendances plus complexes pour lesquelles le modèle CMC-BI peut présenter des faiblesses [54]. Une nouvelle généralisation à été proposée dans [56], qui consiste à introduire un processus auxiliaire noté

$\mathbf{R}$ qui permet, entre autres, de modéliser la non stationnarité des données. Dans ce modèle, appelé chaîne de Markov triplet et noté CMT, le processus triplet $(\mathbf{S}, \mathbf{X}, \mathbf{R})$ est de Markov. Ainsi que nous le verrons par la suite, les CMCouples et les CMT permettent d'étendre des extensions des méthodes bayésiennes utilisées dans les CMC-BI à des fins de segmentation.

La segmentation par des chaînes de Markov peut se faire dans un contexte *supervisé* ou *non supervisé*, c'est-à-dire en connaissant les paramètres du modèle ou non. Dans le cas d'une segmentation non supervisée les paramètres ne sont pas tous connus et on est alors amené à les estimer. Aussi, nous parlons d'estimation en *données complètes* lorsque le couple des processus (caché, observé) sont disponibles, et de *données incomplètes* lorsque nous ne disposons que du processus observé.

Il existe différents algorithmes permettant l'estimation des paramètres. Nous détaillerons plus loin les algorithmes que nous avons utilisés dans notre étude ; à savoir l'algorithme d'Espérance - Maximisation (EM) ou encore des algorithmes récemment proposés tels que l'Estimation Itérative Conditionnelle (ICE), ainsi que la version stochastique de l'algorithme EM (SEM). Toutes ces méthodes seront détaillées dans le cas général ainsi que dans le cas de chaque modèle des chaînes de Markov considéré.

1.3 Séparation et segmentation

Nous visons à établir un lien entre le problème de la séparation de sources et celui de la segmentation. Il est à noter que pour le cas de signaux à états discrets, séparation et segmentation mènent aux mêmes résultats : en effet, séparer les composantes d'un vecteur des observations issues d'un mélange de sources à états discrets revient à estimer la classe à laquelle, à chaque instant t, le vecteur source $\mathbf{s}_t$ appartient ; ce qui revient à faire une segmentation.

Dans cet esprit on propose d'exploiter les techniques de la segmentation bayésienne fondée sur les modèles de chaînes de Markov pour séparer des signaux à nombre fini d'états présentant une structure complexe. Un intérêt est de pouvoir séparer les Q composantes $\mathbf{s}_t^1, \ldots, \mathbf{s}_t^Q$ du vecteur sources $\mathbf{s}_t$ sans qu'elles soient nécessairement statistiquement indépendantes. Ceci permet donc de relâcher l'hypothèse de base des techniques classiques de

la séparation de sources ; en l'occurrence les méthodes de l'analyse en composantes indépendantes qui exigent l'indépendance statistique des sources. Nous nous intéressons particulièrement au problème suivant, soit un vecteur observé $\mathbf{x}_t$ qui résulte d'un mélange de sources $\mathbf{s}_t$ entaché d'un bruit additif tels que :

$$\mathbf{x}_t = \begin{pmatrix} x^1 \\ \vdots \\ x^N \end{pmatrix} \quad \text{et} \quad \mathbf{s}_t = \begin{pmatrix} s^1 \\ \vdots \\ s^Q \end{pmatrix}$$

A chaque instant t, $\mathbf{s}_t$ appartient à l'ensemble de classes $\Omega = \{\omega_1, \ldots, \omega_K\}$. Les signaux sources $\mathbf{s}_t$ sont mélangés par un opérateur $\mathcal{A}$. Aucune hypothèse n'est faite ni sur les dépendances statistiques entre les composantes $\mathbf{s}_t^1, \ldots, \mathbf{s}_t^Q$ du vecteur $\mathbf{s}_t$, ni sur celles entre les $(\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$ du vecteur $\mathbf{x}_t$. Soit $\mathbf{b}$ le vecteur du bruit de taille $N \times T$. Le modèle s'écrit :

$$\mathbf{x}_t = \mathcal{A}(\mathbf{s}_t) + \mathbf{b}_t. \qquad t \in \{1, \ldots, T\} \tag{1.4}$$

On se place dans le cas d'un mélange quelconque bruité. Nous cherchons donc à retrouver les composantes $\mathbf{s}_t^1, \ldots, \mathbf{s}_t^Q$ à partir des $\mathbf{X} = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$, pour tout $t = 1, \ldots, T$, en utilisant les techniques de segmentation par les modèles de chaînes de Markov. Le choix de modélisation par des chaînes de Markov est motivé par la grande efficacité et la grande robustesse des CMC-BI aussi bien dans le cas de la segmentation des chaînes unidimensionnelles que dans le cas des chaînes N-dimensionnelles [57]. De plus, les généralisations récemment proposées ; en l'occurrence les modèles CMCouple et CMT, ont aussi prouvé leur grande efficacité sur des données à structures temporelles encore plus complexes que celles traitées par un modèle CMC-BI. Comme nous allons le voir par la suite, ces modèles, très riches, permettent de réaliser le calcul exact des différentes quantités, liées aux lois *a posteriori*, qui permettent l'application des méthodes bayésiennes de segmentation. Nous avons ainsi une famille de modèles qui permet de décrire de façon plus complète le problème à traiter en choisissant le modèle le plus adapté à chaque situation.

Afin d'illustrer la segmentation bayésienne fondée sur les modèles de chaînes de Markov cachées pour la séparation de sources à nombre d'états

fini, nous proposons, dans ce chapitre, de détailler le modèle ainsi que les techniques de calcul dans le cas simple d'un modèle classique de chaîne de Markov cachée à bruit indépendant. Nous commençons par le cas de la segmentation supervisée, dans laquelle les paramètres de la loi *a priori* des sources et du bruit sont considérés connus. Le signal caché est restauré en utilisant l'estimateur bayésien appelé mode de la marginale *a posteriori* (MPM) que nous détaillerons plus loin. Celui-ci maximise la probabilité marginale *a posteriori* en chaque site de la chaîne de Markov. La probabilité marginale *a posteriori* dans le cas des chaînes de Markov cachées est obtenue grâce aux procédures itératives "Forward-Backward".

Dans le paragraphe suivant, nous rappelons la définition et les caractéristiques des chaînes de Markov ainsi que des chaînes de Markov cachées.

1.3.1 Chaînes de Markov

Une chaîne de Markov, qui sera supposée dans cette thèse à valeurs dans un ensemble fini, est une suite de variables aléatoires $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_t, \ldots, \mathbf{s}_T)$ dont la loi conditionnelle vérifie :

$$p(\mathbf{s}_t|\mathbf{s}_1, \ldots, \mathbf{s}_{t-1}) = p(\mathbf{s}_t \,|\, \mathbf{s}_{t-1}) \quad \forall t = 2, \ldots, T \tag{1.5}$$

Les composantes de la chaîne $\mathbf{S}$ prennent leurs valeurs dans un ensemble d'états au nombre supposé fini et noté $\Omega = \{\omega_1, \ldots, \omega_K\}$. La loi du processus $\mathbf{S}$ est alors déterminée par la loi initiale $\pi = (\pi_1, \ldots, \pi_k)$ et la suite des matrices de transition définies, pour $t = 2, \ldots, T$ par :

$$\pi_i = p(\mathbf{s}_1 = \omega_i)$$

$$a_{ij}^t = p(\mathbf{s}_{t+1} = \omega_i / \mathbf{s}_t = \omega_j);$$

avec $a_{ij}^t \geq 0$ et $\sum_{j \in \Omega} a_{ij}^t = 1 \ \forall t$.
Nous supposerons que

$$c_{ij} = p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j) \tag{1.6}$$

ne dépend pas de t. la chaîne est alors dite "stationnaire". Si a_{ij}^t est de plus indépendante de t : la chaîne est alors dite homogène. Dans le cas stationnaire

et homogène, la loi initiale est donnée par :

$$\pi_i = p(\mathbf{s}_1 = \omega_i) = \sum_{1 \leq j \leq K} c_{ij} \tag{1.7}$$

et les matrices de transition sont données par :

$$a_{ij} = \frac{c_{ij}}{\sum_{1 \leq i \leq K} c_{ij}} \tag{1.8}$$

Dans le cas de chaîne stationnaire, la loi initiale π est un vecteur propre à gauche de la matrice de transition a_{ij} associé à la valeur propre 1, ce qui s'écrit :

$$\pi_i = \sum_{j=1}^{K} \pi_j a_{ij}$$

La loi d'une chaîne $\mathbf{S}$ stationnaire peut ainsi être paramétrée soit par π et $(a_{ij})_{1 \leqslant i,j \leqslant K}$, soit par $(c_{ij})_{1 \leqslant i,j \leqslant K}$. La distribution du processus est alors donnée par :

$$p(\mathbf{s}_1 = \omega_{i_1}, \ldots, \mathbf{s}_T = \omega_{i_T}) = \pi_i \prod_{t=1}^{T-1} a_{i_t i_{t+1}} \tag{1.9}$$

1.3.2 Chaînes de Markov cachées

Une chaîne de Markov cachée est un processus composé de deux processus $\mathbf{S}$ et $\mathbf{X}$ [58]. Le processus $\mathbf{S}$ est une chaîne de Markov et le processus $\mathbf{X}$ sera supposé à valeurs dans $\mathbb{R}^N$. Nous considérons $\mathbf{S}$ à valeurs dans un ensemble fini $\Omega = \{\omega_1, \ldots, \omega_K\}$ et nous nous intéressons à la loi du couple $(\mathbf{S}, \mathbf{X})$. De manière générale cette dernière peut s'écrire

$$p(\mathbf{S}, \mathbf{X}) = p(\mathbf{S}) p(\mathbf{X} \,|\, \mathbf{S}) \tag{1.10}$$

Dans le cas d'une chaîne de Markov cachée, la loi de $\mathbf{S}$ est celle d'une chaîne de Markov telle que définie dans (1.9). Cette loi permet d'introduire les dépendances parmi les $\mathbf{s}_1, \ldots, \mathbf{s}_t$. La loi $p(\mathbf{X} \,|\, \mathbf{S})$ décrit les dépendances entre les observations et les composantes du processus caché. Plus précisément, dans un modèle classique de chaînes de Markov cachées à bruit indépendant, qui sera noté CMC-BI, cette loi vérifie les deux propriétés suivantes :
(i) la loi d'une observation $\mathbf{x}_t$ à l'instant t dépend uniquement de la composante du processus caché à ce même instant, c'est-à-dire nous avons :

$$p(\mathbf{x}_t | \mathbf{s}_1, \ldots, \mathbf{s}_T) = p(\mathbf{x}_t \,|\, \mathbf{s}_t) \tag{1.11}$$

(ii) dans une CMC-BI les observations $\mathbf{x}_t$ sont considérées indépendantes conditionnellement à $\mathbf{s}$. Nous avons :

$$p(\mathbf{X} \,|\, \mathbf{S}) = \prod_{t=1}^{T} p(\mathbf{x}_t \,|\, \mathbf{s}) \tag{1.12}$$

Il s'ensuit finalement que la loi du couple $(\mathbf{S}, \mathbf{X})$ dans le modèle CMC-BI est donnée par :

$$p(\mathbf{S}, \mathbf{X}) = p(\mathbf{s}_1)p(\mathbf{x}_1 \,|\, \mathbf{s}_1) \prod_{t=1}^{T-1} p(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t)p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}). \tag{1.13}$$

Bien que les lois $p(\mathbf{S})$ et $p(\mathbf{x}_t \,|\, \mathbf{s}_t)$ dans une CMC-BI soient simples, ce modèle a prouvé une grande efficacité dans différents domaines d'applications. Il permet très souvent d'estimer les données cachées, ainsi que, éventuellement, des paramètres inconnus du modèle, avec une grande efficacité. L'estimation est réalisée dès que l'on peut calculer les lois $p(\mathbf{S} \,|\, \mathbf{X})$ de $\mathbf{S}$ conditionnellement à $\mathbf{X}$. De manière générale, ces lois sont données à partir de la loi du couple $p(\mathbf{S}, \mathbf{X})$ par

$$p(\mathbf{S} \,|\, \mathbf{X}) = \frac{p(\mathbf{S}, \mathbf{X})}{\sum_{\mathbf{s}_t \in \Omega} p(\mathbf{S}, \mathbf{X})} \tag{1.14}$$

Celles-ci peuvent être calculées à partir de la forme générale (1.10) lorsque le nombre de l'échantillon T est petit. Lorsque T croît, le calcul devient plus complexe, d'où l'intérêt du modèle CMC-BI qui permet le calcul de la loi *a posteriori* pour des échantillons de grande taille grâce à des procédures itératives que nous détaillerons dans le chapitre suivant.

Dans ce qui suit, nous nous limiterons au cas où l'opérateur $\mathcal{A}$ est une matrice carrée notée $\mathbf{A}$, nous avons donc $N = Q$ et, par ailleurs, nous considérons les lois $p(\mathbf{x} \,|\, \mathbf{s})$ gaussiennes c'est à dire que le bruit $\mathbf{b}_t$ est gaussien de loi :

$$p(\mathbf{x}_t \,|\, \mathbf{s}_t) = \frac{1}{(2\pi)^{N/2} |\det \Gamma|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x}_t - \mu)^{\mathrm{T}} \Gamma^{-1} (\mathbf{x}_t - \mu)) \tag{1.15}$$

avec $\mu = \mathbf{A}\mathbf{s}_t$.

1.4 La segmentation par chaînes de Markov cachées

1.4.1 Approche supervisée

Dans une approche supervisée, la segmentation est effectuée en considérant les paramètres du modèle connus.

a Procédure Forward-Backward

La procédure Forward-Backward est une méthode séquentielle de calcul utilisée dans les modèles de Markov cachés. En particulier, elle permet de calculer itérativement les marginales *a posteriori* $p(\mathbf{s}_t = \omega_i \,|\, \mathbf{X})$ des états cachés conditionnellement à toutes les observations $\mathbf{x}_1, \ldots, \mathbf{x}_T$. Ce caractère récursif de la procédure permet le calcul des marginales, même pour les données de très grandes tailles. Considérons les notations classiques suivantes :

$$\alpha_t^*(\omega_i) \triangleq p(\mathbf{s}_t = \omega_i, \mathbf{x}_1, ..., \mathbf{x}_t) \tag{1.16}$$

$$\beta_t^*(\omega_i) \triangleq p(\mathbf{x}_{t+1}, ..., \mathbf{x}_T \,|\, \mathbf{s}_t = \omega_i). \tag{1.17}$$

α_t^* est appelée "probabilité forward" et β_t^* "probabilité Backward" [59, 60]. On montre classiquement que dans une CMC-BI $p(\mathbf{s}_t = \omega_i, \mathbf{X}) = \alpha_t^*(\omega_i)\beta_t^*(\omega_i)$. Ainsi, comme mentionné ci-dessus, les probabilités α_t^* et β_t^* permettent le calcul des marginales *a posteriori* $p(\mathbf{s}_t = \omega_i \,|\, \mathbf{X})$.

Les formules de récurrence pour le calcul des probabilités forward et backward pour le cas simple d'une chaîne de Markov cachée classique à bruit indépendant, sont données par les expressions suivantes [61] :

Forward :

- $\alpha_1^*(\omega_i) = p(\mathbf{s}_1 = \omega_i \,|\, \mathbf{x}_1) = p(\mathbf{x}_1 \,|\, \mathbf{s}_1 = \omega_i)p(\mathbf{s}_1 = \omega_i)$
- pour $t = 1, \ldots, T-1$; calculer :

$$\alpha_{t+1}^*(\omega_j) = \sum_{\omega_i \in \Omega} \alpha_t^*(\omega_i) p(\mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{s}_t = \omega_i) p(\mathbf{x}_t \,|\, \mathbf{s}_t = \omega_i)$$

Backward :

- $\beta_T^*(\omega_i) = 1$

- pour $t = T-1, T-2, \ldots, 1$; calculer :

$$\beta_t^*(\omega_i) = \sum_{\omega_j \in \Omega} \beta_{t+1}^*(\omega_j) p(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t = \omega_i) p(\mathbf{x}_{t+1} | \mathbf{s}_{t+1} = \omega_j)$$

Pour les grands échantillons($\succ$ 1000), le calcul des quantités α^* et β^* se heurte à des difficultés d'ordre numérique.En effet, les quantités figurant dans les expressions de $\alpha_t^*(\omega_i)$ et $\beta_t^*(\omega_i)$ sont très petites et la présence des sommes dans ces expressions ne permet pas d'utiliser le logarithme. Ceci peut être évité en utilisant les expressions normalisées suivantes [62] :

$$\alpha_t(\omega_i) = p(\mathbf{s}_t = \omega_i \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_t) \tag{1.18}$$

$$\beta_t(\omega_i) = \frac{p(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T \,|\, \mathbf{s}_t = \omega_i)}{p(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_t)} \tag{1.19}$$

Les formules de récurrence correspondant aux expressions normalisées des probabilités forward-backward sont les suivantes :

Forward :

- $\alpha_1(\omega_i) = p(\mathbf{s}_1 = \omega_i \,|\, \mathbf{x}_1) = \dfrac{p(\mathbf{x}_1 \,|\, \mathbf{s}_1 = \omega_i) p(\mathbf{s}_1 = \omega_i)}{\sum_{\omega_i \in \Omega} p(\mathbf{x}_1 \,|\, \mathbf{s}_1 = \omega_i) p(\mathbf{s}_1 = \omega_i)}$
- pour $t = 1, \ldots, T-1$; calculer :

$$\alpha_{t+1}(\omega_j) = \frac{p(\mathbf{x}_t \,|\, \mathbf{s}_t = \omega_i) \sum_{\omega_i \in \Omega} \alpha_t(\omega_i) p(\mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{s}_t = \omega_i)}{\sum_{\omega_i, \omega_j \in \Omega^2} p(\mathbf{x}_t \,|\, \mathbf{s}_t = \omega_i) \alpha_t(\omega_i) p(\mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{s}_t = \omega_i)}$$

Backward :

- $\beta_T(\omega_j) = 1$
- pour $t = T-1, T-2, \ldots, 1$; calculer :

$$\beta_t(\omega_i) = \frac{\sum_{\omega j \in \Omega} \beta_{t+1}(\omega_j) p(\mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{s}_t = \omega_i) p(\mathbf{x}_t \,|\, \mathbf{s}_t = \omega_i)}{\sum_{\omega_i, \omega_j \in \Omega^2} \beta_{t+1}(\omega_j) p(\mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{s}_t = \omega_i) p(\mathbf{x}_t \,|\, \mathbf{s}_t = \omega_i)}$$

Nous introduisons des notations abrégées $\zeta_t(\omega_i)$ et $\psi_t(\omega_i, \omega_i)$:

$$\zeta_t(\omega_i) \triangleq p(\mathbf{s}_t = \omega_i \,|\, \mathbf{x})$$

$$\psi_t(\omega_i, \omega_j) \triangleq p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{x})$$

On montre alors que les lois marginales *a posteriori* $p(\mathbf{s}_t = \omega_i \,|\, \mathbf{x})$ et les lois conjointes *a posteriori* $p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{x})$ sont données par les expressions suivantes :

$$\zeta_t(\omega_i) = \alpha_t(\omega_i) \beta_t(\omega_i) \tag{1.20}$$

$$\psi_t(\omega_i,\omega_j) = \frac{\alpha_t(\omega_i)p(\mathbf{s}_{t+1}=\omega_j \,|\, \mathbf{s}_t=\omega_i)p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}=\omega_j)\beta_{t+1}(\omega_i)}{\sum_{\omega_j\in\Omega} p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}=\omega_j) \sum_{\omega_i\in\Omega} \alpha_t(\omega_i)p(\mathbf{s}_{t+1}=\omega_j \,|\, \mathbf{s}_t=\omega_i)} \tag{1.21}$$

b Remarque

La méthode "Mode des Marginales *a posteriori*" (MPM) que nous allons utiliser dans ce travail est une méthode bayésienne, qui minimise une "perte moyenne" correspondant à la fonction de perte L qui pénalise une configuration estimée par rapport à la vraie par le nombre de sites mal classés :

$$\mathrm{L}(\hat{\mathbf{s}}_t, \mathbf{s}_t) = \sum_{t=1}^{T} 1_{\hat{\mathbf{s}}_t \neq \mathbf{s}_t} \tag{1.22}$$

Par définition, l'estimateur bayésien correspondant à cette fonction de perte est obtenu en minimisant l'espérance conditionnelle $\mathbb{E}\{\mathrm{L}(\widehat{\mathbf{s}}(\mathbf{X}), \mathbf{S})/\mathbf{X}\}$. On montre alors classiquement qu'il est bien donné par :

$$\widehat{\mathbf{s}_t} = \mathrm{Arg}\max_{\mathbf{s}_t} p(\mathbf{s}_t \,|\, \mathbf{x}_t) \tag{1.23}$$

L'estimateur MPM maximise ainsi la probabilité marginale *a posteriori* à tout instant. Sa mise en œuvre est possible grâce au fait que ces probabilités sont calculables par les procédures "Forward-Backward".

1.4.2 Estimation des paramètres

Dans une approche non-supervisée de la segmentation bayésienne, l'estimation des paramètres inconnus du modèle considéré est une tâche importante. Il existe différents algorithmes permettant d'effectuer cette tâche. Nous citons ceux que nous allons utiliser dans notre étude à savoir : l'algorithme "Espérance-Maximisation" (EM) [63], l'algorithme "stochastique" EM (SEM) [64] qui en est une version stochastique, et l'algorithme "Estimation Conditionnelle Itérative" (ICE) [61, 65–67].

a L'algorithme EM

L'algorithme d'Espérance-Maximisation (EM) est l'un des algorithmes possibles utilisés pour estimer les paramètres des modèles à variables latentes

[68–70]. Cet algorithme a été proposé au départ pour estimer les paramètres du modèle des chaînes de Markov cachées [63], il a été ensuite généralisé à différents modèles en données incomplètes.

Soit le couple de processus $(\mathbf{S}, \mathbf{X})$ tel que les composantes du processus non observable $\mathbf{S}$ prennent leurs valeurs dans un ensemble d'états discret et $\phi = (\phi_1, \phi_2, \dots)$ l'ensemble des paramètres à estimer. L'objectif de l'algorithme EM est d'approcher le maximum de la vraisemblance par rapport aux paramètres ϕ de façon itérative. En effet, la maximisation de la log-vraisemblance des observations est, dans le cas des modèles à données manquantes, souvent impossible à effectuer analytiquement. Dans l'algorithme EM, on utilise la log-vraisemblance des données complètes, qui sera notée $\mathcal{L}_c(\mathbf{S}, \mathbf{X}; \phi)$, et on on exploite son espérance, conditionnelle aux données observées, par rapport aux données cachées.

L'algorithme EM est une procédure itérative dont le déroulement est le suivant :

1. ϕ est d'abord initialisé à une valeur $\phi^{[0]}$ des paramètres ;
2. à chaque itération $[q]$ on considère deux étapes :
   - **Etape "E" (Espérance)** : on calcule l'espérance conditionnelle

$$\mathbb{E}_{\phi^{[q]}}\{\mathcal{L}_c(\mathbf{S}, \mathbf{x}; \phi) \,|\, \mathbf{X}\}$$

   - **Etape "M" (Maximisation)** : on estime les paramètres par maximisation de l'espérance conditionnelle

$$\widehat{\phi}^{[q+1]} = \operatorname{Arg}\max_{\phi} \mathbb{E}_{\phi^{[q]}}\{\mathcal{L}_c(\mathbf{S}, \mathbf{x}; \phi) \,|\, \mathbf{X}\} \tag{1.24}$$

Nous décrivons brièvement les itérations de l'algorithme EM pour un modèle simple de chaîne de Markov cachée classique à bruit indépendant gaussien. Soit $p(\mathbf{x}_t \,|\, \mathbf{s}_t)$, la densité des observations conditionnellement aux sources. Nous considérons le cas gaussien pour lequel nous avons :

$$p(\mathbf{x}_t \,|\, \mathbf{s}_t) = \frac{1}{(2\pi)^{(N/2)} \det \Gamma^{1/2}} e^{-\frac{1}{2}(\mathbf{x}_t - \mu)^{\mathrm{T}} \Gamma^{-1} (\mathbf{x}_t - \mu)} \tag{1.25}$$

Pour K états, les paramètres à estimer sont, d'une part, les K probabilités initiales π_i correspondant à chaque état ω_i et les K^2 paramètres a_{ij} formant

la matrice de transition. Nous avons à chaque itération $[q]$:

$$\pi_i^{[q+1]} = \frac{1}{T}\sum_{t=1}^{T}\zeta_t^{[q]}(i) \tag{1.26}$$

$$a_{ij}^{[q+1]} = \frac{\sum_{t=1}^{T}\psi_t^{[q]}(i,j)}{\sum_{t=1}^{T}\zeta_t^{[q]}(i)} \tag{1.27}$$

Où $\zeta_t^{[q]}(i)$ et $\psi_t^{[q]}(i,j)$ sont définies par les équations (1.20) et (1.21). D'autre part, on doit également estimer les paramètres définissant les densités de probabilité gaussiennes des observations conditionnellement aux sources, qui sont les K moyennes μ_i et les K matrices de covariance Γ_i. Les ré-estimations de ces paramètres sont données par :

$$\mu_i^{[q+1]} = \frac{\sum_{t=1}^{T}\zeta_t^{[q]}(i)\mathbf{x}_t}{\sum_{t=1}^{T}\zeta_t^{[q]}(i)} \tag{1.28}$$

$$\Gamma_i^{[q+1]} = \frac{\sum_{t=1}\zeta_t(i)(\mathbf{x}_t-\mu_i)(\mathbf{x}_t-\mu_i)^{\mathrm{T}}}{\sum_{t=1}\zeta_t(i)} \tag{1.29}$$

En résumé, l'algorithme se déroule de la manière suivante :

- Initialisation des paramètres $\pi_i^{[0]}$, $a_{ij}^{[0]}$, $\mu_i^{[0]}$, $\Gamma_i^{[0]}$
- Etape E : Calcul des probabilités $\alpha_t^{[q]}(i)$ et $\beta_t^{[q]}(i)$, et déduction de $\psi_t^{[q]}$ à partir de $\alpha_t^{[q]}(i)$ et $\beta_t^{[q]}(i)$ calculée sur la base des paramètres $\pi_i^{[q]}$, $a_{ij}^{[q]}$, $\mu_i^{[q]}$, $\Gamma_i^{[q]}$;
- Etape M : Calcul des paramètres $\pi_i^{[q+1]}$, $a_{ij}^{[q+1]}$, $\mu_i^{[q+1]}$, $\Gamma_i^{[q+1]}$ par les formules (1.26), (1.27), (1.28), (1.29)

La procédure est arrêtée selon un critère d'arrêt choisi, adapté à chaque situation particulière. Ce critère peut se baser, par exemple, sur la convergence de l'un des paramètres estimés.

b L'algorithme ICE

Soit $(\phi_i)_{1\leqslant i\leqslant K}$ les paramètres à estimer. Le principe d'ICE est différent de celui de EM [56, 61, 65–67], la vraisemblance n'y intervient pas obligatoirement. Dans l'algorithme ICE on fait les deux hypothèses suivantes :

1. L'existence d'un estimateur $\widehat{\phi} = \widehat{\phi}(\mathbf{Z})$ à partir des données complètes ;
2. La possibilité de simuler le processus caché $\mathbf{S}$ selon sa loi *a posteriori* $p(\mathbf{S} \mid \mathbf{X})$.

