CONTRIBUTION A LA SEPARATION AVEUGLE DE SOURCES

Ali Mansour

THÈSE

Présentée par

Ali MANSOUR

Pour obtenir le titre de

DOCTEUR
DE L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

(arrêté ministériel du 30 mars 1992)

Spécialité : **Signal, Image, Parole**

CONTRIBUTION A LA SEPARATION AVEUGLE DE SOURCES

Soutenue le 13 Janvier 1997 devant le jury :

Président	Dr.	Odile MACCHI,	Directeur de recherche CNRS
Rapporteurs	Prof.	Philippe LOUBATON,	Prof. université de Marné la vallée
	Dr.	Yannick DEVILLE,	LEP, Philips
Examinateurs	Prof.	Jean-Louis LACOUME,	Prof. INPG
	Prof.	Maurice BELLANGER,	Prof. CNAM
Directeur	Prof.	Christian JUTTEN,	Prof. université de Joseph Fourier

Thèse préparée au sein du laboratoire TIRF

Laboratoire de Traitement d'Images et de Reconnaissance des Formes

A mes parents
A ma femme
A mon frère et sa famille
A ma soeur et son mari

0

Table des matières

Deux choses sont infinies: l'univers et la bêtise humaine; en ce qui concerne l'univers, je n'en ai pas acquis la certitude absolue.

Albert Einstein

Remerciements

Je remercie Monsieur le Prof. C. Jutten, directeur du Laboratoire, et de mes recherches, de m'avoir accueilli et financé tout au long de ma thèse. Je tiens à le remercier aussi bien pour l'encadrement de ma thèse. C'était un bon ami qui m'a encouragé à faire de la recherche et de l'enseignement.

Je remercie chaleureusement l'équipe formidable du TIRF, permanents et thésards sans exception, qui était ma deuxième famille pendant 4 ans, en particulier :

- Monsieur le Prof J. Hérault, pour ses conseils pratiques et rigoureux.
- Je tiens à remercier aussi beaucoup : la secrétaire Marino, les deux ingénieurs systèmes du laboratoire Michel et Danuta qui m'ont aidé beaucoup chaque fois que j'avais besoin d'eux.
- Je tiens à remercier et à souhaiter bonne chance à mes chers amis : C. Aviles (mon premier ami au laboratoire et en DEA), A. Mhani (mon cher ami marocain), R. Chentouf (ma collègue de bureau), C. Croll (un vrai ami français, et je n'oublierai pas les voyages qu'on a fait ensemble), N. Maria (pour son aide appréciable en informatique), G. Cirrincione (c'est un homme franc et grâce à lui, j'ai visite l'Italie et j'ai assisté à tous les mauvais films de cinéma à Grenoble!), N. Alawa (pour sa sympathie).

Je remercie du fond de mon coeur ma femme, pour ses profonds et sincères sentiments envers moi. Et j'apprécie bien son aide, physique et morale, pour que je puisse terminer cette thèse. Je n'oublierai pas les nuits qu'on a passé ensemble dans mon laboratoire et elle avait les yeux presque fermés mais elle voulait rester à côté de moi pour m'aider et m'encourager.

Je n'oublierai pas ma famille, mes parents et la famille de mon oncle Mustapha, qui ont fait l'impossible pour que je puisse finir mes études.

Je tiens à remercier les membres du Jury :

- Monsieur le Prof. M. Bellanger, d'avoir accepté d'être mon examinateur.
- Monsieur le Dr. Y. Deville, d'avoir accepté d'être mon rapporteur.
- Monsieur le Prof. J.-L. Lacoume, d'une part, d'avoir accepté d'être mon examinateur, d'autre part pour son aide, et ses conseils appréciables lorsqu'il était mon professeur et ensuite durant les 3 années suivantes de cette thèse.
- Monsieur le Prof. P. Loubaton pour sa participation en tant que rapporteur à mon jury. J'apprécie beaucoup notre coopération. Je tiens beaucoup à le remercier pour ses conseils, sa gentillesse et les quelques jours qu'il a consacré à notre discussion.
- Madame la Dr. O. Macchi.

Deux parties de cette thèse ont été le fruit de deux collaborations différentes : la première collaboration avec l'université de Granada, la deuxième avec l'Ecole Nationale Supérieure de Télécomunication à Paris et l'université de Marne-la-vallée. Je tiens à remercie les équipes de ces trois universités.

Notations

$cum_q(X) = cum(x_1, x_2, \ldots, x_q)$: auto-cumulant d'ordre q de X.
$\delta(p_x, p_y)$ est la divergence de Kullback-Leibner entre les densités de probabilités $p(x)$ et $p(y)$.
EX = espérance mathématique de X.
$G(t)$ = matrice séparante.
$\mathcal{H}$ = espace de Hilbert.
$\mathcal{H}^D$ = espace dual.
$H(n)$ = matrice du filtre de mélange en temps discret.
$H(t)$ = matrice du filtre de mélange en temps continu.
$h_i(n)$ = ième ligne de $H(n)$.
I = matrice Identité.
I_n = matrice Identité $n \times n$.
$i(p_x, p_y)$ = information mutuelle entre les variables aléatoires x et y.
$J(y)$ = fonction de contraste.
$K(p(x))$ = kurtosis de $p(x)$.
$K_S(x)$ = signe de Kurtosis de $p(x)$.
$Mom(u_1, u_2, \ldots, u_q)$ = moment d'ordre q de u.
Nc = nombre de capteurs.
Ns = nombre de sources.
P = matrice de permutation.
$p(x)$ = densité de probabilité.
$p_x(u)$ = densité de probabilité de la variable aléatoire x appliqué à la valeur u.
$R^\dagger$ = matrice transposée hermitienne de R.
S^1 = tenseur covariant $\in \mathcal{H}$.
S_1 = tenseur contravariant $\in \mathcal{H}^D$.
S^T = vecteur transposé de S.
$S(t)$ = vecteur des sources.
$s_i(t)$ = ième source.
$X(t)$ = sources estimés.
$Y(t)$ = vecteur d'observation (signaux mélangés et filtrés).
$y_i(t)$ = ième signal mélangé.

Δ = matrice diagonale.
δ_{mo}^{np} = symbole de kronecker généralisé.
$\phi_u(v)$ = première fonction caractéristique.
$\Psi_u(v)$ = deuxième fonction caractéristique.
$\bullet$ = produit de contraction.
$\otimes$ = produit tensoriel entre tenseurs.
$\bigotimes$ = produit de Kronecker entre deux matrices.

Abréviations

ACI : Analyse en Composantes Indépendantes.
iid : indépendant, identiquement distribué.
ddp : densité de probabilité.
ODE : Ordinary Differential Equation.

Chapitre 1

Introduction

Le traitement du signal est une science récente qui a permis le développement d'application telle que le radar. Le principe du radar (Radio Detection And Ranging) a été imaginé dès 1911 par l'américain Hugo Gernsback. Des expériences de détection électromagnétique d'avions furent menées avec succès par Pierre David en 1934, et, en 1935, Maurice Ponte et Henri Gutton installaient un détecteur d'obstacles à bord du paquebot Normandie. Toutefois, ce