La valeur suivante $\phi_i^{[q+1]}$ de l'estimée de ϕ_i est donnée à partir de la valeur courante de paramètres $\phi^{[q]}$ à l'itération q et de l'observation $\mathbf{x}$ par :

$$\widehat{\phi_i}^{[q+1]} = \mathbb{E}_{\phi^{[q]}}\{\widehat{\phi_i}(\mathbf{Z})|\mathbf{X}; \phi_i^{[q]}\} \tag{1.30}$$

S'il existe des estimateurs pour lesquels cette espérance n'est pas explicitement calculable, on simule m ($m \in \mathbb{N}^*$ est fixé) réalisations de $\mathbf{S}$ selon sa loi conditionnellement à $\mathbf{X} = \mathbf{x}$ $p(\mathbf{s} \mid \mathbf{X} = \mathbf{x})$:

$$\mathbf{S}^1 = (\mathbf{s}_1^1, \ldots, \mathbf{s}_T^1)$$

$$\vdots$$

$$\mathbf{S}^m = (\mathbf{s}_1^m, \ldots, \mathbf{s}_T^m)$$

et on pose :

$$\widehat{\phi}^{[q+1]} = \frac{1}{m} \sum_{k=1}^{m} \widehat{\phi}(\mathbf{s}^k, \mathbf{X}) \tag{1.31}$$

L'estimation s'effectue sur la base de la moyenne sur l'ensemble des m réalisations de $\mathbf{S}$ simulées. Ceci est justifié par la loi des grands nombres : lorsque m tend vers l'infini, (1.31) tend vers (1.30).

Dans le cas des modèles de chaînes de Markov cachées l'algorithme ICE est applicable. En effet, la distribution de $\mathbf{S}$ conditionnellement à $\mathbf{X}$ est celle d'une chaîne de Markov non stationnaire [61], ce qui permet de simuler les chaînes de Markov $\mathbf{S}^1, \ldots, \mathbf{S}^m$ nécessaires à l'estimation des paramètres. Les transitions *a posteriori* des chaînes de Markov sont données à partir de (1.21) par :

$$\begin{aligned} &p(\mathbf{s}_{t+1} = \omega_j | \mathbf{s}_t = \omega_i, \mathbf{x}) = \\ &\frac{p(\mathbf{s}_{t+1} = \omega_j | \mathbf{s}_t = \omega_i) p(\mathbf{x}_t | \mathbf{s}_t = \omega_j) p(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_1 | \mathbf{s}_t = \omega_j)}{\sum_{\omega_j \in \Omega} p(\mathbf{s}_{t+1} = \omega_j | \mathbf{s}_t = \omega_i) p(\mathbf{x}_t | \mathbf{s}_t = \omega_j) p(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_1 | \mathbf{s}_t = \omega_j)} \end{aligned} \tag{1.32}$$

Nous présentons le déroulement de l'algorithme ICE pour une chaîne de Markov cachée à bruit indépendant dans le cas gaussien que nous considérons :

- Initialisation des paramètres $\pi_i^{[0]}$, $a_{ij}^{[0]}$, $\mu_i^{[0]}$, $\Gamma_i^{[0]}$
- à chaque itération $[q]$:

– ré-estimation des paramètres de la loi de $\mathbf{S}$:

$$a_{ij}^{[q+1]} = \frac{\sum_{t=1}^{T} \psi_t^{[q]}(i,j)}{\sum_{t=1}^{T} \zeta_t^{[q]}(i)} \tag{1.33}$$

$$\pi_i^{[q]} = \frac{1}{T} \sum_{t=1}^{T} \zeta_t^{[q]}(i) \tag{1.34}$$

– ré-estimation des paramètres de la loi de $\mathbf{X}$ conditionnelle à $\mathbf{S}$:
  – Simulation des réalisations de $\mathbf{s}^{m,[p]}$ selon les lois *a posteriori*, nous limitons le nombre de tirage à un seul $m = 1$.
  – Calcul, pour chaque $1 \leq i \leq K$, de μ_i et Γ_i par les formules suivantes :

$$\mu_i^{[q+1]} = \frac{\sum_{t=1}^{T} 1_{\mathbf{s}_t=\omega_i} \mathbf{x}_t}{\sum_{t=1}^{T} 1_{\mathbf{s}_t=\omega_i}} \tag{1.35}$$

$$\Gamma_i^{[q+1]} = \frac{\sum_{t=1} 1_{\mathbf{s}_t=\omega_i} (\mathbf{x}_t - \mu_i)(\mathbf{x}_t - \mu_i)^{\mathrm{T}}}{\sum_{t=1} 1_{\mathbf{s}_t=\omega_i}} \tag{1.36}$$

Notons que dans le contexte des CMC-BI avec du bruit gaussien les deux méthodes EM et ICE donnent des résultats expérimentalement comparables [61]. Par ailleurs et plus récemment, l'algorithme ICE a prouvé sa bonne efficacité dans différentes modélisations complexes [71–74].

c L'algorithme SEM

L'algorithme SEM [64, 75] est une approximation stochastique de l'algorithme EM. Il s'agit en fait d'une combinaison des étapes de l'algorithme EM avec une étape d'apprentissage probabiliste. Le nom SEM vient de "Stochastique" EM. Cet algorithme a été proposé afin de surmonter les limitations de l'algorithme EM suivantes :

– Le nombre de classe K est supposé connu ;
– La solution obtenue dépend fortement de l'initialisation.

L'algorithme SEM est considéré donc comme une amélioration de l'algorithme EM par l'adjonction d'une étape *stochastique* dans laquelle un tirage aléatoire est effectué. Nous avons donc les améliorations suivantes :

– il suffit de connaître une borne supérieure du nombre de classes K ;
– la solution est largement indépendante de l'initialisation.

Les perturbations introduites à chaque itération par les tirages aléatoires sont censées empêcher la convergence de la procédure vers un maximum local de la vraisemblance comme cela peut être le cas pour l'algorithme EM. Le déroulement de l'algorithme SEM est le suivant :

- Initialisation des paramètres à estimer ;
- À chaque itération $[q]$:
  - Simulation d'une seule réalisation x de X selon la loi *a posteriori* basée sur les paramètres courants (même démarche que dans le cas de ICE) ;
  - Re-estimation des paramètres $c_{ij} = p(\mathbf{s}_{t+1} = \omega_j, \mathbf{s}_{t+1} = \omega_i)$:

$$c_{ij}^{[q+1]} = \frac{1}{T-1} 1_{\mathbf{s}_t=\omega_i, \mathbf{s}_{t+1}=\omega_j}$$

    (tout se passe comme si le résultat du tirage était la vraie réalisation de $\mathbf{X}$) ;
  - Re-estimation des paramètres définissant la loi du bruit (μ et Γ)par la même démarche que celle de ICE.

Notons que dans le cas classique des CMC-BI bruitées avec du bruit gaussien l'utilisation du SEM donne des résultats similaires à ceux obtenus avec EM ou ICE [61].

Chapitre 2

Les chaînes de Markov couples et triplets

Dans la modélisation par une chaîne de Markov cachée CMC-BI, on considère les hypothèses suivantes :

- le processus $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_t, \ldots, \mathbf{s}_T)$ est une chaîne de Markov ;
- les observations $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T)$ sont indépendantes conditionnellement aux sources cachées $\mathbf{S} : p(\mathbf{X} \mid \mathbf{S}) = p(\mathbf{x}_1 \mid \mathbf{s}_1) \ldots p(\mathbf{x}_T \mid \mathbf{s}T)$;
- et chaque composante $\mathbf{x}_t$ dépend uniquement de $\mathbf{s}_t : p(\mathbf{x}_t \mid \mathbf{s}) = p(\mathbf{x}_t \mid \mathbf{s}_t)$.

La loi du couple $(\mathbf{S}, \mathbf{X})$ s'écrit alors

$$p(\mathbf{S}, \mathbf{X}) = p(\mathbf{x}_1)p(\mathbf{x}_2 \mid \mathbf{x}_1) \ldots p(\mathbf{x}_T \mid \mathbf{x}_{T-1})p(\mathbf{x}_1 \mid \mathbf{s}_1) \ldots p(\mathbf{x}_T \mid \mathbf{s}_T)$$

avec la loi de $\mathbf{S}$ donnée par

$$p(\mathbf{S}) = p(\mathbf{s}_1)p(\mathbf{s}_2 \mid \mathbf{s}_1) \ldots p(\mathbf{s}_T \mid \mathbf{s}_{T-1})$$

et la loi de $\mathbf{S}$ conditionnelle a $\mathbf{X}$ donnée par

$$p(\mathbf{X} \mid \mathbf{S}) = p(\mathbf{x}_1 \mid \mathbf{s}_1) \ldots p(\mathbf{x}_T \mid \mathbf{s}_T)$$

Les hypothèses ci-dessus rendent la loi $p(\mathbf{S}, \mathbf{X})$ relativement simple ; en effet, une chaîne de Markov est le processus le plus simple permettant l'introduction de la dépendance entre les variables, et la loi $p(\mathbf{S}, \mathbf{X})$ est également très simple. Notons que dire que le modèle est simple revient déjà dire que les hypothèses qu'il vérifie sont restrictives. En effet, la validité des hypothèses ci-dessus peut être facilement critiquée dans certaines situations réelles.

Cependant, malgré sa simplicité, le modèle CMC-BI est extraordinairement robuste et peut donner de bons résultats dans différentes situations complexes. Malgré tout, il existe également des situations complexes dans lesquelles le modèle CMC-BI est mis en défaut.

Dans ce chapitre nous présentons deux généralisations qui permettent de surmonter les limitations des CMC-BI. Une première généralisation est le modèle dit de Chaîne de Markov Couple (CMCouple) introduit par [54, 56]. Dans ce modèle on suppose la markovianité du couple $(\mathbf{S}, \mathbf{X})$ et on relâche l'hypothèse de markovianité du processus caché $\mathbf{S}$. Comme dans les modèles CMC-BI, les CMCouples permettent le calcul des lois *a posteriori* nécessaires à l'estimation des données cachées de manière simple tout en étant plus généraux, et donc plus à même de décrire convenablement les problèmes à traiter. Notons que nous pouvons retrouver une CMC-BI à partir d'une CMCouple si les conditions nécessaires et suffisantes suivantes sont vérifiées [76] :

Proposition 1 *Soit* $\mathbf{Z} = (\mathbf{S}, \mathbf{X})$ *une chaîne de Markov stationnaire et réversible, dont les réalisations sont notées* $\mathbf{z}_t = (\mathbf{s}_t, \mathbf{x}_t)$. *Les trois conditions suivantes :*

(i) $\mathbf{Z}$ *est une chaîne de Markov (c'est-à-dire* $\mathbf{Z} = (\mathbf{S}, \mathbf{X})$ *est une chaîne de Markov cachée),*

(ii) pour tout $2 \leq t \leq T$, $p(\mathbf{x}_t \,|\, \mathbf{s}_t, \mathbf{s}_{t-1}) = p(\mathbf{x}_t \,|\, \mathbf{s}_t)$ *et*

(iii) pour tout $1 \leq t \leq T$, $p(\mathbf{x}_t \,|\, \mathbf{s}) = p(\mathbf{x}_t \,|\, \mathbf{s}_t)$

sont équivalentes.

Après les CMCouples, un modèle triplet a été proposé [55] comme premier niveau de généralisation des CMCouples et deuxième niveau de généralisation pour les CMC-BI. Dans ce modèle un processus aléatoire supplémentaire $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_T)$ est introduit, et on suppose la markovianité du processus triplet $(\mathbf{S}, \mathbf{X}, \mathbf{R})$ sans que le couple $(\mathbf{S}, \mathbf{X})$ soit nécessairement de Markov. Le modèle triplet permet de relâcher la contrainte de markovianité du processus couple $(\mathbf{S}, \mathbf{X})$ et généralise donc les CMCouples.

Ces généralisations offrent de grandes possibilités de modélisation selon la structure des dépendances du phénomène à traiter. En effet, les chaînes considérées dans le modèle classique des chaînes de Markov cachée sont simples le

modèle classique des chaîne de Markov cachées nous allons présenter des modèles plus généraux et plus riches permettant de modéliser des phénomènes plus complexes. Dans ce qui suit nous allons voir les modèles CMCouple et les modèles triplets dans le cas multidimensionnel, c'est-à-dire qu'à tout instant t, $\mathbf{x}_t$ et $\mathbf{s}_t$ seront, respectivement, de tailles N et Q. Nous allons nous intéresser au cas où $N = Q$ et nous allons commencer par décrire le modèle CMCouple de façon générale puis nous exposerons le modèle CMCouple à bruit indépendant que l'on note CMCouple-BI. Nous détaillerons les calculs des lois *a posteriori* ainsi que les algorithmes d'estimations de paramètres du modèle CMCouple-BI pour des données N-dimensionnelles. Nous allons ensuite nous intéresser au modèle triplet et nous détaillerons le modèle particulier que nous notons CMT et que nous considérerons par la suite dans nos simulations. On précise que $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_T)$ et $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ sont de taille $N \times T$, chaque vecteur $\mathbf{x}_1$ est de taille $N \times 1$ Triplet pour un signal multidimensionnel. Ces modèles sont basés sur la markovianité du processus joint $\mathbf{Z} = (\mathbf{S}, \mathbf{X})$ pour la chaîne de Markov couple, et du processus triplet $\mathbf{Y} = (\mathbf{R}, \mathbf{S}, \mathbf{X})$ caractérisé par un processus auxiliaire $\mathbf{R}$ gouvernant la changement du régime de la chaîne cachée $\mathbf{S}$.

2.1 Chaînes de Markov Couples

Soient $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ et $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_T)$ deux suites de vecteurs aléatoires. À tout instant t, la composante $\mathbf{s}_t$ de la chaîne cachée $\mathbf{S}$ est un vecteur de Q éléments :
$\mathbf{s}_t = \begin{pmatrix} s^1 \\ \vdots \\ s^Q \end{pmatrix}$, la composante $\mathbf{x}_t$ est un vecteur de N éléments : $\mathbf{x}_t = \begin{pmatrix} x^1 \\ \vdots \\ x^N \end{pmatrix}$.
Chaque vecteur caché $\mathbf{s}_t$ prend ses valeurs dans un ensemble d'états fini $\Omega = \{\omega_1, \ldots, \omega_K\}$, chaque vecteur d'état ω_i est composé de Q éléments : $\omega_i = \begin{pmatrix} \omega_i^1 \\ \vdots \\ \omega_i^Q \end{pmatrix}$.
Les composantes $\mathbf{x}_t$ du vecteur des observations $\mathbf{X}$ prennent leurs valeurs

dans l'ensemble des vecteurs réels $\mathbb{R}^N$.

Soit $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)$ un processus aléatoire markovien, avec $\mathbf{Z} = (\mathbf{X}, \mathbf{S})$. La distribution du processus $\mathbf{Z}$ est donnée par :

$$p(\mathbf{Z}) = p(\mathbf{z}_1) \prod_{t=1}^{T-1} p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t). \tag{2.1}$$

Les transitions du modèle peuvent être explicitées de deux manières :

$$p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t) = p(\mathbf{x}_{t+1} \,|\, \mathbf{z}_t) p(\mathbf{s}_{t+1} \,|\, \mathbf{x}_{t+1}, \mathbf{z}_t); \tag{2.2}$$

ou encore

$$p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t) = p(\mathbf{s}_{t+1} \,|\, \mathbf{z}_t) p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}, \mathbf{z}_t). \tag{2.3}$$

Nous supposerons dans toute la suite que le processus $\mathbf{Z}$ est stationnaire. Nous remarquons que nous retrouvons une CMC-BI à partir d'une CMCouple si :

- $p(\mathbf{s}_{t+1} \,|\, \mathbf{z}_t) = p(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t)$
- $p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}, \mathbf{z}_t) = p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1})$

En effet, si on intègre (2.1), vérifiant les propriétés ci-dessus, par rapport aux $\mathbf{x}_1, \ldots, \mathbf{x}_T$ on obtient la distribution de $p(\mathbf{S})$, qui s'écrit alors

$$p(\mathbf{S}) = p(\mathbf{s}_1) \prod_{t=1}^{T-1} p(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t),$$

et qui est donc une distribution markovienne. Dans ce cas et selon la proposition 1, les processus $(\mathbf{S}, \mathbf{X})$ forment une CMC-BI.

2.1.1 CMCouple : approche supervisée

Le calcul de la loi *a posteriori* et des diverses quantités qui y sont liées est l'étape clé d'une segmentation par le modèle des CMCouples. On montre que les probabilités "Forward" α_t^* et "Backward" β_t^* définies dans le chapitre précédent pour les CMC-BI et étendues au cas des CMCouples, permettent, par une procédure analogue à celle utilisée dans les CMC-BI, le calcul analytique de la loi *a posteriori*. On pose $\mathbf{z}_t = (\mathbf{s}_t = \omega_i, \mathbf{x}_t)$ et $\mathbf{z}_{t+1} = (\mathbf{s}_{t+1} = \omega_j, \mathbf{x}_{t+1})$, soit :

$$\alpha_t^*(\omega_i) \triangleq p(\mathbf{z}_t, \mathbf{x}_1, ..., \mathbf{x}_{t-1}) \tag{2.4}$$

On peut remarquer que la probabilité "Forward" d'une CMCouple qui est identique à celle des CMC-BI. La probabilité "Backward" est définie par :

$$\beta_t^*(\omega_i) \triangleq p(\mathbf{x}_{t+1}, ..., \mathbf{x}_T \,|\, \mathbf{z}_t) \tag{2.5}$$

On peut alors voir que lorsque la CMCouple considérée est une CMC-BI, on retrouve les probabilités "Forward" et "Backward" classiques. Comme dans le cas des CMC-BI, on montre que ces probabilités sont calculables par récurrence de la manière suivante :

Forward :

- $\alpha_1^*(\omega_i) = p(\mathbf{z}_1) = p(\mathbf{s}_1 = \omega_i, \mathbf{x}_1)$
- pour $t = 1, \ldots, T-1$; calculer :

$$\alpha_{t+1}^*(\omega_j) = \sum_{\omega_i \in \Omega} \alpha_t(\omega_i) p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t)$$

Backward :

- $\beta_T^*(\omega_j) = 1$
- pour $t = T-1, T-2 \ldots, 1$; calculer :

$$\beta_t^*(\omega_i) = \sum_{\omega_j \in \Omega} \beta_{t+1}(\omega_j) p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t)$$

Comme ci-dessus, on peut voir que lorsque la CMCouple considérée est une CMC-BI, on retrouve les procédures classiques du calcul des probabilités "Forward" et "Backward" classiques. Les expressions normalisées de α_t^* et β_t^*, notées α_t et β_t, sont définies par :

$$\alpha_t(\omega_i) = p(\mathbf{s}_t = \omega_i \,|\, \mathbf{x}_1, ..., \mathbf{x}_t) \tag{2.6}$$

$$\beta_t(\omega_i) = \frac{p(\mathbf{x}_{t+1}, ..., \mathbf{x}_T \,|\, \mathbf{s}_t = \omega_i, \mathbf{x}_t)}{p(\mathbf{x}_{t+1}, ..., \mathbf{x}_T \,|\, \mathbf{x}_1, ..., \mathbf{x}_t)} \tag{2.7}$$

Ces probabilités renormalisées ne souffrent plus de problème au niveau numérique. Comme dans le cas des CMC-BI elles sont calculables à partir des formules de récurrence suivantes, qui généralisent les formules classiques utilisées dans les CMC-BI :

Forward :

- $\alpha_1(\omega_i) = p(\mathbf{s}_1 = \omega_i \,|\, \mathbf{x}_1) = \dfrac{p(\mathbf{s}_1 = \omega_i, \mathbf{x}_1)}{\sum_{\mathbf{s}_1} p(\mathbf{s}_1 = \omega_i, \mathbf{x}_1)}$

- pour $t = 1, \ldots, T-1$; calculer :

$$\alpha_{t+1}(\omega_j) = \frac{\sum_{\omega_i \in \Omega} \alpha_t(\omega_i) p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t)}{\sum_{\omega_i, \omega_j \in \Omega^2} \alpha_t(\omega_i) p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t)}$$

Backward :

- $\beta_T(\omega_i) = 1$
- pour $t = T-1, T-2, \ldots, 1$; calculer :

$$\beta_t(\omega_i) = \frac{\sum_{\omega_j \in \Omega} \beta_{t+1}(\omega_j) p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t)}{\sum_{\omega_i, \omega_j \in \Omega^2} \beta_{t+1}(\omega_j) p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t)}$$

La loi de probabilité marginale *a posteriori* $\zeta_t(\omega_i) = p(\mathbf{s}_t = \omega_i \,|\, \mathbf{X})$ et la loi de probabilité *a posteriori* marginale jointe $\psi_t(\omega_i, \omega_j) = p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{X})$ sont exprimées par :

$$\zeta_t(\omega_i) = \alpha_t(\omega_i)\beta_t(\omega_i) \tag{2.8}$$

et,

$$\psi_t(\omega_i, \omega_j) = \frac{\alpha_t(\omega_i) p(\mathbf{s}_{t+1} = \omega_j, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t = \omega_i, \mathbf{x}_t) \beta_{t+1}(\omega_j)}{\sum_{\omega_i, \omega_j \in \Omega^2} \alpha_t(\omega_i) p(\mathbf{s}_{t+1} = \omega_j, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t = \omega_i, \mathbf{x}_t) \beta_{t+1}(\omega_j)} \tag{2.9}$$

Comme dans le cas des CMC-BI on montre que la loi *a posteriori* $p(\mathbf{S} \,|\, \mathbf{X})$ de $\mathbf{S}$ est markovienne, avec les transitions sont données à partir (2.8) et (2.9) par [54] :

$$p(\mathbf{s}_{t+1} = \omega_j \,|\, \mathbf{s}_t = \omega_i, \mathbf{X}) = \frac{\psi_t(\omega_i, \omega_j)}{\zeta_t(\omega_i)} \tag{2.10}$$

2.1.2 CMCouple-BI : le modèle

Dans le modèle de CMCouple-BI, la loi directe du couple $p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1})$ peut être décomposée en produit de deux probabilités :

$$p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) = p(\mathbf{x}_t \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) \tag{2.11}$$

Les probabilités de transitions prennent la forme suivante :

$$p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t) = p(\mathbf{s}_{t+1} \,|\, \mathbf{z}_t) p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) \tag{2.12}$$

La différence entre une CMCouple et une CMCouple-BI est que dans une CMCouple-BI, les observations sont indépendantes conditionnellement au processus $\mathbf{S}$:

$$p(\mathbf{X}|\mathbf{S}) = \prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{S}) \tag{2.13}$$

Bien que les observations $\mathbf{x}_1, \ldots, \mathbf{x}_T$ soient indépendantes conditionnellement à $\mathbf{S}$, la structure du bruit dans une CMCouple est assez complexe et bien différente de celle dans une CMC-BI. En effet, considérons la loi $p(\mathbf{x}_t|\mathbf{s}_t)$. Dans une CMCouple-BI nous avons

$$p(\mathbf{x}_t|\mathbf{s}_t) = \sum_{\substack{\mathbf{s}_1,\ldots,\mathbf{s}_{t-1},\\ \mathbf{s}_{t+1},\ldots,\mathbf{s}_T}} p(\mathbf{x}_t, \mathbf{s}_1, \ldots, \mathbf{s}_{t-1}, \mathbf{s}_{t+1}, \ldots, \mathbf{s}_T|\mathbf{s}_t) \tag{2.14}$$

$$p(\mathbf{x}_t|\mathbf{s}_t) = \sum_{\substack{\mathbf{s}_1,\ldots,\mathbf{s}_{t-1},\\ \mathbf{s}_{t+1},\ldots,\mathbf{s}_T}} p(\mathbf{s}_1, \ldots, \mathbf{s}_{t-1}, \mathbf{s}_{t+1}, \ldots, \mathbf{s}_T|\mathbf{s}_t) p(\mathbf{x}_t|\mathbf{s}_1, \ldots, \mathbf{s}_{t-1}, \mathbf{s}_t, \mathbf{s}_{t+1}, \ldots, \mathbf{s}_T) \tag{2.15}$$

En fait, cette somme n'est pas calculable et donc la loi $p(\mathbf{x}_t|\mathbf{s}_t)$ n'est pas connue dans une CMCouple-BI (sauf si cette dernière est une CMC-BI, auquel cas on retrouve la loi classique $p(\mathbf{x}_t|\mathbf{s}_t)$. Par exemple, si on se place dans le cas gaussien, $p(\mathbf{x}_t|\mathbf{s}_t)$ n'est pas une simple distribution gaussienne (sauf si la CMCouple-BI considérée est une CMC-BI), mais un mélange d'un grand nombre de distributions gaussiennes. Le modèle du bruit est alors plus riche et général, permettant ainsi la modélisation de situations réelles dans lesquelles les dépendances entre données cachées et observées seraient complexes. Par exemple, soit une image d'une zone géographique dans laquelle une forêt est délimitée par l'eau d'un côté et le désert de l'autre côté. La végétation au voisinage de l'eau est naturellement plus verte que celle proche du désert qui a tendance à avoir une coloration jaunâtre. L'aspect visuel d'un pixel observé donné est alors différent selon que si le pixel est au voisinage de l'eau ou de la foret. Cette situation ne peut être modélisée par une CMC-BI qui ne tient compte que des dépendances simples entre les processus caché et observé. Cependant le fait de prendre en compte des dépendances de chaque observation $\mathbf{x}_t$ de l'ensemble des composantes du processus caché permet de modéliser le problème de manière plus complète. On note que dans une

CMCouple-BI, le processus $\mathbf{S}$ n'est pas nécessairement une chaîne de Markov. En effet, on peut montrer que lorsque la CMCouple est stationnaire et réversibles alors $\mathbf{S}$ est de Markov si et seulement si $p(\mathbf{s}_t|\mathbf{X}) = p(\mathbf{s}_t|\mathbf{x}_t)$ [76]. Finalement, nous considérons un processus couple doublement original par rapport aux CMC-BI classiques. D'une part, il est impossible d'écrire les lois conditionnelles $p(\mathbf{s}_t|\mathbf{x}_t)$,qui sont très générales. L'intérêt de cette généralité a été discutée ci-dessus. D'autre part, le processus caché n'est pas nécessairement de Markov (comme il peut l'être, auquel cas on retrouve une CMC-BI, on a une situation plus générale). Cette deuxième propriété pourrait probablement être utile dans les cas de processus cachés qui ne sont manifestement pas markoviens. Comme mentionné ci-dessus, malgré ces deux différences notables, les mêmes traitements que dans les CMC-BI classiques sont possibles.

On note qu'à partir d'une CMCouple-BI nous pouvons retrouver une CMC-BI, selon la proposition 1, si on considère les équations suivantes :

- $p(\mathbf{s}_{t+1}|\mathbf{z}_t) = p(\mathbf{s}_{t+1}|\mathbf{s}_t)$
- $p(\mathbf{x}_{t+1}|\mathbf{s}_{t+1}, \mathbf{s}_t) = p(\mathbf{x}_{t+1}|\mathbf{s}_{t+1})$

2.1.3 Estimation des paramètres

La modélisation par CMCouple-BI permet la segmentation du processus caché aussi bien dans le cas supervisé que dans le cas non supervisé. Dans ce dernier cas, l'estimation de paramètres est requise. L'un des principaux avantages des CMCouples-BI est qu'ils permettent d'appliquer les algorithmes des estimations de paramètres utilisés pour les CMC-BI. En particulier les algorithmes EM et ICE ont été étendus avec succès aux CMCouples [77]. Nous détaillerons, dans ce qui suit, ces deux algorithmes pour les CMCouples-BI N-dimensionnelles. Nous nous sommes intéressés particulièrement aux algorithmes EM et ICE présentés dans le chapitre précédent. Nous allons alors présenter ces algorithmes dans le contexte des CMCouples. Le processus $\mathbf{Z}$ est supposé stationnaire. Sa loi est alors donnée par $p(\mathbf{z}_t, \mathbf{z}_{t+1})$, qui peut s'écrire :

$$p(\mathbf{z}_t, \mathbf{z}_{t+1}) = p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) p(\mathbf{s}_t, \mathbf{s}_{t+1}) \tag{2.16}$$

Nous supposons dans ce paragraphe, que les lois conditionnelles $p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1})$ sont des distributions gaussiennes. La décomposition (2.11) de $p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1})$

permet de réduire de manière notable le nombre de paramètres à estimer. En effet, dans un modèle CMCouple nous avions K^2 éléments correspondant à la probabilité conjointe $p(\mathbf{s}_t, \mathbf{s}_{t+1})$, $N \times K^2$ pour chacun des deux vecteurs des moyennes,$2 \times K^2$ matrices $N \times N$ de covariance et $N \times K^2$ coefficients de corrélation. Alors que dans une CMCouple-BI il suffit d'estimer les K^2 éléments de la densité $p(\mathbf{s}_t, \mathbf{s}_{t+1})$ les $2 \times K^2$ vecteurs moyenne de taille $N \times 1$ et $2 \times K^2$ matrices carré de taille $N \times N$ de covariance correspondant à $p(\mathbf{x}_t / \mathbf{s}_t, \mathbf{s}_{t+1})$ et à $p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1})$. Nous présentons dans ce qui suit le déroulement des algorithmes EM et ICE adaptés au modèle CMCouple-BI.

a Estimation EM

L'estimation par l'algorithme EM [69] est communément utilisée dans le cas des chaînes de Markov cachées classiques. Comme nous l'avons vu dans le chapitre précédent, cet algorithme fonctionne en deux étapes. Nous avons à chaque itération $[q]$: étape "E" (Espérance) dans laquelle on calcule l'espérance de la vraisemblance conditionnelle à $\mathbf{X}$, et l'étape "M" (Maximisation) dans laquelle on maximise l'espérance conditionnelle calculée à l'étape "E" par rapport aux paramètres ϕ à estimer. Nous décrivons les étapes de l'algorithme EM dans le cas du modèle CMCouple-BI. Soit $\mathcal{Q}(\phi/\phi^{[q]})$ l'espérance conditionnelle de la vraisemblance en $\mathbf{X}$ sous le paramètre ϕ.

$$\mathcal{Q}(\phi/\phi^{[q]}) = \mathbb{E}_{\phi^{[q]}}\{\mathcal{L}_c(\mathbf{S}, \mathbf{X}; \phi) \,|\, \mathbf{X}\}$$

Calcul de la fonction $\mathcal{Q}(\phi/\phi^{[q]}) \Rightarrow$ Étape E :

Nous décomposons la loi de $\mathbf{Z}$ de la manière suivante :

$$\begin{aligned}
p(\mathbf{z}) &= p(\mathbf{z}_1) \prod_{t=2}^{T-1} p(\mathbf{z}_{t+1} \,|\, \mathbf{z}_t) \\
&= p(\mathbf{z}_1) \prod_{t=2}^{T-1} p(\mathbf{s}_{t+1}, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{x}_t) \\
&= p(\mathbf{z}_1) \prod_{t=2}^{T-1} \frac{p(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{s}_t, \mathbf{s}_{t+1})}{p(\mathbf{z}_t)} \\
&= p(\mathbf{z}_1) \prod_{t=2}^{T-1} \frac{p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) p(\mathbf{s}_t, \mathbf{s}_{t+1})}{p(\mathbf{z}_t)}
\end{aligned}$$

nous obtenons alors la formule suivante :

$$p(\mathbf{z}) = \frac{1}{\prod_{t=2}^{T-1} p(\mathbf{z}_t)} \times \prod_{t=1}^{T-1} p(\mathbf{x}_{t+1}, \mathbf{x}_t \,|\, \mathbf{s}_{t+1}, \mathbf{s}_t) \times \prod_{t=1}^{T-1} p(\mathbf{s}_{t+1}, \mathbf{s}_t) \qquad (2.17)$$

La log-vraisemblance est alors donnée par :

$$\mathcal{L}_c(\phi) = \sum_{t=1}^{T-1} \log p(\mathbf{s}_{t+1}, \mathbf{s}_t) - \sum_{t=2}^{T-1} \log p(\mathbf{z}_t) + \sum_{t=1}^{T-1} \log p(\mathbf{x}_{t+1}, \mathbf{x}_t \,|\, \mathbf{s}_{t+1}, \mathbf{s}_t) \qquad (2.18)$$

Nous obtenons l'expression de la fonction $\mathcal{Q}(\phi \,|\, \phi^{[q]})$ en utilisant la définition de l'espérance conditionnelle :

$$\mathbb{E}_{\phi^{[q]}}\{\mathcal{L}_c(\mathbf{Z}; \phi^{[q]}) \,|\, \mathbf{X}\} = \sum_{\mathbf{s}} \mathcal{L}_c(\mathbf{z}; \phi^{[q]}) p(\mathbf{s} \,|\, \mathbf{X})$$

Pour simplifier, nous introduisons les notations suivantes :

$$\psi_{t-1}^{[q]}(i,j) = p(s_{t-1} = \omega_i, s_t = \omega_j \,|\, \mathbf{x}, ; \phi^{[q]})$$

$$c_{ij}^{[q]} = p(s_{t-1} = \omega_i, s_t = \omega_j)$$

$$\zeta_t^{[q]} = p(s_t = \omega_j \,|\, \mathbf{x}; \phi^{[q]})$$

$$b_j^{[q]} = p(s_t = \omega_j, \mathbf{x}_t)$$

$$f_{i,j} = p(\mathbf{x}_{t-1}, \mathbf{x}_t \,|\, s_{t-1} = \omega_i, s_t = \omega_j)$$

Nous déduisons alors de (18) l'expression de $\mathcal{Q}(\phi/\phi^{[q]})$:

$$\mathcal{Q}(\phi \,|\, \phi^{[q]}) = \sum_{t=2}^{T} \sum_{(i,j)\in\Omega^2} \psi_{t-1}^{[q]}(i,j) \log c_{ij} - \sum_{t=2}^{T-1} \sum_{j\in\Omega} \zeta_t^{[q]} \log b_j + \sum_{t=2}^{T} \sum_{(i,j)\in\Omega^2} \psi_{t-1}^{[q]}(i,j) \log f_{i,j}$$

Les probabilités *a posteriori* $\zeta_t^{[q]}(\omega_i)$ et $\psi_{t-1}^{[q]}(\omega_i, \omega_j)$ sont calculées, à chaque itération, en utilisant la procédure Forward-Backward.

Maximisation de la fonction $\mathcal{Q}(\phi \,|\, /\phi^{[q]}) \Rightarrow$ Étape M :

Dans l'étape M de l'algorithme EM on maximise la fonction $\mathcal{Q}(\phi \,|\, \phi^{[q]})$ (par rapport a ϕ) en utilisant la méthode des multiplicateurs de Lagrange sous les trois contraintes suivantes : $\sum_{(i,j)\in\Omega^2} c_{ij} = 1$, $\sum_{(i,l)\in\Omega\times\mathbb{R}} c_{ij} f_{i,j}(l,m) = b_j(m)$, et $\sum_{(i,l)\in\mathbb{R}^2} f_{i,j}(l,m) = 1$.

Dans le cas d'observations réelles avec des densités conditionnelles gaussiennes, les formules de ré-estimation ont été démontrées dans [77]. Dans le cas N-dimensionnelles, elle sont données par :

$$c_{i,j}^{[q+1]} = \frac{1}{T-1} \sum_{t=1}^{T} \psi_t^{[q]}(i,j)$$

$$\widehat{\mu}_1^{i,j[q+1]} = \frac{\sum_{t=1}^{T-1} \psi_t(i,j)\mathbf{x}_t}{\sum_{t=1}^{T-1} \psi_t(i,j)}$$

$$\widehat{\mu}_2^{i,j[q+1]} = \frac{\sum_{t=1}^{T-1} \psi_t(i,j)\mathbf{x}_{t+1}}{\sum_{t=1}^{T-1} \psi_t(i,j)}$$

$$\widehat{\Gamma}_1^{i,j[q+1]} = \frac{\sum_{t=1}^{T-1} \psi_t(i,j) \left(\mathbf{x}_t - \widehat{\mu}_1^{i,j}\right) \left(\mathbf{x}_t - \widehat{\mu}_1^{i,j}\right)^{\mathrm{T}}}{\sum_{t=1}^{T-1} \psi_t(i,j)}$$

$$\widehat{\Gamma}_2^{i,j[q+1]} = \frac{\sum_{t=1}^{T-1} \psi_t(i,j) \left(\mathbf{x}_{t+1} - \widehat{\mu}_2^{i,j}\right) \left(\mathbf{x}_{t+1} - \widehat{\mu}_2^{i,j}\right)^{\mathrm{T}}}{\sum_{t=1}^{T-1} \psi_t(i,j)}$$

$\mu_1^{i,j[q+1]}$ et $\mu_2^{i,j[q+1]}$ des vecteurs de taille $N \times K$.

b Estimation ICE

L'estimation par ICE consiste à prendre un estimateur $\widehat{\phi}$ défini à partir des données complètes $\mathbf{z}$ et de l'approcher, a chaque itération en utilisant les paramètres courants, par l'espérance conditionnelle à $\mathbf{X}$. Contrairement à EM, l'estimation par ICE ne contient pas de phase de "maximisation" de vraisemblance. Cependant, comme dans le cas des CMC-BI, l'estimateur du maximum de vraisemblance peut y intervenir. En effet, dans notre application de l'ICE aux CMCouples-BI, l'estimateur $\widehat{\phi}$ considéré est celui du

maximum de vraisemblance. Bien entendu, rechercher le maximum de vraisemblance sur les données complètes est, en général, bien plus facile que de le chercher sur les données incomplètes, et l'estimateur du maximum de vraisemblance $\widehat{\phi}$ utilisé est obtenu classiquement.
Les formules de ré-estimation dans le cas des densités conditionnelles gaussiennes sont données par :

- Ré-estimation des paramètres $(c_{i,j})$ par

$$c_{i,j}^{[q+1]} = \frac{1}{T-1} \sum_{t=1}^{T} \psi_t^{[q]}(i,j)$$

  Notons que les paramètres $(c_{i,j})$ ne définissent pas nécessairement la loi de $\mathbf{S}$; cela n'est vrai que si $\mathbf{S}$ est markovienne, et donc la CMCouple-BI est une CMC-BI classique ;
- Ré-estimation des paramètres de la loi de $\mathbf{X}$ conditionnellement à $\mathbf{S}$:
  - Simulation des réalisations de $\mathbf{s}^{m,[q]}$ selon la loi *a posteriori*, en utilisant les transition *a posteriori*. Nous limitons le nombre de tirages à un seul $(m = 1)$.
  - Calcul, pour chaque $1 \leq i \leq K$, de μ_i et Γ_i par les formules suivantes :

$$\widehat{\mu}_1^{i,j[q+1]} = \frac{1}{Card(A^{i,j})} \sum_{t=1}^{T-1} 1_{A^{i,j}} . \mathbf{x}_t \tag{2.19}$$

$$\widehat{\mu}_2^{i,j[q+1]} = \frac{1}{Card(A^{i,j})} \sum_{t=1}^{T-1} 1_{A^{i,j}} . \mathbf{x}_{t+1} \tag{2.20}$$

$$\widehat{\Gamma}_1^{i,j[q+1]} = \frac{\sum_{t=1}^{T-1} 1_{A^{i,j}} . \left(\mathbf{x}_t - \widehat{\mu}_1^{i,j}\right) \left(\mathbf{x}_t - \widehat{\mu}_1^{i,j}\right)^{\mathrm{T}}}{Card(A^{i,j})} \tag{2.21}$$

$$\widehat{\Gamma}_2^{i,j[q+1]} = \frac{\sum_{t=1}^{T-1} 1_{A^{i,j}} . \left(\mathbf{x}_{t+1} - \widehat{\mu}_2^{i,j}\right) \left(\mathbf{x}_{t+1} - \widehat{\mu}_2^{i,j}\right)^{\mathrm{T}}}{Card(A^{i,j})} \tag{2.22}$$

Où $A^{i,j}$ est l'ensemble des éléments dans l'échantillon simulé $\mathbf{s}^{1,[q]}$ pour lesquels nous avons $(\mathbf{s}_t, \mathbf{s}_{t+1}) = (\omega_i, \omega_j)$. Notons que, bien qu'obtenue par des principes généraux différents, la ré-estimation des c_{ij} est identique dans les cas de EM et ICE.
Nous observons que la méthode EM est déterministe et la méthode ICE contient un aspect stochastique.

2.2 Chaînes de Markov Triplets

2.2.1 CMT : le modèle

Une extension possible des CMCouples consiste à introduire un processus auxiliaire $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_T)$ et à considérer que le triplet $(\mathbf{R}, \mathbf{S}, \mathbf{X})$ est markovien [56, 78]. Dans le présent travail, chaque élément de $\mathbf{R}$ prend ses valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \ldots, \lambda_M\}$, et on suppose donc que le processus $\mathbf{Y} = (\mathbf{R}, \mathbf{S}, \mathbf{X})$ est une chaîne de Markov ; un tel modèle sera dit "Chaînes de Markov Triplets" (CMT). Sa loi est donc définie par :

$$p(\mathbf{Y}) = p(\mathbf{y}_1) \prod_{t=1}^{T-1} p(\mathbf{y}_{t+1} \,|\, \mathbf{y}_t) \tag{2.23}$$

Les transitions de la chaîne $\mathbf{Y}$ peuvent se factoriser de différentes façons. Dans la suite, nous utiliserons la factorisation suivante :

$$\begin{aligned} p(\mathbf{y}_{t+1} \,|\, \mathbf{y}_t) &= p(\mathbf{r}_{t+1}, \mathbf{s}_{t+1}, \mathbf{x}_{t+1} \,|\, \mathbf{r}_{t+1}, \mathbf{s}_t, \mathbf{x}_{t+1}) \\ &= p(\mathbf{r}_{t+1}|\mathbf{y}_t) p(\mathbf{s}_{t+1}|\mathbf{y}_t, \mathbf{r}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{y}_t, \mathbf{s}_{t+1}, \mathbf{r}_{t+1}) \end{aligned} \tag{2.24}$$

Ce modèle est strictement plus général que les CMCouples. En effet, en posant $\mathbf{Z} = (\mathbf{S}, \mathbf{X})$, le couple $(\mathbf{R}, \mathbf{Z})$ est markovien. Cependant, $\mathbf{Z}$ ne l'est pas nécessairement, et donc $(\mathbf{S}, \mathbf{X})$ n'est pas nécessairement une CMCouple. Il est possible de construire des CMT relativement simples telles que $(\mathbf{S}, \mathbf{X})$ n'est pas markovienne.
Posons $\mathbf{V} = (\mathbf{R}, \mathbf{S})$. Puisque $(\mathbf{V}, \mathbf{X})$ est une chaîne de Markov, c'est une CMCouple. Ceci nous permet alors d'utiliser les techniques de restauration (la méthode du MPM) et les algorithmes d'estimation des paramètres (en particulier l'algorithme ICE et SEM) que nous avons utilisé dans dans la segmentation ainsi que l'estimation des paramètres des CMC-BI et CMCouples-BI.

Comme dans le cas des CMC-BI et CMCouples-BI, les CMT offrent une grande simplicité de calcul des lois *a posteriori*. En effet, les lois marginales *a posteriori* sont calculables du moment que $\mathbf{Y} = (\mathbf{V}, \mathbf{X})$ est une CMCouple.

$$p(\mathbf{v}_t \,|\, \mathbf{X}) = p(\mathbf{r}_t, \mathbf{s}_t \,|\, \mathbf{X}) \tag{2.25}$$

Les lois $p(\mathbf{s}_t \,|\, \mathbf{X})$ sont obtenus par une simple sommation par rapport aux $\mathbf{r}_t$:

$$p(\mathbf{s}_t \,|\, \mathbf{X}) = \sum_{\mathbf{r}_t \in \Lambda} p(\mathbf{r}_t, \mathbf{s}_t \,|\, \mathbf{X}) \tag{2.26}$$

Nous avons alors une famille de modèles très riche qui donne lieu à plusieurs modèles particuliers dont celui de que nous allons considérer dans ce qui suit et que nous notons CMT.

Soit $\mathbf{Y} = (\mathbf{R}, \mathbf{S}, \mathbf{X})$ une CMT. $\mathbf{X}$ et $\mathbf{R}$ prennent donc leurs valeurs respectives à partir des ensembles d'états discrets Ω et Λ. Le processus $\mathbf{Y}$ est donc une chaîne de Markov dont la loi est donnée par la loi initiale $p(\mathbf{r}_1, \mathbf{s}_1, \mathbf{x}_1)$ et les transitions. Considérons les transitions de la forme générale suivante :

$$p(\mathbf{y}_{t+1} \,|\, \mathbf{y}_t) = p(\mathbf{r}_{t+1} \,|\, \mathbf{r}_t) p(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t, \mathbf{r}_{t+1}) p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}) \tag{2.27}$$

Ceci implique que :

– Le processus $\mathbf{R}$ est de Markov, et donc :

$$p(\mathbf{R}) = p(\mathbf{r}_1) \prod_{t=1}^{T-1} p(\mathbf{r}_t \,|\, \mathbf{r}_{t-1}) \tag{2.28}$$

En effet, on le constate en intégrant la probabilité conjointe

$$p(\mathbf{r}, \mathbf{s}, \mathbf{x}) = p(\mathbf{r}_1, \mathbf{s}_1, \mathbf{x}_1) \prod_{t=1}^{T-1} p(\mathbf{r}_t, \mathbf{s}_t, \mathbf{x}_t | \mathbf{r}_{t-1}, \mathbf{s}_{t-1}, \mathbf{x}_{t-1}) \tag{2.29}$$

par rapport à $\mathbf{x}_T$, $\mathbf{s}_T$, $\mathbf{x}_{T-1}$, $\mathbf{s}_{T-1}$, $\ldots$, $\mathbf{x}_1$, $\mathbf{s}_1$.

– On montre qu'aucun des processus $\mathbf{S}$, $\mathbf{X}$ $(\mathbf{S}, \mathbf{X})$ n'est nécessairement un processus de Markov.

– On note également les processus $\mathbf{S}$ et $\mathbf{X}$ sont indépendants conditionnellement à $\mathbf{R}$.

Bien entendu, le troisième point ne signifie pas que les processus $\mathbf{S}$ et $\mathbf{X}$ sont indépendants. Cette hypothèse, qui est simplificatrice et et que nous faisons pour limiter le nombre de paramètres, peut paraître surprenante. Cependant, la même hypothèse a été faite dans le cadre de la segmentation d'images dans [76] et les résultats obtenus sont très satisfaisants.

D'autre part, dans le modèle CMT considéré nous avons :

$$p(\mathbf{S} \,|\, \mathbf{R}) = p(\mathbf{s}_1 \,|\, \mathbf{r}_1) \prod p(\mathbf{s}_t \,|\, \mathbf{s}_{t-1}, \mathbf{r}_t) \tag{2.30}$$

$$p(\mathbf{X} \mid \mathbf{S}) = \prod_{t=1}^{T} p(\mathbf{x}_t \mid \mathbf{s}_t) \tag{2.31}$$

Notons que le modèle considéré est relativement simple et ses extensions, qui impliqueraient l'augmentation du nombre des paramètres, pourrait éventuellement améliorer les résultats présentés dans la suite.

2.2.2 Approche supervisée

Nous avons vu que le modèle CMT considéré peut être vu comme une CMCouple, avec le processus caché discret $\mathbf{V} = (\mathbf{R}, \mathbf{S})$. La procédure Forward-Backward exposée ci-dessus permet donc le calcul direct des probabilités marginales *a posteriori* $p(\mathbf{v}_t|\mathbf{X})$ du processus $\mathbf{V}$. Les lois marginales selon $\mathbf{S}$ se déduisent en sommant l'équation (2.26) sur tous les états λ_i du processus $\mathbf{R}$. Cependant, rappelons la procédure forward-backward normalisée adaptée aux CMT considérés ; elle est décrite par :

Forward :

- $\alpha_1(\omega_i) = p(\mathbf{v}_1 \mid \mathbf{x}_1) = \dfrac{p(\mathbf{x}_1 \mid \mathbf{s}_1)p(\mathbf{s}_1 \mid \mathbf{r}_1)p(\mathbf{r}_1)}{\sum_{\mathbf{v}_1 \in \Lambda \times \Omega} p(\mathbf{x}_1 \mid \mathbf{s}_1)p(\mathbf{s}_1 \mid \mathbf{r}_1)p(\mathbf{r}_1)}$
- pour $t = 1, \ldots, T-1$; calculer :

$$\alpha_{t+1}(\omega_j) = \frac{\sum_{\mathbf{v}_t \in \Omega \times \Lambda} \alpha_t(\omega_i) p(\mathbf{y}_{t+1} \mid \mathbf{y}_t)}{\sum_{\mathbf{v}_t, \mathbf{v}_{t+1} \in (\Omega \times \Lambda)^2} \alpha_t(\omega_i) p(\mathbf{y}_{t+1} \mid \mathbf{y}_t)}$$

Backward :

- $\beta_T(\omega_j) = 1$
- pour $t = T-1, T-2, \ldots, 1$; calculer :

$$\beta_t(\omega_i) = \frac{\sum_{\mathbf{v}_{t+1} \in \Omega \times \Lambda} \beta_{t+1}(\omega_j) p(\mathbf{t}_{t+1} \mid \mathbf{t}_t)}{\sum_{\mathbf{v}_t, \mathbf{v}_{t+1} \in (\Omega \times \Lambda)^2} \beta_{t+1}(\omega_j) p(\mathbf{y}_{t+1} \mid \mathbf{y}_t)}$$

Les probabilités *a posteriori* et les probabilités conjointes *a posteriori* du processus $\mathbf{V}$ sont données par :

$$\psi_t(\omega_i, \omega_j) = \frac{\alpha_t(\omega_i) p(\mathbf{v}_{t+1}, \mathbf{x}_{t+1} \mid \mathbf{v}_t, \mathbf{x}_t) \beta_t(\omega_i)}{\sum_{\mathbf{v}_t, \mathbf{v}_{t+1} \in (\Omega \times \Lambda)^2} \alpha_t(\omega_i) p(\mathbf{s}_{t+1}, \mathbf{x}_{t+1} \mid \mathbf{s}_t, \mathbf{x}_t)\ mathbf{f}\beta_t(\omega_j)}, \tag{2.32}$$

et

$$\zeta_t(\omega_i) = \alpha_t(\omega_i)\beta_t(\omega_i). \tag{2.33}$$

Les lois marginales *a posteriori* du processus caché sont déduites de la manière suivante :

$$p(\mathbf{s}_t = \omega_i \,|\, \mathbf{x}) = \sum_{\mathbf{r}_t \in \Lambda} p(\mathbf{v}_t \,|\, \mathbf{x}) = \sum_{\mathbf{r}_t \in \Lambda} \zeta_t(\omega_i) \tag{2.34}$$

Notons que nous pouvons déduire les lois marginales *a posteriori* du processus auxiliaire $\mathbf{R}$ de la même manière que celles du processus $\mathbf{S}$:

$$p(\mathbf{r}_t = \lambda_i|\mathbf{X}) = \sum_{\mathbf{s}_t \in \Omega} p(\mathbf{v}_t|\mathbf{X}) = \sum_{\mathbf{s}_t \in \Omega} \zeta_t(\omega_i) \tag{2.35}$$

ce qui permet, par exemple, la recherche des "stationnarités" en imagerie [76].

2.2.3 Estimation des paramètres

Ainsi que discuté ci-dessus, en considérant $\mathbf{y}_t = (\mathbf{v}_t, \mathbf{x}_t)$ avec $\mathbf{v}_t = (\mathbf{r}_t, \mathbf{x}_t)$, nous passons d'un modèle triplet à un modèle CMCouple. Pour simplification, la loi $p(\mathbf{x} \,|\, \mathbf{v})$ de $\mathbf{x}$ conditionnellement à $\mathbf{V}$ vérifie :

$$p(\mathbf{x}|\mathbf{v}) = p(\mathbf{x}|\mathbf{s}) = \prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{s}_t)$$

Il est à noter que malgré la simplicité de $p(\mathbf{x} \,|\, \mathbf{v})$, le processus global $\mathbf{y}_t$ n'est pas un modèle de CMC-BI. Aucune des propriétés de la proposition définissant une CMC-BI : (i) $\mathbf{S}$ est une chaîne de Markov, (ii) les variables aléatoires $\mathbf{x}_1, \ldots, \mathbf{x}_T$ sont indépendantes conditionnellement à $\mathbf{S}$ et (iii) $p(\mathbf{x}_t|\mathbf{S}) = p(\mathbf{x}_t|\mathbf{s}_t)$, n'est vérifiée. En effet, dans le modèle que nous considérons $\mathbf{S}$ n'est pas une chaîne de Markov, les observations $\mathbf{x}_1, \ldots, \mathbf{x}_T$ sont indépendantes conditionnellement au couple $\mathbf{V} = (\mathbf{S}, \mathbf{R})$ et $p(\mathbf{x}_t|\mathbf{s}_t, \mathbf{r}_t) \neq p(\mathbf{x}_t|\mathbf{s}_t)$.

Pour le cas gaussien, nous avons à estimer K vecteurs moyenne, K matrice de covariance, ainsi que les $(K \times M)^2$ élément $p_{ij} = p(\mathbf{v}_t = \omega_i, \mathbf{v}_{t+1} = \omega_j)$. Nous allons utiliser l'algorithme ICE et l'algorithme SEM. Pour les deux algorithmes, un échantillon $\mathbf{v}$ est simulé selon les lois conditionnelles aux observations suivantes :

$$\psi_t(\omega_i, \omega_j) = p(\mathbf{v}_{t+1}, \mathbf{v}_t \,|\, \mathbf{x}, \phi^{[q]}),$$

les lois marginales *a posteriori* de $\mathbf{V}$ sont données par :

$$\zeta_t(\omega_i) = p(\mathbf{v}_t \,|\, \mathbf{x}, \phi^{[q]}) = p(\mathbf{r}_t, \mathbf{s}_t = \omega_i \,|\, \mathbf{x}, \phi^{[q]})$$

Les formules de ré-estimation dans le cas des CMT gaussiennes sont :

$$c_{i,j_{SEM}}^{[q+1]} = \frac{1}{T-1}\sum_{t=2}^{T} 1_{[\mathbf{v}_{t-1},\mathbf{v}_t]} \qquad a_{i,j_{ICE}}^{[q+1]} = \frac{\sum_{t=1}^{T}\psi_t^{[q]}(\omega_i,\omega j)}{\sum_{t=1}^{T}\zeta_t^{[q]}(\omega_i)} \tag{2.36}$$

$$\mu_{k_{SEM,ICE}}^{[q+1]} = \frac{\sum_{t=1}^{T}\mathbf{x}_t 1_{[\mathbf{s}_t=\omega_k]}}{\sum_{t=1}^{T}1_{[\mathbf{s}_t=\omega_k]}} \tag{2.37}$$

$$\Gamma_{k_{SEM,ICE}}^{[q+1]} = \frac{\sum_{t=1}^{T}(\mathbf{x}_t-\mu_k^{q+1})(\mathbf{x}_t-\mu_k^{q+1})^{\mathrm{T}}1_{[\mathbf{s}_t=\omega_k]}}{\sum_{t=1}^{T}1_{[\mathbf{s}_t=\omega_k]}} \tag{2.38}$$

Chapitre 3

Application des CMCouple-BI et CMT à la séparation de sources

Les modélisations markoviennes ont prouvé leur grande efficacité dans différentes applications. En particulier, les CMC-BI ont été appliquées avec un grand succès dans l'économétrie et les finances [79, 80], les biosciences [81–83], les communications [84], le traitement des signaux et des images [57, 85, 86].

Dans la segmentation d'images, la modélisation par les CMC-BI est efficace lorsque le bruit n'est pas trop complexe, chaque pixel étant bruité indépendamment des pixels voisins [61]. Or, un pixel donné appartenant à une certaine classe pourrait avoir un aspect visuel différent s'il est proche de la bordure où s'il appartenait à une zone où les pixels sont de la même classe. Dans ce type de problème les dépendances entre les observations et les sources sont souvent plus complexes, et ne peuvent être décrites convenablement par un modèle CMC-BI. La généralisation du modèle CMC-BI au modèle CMCouple-BI [54, 56, 77] permet une modélisation plus adéquate. En effet, dans un modèle CMCouple-BI, le processus couple (Caché $\mathbf{S}$, Observé $\mathbf{X}$) est de Markov et le processus caché $\mathbf{S}$ ne l'est pas nécessairement. Ce modèle est plus général que le modèle CMC-BI et les expérimentations effectuées sur les CMCouples ont abouti à des résultats satisfaisant dans la

segmentation supervisée et non supervisée d'images [54, 55], montrant que cette plus grande généralité pouvait se traduire par une plus grande efficacité.

D'autre part, dans la segmentation d'images comportant un nombre fini de stationnarités, l'intérêt du modèle CMT a été validé par différentes expérimentations [77, 78]. En effet, dans le modèle CMT un troisième processus $\mathbf{R}$ modélisant les différentes stationnarités est introduit, et le processus $(\mathbf{S}, \mathbf{X}, \mathbf{R})$ est alors markovien. Ce modèle est plus général que le modèle CMCouples car le processus couple $(\mathbf{S}, \mathbf{X})$ n'est pas nécessairement de Markov. CMT est très riche et donne naissance à plusieurs modèles particuliers qui ont été appliqués dans différentes modélisations complexes comme chaînes semi-markoviennes cachés ou chaînes semi-markoviennes cachés non-stationnaires [73, 74].

Nous proposons d'exploiter les propriétés intéressantes des modèles CMCouples-BI et CMT dans un contexte de séparation de processus à états discrets. Nous avons un nombre N d'images observées issues de mélanges bruitées de Q images binaires que nous désirons retrouver. Chaque image est balayée par un parcours de Hilbert-Peano. Nous obtenons alors un processus observé $\mathbf{X}$ dont chaque vecteur $\mathbf{x}_t$ est composé de N composantes correspondant à N images à séparer. Aucune hypothèse sur les dépendances entre les composantes des vecteurs $\mathbf{s}_t$, qui constituent le processus caché, n'est considérée. L'hypothèse d'indépendance statistique des sources, exigée par les méthodes classiques de séparation de sources, est alors relâchée, ce que va nous permettre de séparer des images où les dépendances sont plus complexes.

Nous avons présenté dans le chapitre précédent les différentes techniques et algorithmes utilisés dans la segmentation par les CMCouples-BI et CMT dans le cas de processus multidimensionnel. Dans ce chapitre nous proposons d'appliquer ces algorithmes, dans un premier temps, à la séparation d'images réelles bruitées synthétiquement. Dans un deuxième temps, nous allons considérer des images réelles de manuscrits scannés affectées d'un effet de transparence et de diffusion d'encre. Nous réaliserons une segmentation non supervisée fondée sur les modèles CMCouples-BI et CMT que nous comparerons avec les résultats de segmentation fondée sur le modèle CMC-BI.

3.1 Modèle du mélange

On considère que l'on a accès à une version mélangée $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ de taille $N \times T$ de données inconnues $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_T)$, soit $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_T)$ un bruit additif. A tout instant t les observations sont données par :

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{b}_t \tag{3.1}$$

Le mélange correspond à un modèle linéaire instantané où $\mathbf{A}$ est une matrice de taille $N \times Q$. N est la taille du vecteur observation $\mathbf{x}_t$ à tout instant t, et Q la taille du vecteur $\mathbf{s}_t$ à tout instant t.

3.2 Parcours de Hilbert-Peano

L'application des modèles de chaînes de Markov dans la segmentation ou la séparation d'images nécessite la transformation d'une image (bidimensionnelle) en une chaîne monodimensionnelle, ce qui peut être réalisé avec un parcours de Hilbert—Peano représenté sur la figure 3.1. Ce parcours fractal

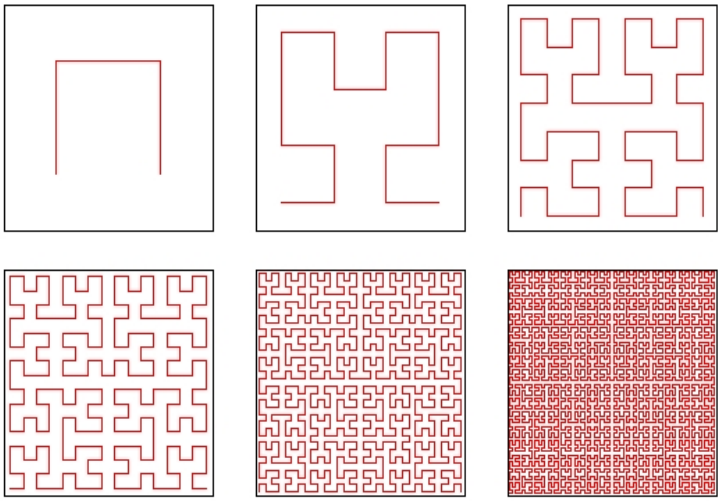


FIGURE 3.1 – Parcours Hilbert-Peano

est obtenu en reproduisant un élément de base, dit "générateur". Conformément à la figure 3.1, le premier instant de la chaîne de Markov correspond au pixel de la dernière ligne et première colonne, le deuxième instant correspond au pixel de l'avant dernière ligne et première colonne, le troisième

instant correspond au pixel de l'avant dernière ligne et deuxième colonne et ainsi de suite. Nous rappelons qu'il existe d'autres types de parcours, à titre d'exemple, le parcours "ligne par ligne" ou "colonne par colonne" qui sont aussi simples et faciles à réaliser. Néanmoins, le parcours d'Hilbert-Peano donne des meilleurs résultats, en segmentation d'images. En effet, avec une courbe d'Hilbert-Peano, les pixels appartenant à la même région de l'image restent, globalement, voisins dans la chaîne [61, 87], ce qui constitue l'intérêt principal de ce parcours par rapport aux parcours simples de type "ligne par ligne" ou "colonne par colonne". Chaque image est balayée et transformée en chaîne monodimensionnelle. L'ensemble des chaînes obtenues est considéré comme une réalisation multidimensionnelle bruitée d'une chaîne de Markov.

Dans ce qui suit, nous proposons donc de comparer les performances des algorithmes fondés sur les modèles CMCouples-BI et CMT avec celles d'un algorithme basé sur le modèle CMC-BI. Nous considérons en premier, des chaînes simulées, nous enchaînerons avec des images réelles bruitées synthétiquement. Nous allons ensuite appliquer nos algorithmes aux images issues de manuscrits scannées. Ces simulations valident l'efficacité de nos algorithmes par rapport à ceux fondé sur un modèle CMC-BI.

3.3 Applications des CMCouples

Nous considérons un processus caché à quatre classes $\Omega = (\omega_1, \ldots, \omega_4)$. Les vecteurs cachés et observés étant constitués de deux éléments $N = Q = 2$, chaque élément du vecteur $\omega_i = (\omega_i^1 \quad \omega_i^2)^{\mathrm{T}}$ est supposé prendre une valeur binaire ± 1. Dans ce cadre multidimensionnel, nous allons comparer les performances de la modélisation par les CMCouples-BI par rapport au modèle CMC-BI. On note que l'efficacité du modèle CMCouple a été validée dans le cas de la segmentation de données monodimensionnelles [54]. Nous visons alors à valider l'efficacité de ce modèle dans le contexte de séparation de données à nombre d'états fini et à dépendances complexes, pour lesquelles le modèle CMC-BI pourrait avoir des faiblesses. Dans ce qui suit, nous allons réaliser la segmentation sur des données multidimensionnelles avec différentes structures temporelles pour les sources, ainsi que différents modèles de bruit.

3.3.1 Application aux chaînes synthétiques

Afin de valider l'intérêt de la segmentation par un modèle CMCouple-BI dans un problème de séparation de données à états discrets, nous allons comparer les performances de la segmentation par modèle CMCouple-BI par rapport au modèle CMC-BI. En effet, nous générons des processus synthétiques avec différentes structures temporelles et différents types de dépendances liant les données cachées $\mathbf{S}$ aux observations $\mathbf{X}$. L'idée est de mettre en évidence l'apport d'une modélisation par CMCouples-BI par rapport à un modèle CMC-BI sur des cas où le modèle CMC-BI pourrait être insuffisant. Nous allons alors générer des données qui ne vérifient pas nécessairement le modèle CMCouple-BI ni le modèle CMC-BI. Dans nos chaînes simulées les processus cachés $\mathbf{S}$ sont générés en premier, ils sont i.i.d ou de Markov. Ces sources sont ensuite mélangées par un opérateur linaire $\mathbf{A}$, et bruitées par deux bruits de structures différentes.

Modèle du Mélange

Nous générons les chaînes avec les structures temporelles et les bruits décrits ci-dessous avec $Q = N = 2$, nous avons pour chaque expérimentation une chaîne de taille $2 \times T$. Nous mélangeons les deux sous chaînes $\mathbf{S}^1$ et $\mathbf{S}^2$ selon l'équation (3.1) ($\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$) avec $\mathbf{A}$ une matrice carrée de mélange :

$$\mathbf{A} = \begin{pmatrix} 0.8 & 0.7 \\ 0.7 & 0.8 \end{pmatrix}$$

Ensuite nous rajoutons, indépendamment du mélange, soit le bruit i.i.d ou le bruit dit "bruit CMCouple-BI" que nous définissons dans ce qui suit. Nous avons, pour le processus global $\mathbf{Z} = (\mathbf{S}, \mathbf{X})$, quatre modèles de données que nous traiterons par nos algorithmes :

a Données 1 : sources i.i.d et bruit i.i.d

Dans une première expérimentation on considère un modèle de données tel que les sources $\mathbf{S}$ sont temporellement indépendantes et le bruit $\mathbf{B}$ est i.i.d. La loi du processus $\mathbf{Z}$ est alors donnée par :

$$p(\mathbf{Z}) = p(\mathbf{S}, \mathbf{X})$$

$$= p(\mathbf{S})p(\mathbf{X}|\mathbf{S})$$
$$= \prod_{t=1}^{T} p(\mathbf{s}_t)p(\mathbf{x}_t|\mathbf{s}_t).$$

b Données 2 : sources de Markov et bruit i.i.d

Dans une deuxième expérimentation, les sources sont générées selon une distribution de Markov. Le bruit est i.i.d, nous avons :

$$p(\mathbf{Z}) = p(\mathbf{s}_1)p(\mathbf{x}_1|\mathbf{s}_1)\prod_{t=2}^{T} p(\mathbf{s}_t|\mathbf{s}_{t-1})p(\mathbf{x}_t|\mathbf{s}_t) \tag{3.2}$$

Selon la proposition définissant les CMC-BI, ce modèle correspond bien à une CMC-BI : $\mathbf{S}$ est de Markov et $p(\mathbf{x}_t|\mathbf{S}) = p(\mathbf{x}_t|\mathbf{s}_t)$. Notons que pour les données 1 et 2 la loi $p(\mathbf{x}_t|\mathbf{s}_t)$ sera supposée gaussienne et sera donc caractérisée par une moyenne μ et une covariance Γ.

c Données 3 : sources i.i.d et bruit CMCouple-BI

Nous parlons de "bruit CMCouple-BI" lorsque la loi $p(\mathbf{X}|\mathbf{S})$ vérifie :

$$p(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1}) = p(\mathbf{x}_t \,|\, \mathbf{s}_t, \mathbf{s}_{t+1})p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_t, \mathbf{s}_{t+1})$$

La loi du bruit à tout instant t est alors donnée par :

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{s}_{t-1}, \mathbf{s}_t) = \frac{p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{s}_{t-1}, \mathbf{s}_t)}{\sum_{x_t} p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{s}_{t-1}, \mathbf{s}_t)} \tag{3.3}$$

La structure du "bruit CMCouple-BI" est assez complexe : lorsque l'on se place dans le cas gaussien, la loi du bruit $p(\mathbf{x}_t|\mathbf{s}_t)$ est un mélange d'un très grand nombre de distributions gaussiennes. Le modèle global obtenu est alors donné par :

$$p(\mathbf{Z}) = p(\mathbf{s}_1)\prod_{t=2}^{T} p(\mathbf{s}_t)p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{s}_t, \mathbf{s}_{t-1}) \tag{3.4}$$

Ce modèle est assez particulier dans lequel la dépendance temporelle entre sources n'est pas explicite. Les liens entre les $\mathbf{s}_t$ sont assurés à travers le bruit. On remarque que lorsqu'à tout instant t la loi $p(\mathbf{x}_t|\mathbf{s}_t)$ modélisant le bruit est une simple distribution gaussienne, nous retrouvons le modèle 1.

d Données 4 : sources de Markov et bruit CMCouple-BI

Dans la quatrième expérimentation, la loi des sources $\mathbf{S}$ est markovienne et le bruit est un "bruit CMCouple-BI". La distribution du processus $\mathbf{Z}$ est la suivante :

$$p(\mathbf{Z}) = p(\mathbf{s}_1) \prod_{t=2}^{T} p(\mathbf{s}_t|\mathbf{s}_{t-1}) p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{s}_t, \mathbf{s}_{t-1}) \tag{3.5}$$

Ce modèle de données est assez difficile à traiter, la structure de données ainsi que le bruit étant relativement complexes. Nous rappelons que pour le cas gaussien que nous traitons, le bruit à tout instant t $p(\mathbf{x}_t|\mathbf{s}_t)$ est un mélange d'un grand nombre gaussiennes. Lorsque le bruit se réduit à une seule gaussienne on retrouve le modèle 2 correspondant à une CMC-BI.

e Chaînes synthétiques : séparation non supervisée

Notre méthode de séparation pour le cas de données à états discrets consiste à effectuer une segmentation non supervisée par un algorithme basé sur le modèle CMCouple-BI. Nous comparons ses performances par rapport à l'algorithme basé sur le modèle CMC-BI. Pour les deux algorithmes la segmentation est réalisée au sens du MPM. L'estimation des paramètres est réalisée avec les algorithmes EM et ICE adaptés à chaque modèle. Nous présentons les résultats de la manière suivante : pour chaque structure temporelle des sources nous comparons les performances des modèles CMCouples-BI et CMC-BI selon la nature du bruit.

Dans le tableau 3.1, nous considérons d'une part, une chaîne vérifiant le modèle 1 : la chaîne est i.i.d, le bruit est gaussien indépendant. Les observations sont indépendantes temporellement conditionnellement aux sources ; de plus, à tout instant t le vecteur observation $\mathbf{x}_t$ dépend seulement de $\mathbf{s}_t$. Et d'autre part, une chaîne vérifiant le modèle 3 : les sources sont i.i.d tandis que le bruit est un "bruit CMCouple-BI".

Dans le tableau 3.2, les sources sont générées selon une distribution markovienne, les deux types de bruits (i.i.d et “bruit CMCouple-BI") sont considérés. Notons que malgré la structure du “bruit CMCouple" inspiré du modèle des CMCouples, le processus $\mathbf{Z}$ généré n'est pas une CMCouple-BI. En effet, dans une CMCouple-BI le processus $\mathbf{S}$ ne doit pas être une chaîne de

Markov. Or, dans nos données considérées le processus **S** est soit de Markov soit i.i.d qui est un cas particulier de chaînes de Markov.

Nous allons, dans ce qui suit, comparer d'une part, les performances de l'algorithme EM par rapport à l'algorithme ICE dans le cadre du modèle de CMCouple-BI. Et d'autre part, nous comparerons les taux de composantes mal segmentées dans le modèle CMCouple-BI par rapport au modèle CMC-BI. La restauration est complètement non supervisée, tous les paramètres sont considérés inconnus. Les résultats sont présentés dans les tableaux 3.1 et 3.2 pour des échantillons de taille $T = 2000$ et pour 25 itérations. Les paramètres des modèles ont été choisis aléatoirement.

Méthode	*Données 1* (s^1, s^2)	*Données 3* (s^1, s^2)
CMCouple-BI-ICE	$(22.7, 22.9)$	$(34.7, 32.4)$
CMCouple-BI-EM	$(20.0, 19.4)$	$(34.6, 32.5)$
CMC-BI-ICE	$(20, 8, 20.4)$	$(35.5, 36.4)$

TABLE 3.1 – Taux de sites mal classés pour une chaîne iid en %

Pour les *Données 1*, les taux de sites mal classés obtenus avec un modèle CMCouple-BI et CMC-BI sont comparables (environ 20%). En effet, les sources et le bruit sont i.i.id, aucune corrélation temporelle n'est présente. Nous remarquons que malgré la complexité de l'étape de l'estimation de paramètres dans la segmentation avec les CMCouples-BI par rapport aux CMC-BI (le nombre de paramètres à estimer pour les CMCouples-BI est plus important que dans les CMC-BI), le modèle CMCouple-BI permet une bonne estimation de paramètres. Il est ainsi plus riche que les CMC-BI. D'autre part, dans l'expérimentation avec les *Données 3* le bruit est "CMCouple-BI". La segmentation par les CMCouple-BI est plus avantageuse (environ 32%) par rapport aux CMC-BI (environ 36%) qui ne permet pas de prendre compte de la complexité du bruit. Notons que les taux de sites mal classés dans les *Données 3* sont globalement plus élevés que pour un bruit i.i.d, ceci est dû à la nature complexe du bruit : nous rappelons que le "bruit CMCouple-BI" dans le cas gaussien est un mélange d'un grand nombre de gaussiennes. Nous remarquons aussi les performances des algorithmes basés sur le modèle CMCouple-BI avec l'algorithme EM et ICE sont comparables.

Méthode	*Données 2* (s^1, s^2)	*Données 4* (s^1, s^2)
CMCouple-BI-ICE	$(12.3, 11.8)$	$(29.6, 31.9)$
CMCouple-BI-EM	$(11.3, 11.7)$	$(29.7, 31.3)$
CMC-BI	$(12.2, 11.3)$	$(35.6, 38.2)$

TABLE 3.2 – Taux de sites mal classés pour une chaîne de Markov cachée en %

Pour les résultats présentés dans le tableau 3.2, les sources sont de Markov. Et pour les *Données 2* le bruit est i.i.d, le processus simulé est une CMC-BI. Nous remarquons que les taux de sites mal classés obtenus par les modèles CMCouple-BI et CMC-BI sont quasiment identiques (environ 11%) malgré le nombre important de paramètres à estimer pour les CMCouples-BI. Ce modèle est alors riche et robuste, il généralise le modèle des CMC-BI [54] et permet une bonne estimation de paramètres ainsi qu'une bonne qualité de segmentation.

La richesse du modèle CMCouple-BI par rapport aux CMC-BI se manifeste remarquablement lorsque le bruit n'est plus simple. En effet, pour les *Données 4* dans lesquels nous avons un "bruit CMCouple-BI" le taux de sites mal classés avec un modèle CMC-BI est plus important (environ 38%) qu'avec le modèle CMCouple-BI (environ 31%), qui permet une meilleure modélisation.

3.3.2 Séparation d'images réelles bruitées synthétiquement

Cette section est consacrée à la séparation d'images réelles (noir et blanc) mélangées et bruitées synthétiquement. Les pixels de chaque image sont supposés prendre une valeur ± 1 représentant les deux couleurs noir et blanc. Nous avons alors des processus à états discrets dont la séparation revient à faire une segmentation. Nous allons réaliser une segmentation non supervisée par un modèle CMCouple-BI gaussien. De la même manière que dans [54, 61, 78], les images sont transformées en chaînes en utilisant le parcours de Hilbert-Peano.

Nous considérons deux images initiales en noir et blanc (figure 3.2) issues de documents réels numérisés. Les images sont alors transformées en chaînes, et les deux chaînes obtenues représentent les sources $\mathbf{S}$. Elles sont ensuite

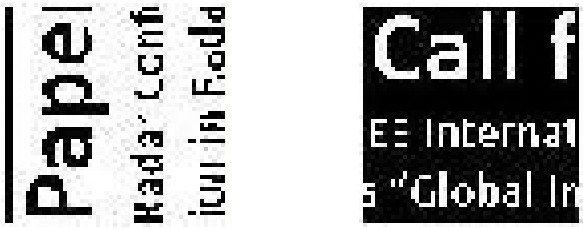


FIGURE 3.2 – Images originales (128×128 pixels)

mélangées par la même matrice $\mathbf{A}$ que dans la section précédente :

$$\mathbf{A} = \begin{pmatrix} 0.8 & 0.7 \\ 0.7 & 0.8 \end{pmatrix}$$

L'équation du mélange est donnée par l'équation (3.1) ($\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{b}_t$), où $\mathbf{b}$ est donc un bruit additif. Nous réalisons deux expériences avec deux bruits différents : dans la première expérience nous considérons un "bruit CMCouple-BI" défini précédemment dans ce chapitre, et dans la deuxième expérience nous bruitons les images préalablement mélangées et avec un "bruit corrélé" que nous définissons plus loin. Nous comparons les performances du modèle CMCouple-BI par rapport au modèle CMC-BI. Les sources sont restaurées avec la technique du MPM, et les paramètres sont estimés avec l'algorithme ICE.

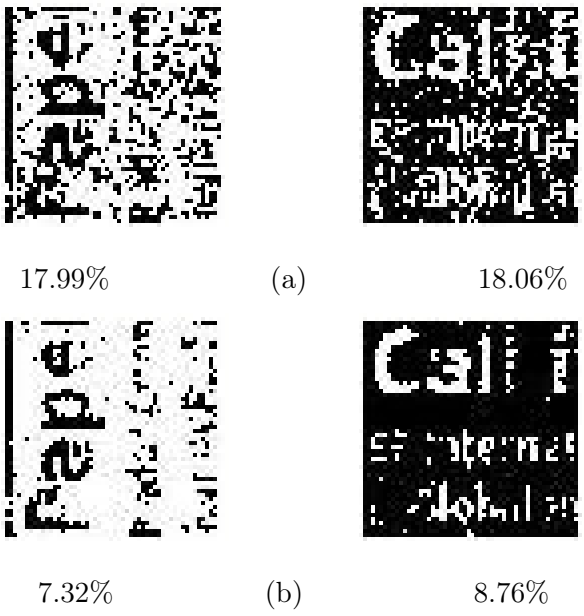
a Séparation de deux images affectées par le "bruit CMCouple-BI"

Les images originaux (figure 3.2) sont d'abord transformées en chaînes à l'aide du parcours Hilbert-Peano. Les chaînes obtenues sont mélangées avec la matrice $\mathbf{A}$ et affectées d'un "bruit CMCouple-BI" . La figure 3.3 représente les observations obtenues suite au mélange et au bruitage. Nous reconstituons les images 2D par le parcours inverse de la courbe d'Hilbert-Peano.

La segmentation non supervisée par le modèle CMCouple-BI offre une bonne restauration des images initiales avec un taux de sites mal classés de seulement 7%, tandis qu'avec le modèle CMC-BI, il est de plus de 17%. visuellement aussi la qualité de séparation avec un modèle CMCouple-BI (figure 3.4(b)) est nettement meilleure que celle obtenue par le modèle CMC-BI

FIGURE 3.3 – Mélange affecté d'un bruit CMCouple-BI

(figure 3.4(a)). Les résultats de segmentation montre la grande efficacité de la modélisation par les CMCouples-BI par rapport au CMC-BI lorsque la structure du bruit est assez complexe. En effet, le modèle CMC-BI n'exploite pas toute l'information sur les dépendances complexes entre sources et bruit ce qui se traduit par le taux élevé de sites mal classés, alors que les CMCouples-BI tiennent mieux compte des liens probabilistes complexes entre les pixels et entre les pixels et le bruit.



17.99% (a) 18.06%

7.32% (b) 8.76%

FIGURE 3.4 – Images séparées et taux de sites mal classés avec (a) ICE pour CMC-BI, (b) ICE pour CMCouple-BI.

b Séparation de deux images affectées de "bruit corrélé"

Nous souhaitons à présent tester l'efficacité de l'algorithme proposé sur un mélange caractérisé par une structure complexe mais ne vérifiant pas le modèle CMCouple-BI. Nous avons choisi donc de considérer le même

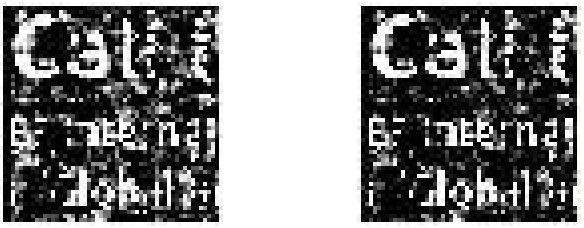
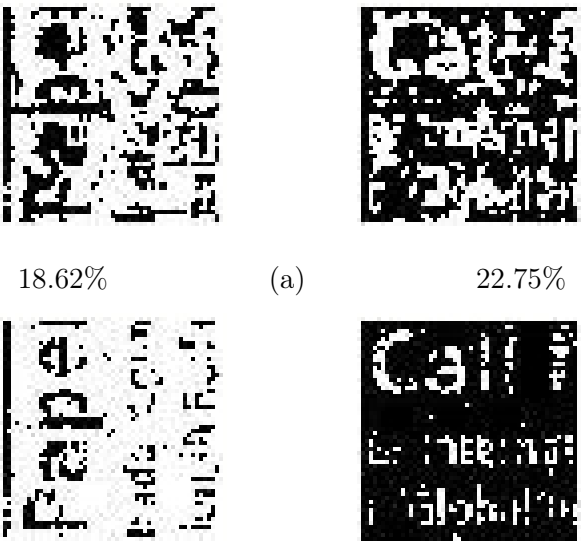


FIGURE 3.5 – Images mélangées et affectées d'un bruit corrélé, a = 0.7



18.62% (a) 22.75%

7.10% (b) 7.49%

FIGURE 3.6 – Images séparées et taux de sites mal classés avec (a) ICE pour CMC-BI, (b) ICE pour CMCouple-BI.

mélange que dans la précédente expérience, mais cette fois le bruit, corrélé, est (figure 3.5) généré selon l'expression suivante :

$$b(r) = \frac{1}{1+4a}[\varepsilon(r) + a(\sum_{i=1}^{4} \varepsilon(r_i))]$$

où $\varepsilon(r)$ est bruit gaussien iid, $\varepsilon(r_i)$ $i = 1, \ldots 4$ est la valeur du bruit des 4 plus proches voisins de r, et a est un paramètre donné.

Cette fois encore la segmentation par un modèle CMCouple-BI (figure 3.6(b)) est plus avantageuse avec un taux d'erreur de l'ordre de 7%. En effet, le taux de sites mal classé obtenu avec l'algorithme basé sur le modèle CMC-BI (figure 3.6(a)) est beaucoup plus élevé et se situe autour de 20%. Une nouvelle fois nous constatons donc que le modèle CMC-BI est bien moins efficace que le modèle CMCouple-BI pour prendre en compte des bruits complexes. La différence entre les performances des deux modèles nous paraît d'autant plus intéressante que les données sont très complexes et ne correspondent à aucun des deux modèles.

3.4 Applications des CMT aux chaînes synthétiques

Ainsi que nous l'avons introduit dans le chapitre précédent, le modèle des chaînes de Markov triplets CMT, qui est une généralisation du modèle CMCouple-BI, permet, entre autres, de prendre en compte les changements de régime dans le cas général du couple (caché $\mathbf{S}$, observé $\mathbf{X}$). En effet, ces changements peuvent être modélisés par l'introduction d'un processus auxiliaire $\mathbf{R}$. Pour illustration nous considérons que chaque élément du processus $\mathbf{R}$ prend ses valeurs dans l'ensemble des trois états $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$. Les quatre états du processus $\mathbf{S}$ prennent leurs valeurs dans l'ensemble $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, chaque état ω_i est un vecteur contenant deux éléments :

$$\omega_i = \begin{pmatrix} \omega_i^1 \\ \omega_i^2 \end{pmatrix}, \text{tel que } \omega_i^l \text{ prend la valeur } \pm 1$$

Nous avons donc quatre classes et trois différentes stationnarités pour la chaîne de Markov $\mathbf{S}$. Notons que la loi du processus joint $\mathbf{V} = (\mathbf{R}, \mathbf{S})$ est, dans le cas général, une loi marginale d'une chaîne de Markov $(\mathbf{R}, \mathbf{S}, \mathbf{X})$, elle peut donc être de Markov ou pas. De même, la chaîne $\mathbf{R}$ peut être de Markov ou pas.

Nous rappelons que le modèle CMT que nous considérons est un modèle particulier simple des chaînes de Markov triplets. Ce modèle est défini, comme nous l'avons vu dans le chapitre précédent, par des dépendances

relativement simples entres les processus $\mathbf{R}$, $\mathbf{S}$ et $\mathbf{X}$. En particulier, les processus $\mathbf{R}$ et $\mathbf{V} = (\mathbf{R}, \mathbf{S})$ sont des chaînes de Markov. Le processus global $\mathbf{Y} = (\mathbf{R}, \mathbf{S}, \mathbf{X})$ est généré de la manière suivante. Le processus auxiliaire $\mathbf{R}$ est généré en premier, il est donné par une distribution de Markov à trois états caractérisée par une loi initiale $\pi_{\mathbf{R}} = (1/3, 1/3, 1/3)$ et les transitions suivantes :

$$\mathbf{M_R} = \begin{pmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.99 & 0.01 \\ 0.01 & 0 & 0.99 \end{pmatrix}$$

La chaîne cachée $\mathbf{S}$ est simulée conditionnellement au processus auxiliaire $\mathbf{R}$ à partir de :

$$p(\mathbf{s}_1/\mathbf{r}_1) = (1/4, 1/4, 1/4, 1/4) \tag{3.6}$$

$$p(\mathbf{s}_{t+1}/\mathbf{s}_t, \mathbf{r}_{t+1} = \lambda_i) = \mathbf{M}_{\lambda_i} \tag{3.7}$$

avec :

- $\mathbf{M}_{\lambda_1} = \begin{pmatrix} 0.5000 & 0.1667 & 0.1667 & 0.1667 \\ 0.1667 & 0.5000 & 0.1667 & 0.1667 \\ 0.1667 & 0.1667 & 0.5000 & 0.1667 \\ 0.1667 & 0.1667 & 0.1667 & 0.5000 \end{pmatrix}$

- $\mathbf{M}_{\lambda_2} = \begin{pmatrix} 0.8000 & 0.0667 & 0.0667 & 0.0667 \\ 0.0667 & 0.8000 & 0.0667 & 0.0667 \\ 0.0667 & 0.0667 & 0.8000 & 0.0667 \\ 0.0667 & 0.0667 & 0.0667 & 0.8000 \end{pmatrix}$

- $\mathbf{M}_{\lambda_3} = \begin{pmatrix} 0.0667 & 0.8000 & 0.0667 & 0.0667 \\ 0.8000 & 0.0667 & 0.0667 & 0.0667 \\ 0.0667 & 0.0667 & 0.0667 & 0.8000 \\ 0.0667 & 0.0667 & 0.8000 & 0.0667 \end{pmatrix}$

Les observations sont échantillonnées par rapport à $p(\mathbf{x}_t/\mathbf{s}_t)$ où les densités $p(\mathbf{x}_t/\mathbf{s}_t = \omega_i)$ sont des gaussiennes.

La restauration non supervisée du signal caché est effectuée à l'aide de l'algorithme de reconnaissance MPM en utilisant les paramètres estimés par les

algorithme ICE et SEM initialisés à partir d'un algorithme de K-moyennes. Nous nous intéressons au cas dans lequel $\mathbf{V}$ est une chaîne de Markov, $\mathbf{S}$ correspond à une chaîne non stationnaire. La loi de $\mathbf{V} = (\mathbf{R}, \mathbf{S})$ est donnée par : $p(\mathbf{v}) = p(\mathbf{v}_1) \prod_{t=1}^{t-1} p(\mathbf{v}_{t+1}/\mathbf{v}_t)$ telle que :

$$P(\mathbf{v}_{t+1}/\mathbf{v}_t) = p(\mathbf{s}_{t+1}/\mathbf{s}_t, \mathbf{r}_{t+1}) p(\mathbf{r}_{t+1}/\mathbf{r}_t) \tag{3.8}$$

Ensuite $\mathbf{S}$ et $\mathbf{X}$ sont échantillonnés à partir des équations (3.6) et (3.7). Les processus cachés $\mathbf{S}$ et $\mathbf{X}$ sont représentés sur les figures 3.7 et 3.8 ci-après.

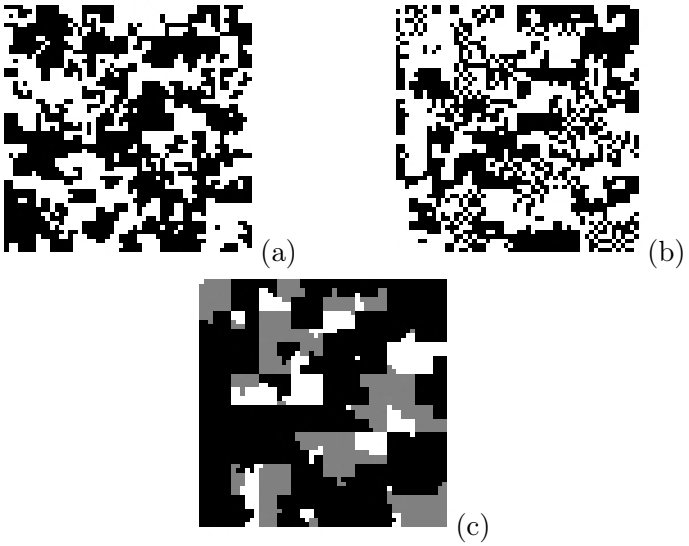


FIGURE 3.7 – (a) Simulation de la première composante de $\mathbf{S}$, (b) simulation de la deuxième composante de $\mathbf{S}$,(c) simulation du processus auxiliaire $\mathbf{R}$

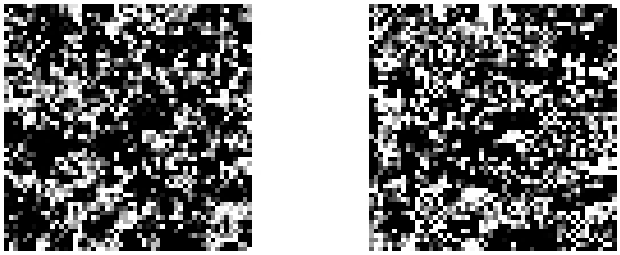


FIGURE 3.8 – Simulation de $\mathbf{X}$ à partir de distributions gaussiennes

La restauration du processus caché est réalisée au sens du MPM avec l'algorithme basé sur le modèle CMC-BI (figure 3.9), et avec l'algorithme

basé sur le modèle CMT (figure 3.10). Les paramètres sont estimés par l'algorithme SEM initialisé à partir d'un algorithme de K-moyennes.

FIGURE 3.9 – Restauration de **S** en considérant le modèle CMC-BI

FIGURE 3.10 – Restauration de **S** en considérant le modèle CMT

Nous remarquons qu'avec un modèle CMT nous gagnons en qualité de séparation. En effet, avec le modèle CMC-BI le taux moyen de sites mal classés est alentour de 8%, alors qu'avec le modèle CMT le taux de sites mal classés est autour de 5%.

3.5 Séparation des images réelles

Dans cette partie, nous nous intéressons à la restauration non supervisée d'images issues de documents manuscrits scannés ou encore d'images issues de photographies. En effet, suite à une dégradation due à l'effet de transparence ou effet de mélange, il est souvent difficile d'exploiter ces images. Nous nous intéressons, en particulier, aux images en noir et blanc. Le traitement de tels types d'images est assez complexe du fait du manque d'informations *a priori* et d'images initiales avant dégradation. La modélisation de ce problème nécessite la prise en compte de la complexité des dépen-

dances entre pixels voisins d'une part, et des liens entre les deux images (recto/verso), d'autre part. Une telle situation ne peut être modélisée par un modèle simple (en l'occurrence le modèle CMC-BI) ; néanmoins, comme nous l'avons vu dans les sections précédentes, le modèle CMCouple permet une modélisation assez complète pour des données à structures complexes. Par ailleurs, une deuxième solution serait de considérer les images comme des données non-stationnaires, ce qui peut être traité grâce au modèle CMT. À partir de là nous proposons d'appliquer nos algorithmes fondés sur les modèles CMCouple et CMT pour rendre exploitable des images ayant subi des dégradations par effet de mélange, de transparence ou de diffusion d'ancre. Notre méthode est un outil automatique (aucun paramètre n'est fixé manuellement), et relativement rapide par rapport aux algorithmes basés sur les méthodes MCMC (Markov Chain Monte Carlo). Le déroulement de notre méthode est le suivant :

1. chaque image (noir et blanc) est transformée en chaîne à états discrets (chaque élément de la chaîne prend la valeur ± 1) grâce au parcours Hilbert-Peano,
2. les paramètres des modèles sont initialisés par les valeurs indiquées ci-dessous,
3. la segmentation est effectuée par la méthode du MPM,
4. les paramètres sont estimés par l'algorithme ICE.

Initialisations

Les paramètres à estimer sont initialisés aux valeurs suivantes :

– **CMC-BI :**

$$\mu = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

– **CMCouple-BI** :

$$\mu^1 \Rightarrow \mu^1_{1:4} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \mu^1_{5:8} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}; \mu^1_{9:12} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}; \mu^1_{13:16} = \begin{bmatrix} -1 \\ -1 \end{bmatrix};$$

$$\mu^2 \Rightarrow \mu^1_{1:4} = \mu^1_{5:8} = \mu^1_{9:12} = \mu^1_{13:16} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

$$\Gamma^1 = \Gamma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

– **CMT** :

$$\mu \Rightarrow \mu^1_{1:2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \mu^1_{3:4} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}; \mu^1_{5:6} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}; \mu^1_{7:8} = \begin{bmatrix} -1 \\ -1 \end{bmatrix};$$

$$\Gamma_{1:8} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Nous allons, dans ce qui suit, comparer les résultats obtenus en appliquant les algorithmes fondés sur les CMCouples-BI et CMT avec ceux obtenus par un modèle CMC-BI. Nous allons dans un premier temps traiter des images issues de documents manuscrits, et nous présenterons ensuite les résultats de séparation d'images issues de photographies par le modèle CMCouple-BI.

3.5.1 Séparation de documents manuscrits

Dans cette expérimentation, nous considérons des images réelles issues d'un manuscrit scanné présentées sur les figures 3.11 et 3.13 avec $T = 128 \times 128$. Le manuscrit est affecté d'un effet de transparence, les écritures sur chacune des faces du manuscrit est visible sur l'autre face. Nous prenons deux échantillons avec des textes différents. Chaque image (représentant une face du document) est transformée en chaîne à l'aide du parcours Hilbert-Peano. La restauration est effectuée au sens du MPM. Nous comparons

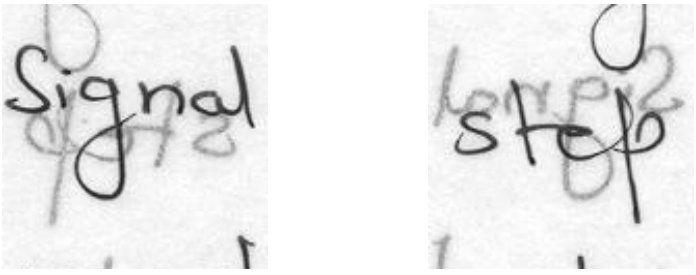


FIGURE 3.11 – **Expérience 1** : Documents manuscrit scannés(//www.site.uottawa.ca/ edubois/documents).

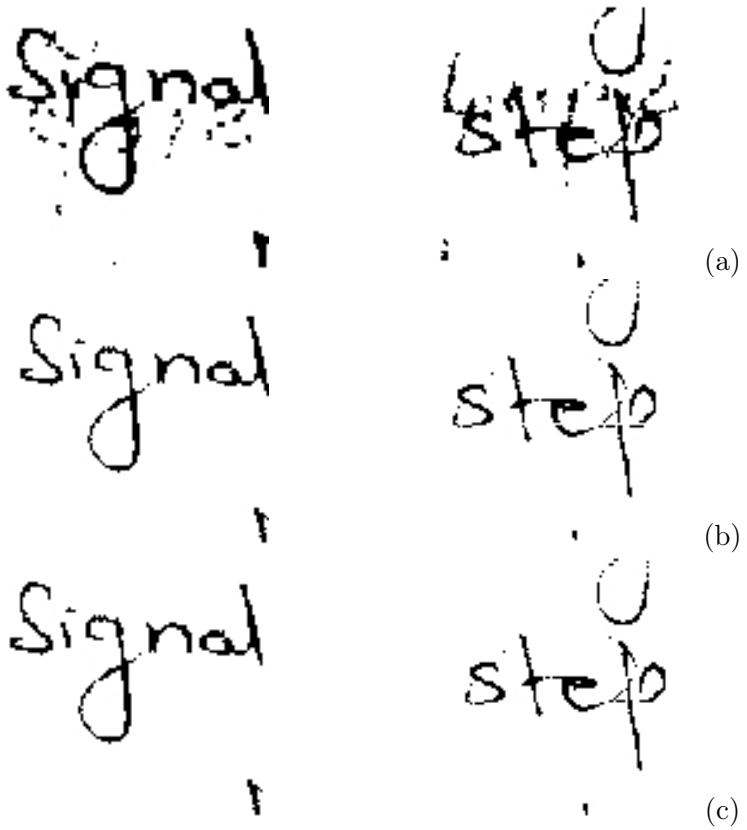


FIGURE 3.12 – Images séparées avec (a) le modèle de CMC-BI , (b) le modèle de CMCouple-BI et (c) le modèle de CMT

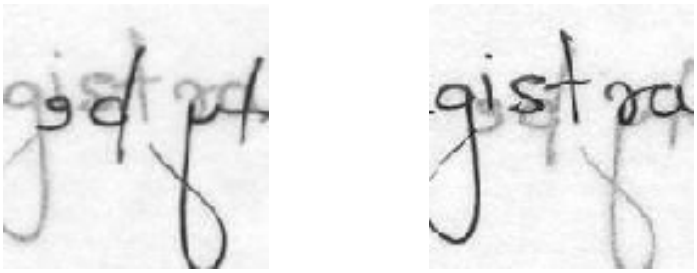
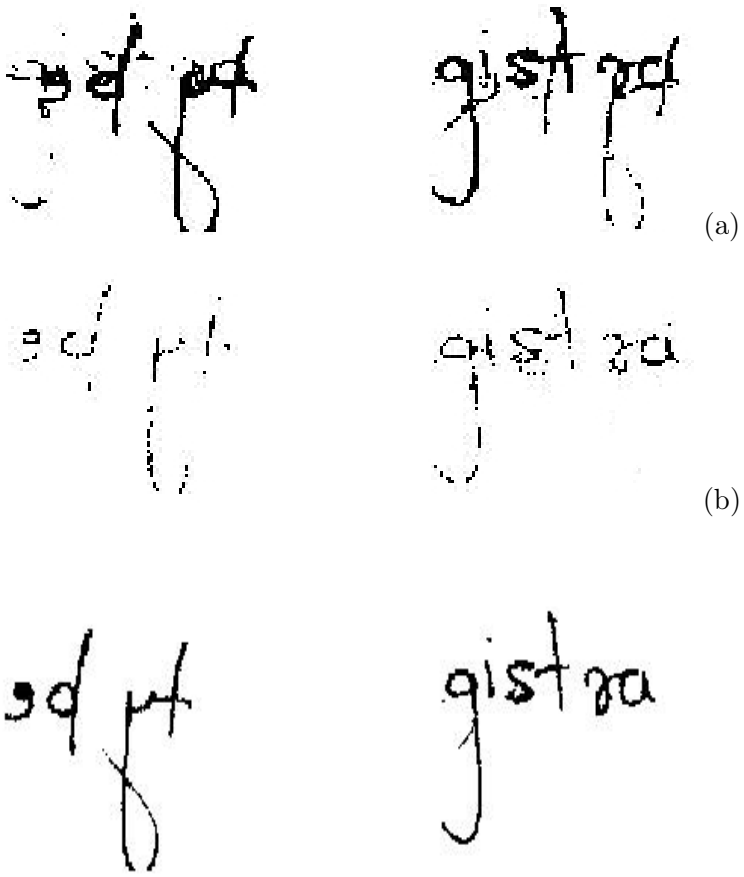


FIGURE 3.13 – **Expérience 2** : Documents manuscrit scannés (://www.site.uottawa.ca/ edubois/documents).



(a)

(b)

(c)

FIGURE 3.14 – Images séparées avec (a) le modèle de CMC-BI, (b) le modèle de CMCouple-BI et (c) le modèle de CMT

l'efficacité des algorithmes basés sur les modèles CMC-BI, CMCouple-BI et CMT pour les deux échantillons.

Dans les algorithmes utilisés nous effectuons une segmentation sur quatre états $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4)$, chaque état prend une valeur binaire ± 1, le nombre d'itérations des algorithmes est fixée à 4 itérations.

Nous remarquons, à partir des figures 3.12(a), (b) et (c) et 3.14(a), (b),(c), qu'avec la modélisation par les CMCouples-BI permet une bonne restauration du texte contrairement au modèle CMC-BI. La prise en compte des dépendances entre pixels voisins offre une bonne segmentation des données. Nous remarquons aussi que le modèle CMT est aussi efficace dans cette situation. En effet, nous pouvons exploiter la modélisation par non-stationnarité pour obtenir une bonne qualité de restauration.

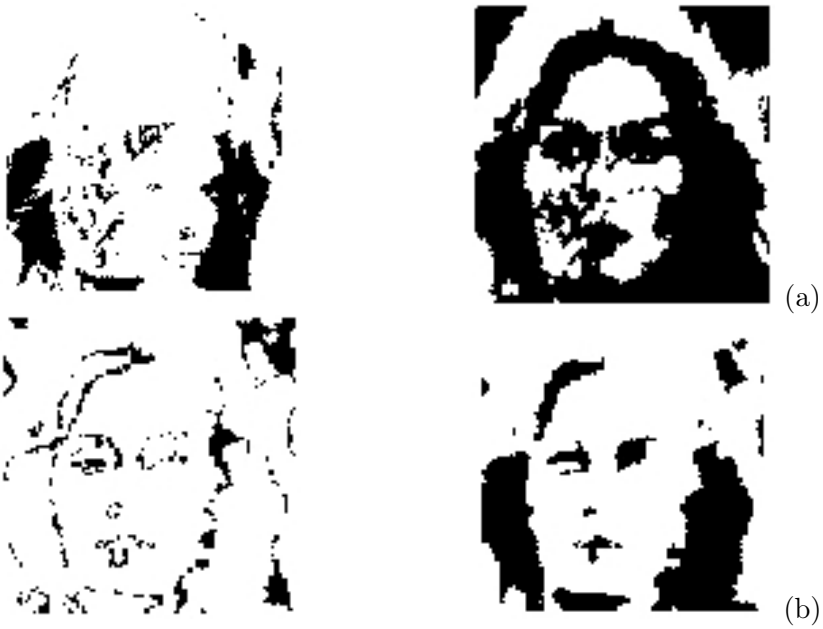
3.5.2 Séparation d'images photographiques

Une troisième expérimentation concerne la segmentation non supervisée d'images issues de photographies 3.15(a), les images mélangées sont représentées sur la figure 3.15(b).

FIGURE 3.15 – *Lena* et *Lynda* images originales

Nous comparons les résultats de séparation obtenus par l'algorithme basé sur le modèle CMC-BI avec ceux obtenus par l'algorithme basé sur le modèle CMCouple-BI. Les images sont en effet en niveau de gris. Nous ne souhaitons pas, à travers cette expérience, tirer de conclusions mais le but était de montrer la capacité d'une modélisation par le modèle CMCouple-BI de traiter des données ne vérifiant pas le modèle. Les paramètres sont initialisés

FIGURE 3.16 – *Lena* et *Lynda* images mélangées.



(a)

(b)

FIGURE 3.17 – Images séparées avec (a) le modèle CMCouple-BI, (b) CMC-BI

aux mêmes valeurs que dans les expériences précédentes.

Les résultats de segmentation sont présentés sur la figure 3.17. Nous remarquons que le modèle CMCouple-BI est plus riche et robuste que le modèle CMC-BI. Il permet de distinguer les deux visages 3.17(a) : le visage de Lena sur la première image et celui de Lynda sur la deuxième. En revanche la segmentation basée sur le modèle CMC-BI ne permet pas de distinguer les deux visages 3.17(a) : nous retrouvons le visage de Lynda sur les deux images.

Chapitre 4

Extension du Modèle d'Analyse en Composantes Indépendantes

Dans ce chapitre est exposée une méthode de séparation de sources permettant de généraliser le modèle d'analyse en composantes indépendantes. Notre but est de relâcher l'une des hypothèses du modèle, à savoir l'hypothèse d'indépendance mutuelle des sources. En effet, nous considérons ici que seul un sous-ensemble des échantillons des sources est effectivement généré de sorte que les composantes soient indépendantes. En d'autres termes, les sources sont générées selon une loi de mélange de distributions : l'un des termes de ce mélange correspond au cas où les données sont indépendantes, tandis que l'autre terme correspond aux données mutuellement dépendantes. Notons que les méthodes classiques de séparation de sources basées sur le modèle d'analyse en composantes indépendantes ne peuvent pas *a priori* séparer des sources avec de telles caractéristiques. Néanmoins, nous allons montrer dans ce chapitre que nous pouvons rendre possible cette séparation en nous basant sur les échantillons à des instants où les hypothèses d'analyse en composantes indépendantes sont vérifiées. Pour ce faire, nous allons utiliser une nouvelle méthode qui combine un algorithme classique d'analyse en composantes indépendantes, et un algorithme d'estimation du processus à variables latentes. Ce dernier repose sur une procédure d'estimation conditionnelle itérative.

4.1 Variables latentes

L'idée de la méthode proposée consiste à introduire un processus caché $(r_t)_{t=1\cdots T}$ qui, selon sa valeur à l'instant t, contrôle l'indépendance ou la dépendance des composantes de $\mathbf{s}_t$. Soit $\mathbf{r} \triangleq (r_1, \ldots, r_t)$ tel que chaque r_t prend ses valeurs dans $\{0, 1\}$. On suppose :

A1. *Conditionnellement à* $\mathbf{r}$*, les variables aléatoires vectorielles* $\mathbf{s}_1, \ldots, \mathbf{s}_t$ *à différents instants sont indépendantes.*

A2. *Conditionnellement à* r_t*, lorsque* $r_t = 0$*, les composantes de* $\mathbf{s}_t$ *sont mutuellement indépendantes et non gaussiennes, sauf éventuellement une au plus ;*

A3. *Conditionnellement à* r_t*, lorsque* $r_t = 1$*, les composantes de* $\mathbf{s}_t$ *sont dépendantes.*

Remarquons que conditionnellement à $r_t = 0$ les composantes de $\mathbf{s}_t$ à un instant t vérifient les hypothèses habituellement exigées par le modèle ACI. On pourrait alors, si le processus $\mathbf{r}$ était connu, appliquer les techniques d'ACI en ignorant les échantillons à des instants où les sources sont dépendantes. Plus précisément, soit :

$$\mathcal{I}_0 \triangleq \{t \in \{1, \ldots, T\} \mid r_t = 0\} \tag{4.1}$$

$$\text{et} \quad \mathbf{X}_0 \triangleq (\mathbf{x}_t)_{t \in \mathcal{I}_0} \tag{4.2}$$

$\mathcal{I}_0$ est l'ensemble des instants où les composantes de $\mathbf{s}_t$ sont indépendantes. Alors, le sous-ensemble $\mathbf{X}_0$ de l'ensemble des observations $\mathbf{X}$ satisfait les hypothèses habituellement exigées par les techniques d'ACI. L'idée de notre travail consiste à effectuer alternativement et de manière itérative une estimation de $\mathbf{B}$ (correspondant à $\mathbf{A}^{-1}$) et des données cachées $\mathbf{r}$.

4.1.1 Modèle des sources étudiées

Dans cette section, nous considérons des sources qui satisfont les hypothèses A1, A2 et A3. Nous proposons d'illustrer ces hypothèses sur un échantillon de vecteurs $\mathbf{s}_t$ contenant deux sources ($N = 2$). Nous présentons nos choix particuliers pour les hypothèses A2 et A3 que nous noterons A2', A3' et A3" et que nous considérerons par la suite dans nos simulations. Soit le cas particulier de l'hypothèse A2, noté A2' et défini par :

A2'. *Lorsque* $r_t = 0$, *les composantes* (s_t^1, s_t^2) *de* $\mathbf{s}_t$ *sont mutuellement indépendantes et uniformément distribuées sur le segment* $[-\sqrt{3}, \sqrt{3}]$ *avec une moyenne nulle et variance unité.*

a Modèle de données : *Exemple 1*

Ce premier exemple est construit à partir de signaux binaires BPSK (Binary Phase Shift Keying). Ces signaux prennent, par définition, les valeurs $s = +1$ ou $s = -1$ de manière équiprobable. Les sources considérées sont un cas particulier de l'hypothèse A3 vérifiant les spécifications suivantes :

A3'. *Lorsque* $r_t = 1$, *les composantes sont telles que :*

$$s_t^1 = a_t \quad et \quad s_t^2 = \epsilon_t a_t$$

où ϵ_t *est une variable aléatoire binaire BPSK et* a_t *est une variable aléatoire à valeur réelle de moyenne nulle et de variance unité.*

On notera que dans ce cas, (s_t^1, s_t^2) est situé sur les bissectrices des axes s_t^1 et s_t^2. Dans les simulations, nous avons choisi pour a_t un mélange de gaussiennes dont la distribution est donnée par :

$$a \sim \frac{1}{2}\mathcal{N}(\mu, \sigma^2) + \mathcal{N}(-\mu, \sigma^2) \tag{4.3}$$

où μ prend une valeur dans $[0, 1]$ et $\sigma^2 = 1 - \mu^2$, tel que :

$$\mathbb{E}\{(s^1)^2\} = \mathbb{E}\{(s^2)^2\} = 1$$

Un exemple illustratif de réalisation de données simulées, et dont les distributions vérifient les hypothèses A1, A2' et A3', est présenté dans la figure 4.1. On peut montrer que les algorithmes, comme l'algorithme CoM2, basés sur maximisation du cumulant d'ordre quatre ne permettent pas la séparation de ce type de données [88, 89]. Néanmoins, si nous considérons le sous-ensemble $\mathbf{X}_0$ ou en d'autre termes, si nous enlevons le nuage de points dépendants dans la figure 4.1, et nous utilisons les données restantes dans une méthode basée sur l'analyse en composantes indépendantes classique, la séparation doit être possible.

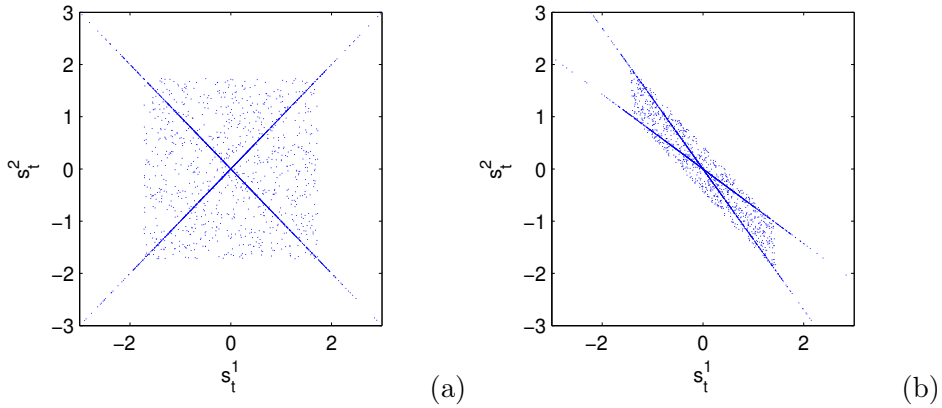


FIGURE 4.1 – (a) Réalisation des sources $\mathbf{s}$ de l'*Exemple 1*. (b) Observations $\mathbf{x} = \mathbf{As}$ correspondantes.

b Modèle de données : *Exemple 2*

L'exemple précédent est un cas extrême, où conditionnellement à $r_t = 1$, nous avons l'une des deux situations :

$$\begin{cases} s_t^1 = s_t^2 \quad \text{ou} \\ s_t^1 = -s_t^2 \end{cases}$$

Ainsi dans l'*Exemple 1*, l'ensemble des échantillons dont les composantes s_t^1 et s_t^2 sont dépendantes se trouve sur les deux bissectrices du plan.

Dans ce nouvel exemple (*Exemple 2*) nous prenons, lorsque $r_t = 1$, un vecteur dépendant $(s_t^1, s_t^2)^{\mathrm{T}}$ distribué selon une densité de probabilité continue. Soit le vecteur aléatoire $\mathbf{u}$ défini par :

$$\mathbf{u} \triangleq \frac{1}{\sqrt{2}} \begin{pmatrix} s_t^1 + s_t^2 \\ s_t^1 - s_t^2 \end{pmatrix} \tag{4.4}$$

Dans cet exemple, l'hypothèse A3 se particularise ainsi :

A3". *Lorsque* $r_t = 1$, $\mathbb{P}(\mathbf{s}_t \,|\, r_t = 1)$ *est telle que les deux composantes* u_i $(i = 1, 2)$ *du vecteur* $\mathbf{u}$ *dans (4.4) sont distribuées selon :*

$$\begin{cases} u_1 \sim \mathcal{N}(0, \sigma) \\ u_2 \sim \mathcal{L}(\lambda) \end{cases}$$

où λ *est un paramètre positif et* $\sigma_\lambda^2 = 2(1 - \frac{1}{\lambda^2})$.

Nous pouvons vérifier que le choix de σ_λ est tel que conditionnellement à $r_t = 1$, les sources sont de variance unité. Par conséquent les données dépendantes et les données indépendantes se superposent en se répartissant sur des domaines de valeurs identiques. De plus, pour $r_t = 1$, chaque composante de $\mathbf{u}$ est modélisée par un mélange de densités de probabilité gaussienne et de Laplace. Un exemple illustratif de cette distribution est donné par des valeurs simulées dans la figure 4.2(a) avec $\lambda = 5$, et dans la figure 4.2(b) avec $\lambda = \sqrt{2}$. Nous rappelons que considérer $\mathbf{X}_0$ revient à enlever le nuage des points dépendants dans les distributions. Visuellement, ceci semble clairement possible au 4.2(a) et moins évident au 4.2(b).

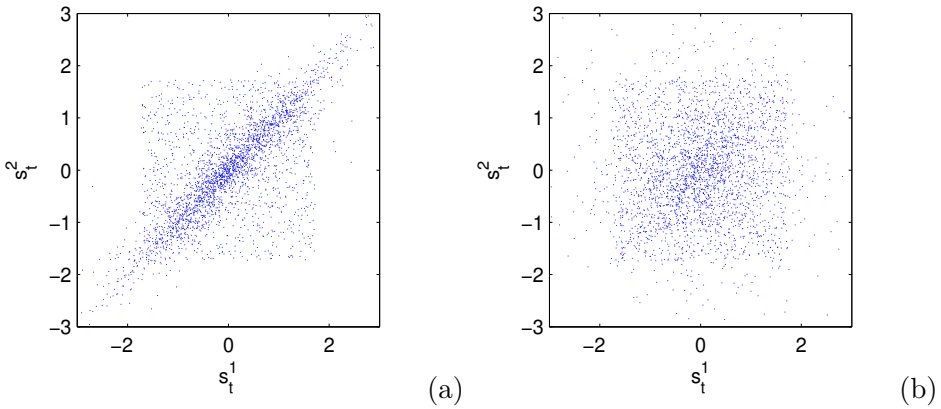


FIGURE 4.2 – Réalisation des sources $\mathbf{s}$ de l'*Exemple 2*. (a) $\lambda = 5$ et (b) $\lambda = \sqrt{2}$.

4.2 Méthode de séparation, le modèle d'ACI étendu

4.2.1 Données complètes et incomplètes

Notons ϕ les paramètres à estimer, c'est-à-dire la matrice de séparation $\mathbf{B}$ et le vecteur de paramètres η qui caractérisent la distribution de $\mathbf{r}$. L'ensemble $(\mathbf{r}, \mathbf{X})$ constitue les données *complètes*, tandis que $\mathbf{X}$ constitue les données *incomplètes* (voir la partie 1.2). On précise que l'adjectif *aveugle* est utilisé pour indiquer que $\mathbf{S}$ n'est pas disponible, tandis que *incomplète* indique que $\mathbf{r}$ est inconnu.

4.2.2 Estimation conditionnelle itérative

L'estimation des paramètres dont η est une étape clé de la méthode que nous proposons. Pour ce faire, nous utilisons l'algorithme d'estimation conditionnelle itérative présenté dans la section 1.4.2.b. Nous rappelons que cet algorithme, noté ICE, est une méthode stochastique utilisée dans le contexte de données incomplètes, ne faisant pas intervenir la maximisation de la vraisemblance. Nous allons montrer, dans le paragraphe suivant, que cette méthode peut parfaitement s'appliquer dans notre cas et que les conditions de son application sont bien vérifiées. Nous détaillerons dans les paragraphes suivants les formules de ré-estimation des paramètres des modèles des données étudiées.

4.2.3 Applicabilité d'ICE et distributions supposées

Deux hypothèses sont nécessaires pour appliquer l'algorithme ICE. Comme nous l'avons présenté dans le paragraphe 1.4.2.b, il doit exister un estimateur en données complètes des paramètres à estimer, et le calcul de la distribution *a posteriori* du processus caché doit être possible. Ces hypothèses constituent les conditions pour appliquer la méthode ICE. Comme nous l'avons expliqué dans la section 4.1, si nous connaissons le processus $\mathbf{r}$, l'estimation de la matrice de séparation est alors facile. Il suffirait d'utiliser le sous-ensemble $\mathbf{X}_0$ défini par l'équation (4.1) et lui appliquer l'une des méthodes classiques d'ACI : ceci permet de disposer d'un estimateur en données complètes. D'autre part, la deuxième hypothèse d'applicabilité de la méthode ICE est la possibilité de calculer la loi de $\mathbf{r} \,|\, \mathbf{X}; \eta$. Nous vérifions que celle-ci est facilement calculable du moment que $\mathbf{X} = \mathbf{AS}$. En effet puisque $\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} = \mathbf{BX}$, nous avons :

$$\mathbb{P}(\mathbf{r} \,|\, \mathbf{X}; \eta) = \frac{\mathbb{P}(\mathbf{r}, \mathbf{X}; \eta)}{\mathbb{P}(\mathbf{X}; \eta)} = \frac{\mathbb{P}(\mathbf{r}, \mathbf{S}; \eta)}{\mathbb{P}(\mathbf{S}; \eta)} \tag{4.5}$$

La loi $\mathbf{r} \,|\, \mathbf{X}; \eta$ est donc identique à la loi $\mathbf{r} \,|\, \mathbf{S}; \eta$ et elle peut s'exprimer explicitement lorsqu'un modèle pour $\mathbb{P}(\mathbf{S}, \mathbf{r}; \eta)$ est posé. Nous tenons à préciser que le modèle supposé pour modéliser les données dans l'algorithme ICE peut être différent du modèle effectivement utilisé pour générer les données réelles. Ceci est important en pratique car la distribution de $\mathbf{S}$ est en général

supposée inconnue dans le cadre aveugle. En particulier, nous choisissons ici un modèle qui suit les hypothèses A1, A2 et A3. Celui-ci diffère du modèle particulier considéré dans les hypothèses A2', A3' et A3" de la section 4.1.1. La distribution conjointe de $(\mathbf{r}, \mathbf{S})$ sous le paramètre η est donnée par :

$$\mathbb{P}(\mathbf{r}, \mathbf{S}; \eta) = \mathbb{P}(\mathbf{r}; \eta) \prod_{t=1}^{T} \mathbb{P}(\mathbf{s}_t \,|\, r_t; \eta) \tag{4.6}$$

Cette équation correspond à la fois à la distribution supposée dans l'algorithme d'estimation ICE et à la distribution des données réelles. En revanche les expressions de $\mathbb{P}(\mathbf{r}; \eta)$ et $\mathbb{P}(\mathbf{s}_t \,|\, r_t; \eta)$ supposées dans ICE, sont différentes de celles des données réelles et précisées dans le paragraphe suivant.

a Distribution $\mathbb{P}(\mathbf{s}_t \,|\, r_t)$ supposée dans ICE

Comme dans la distribution des données réelles, la loi $\mathbb{P}(\mathbf{s}_t \,|\, r_t; \eta)$ supposée dans ICE ne dépend pas des paramètres η, et elle sera écrite $\mathbb{P}(\mathbf{s}_t|r_t)$. Expérimentalement, et pour des raisons de robustesse, la distribution $\mathbb{P}(\mathbf{s}_t \,|\, r_t)$ ne doit pas avoir de support borné. Pour cette raison nous avons choisi les distributions suivantes :

- la loi $\mathbf{s}_t \,|\, (r_t = 0) \sim \mathcal{N}(0, 1)$ et,
- la loi $\mathbf{s}_t \,|\, (r_t = 1)$ est telle que les composantes de $\mathbf{u}$ définies dans (eq.4.4) sont distribuées selon :

$$\frac{1}{2}\mathcal{L}(\lambda) + \frac{1}{2}\mathcal{N}(0, \sigma_\lambda^2)$$

Un exemple de réalisation des distributions supposées pour le vecteur $(s_t^1, s_t^2)^{\mathrm{T}}$ dans l'algorithme ICE est donné dans la figure 4.3. Cette distribution est différente de celle des données simulées représentées dans les figures 4.1 et 4.2. Néanmoins, les résultats des simulations montrerons la pertinence et la robustesse de ce choix.

Conditionnellement à $r_t = 0$, la distribution du couple (s_t^1, s_t^2) est supposée gaussienne, indépendante, de moyenne nulle et de variance unité : nous avons choisi cette distribution car elle est une loi moins informative ayant une densité à support non bornée. Par ailleurs, conditionnellement à $r_t = 1$, chaque composante du vecteur $\mathbf{u}$ dans la distribution supposée dans l'algorithme d'estimation est un mélange de loi de Laplace et de loi gaussienne.

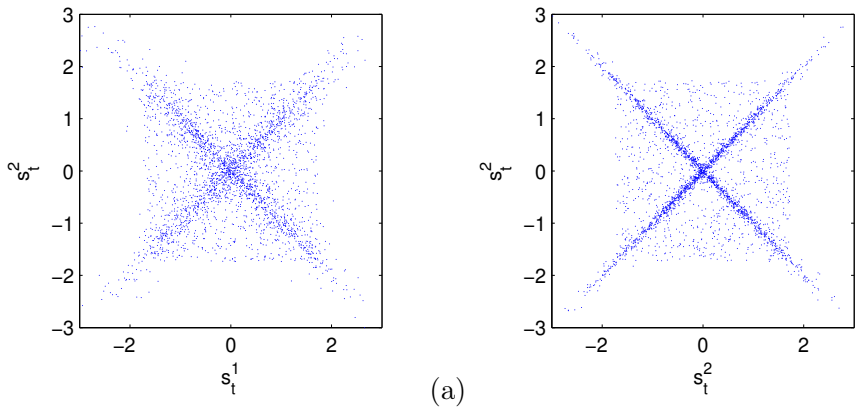


FIGURE 4.3 – Réalisation typique des sources supposées dans ICE avec : (a) $\lambda = 7$ et $\lambda = 25$.

Expérimentalement, et en observant les figures 4.1 et 4.3, ce choix semble bien approcher le modèle de données de l'*Exemple 1*. Le paramètre λ doit réaliser un compromis entre une modélisation correcte des données (λ grand) et d'autre part une bonne robustesse au niveau de l'algorithme d'estimation (λ petit). En outre, la distribution conditionnelle supposée $\mathbf{s}_t \,|\, r_t = 1$ est invariante par rapport aux changements de permutations et de signe. Cette symétrie est nécessaire dans notre méthode et permet d'éviter tout problème lié aux ambiguïtés des algorithmes basés sur l'analyse en composantes indépendantes. Pour cette raison, nous avons considéré la même distribution pour modéliser les sources générées selon l'*Exemple 2* du paragraphe 4.1.1.b.

b Distribution $\mathbb{P}(r_t; \eta)$ supposée dans ICE

Nous proposons deux modèles différents pour le processus latent $\mathbf{r}$:

b.1 Processus latent i.i.d

Dans un premier temps, nous considérons le cas le plus simple de processus latent. $\mathbf{r}$ est un processus de Bernoulli i.i.d donné par :

$$\mathbb{P}(\mathbf{r}; \eta) = \prod_{t=1}^{T} \mathbb{P}(r_t; \eta) \tag{4.7}$$

η est un paramètre scalaire dans $[0, 1]$, et pour tout $t \in \{1, \ldots, T\}$:

$$\mathbb{P}(r_t = 0; \eta) = \eta \quad et \quad \mathbb{P}(r_t = 1; \eta) = 1 - \eta \tag{4.8}$$

b.2 Processus latent de Markov

Une autre possibilité consiste à considérer une distribution markovienne pour le processus latent. $\mathbf{r}$ est une chaîne de Markov stationnaire définie par :

$$\mathbb{P}(\mathbf{r}) = \mathbb{P}(r_1) \prod_{t=2}^{T} \mathbb{P}(r_t \,|\, r_{t-1}) \tag{4.9}$$

où $\mathbb{P}(r_t \,|\, r_{t-1})$ est donnée par une matrice de transition indépendante de t. L'intérêt principal de la modélisation markovienne est que tout en prenant en compte des dépendances de $\mathbf{r}$, elle permet le calcul exact de la loi $\mathbb{P}(\mathbf{r} \,|\, \mathbf{X}; \eta)$ grâce à l'algorithme itératif Forward-Backward, ce qui rend applicable l'algorithme ICE.

4.3 Algorithme de séparation

Dans ce paragraphe, nous détaillerons d'une part, notre méthode de séparation qui combine le modèle d'ACI à l'algorithme d'ICE. D'autre part, nous établissons un lien entre l'algorithme ICE et la génération de variables aléatoires.

4.3.1 Algorithme ACI/ICE

Notons `ICA` une des méthodes basées sur le modèle d'analyse en composantes indépendantes, comme par exemple JADE [39], CoM2 [21], ou FastICA [40]. Soit $\mathbf{B} = \texttt{ICA}(\mathbf{X})$ la matrice de séparation estimée à partir d'un ensemble de données observées $\mathbf{X}$. En données complètes, un estimateur de $\mathbf{B}$ est donné par $\mathbf{B} = \texttt{ICA}(\mathbf{X}_0)$, où $\mathbf{X}_0$ est tel que défini dans l'équation (4.1). Notre algorithme consiste en une estimation par l'algorithme ICE des paramètres $\phi = (\mathbf{B}, \eta)$. Les paramètres η caractérisant la distribution de $\mathbf{r}$ sont estimés par l'espérance conditionnelle définie dans la section 1.4.2.b et donnée dans ce cas par l'équation (1.30). Notons que le couple $\mathbf{Z}$ dans l'équation (1.30) correspond ici au couple $(\mathbf{r}, \mathbf{X})$. La matrice de séparation $\mathbf{B}$ est

estimée à partir de l'équation (1.31) avec $K = 1$. Nous décrivons ci-dessous l'algorithme proposé :

Initialiser les paramètres $\hat{\phi}^{[0]} = (\hat{\mathbf{B}}^{[0]}, \hat{\eta}^{[0]})$.
Pour $q = 1, 2, \ldots, q_{\max}$, répéter :

1. calculer $\mathbb{P}(\mathbf{r} \,|\, \mathbf{X}; \hat{\phi}^{[q]})$ et tirer $\hat{\mathbf{r}}^{[q]}$ selon cette distribution,
2. définir : $\hat{\mathcal{I}}_0^{[q]} = \{t \,|\, \hat{r}_t^{[q]} = 0\}$ et $\hat{\mathbf{X}}_0^{[q]} = (x_t)_{t \in \hat{\mathcal{I}}_0^{[q]}}$
3. poser $\hat{\mathbf{B}}^{[q+1]} = \texttt{ACI}(\hat{\mathbf{X}}_0^{[q]})$
4. mettre à jour les paramètres η du processus $\mathbf{r}$.

Les étapes 1 et 3 de l'algorithme ci-dessus seront détaillées dans les paragraphes suivants. Notons que lorsque les paramètres η sont connus, l'étape 4 n'est plus nécessaire et leurs vraies valeurs peuvent être utilisées.

4.3.2 Précisions dans le cas d'un processus latent i.i.d

Nous détaillons les trois étapes de notre algorithme dans le cas où le processus latent $\mathbf{r}$ est i.i.d. :

étape 1

à partir des équations (4.5), (4.6) et (4.8), pour $\mathbf{r}$ i.i.d nous obtenons les expressions suivantes :

$$\begin{aligned}\mathbb{P}(\mathbf{r} \,|\, \mathbf{X}) &= \prod_{t=1}^{T} \frac{\mathbb{P}(r_t; \eta)\mathbb{P}(\mathbf{s}_t | r_t)}{\mathbb{P}(\mathbf{s}_t)} \\ &= \prod_{t=1}^{T} \mathbb{P}(r_t | \mathbf{s}_t; \eta)\end{aligned}$$

où la probabilité $\mathbb{P}(r_t; \eta)$ est donnée dans l'équation (4.8) et $\mathbb{P}(\mathbf{s}_t \,|\, r_t)$ est décrite dans la partie 4.2.3.b.1.
Pour des raisons de clarté, nous définissons le paramètre α_t :

$$\begin{aligned}\alpha_t &= \mathbb{P}(r_t = 0 \,|\, \mathbf{X}; \eta) \\ 1 - \alpha_t &= \mathbb{P}(r_t = 1 \,|\, \mathbf{X}; \eta)\end{aligned}$$

Nous déduisons de l'équation ci-dessus que dans l'étape 1 de l'algorithme, toutes les réalisations de $\hat{\mathbf{r}}^{[q]}$ sont indépendantes et telles que :

$$\begin{cases} \hat{r_t}^{[q]} = 0 & \text{avec la probabilité} \quad \alpha_t \\ \hat{r_t}^{[q]} = 1 & \text{avec la probabilité} \quad 1 - \alpha_t \end{cases}$$

Méthode d'acceptation-rejet

Lorsque $\mathbf{X}_0$ est déterminé, seule une partie de l'échantillon est gardée alors que le reste est ignoré. Ceci est étroitement similaire à la génération de variables aléatoires par la méthode d'acceptation-rejet. En effet, lorsque le processus $\mathbf{r}$ est i.i.d vérifiant l'équation (4.8), la distribution des sources $\mathbf{s}_t$ est donnée, selon l'équation (4.6), par :

$$\eta \mathbb{P}_0(\mathbf{s}_t) + (1 - \eta)\mathbb{P}_1(\mathbf{s}_t)$$

où l'on a noté $\mathbb{P}_0(\mathbf{s}_t) = \mathbb{P}(\mathbf{s}_t \,|\, r_t = 0)$ et $\mathbb{P}_1(\mathbf{s}_t) = \mathbb{P}(\mathbf{s}_t \,|\, r_t = 1)$.
Soit la remarque suivante :

Remarque 1 *Soit* $\mathbf{s}$ *une variable aléatoire (ou un vecteur aléatoire), dont la distribution de probabilité est donnée par :*

$$\mathbf{s} \sim \eta \mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})$$

Soit $\hat{r}$ *une variable aléatoire binaire* $\in \{0, 1\}$ *dont la distribution est exprimée par les formules suivantes :*

$$\mathbb{P}(\hat{r} = 0 \,|\, \mathbf{s}) = \frac{\eta \mathbb{P}_0(\mathbf{s})}{\eta \mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})} \tag{4.10}$$

$$\mathbb{P}(\hat{r} = 1 \,|\, \mathbf{s}) = \frac{\eta \mathbb{P}_1(\mathbf{s})}{\eta \mathbb{P}_1(\mathbf{s}) + (1 - \eta)\mathbb{P}_0(\mathbf{s})} \tag{4.11}$$

La distribution conditionelle $\mathbf{s} \,|\, \hat{r} = 0$ *est alors* $\mathbb{P}_0(\mathbf{s})$.

Preuve : la distribution de probabilité conjointe peut être écrite de la manière suivante :

$$\mathbb{P}(\mathbf{s}, \hat{r}) = \eta 1_{\hat{r}=0} \mathbb{P}_0(\mathbf{s}) + (1 - \eta) 1_{\hat{r}=1} \mathbb{P}_1(\mathbf{s})$$

La loi de $\mathbf{s} \,|\, \hat{\mathbf{r}} = 0$ est obtenue alors par conditionnement.

La remarque ci-dessus fait la liaison entre notre algorithme et la méthode de génération de variables aléatoires par acceptation-rejet. En effet, tirer $\hat{r}$

et conditionner par $\hat{r} = 0$ pour obtenir la loi de $\mathbf{s} \,|\, \hat{r} = 0$ est équivalent à ne garder que les échantillons de $\mathbf{s}$ où $\hat{r} = 0$ et rejeter les échantillon où $\hat{r} = 1$. En d'autres termes, dans l'algorithme ICE la distribution $\mathbb{P}_0(\mathbf{s})$ est générée en utilisant la distribution instrumentale suivante :

$$g(\mathbf{s}) = \eta \mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})$$

Nous obtenons alors un ensemble de données distribuées selon $\mathbb{P}_0(\mathbf{s})$: celle-ci correspond précisément à la distribution permettant d'appliquer l'algorithme ACI.

Remarque 2 *La distribution de probabilité du processus* $\mathbf{S}$ *dans la* ***Remarque 1*** *peut être considérée comme la marginale du couple* $(\mathbf{r}, \mathbf{S})$*, dans le cas où* $\mathbf{r}$ *est un processus de Bernoulli et la loi de* $\mathbf{S}$ *conditionnellement à* $\mathbf{r}$ *sont données par* $\mathbb{P}_0$ *et* $\mathbb{P}_1$*. Cependant,* $\hat{\mathbf{r}}$ *est obtenu indépendamment de* $\mathbf{r}$*; en particulier* $\hat{\mathbf{r}}$ *est différent de* $\mathbf{r}$*. C'est-à-dire que dans l'algorithme proposé, le processus latent original* $\mathbf{r}$ *et le processus* $\hat{\mathbf{r}}^{[q]}$ *obtenu par ICE peuvent être très différents bien qu'elles soient générés selon* $\mathbb{P}_0$*. Ceci sera illustré dans le paragraphe 4.4.1.b*

étape 3

Un estimateur en données complètes du paramètre η est donné par la fréquence empirique suivante lorsque r_t est disponible :

$$\hat{\eta} = \frac{1}{T} \sum_{t=1}^{T} 1_{\hat{r}_t = 0}$$

A partir de l'équation (1.30), la formule de ré-estimation du paramètre η est donnée par :

$$\hat{\eta}^{[q+1]} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{P}(r_t = 0 \,|\, \mathbf{x}_t, \hat{\phi}^{[q]}) \tag{4.12}$$

4.3.3 Processus latent de Markov

Dans le cas où le processus $\mathbf{r}$ est une chaîne de Markov, le vecteur de paramètres à estimer η correspond à la loi initiale

$$\pi_i = \mathbb{P}(r_1 = \omega_i)$$

avec $\omega_i = \{0, 1\}$, et aux composantes de la matrice de transition

$$a_{ij} = \mathbb{P}(r_t = \omega_i \,|\, r_{t-1} = \omega_j)$$

Les formules ré-estimations de ces paramètres sont identiques à celles données dans les équations (1.34) et (1.33) du chapitre 1. Elles sont données par :

$$\hat{a}_{ij}^{[q+1]} = \frac{\sum_{t=1}^{T} \mathbb{P}(r_t = \omega_i, r_{t-1} = \omega_j \,|\, \mathbf{x}; \phi^{[q]})}{\sum_{t=1}^{T} \mathbb{P}(r_{t-1} = \omega_j \,|\, \mathbf{x}; \phi^{[q]})} \tag{4.13}$$

$$\hat{\pi}_i^{[q+1]} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{P}(r_t = \omega_i \,|\, \mathbf{x}; \phi^{[q]}) \tag{4.14}$$

Les lois *a posteriori* $\mathbb{P}(r_t = \omega_i, r_{t-1} = \omega_j \,|\, \mathbf{s})$ et $\mathbb{P}(r_{t-1} = \omega_j \,|\, \mathbf{s})$ nécessaires à la ré-estimation des paramètres η est calculée grâce à la procédure itérative Forward-Backward définie dans la partie 1.4.1.

4.4 Résultats de simulations

Nous avons réalisé plusieurs simulations sur des échantillons de différentes tailles (respectivement $T = 1000$, $T = 5000$ et $T = 10000$). Nous évaluons la qualité de séparation par le critère d'Erreur Quadratique Moyenne (EQM) : les valeurs présentées sont une moyenne de l'erreur quadratique moyenne sur toutes les sources. De plus, nous considérons le taux de bonne segmentation empirique obtenu par l'algorithme ICE exprimé par :

$$\Upsilon_{ICE} = \frac{1}{T} \sum_{t=1}^{T} 1_{\hat{r}_t^{[q_{max}]} = r_t}$$

Ce taux correspond au nombre d'instants où les éléments de l'échantillon sont correctement classifiés et correspondent bien aux sources indépendantes ($r_t = 0$) ou aux sources dépendantes ($r_t = 1$). Dans toutes nos simulations, les signaux sources ainsi que la matrice de mélange sont générés aléatoirement, et les résultats proviennent de 1000 réalisations de Monte-Carlo. L'algorithme de séparation utilisé dans les simulations, noté `ACI` dans la section 4.3, est CoM2. Afin de comparer les performances de notre algorithme, nous avons appliqué l'algorithme CoM2 aux données générées avec des composantes dépendantes et indépendantes : d'une part en ignorant la violation

des hypothèses du modèle ACI, et d'autre part en considérant le cas des données complètes où le processus latent $\mathbf{r}$ est connu. Et la séparation est effectuée en se basant sur le sous-ensemble $\mathbf{X}_0$. Les données testées dans nos simulations sont générées selon les deux modèles suivants :

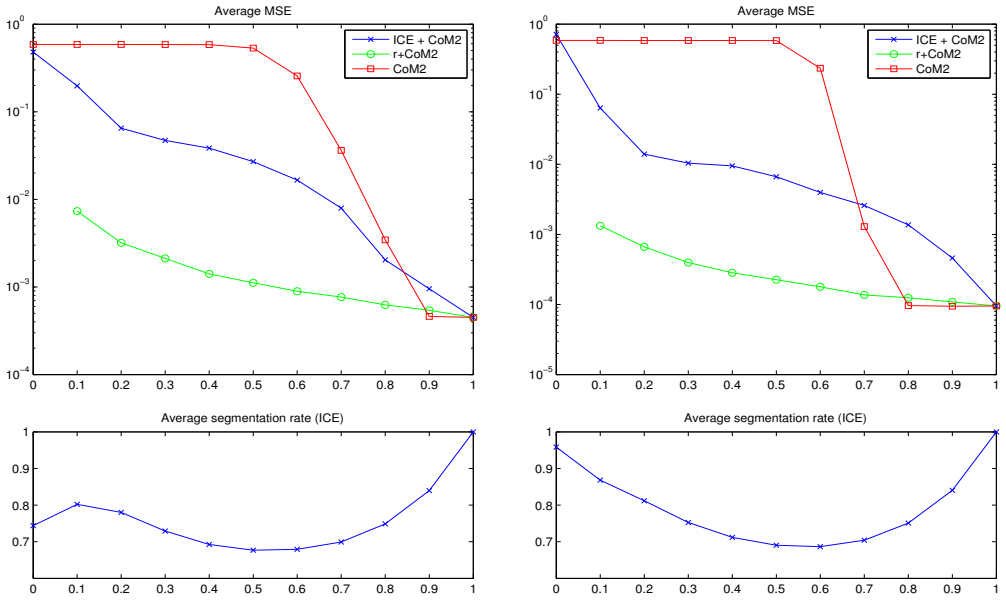
- Dans le premier modèle, les sources vérifient l'*Exemple 1* du paragraphea telles que le processus a_t, défini dans l'équation (4.3), est un mélange de gaussiennes avec $\mu = 0,5$. Et comme indiqué dans le paragraphe 4.1.1.a, l'algorithme CoM2 ne peut pas séparer de telles sources contenant des vecteurs dépendants.
- Dans le deuxième modèle, les sources vérifient l'*Exemple 2* présenté dans la section 4.1.1.b avec $\lambda = \sqrt{2}$.

4.4.1 Processus latent i.i.d et η connu

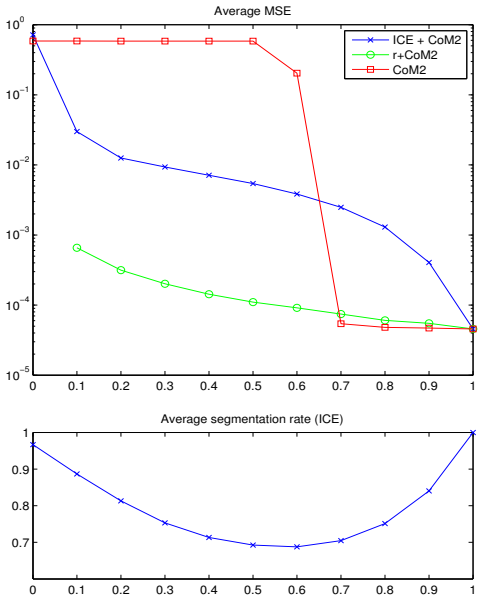
Nous avons d'abord réalisé des expérimentations avec le processus $\mathbf{r}$ i.i.d (équation (4.7)) et les paramètres η connus.

a Résultats de l'*Exemple 1*

Les valeurs de l'EQM pour $T = 1000$, $T = 5000$ et $T = 10000$ sont représentées respectivement dans les figures 4.4(a), 4.4(b) et 4.4(c) en fonction de $\eta \in [0, 1]$. Notons que pour $\eta = 0$, tous les échantillons des sources sont dépendants et la séparation en données complètes est donc impossible. De manière similaire, lorsque $\eta = 0$ dans l'algorithme ICE, tous les échantillons sont nécessairement estimés comme dépendants ce qui rend notre méthode inapplicable. Pour cette raison, lorsque η est nulle dans les données simulées, nous avons forcé sa valeur à $0,1$ dans l'algorithme d'estimation. Nous comparons la qualité de séparation des trois méthodes : (CoM2) appliqué sur $\mathbf{X}$, notre méthode notée (ICE+CoM2) et enfin CoM2 appliquée sur $\mathbf{X}_0$ pour $\mathbf{r}$ connu (r+CoM2). Ce dernier cas correspond à la séparation en données complètes donnant, bien évidemment, les meilleurs résultats. Nous remarquons que lorsque η tend vers zéro le nombre d'échantillons dans $\mathbf{X}_0$ diminue, ceci explique que l'EQM augmente lorsque η diminue pour les résultats de (r+CoM2) pour des valeurs de η petites. Par ailleurs, ignorer la dépendance des sources donne des mauvais résultats de séparation pour toute valeur de η



(a) (b)



(c)

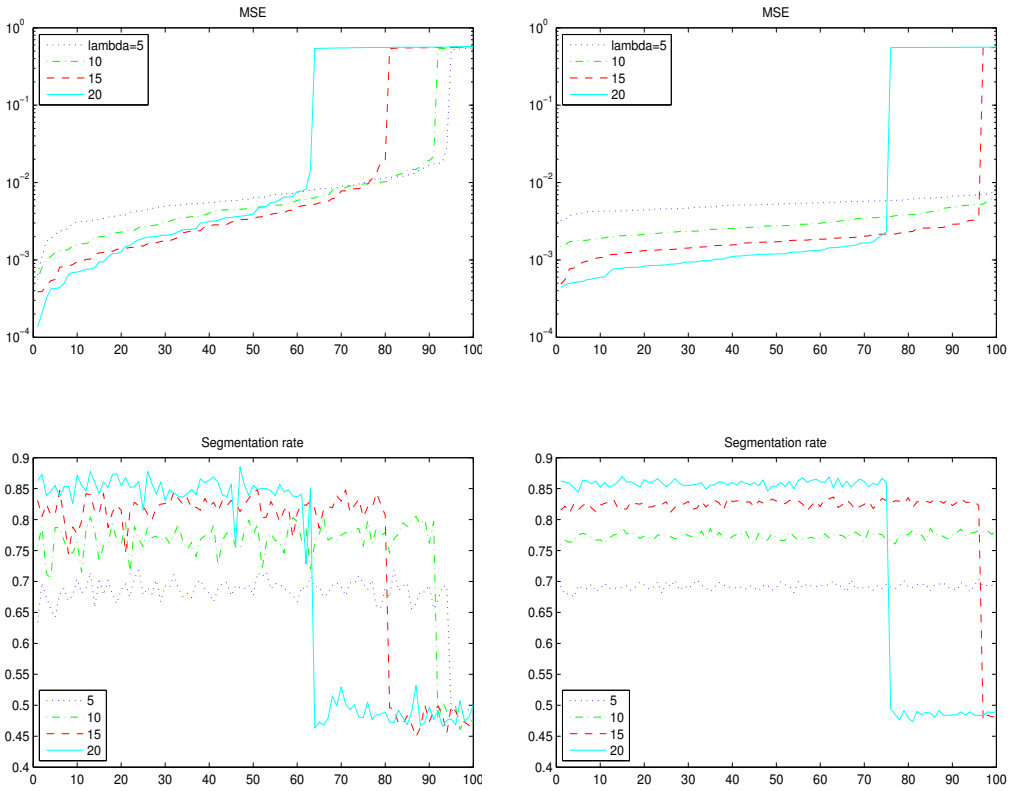
FIGURE 4.4 – Valeurs moyennes de l'EQM et taux de bonne segmentation en fonction de η pour des sources générées selon *Exemple 1*, avec $\lambda = \sqrt{2}$ et η connu. (a) $T = 1000$ échantillons, (b) $T = 5000$ échantillons et (c) $T = 10000$ échantillons.

inférieure à la valeur approximative de $0,7$. L'algorithme CoM2 semble pouvoir séparer les sources lorsque la valeur η est supérieure à environ $0,7$ (la proportion des données indépendantes étant dominante), ceci n'est pas le cas lorsque la valeur de η est inférieure à $0,7$ environ. D'autre part, lorsque η est inférieur à $0,7$ notre méthode (ICE+CoM2) apporte une nette amélioration à la qualité de séparation par rapport à celle obtenue en ignorant totalement la dépendance (CoM2). Nous avons tracé aussi, dans la figure 4.4, le taux de bonne segmentation obtenue par l'algorithme ICE, nous remarquons que les échantillons de $\mathbf{s}$ dépendants/indépendants sont bien classifiés, ce qui traduit par une valeur de Υ_{ICE} supérieure à $0,7$ environ.

En outre, nous avons étudié l'influence du paramètre λ pour une valeur de η fixée à $0,5$. Les valeurs de l'EQM pour 100 réalisations de Monte-Carlo sont représentées pour $T = 1000$ et $T = 10000$ dans la figure 4.5. Le taux de bonne segmentation est aussi représenté dans cette figure. Nous remarquons que les résultats de séparation des réalisations de Monte-Carlo peuvent être séparés en deux groupes :

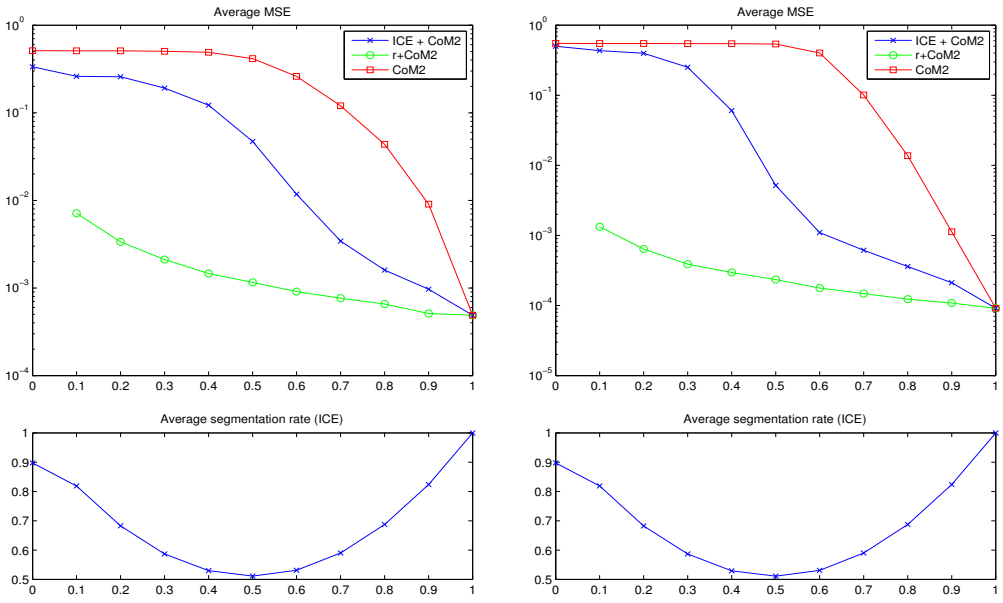
- Pour une minorité de cas, le taux de bonne segmentation Υ_{ICE} est proche de $0,5$. Ceci signifie que la méthode d'estimation ne permet pas de segmenter correctement le processus $\mathbf{r}$, et dans de telles situations, les valeurs de l'EQM sont assez élevées et le séparation n'aboutit pas. Ce comportement est plus fréquent pour des valeurs de λ grandes, qui impliquent une perte de robustesse de notre méthode.
- Dans la majorité des cas, le taux de bonne segmentationΥ_{ICE} est supérieur à $0,5$. Nous avons alors une bonne segmentation du processus latent $\mathbf{r}$, et une bonne qualité de séparation ce qui se traduit par de faibles valeurs de l'EQM. On remarque que pour des valeurs de λ grandes le taux de bonne segmentation est élevé tandis que les valeurs de l'EQM sont petites.

Nous déduisons que le choix du paramètre λ dans l'algorithme d'estimation doit être un compromis entre une bonne segmentation et une bonne qualité de séparation.



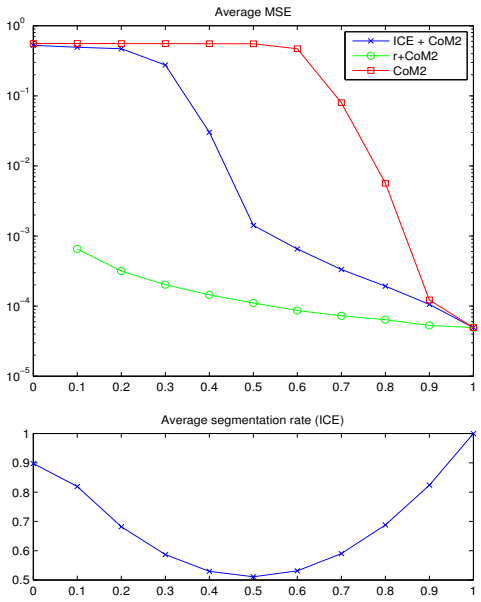
(a) (b)

FIGURE 4.5 – Valeurs de l'EQM présentés en incrémentant les réalisations Monte-Carlo et taux de bonne segmentation correspondant. Avec $\lambda = 0.5$, (a) $T = 1000$ et (b) $T = 10000$.



(a)

(b)



(c)

FIGURE 4.6 – Valeurs moyennes de l'EQM et taux de bonne segmentation pour des sources générées selon *Exemple 2*, avec $\lambda = 5$ et η connu. (a) $T = 1000$ échantillons, (b) $T = 5000$ échantillons et (c) $T = 10000$ échantillons.

b Résultats de l'*Exemple 2*

Dans la figure 4.6(a) sont représentées les valeurs de l'EQM obtenues pour les données de l'*Exemple 2* décrites dans la section 4.2.2 avec $\lambda = \sqrt{2}$. L'estimation en données complètes donne les meilleurs résultats de séparation, tandis qu'en ignorant la dépendance nous obtenons de mauvais résultats de séparation. Néanmoins, avec notre méthode nous arrivons à avoir une remarquable amélioration, en particulier, lorsque η est compris approximativement entre $0,4$ et $0,8$. Plus intéressant encore, la figure 4.6(b) montre que nous pouvons obtenir une bonne qualité de séparation bien que le taux de bonne segmentation soit faible. Cette remarque est très importante et montre qu'une bonne segmentation du processus latent est une condition suffisante mais pas nécessaire pour une bonne estimation au sens d'ACI. En effet, selon la ***Remarque 1***, l'étape basée sur l'estimation par ICE dans notre algorithme sélectionne un ensemble de l'échantillon $\mathbf{X}_0$ tel que sa distribution remplit les conditions habituelles d'ACI même dans le cas de faibles taux de bonne segmentation.

4.4.2 Processus latent i.i.d et η inconnu

Nous considérons à présent le cas où le processus $\mathbf{r}$ est i.i.d et le paramètre η de l'équation (4.8) est inconnu. Nous estimons le paramètre η par l'algorithme ICE. Les résultats sont moyennés sur 1000 réalisations de Monte-Carlo, et présentés dans le tableau 4.1 pour les sources de l'*Exemple1* et de l'*Exemple 2*. Pour le cas des sources obtenues à partir de l'*Exemple 1*, nous avons pris dans les distributions supposées $\lambda = 5$ tandis que $\lambda = \sqrt{2}$ pour les sources de l'*Exemple 2*. Les valeurs de l'EQM obtenues par CoM2 sont aussi données afin de comparer les résultats de notre méthodes aux résultats obtenus en ignorant totalement les dépendances des sources. Les résultats montrent que notre méthode est valide même dans le cas où η est estimé.

T	$T = 1000$			$T = 5000$			$T = 10000$		
η	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
Exemple1, η connu	7.40 e^{-1}	2.70e^{-2}	7.94 e^{-3}	1.04 e^{-2}	6.66 e^{-3}	2.59 e^{-3}	9.34 e^{-3}	5.42 e^{-3}	2.48 e^{-3}
Exemple1, η estimé	4.35 e^{-1}	8.17 e^{-2}	9.46 e^{-3}	5.36 e^{-1}	1.59 e^{-2}	4.59 e^{-3}	5.73 e^{-1}	8.50 e^{-3}	4.38 e^{-3}
Exemple1, CoM2	5.86 e^{-1}	5.33 e^{-1}	3.62 e^{-2}	5.86 e^{-1}	5.84 e^{-1}	1.29 e^{-3}	5.86 e^{-1}	5.85 e^{-1}	5.43 e^{-5}
Exemple2, η connu	1.91 e^{-1}	4.71e^{-2}	3.43 e^{-3}	2.51 e^{-1}	5.16 e^{-3}	6.15 e^{-4}	2.76 e^{-1}	1.42 e^{-3}	3.33 e^{-4}
Exemple2, η estimé	3.85 e^{-1}	6.66 e^{-2}	3.53 e^{-3}	5.43 e^{-1}	6.34 e^{-3}	6.60 e^{-4}	5.59 e^{-1}	1.77e^{-3}	3.33e^{-4}
Exemple2, CoM2	5.04 e^{-1}	4.16 e^{-1}	1.20 e^{-1}	5.47 e^{-1}	5.38 e^{-1}	1.01 e^{-1}	5.57 e^{-1}	5.54 e^{-1}	5.99 e^{-2}

TABLE 4.1 – Valeurs moyennes de l'EQM. Les sources **s** sont générées selon *Exemple 1* avec $\lambda = 5$ et *Exemple 2* avec $\lambda = \sqrt{2}$, **r** i.i.d

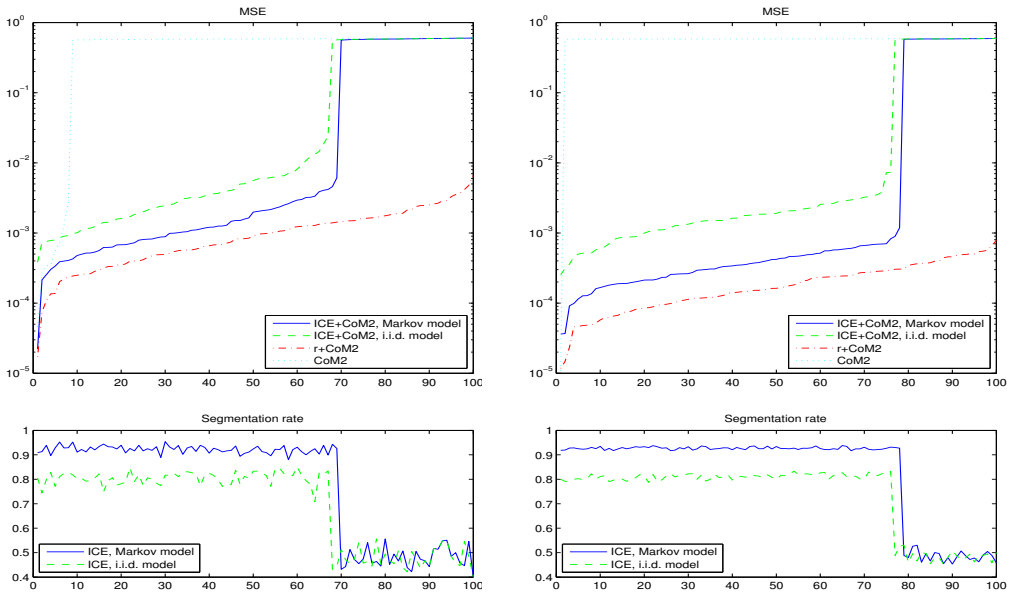
4.4.3 Processus latent de Markov

Comme nous l'avons expliqué dans le premier chapitre, l'algorithme d'estimation ICE peut modéliser une dépendance markovienne du processus latent **r**. Dans cet exemple **r** est une chaîne de Markov stationnaire caractérisée par la matrice de transition a_{ij} suivante :

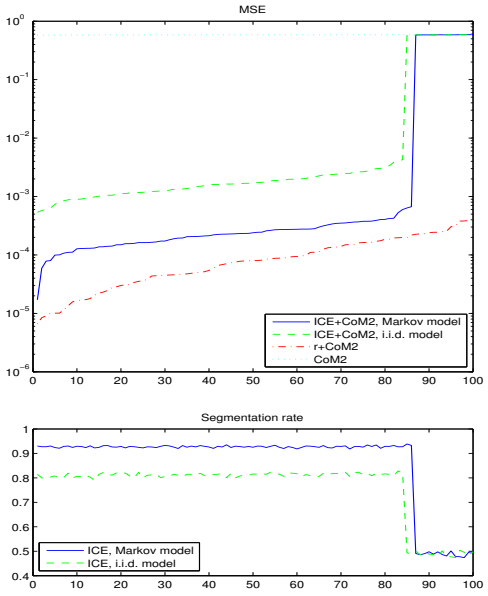
$$a_{ij} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

Nous avons appliqué notre méthode de séparation sur le même ensemble de données dans le cas où **r** est modélisée par une distribution i.i.d et par une chaîne Markov. Dans la figure 4.7 sont présentés les résultats pour les sources de l'*Exemple 1* (avec $\lambda = 15$ dans ICE) et dans la figure 4.8 les résultats pour les sources de l'*Exemple 2* (avec $\lambda = \sqrt{2}$ dans ICE).

Les résultats de séparation avec les données complètes et avec CoM2 ignorant les dépendances des sources, sont aussi tracés. Ces derniers correspondent évidemment aux cas extrêmes où l'algorithme donne les meilleurs et les pires résultats. Dans la figure 4.7 (en haut de la figure), on remarque que pour les sources de l'*Exemple 1* la qualité de séparation est améliorée lorsque nous tenons compte de la markovianité du processus **r** (courbe ICE+CoM2 Markov model). Nous observons (à partir des valeurs de l'EQM) que pour la majorité des réalisations, l'hypothèse de markovianité a significativement amélioré la qualité de séparation. Cependant, la propriété de markovianité n'apporte aucune amélioration à la robustesse de notre méthode puisque le

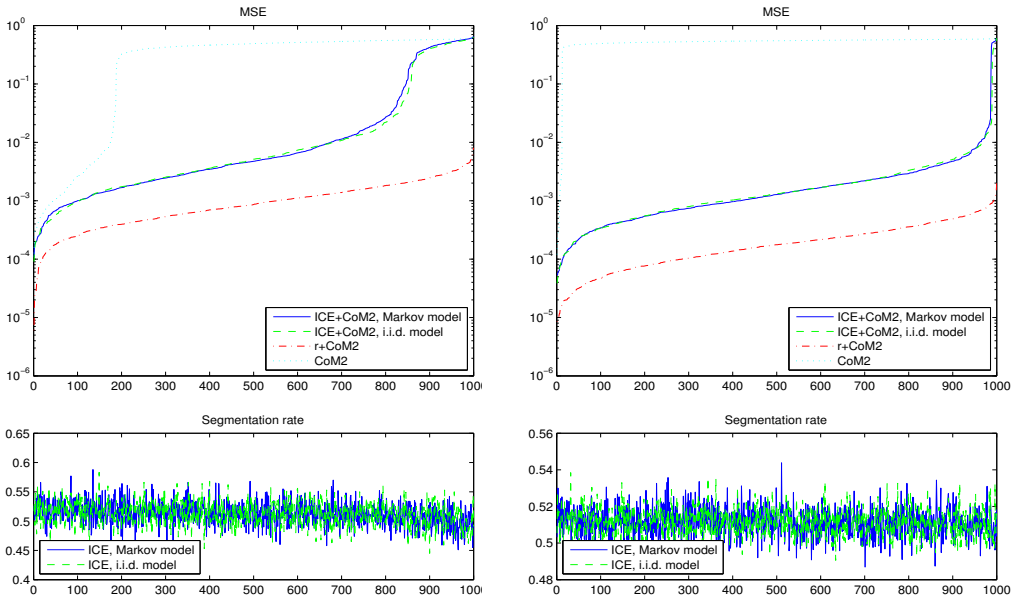


(a) (b)

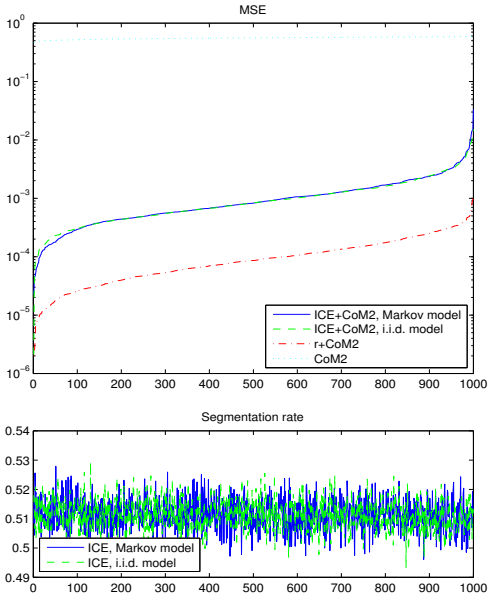


(c)

FIGURE 4.7 – Valeur l'EQM et taux de bonne segmentation (a) $T = 1000$, (b) $T = 5000$ et (c) $T = 10000$. Les sources $\mathbf{s}$ sont générées selon *Exemple 1*, $\mathbf{r}$ de Markov.



(a) (b)



(c)

FIGURE 4.8 – Valeur l'EQM et taux de bonne segmentation (a) $T = 1000$, (b) $T = 5000$ et (c) $T = 10000$. Les sources $\mathbf{s}$ sont générées selon *Exemple 2*, $\mathbf{r}$ de Markov.

nombre de réalisations où la séparation n'aboutit pas est identique que ce soit avec le modèle i.i.d ou de Markov. Par ailleurs dans la figure 4.8, nous remarquons différents comportements avec des données générées selon l'*Exemple 2*. En effet, la performance de l'algorithme ne s'améliore pas lorsque $\mathbf{r}$ est modélisée par une chaîne de Markov. Intuitivement, ceci pourrait provenir du fait qu'il n'est pas évident de distinguer entre les deux composantes du mélange de probabilité dans la distribution des sources.

Chapitre 5

Conclusion

Dans cette thèse, nous avons traité le problème de restauration dans deux contextes différents. D'une part, nous nous sommes intéressés à cette problématique dans le contexte de la segmentation de données à nombre fini d'états. En effet, pour des données ayant subi une transformation quelconque et en présence de bruit, nous avons exploité les propriétés des modèles de chaînes de Markov couples et triplets, récemment proposés, dans une restauration non supervisée. D'autre part, nous avons traité le problème de séparation aveugle de sources dépendantes par le modèle d'analyse en composantes indépendantes. Dans ce cas, pour des données à dépendance particulière, nous avons proposé un modèle étendu du modèle ACI en introduisant un processus latent.

D'abord, nous avons présenté les différents modèles et techniques qui ont été utilisés dans nos études. En particulier, nous avons détaillé la modélisation par les chaînes de Markov cachées classiques tout en mettant en relief les principales méthodes de segmentation et d'estimation de paramètres. Ensuite, nous avons présenté le modèle d'analyse en composantes indépendantes et nous avons décrit brièvement les différents modèles de mélanges, et quelques algorithmes de séparation.

Ensuite, nous avons présenté les modèles récents des chaînes de Markov couples et triplets. Nous avons alors proposé leurs extensions aux données multidimensionnelles, et avons présenté les algorithmes et les méthodes qui sont utilisés. Des résultats de séparation, en utilisant les modèles de Markov

couples et triplets, réalisés sur des données synthétiques ainsi que des images réelles dégradées par effet de transparence et de diffusion d'encre ont été présentés par la suite dans ce travail. Les différents résultats obtenus montrent la robustesse et l'efficacité de ces modèles qui portent systématiquement une amélioration à la segmentation par rapport au modèle de chaînes de Markov cachées classique.

Dans la dernière partie de ce travail, nous avons introduit un modèle de sources qui consiste en un mélange de deux distributions. Pour ces sources, seule l'une des composantes de ce mélange satisfait les hypothèses du modèle d'analyse en composantes indépendantes. Nous avons alors exploité cette information pour réaliser la séparation. Notre méthode permet la séparation d'autres types de sources dépendantes si un modèle de distributions adéquat peut être donné. Finalement, nous avons présenté des expérimentations dont les résultats valident la méthode proposée.

Bibliographie

[1] P. Comon and C. Jutten, *"Séparation de sources, tome 1 : concepts de base et analyse en composantes indépendantes"*, ch. Mélanges convolutifs. IC2, Hermès, 2007.

[2] P. Comon and C. Jutten, eds., *"Handbook of Blind Source Separation, Independent Component Analysis and Applications"*. Academic Press, 2010.

[3] I. Shiessel, M. Stetter, J. E. W. Mayhew, N. McLoughin, J. S. Lund, and K. Obermayer, "Blind signal separation from optical imaging recordings with extended spatial decorrelation," *IEEE Trans. on Biomedical Engineering*, vol. 47, no. 5, pp. 573–577, 2000.

[4] I. Shiessel, M. Stetter, S. Skew, J. E. W. Mayhew, N. McLoughin, J. B. Levitt, J. S. Lund, and K. Obermayer, "Blind separation of spatial signal patterns from optical imaging records," *International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, France, 1999.

[5] V. Zarzoso, A. K. Nandi, and E. Bacharakis, "Maternal and foetal ecg separation using blind source separation methods," *IMA Journal of Mathematics Applied in Medicine and Biology*, vol. 14, pp. 207–225, 1997.

[6] N. Charkani and Y. Deville, "Self-adaptive separation of convolutively mixed signals with a recursive structure. part ii : Theoretical extensions and application to synthetic and real signals," *Signal Processing*, vol. 75, no. 2, pp. 117–140, 1999.

[7] Y. Deville, J.Damour, and N.Charkani, "Multi-tag radio-frequency identification systems based on new blind source separation neural networks," *Neurocomputing*, vol. 49, no. 1-4, pp. 369–388, 2002.

[8] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, 2000.

[9] K. Rahbar and P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5, pp. 832–844, 2005.

[10] D. Schobben, K. Torkkola, and P.Smaragdis, "Evaluation of blind signal separation methods," *First International Workshop on Independent Component Analysis and Blind Signal Separation*, France, Jan 1999.

[11] N. Charkani, "Séparation auto-adaptative de sources pour les mélanges convolutifs. application à la téléphonie mains-libres dans les voitures," *PhD thesis*, 1996.

[12] Y. Deville and L. Danry, "Application of blind source separation techniques to multi-tag contactless identification systems," *International Symposium on Nonlinear Theory and Its Applications (NOLTA '95)*, pp. 73–78, Las Vegas, U.S.A, December 1995.

[13] Y. Deville and L. Danry, "Système d'échange de données comportant une pluralité de porteurs de données," *French Patent no. 95 10444*, 1995.

[14] S. Hosseini and Y. Deville, "Blind separation of linear-quadratic mixtures of real sources using a recurrent structure," *In Proc. of the 7th Int. Work-Conference on Artificial and Natural Neural Networks (IWANN2003), Lecture Notes in Computer Science*, vol. 2686, pp. 241–248, 2003.

[15] Y. Deville and S. Hosseini, "Recurrent networks for separating extractable-target nonlinear mixtures, part i : Non-blind configurations," *Signal Processing*, vol. 89, no. 4, pp. 378–393, 2009.

[16] J. Hérault and B. Ans, "Circuits neuronaux à synapses modifiables : décodage de messages composites par apprentissage non supervisé," *C.R. de l'académie des sciences*, vol. 299, no. III-13, pp. 525–528, 1984.

[17] C. Jutten, "Calcul neuromimétique et traitement de signal : analyse en composantes indépendantes," *Thèse d'état ès sciences physiques, UJF-INP*, 1987.

[18] J. Hérault, C. Jutten, and B. Ans, "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé," *In Actes du Xème colloque GRETSI*, 1985.

[19] C. Jutten and J. Hérault, "Blind separation of sources, part i : An adaptative algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[20] P. Comon, C. Jutten, and J. Hérault, "Blind separation of sources, part ii : problem statement," *Signal Processing*, vol. 24, no. 1, pp. 11–20, 1991.

[21] P. Comon, "Independent Component Analysis, A new concept," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[22] C. Jutten and J. Hérault, "ICA versus PCA," *Eusipco88-signal Processing IV : Theories and Application ; J.L Lacoume, A. Chehikian, N. Martin, J.Malbos (Eds)*, 1988.

[23] D. T. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. on Signal Processing*, vol. 45, no. 7, pp. 1712–1725, 1997.

[24] J. C. Pesquet and E. Moreau, "Cumulant-based independence measures for linear mixtures," *IEEE Trans. on Information Theory*, vol. 47, no. 5, pp. 1947–1956, 2000.

[25] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1712–1725, 1995.

[26] J. Cardoso, "Blind signal separation : statistical principles," *Proceedings of the IEEE. Special issue on blind identification and estimation*, vol. 9, no. 10, pp. 2009–2026, 1998.

[27] P. Comon and L. Rota, "Blind separation of independent sources from convolutive mixtures," *IEICE Trans. on Fundamentals of Electronics, Communications, and Computer Sciences*, vol. E86-A, no. 3, pp. 542–549, 2003.

[28] M. Castella and E. Moreau, "Generalized identifiability conditions for blind convolutive MIMO separation," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2846–2852, 2009.

[29] M. Castella, "Inversion of polynomial systems and separation of nonlinear mixtures of finite-alphabet sources," *IEEE Trans. Signal Processing*, vol. 56, no. 8, Part 2, pp. 3905–3917, 2008.

[30] M. Castella, S. Rhioui, E. Moreau, and J. C. Pesquet, "Quadratic higher-order criteria for iterative blind separation of a MIMO convolutive mixture of sources," *IEEE Trans. Signal Processing*, vol. 55, no. 1, pp. 218–232, 2007.

[31] M. Castella, P. Bianchi, A. Chevreuil, and J. C. Pesquet, "A blind source separation framework for detecting CPM sources mixed by a convolutive MIMO filter," *Signal Processing*, vol. 86, no. 8, pp. 1950–1967, 2006.

[32] M. Castella, J. C. Pesquet, and A. P. Petropulu, "A family of frequency- and time-domain contrasts for blind separation of convolutive mixtures of temporally dependent signals," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 107–120, 2005.

[33] M. Castella, A. Chevreuil, and J. C. Pesquet, *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, ch. Convolutive Mixtures. Academic Press, 2010.

[34] A. Hyvärinen and P. Pajunen, "Nonlinear Independent Component Analysis : Existence and uniqueness results," *Neural Networks*, vol. 12, no. 3, pp. 429–439, 1999.

[35] C. Jutten and J. Karhunen, "Advances in Blind Source Separation BSS and Independent Component Analysis ICA for nonlinear mixtures," *International Journal of Neural Systems*, vol. 14, no. 5, pp. 267–292, 2004.

[36] J. F. Cardoso, "Analyse en Composantes Indépendantes," *Actes des XXXIVèmes Journées de Statistique JSBL*, Belgium, 2002.

[37] G. Darmois, "Analyse générale des liaisons stochastiques," *Review of the International Statistical Institute*, vol. 21, no. 1/2, pp. 2–8, 1953.

[38] A. Taleb and C. Jutten, "On underdetermined source separation," *In Proceedings of the Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1445–1448, 1999.

[39] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE-Proceeding-F*, vol. 140, no. 6, pp. 362–370, 1993.

[40] A. Hyvärinen and E. Oja, "Independent Component Analysis : Algorithms and application," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[41] A. Mohammed-Djafari, "Inverse Problems in Vision and 3D Tomography," *ISTE - WILEY*, December, 2009.

[42] S. Moussaoui, A. Mohammad-Djafari, D. Brie, O. Caspary, and B. Humbert, "A Bayesian method for positive source separation : Application to mixture analysis in spectroscopy," *IAR'2003*, Novembre, 2003.

[43] A. Mohammad-Djafari, "A Bayesian approach to source separation," *proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'99)*, pp. 221–224, USA, 1999.

[44] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images," *Journal of Electronic Imaging*, vol. 13, pp. 349–361, April 2004.

[45] H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," *VLSI Signal Processing Systems*, vol. 37, pp. 263–279, June - July 2004.

[46] H. Snoussi and A. Mohammad-Djafari, "Estimation of Structured Gaussian Mixtures : The Inverse EM Algorithm," *IEEE Trans. on Image Processing*, vol. 55, pp. 3185–3191, July 2007.

[47] N. Bali and A. Mohammad-Djafari, "Hierarchical Markovian Models for joint classification, Segmentation and Data Reduction of Hyperspectral Images," *ESANN 2006*, pp. 4–8, September 2006.

[48] N. Bali and A. Mohammad-Djafari, "A variational Bayesian Algorithm for BSS problem with Hidden Gauss-Markov Models for the sources in : Independent Component Analysis," *ICA 2007 and Signal Separation, Edited by :M.E. Davies, Ch.J. James, S.A. Abdallah, M.D. Plumbley*, pp. 137–144 Springer (LNCS 4666), 2007.

[49] N. Bali and A. Mohammad-Djafari, "Bayesian Approach with Hidden Markov Modeling and Mean Field Approximation for Hyperspectral

Data Analysis," *IEEE Trans. on Image Processing*, vol. 17, pp. 1887–1899, Feb 2008.

[50] M. M. Ichir and A. Mohammad-Djafari, "Hidden Markov models for wavelet-based blind source separation," *IEEE Trans. on Image Processing*, vol. 15, pp. 1887–1899, July 2006.

[51] F. Su and A. Mohammad-Djafari, "An Hierarchical Markov Random Field Model for Bayesian Blind Image Separation," *International Congress on Image and Signal Processing (CISP2008)*, May 2008.

[52] S. Moussaoui, C. Carteret, D. Brie, and A. Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods sampling," *Chemometrics and Intelligent Laboratory Systems*, vol. 81, no. 2, pp. 137–148, 2006.

[53] A. Belouchrani and J. F. Cardoso, "Maximum likelihood source separation for discrete sources," *EUSIPCO*, Scotland, 1994.

[54] S. Derrode and W. Pieczynski, "Signal and Image Segmentation Using Pairwise Markov Chains," *IEEE Trans. on Signal Processing*, vol. 52, no. 9, pp. 2477-2489, 2004.

[55] W. Pieczynski, "Pairwise Markov Chains," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 634–639, 2003.

[56] W. Pieczynski, C. Hulard, and T. Veit, "Triplet Markov Chains in hidden signal restoration," *SPIE's International Symposium on Remote Sensing*, September 2002.

[57] N. Giordana and W. Pieczynski, "Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 465–475, 1997.

[58] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, vol. 37, no. 6, pp. 1554–1563, 1966.

[59] L. E. Baum, "An inequality and associeted maximization technique in statistical estimation for probabilistic functions of Markov processes," *Inequalities, III (Proc. 3rd Symp., Univ. Calof., Los Angeles)*, pp. 1–8, 1972.

[60] L. E. Baum, T. Petrie, G. Soules, and N.Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann., Math., Statistic.*, vol. 41, no. 1, pp. 164–171, 1970.

[61] B. Benmiloud and W. Pieczynski, "Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images," *Traitement du signal*, vol. 12, no. 5, pp. 433–454, 1995.

[62] P. A. Devijver, "Baum's forward-backward algorithm revisited," *Pattern Recognition Letters*, vol. 3, pp. 369–373, 1985.

[63] L. E. Baum, T. Peterie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chaines," *The Annals of mathematical statistics*, vol. 41, pp. 164–171, 1970.

[64] G. Celeux and J. Diebolt, "The sem algorithm : A probabilistic teacher algorithm derived from the em algorithm for the mixture problem," *Computational Statistics Quarterly*, vol. 2, no. 1, pp. 73–82, 1985.

[65] W. Pieczynski, "Statistical image segmentation," *Machine Graphics & Vision*, vol. 1, no. 1/2, pp. 262–268, 1992.

[66] W. Pieczynski, "Champs de Markov cachés et estimation conditionnelle itérative," *Traitement du signal*, vol. 11, no. 2, pp. 141–153, 1994.

[67] A. Peng and W. Pieczynski, "Adaptive mixture estimation and unsupervised local bayesian image segmentation," *Graphical Models and Image Processing*, vol. 57, no. 5, pp. 389–399, 1995.

[68] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.

[69] G. J. McLachlan and T. Krishnan, "Em algorithm and extensions," *Wiley, Series in Probabilities and Statistics*, 1997.

[70] C. Wu, "On the convergence properties of the em algorithm," *Annals of Statistics*, vol. 11, pp. 59–103, 1983.

[71] D. Benboudjema, *Champs de Markov triplets et segmentation bayésiennes non supervisée d'images.* PhD thesis, 2005.

[72] D. Benboudjema and W. Pieczynski, "Unsupervised image segmentation using triplet Markov fields," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 476–498, 2005.

[73] J. L. Lahorgue and W. Pieczynski, "Unsupervised segmentation of hidden semi-Markov non stationary chains," *Signal Processing*, vol. 92, no. 1, pp. 29–42, 2012.

[74] J. L. Lahorgue and W. Pieczynski, "Partially Markov models and unsupervised segmentation of semi-Markov chains with long-depndent noise," *International Symposium on Applied Stochastic Models and Data Analysis ASMDA*, Greece 2007.

[75] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the em algorithm : an experiemental study," *Journal of Statistical Computation and Simulation*, no. 55, pp. 287–314, 1996.

[76] P. Lanchantin, J. L. Lahorgue, and W. Pieczynski, "Unsupervised segmentation of randomly switching data hidden with non-gaussian correlated noise," *Signal Processing*, vol. 91, no. 2, pp. 163–175, 2011.

[77] P. Lanchantin, *Chaînes de Markov Triplets et segmentation non supervisée des signaux*. PhD thesis, 2006.

[78] P. Lanchantin and W. Pieczynski, "Unsupervised non stationary image segmentation using triplet Markov chains," *Advanced Concepts for Intelligent Vision Systems (ACVIS.04)*, Sept. 2004.

[79] S. Grégoir and F. Lenglart, "Measuring the probability of a business cycle turning point by using a multivariate qulitative hiden Markov model," *Journal of forecasting*, vol. 19, no. 2, pp. 81, 2000.

[80] L. Thomas, D. Allen, and N. Morkel-kingsbury, "A hiden Markov chain model for the term structure of bond credit risk spreads," *International Review of financial Analysis*, vol. 11, no. 3, pp. 311–329, 2002.

[81] T. Koski, *"Hidden Markov models for bioinformatics"*. Kluwer Academic Publishers, 2001.

[82] P. Nicolas, L. Bize, F. Muri, M. Hoebeke, F. Rodolphe, S. D. Ehrlich, B. Prum, and P. Bessières, "Mining bacillus subtilis chromosome heterogeneities using hidden Markov models," *Nucleic Acid Research*, vol. 30, 2002.

[83] G. Nuel and B. Prum, *"Analyse statistique des séquences biologiques ; modélisation Markovienne, alignements et motifs"*. Hermes, Collection Bioinformatique, 2007.

[84] O. Cappé, E. Moulines, and T. Ryden, "Inference in hidden Markov models," *Springer, Series in Statistics*, 2005.

[85] H. Maître, "Traitement des images RSO," *Hermes collection IC2*, 2001.

[86] C. Raphael, "Automatic segmentation of acoustic musical signals using Hidden Markov Models," *IEEE Trans. on Pattern Analysis and Machine intelligence*, vol. 21, no. 4, pp. 360, 1999.

[87] F. Salzenstein and W. Pieczynski, "Sur le choix de méthode de segmentation statistique d'images," *Traitement de Signal*, vol. 15, no. 2, pp. 120–127, 1998.

[88] M. Castella and P. Comon, "Séparation aveugle de sources dépendantes," *Actes 21ème Colloque GRETSI*, September 2007.

[89] M. Castella, S. Rafi, P. Comon, and W. Pieczynski, "Separation of instantaneous mixtures of dependent sources using classical ICA methods," *EURASIP Journal on Advances in Signal Processing*, submitted.

Table des figures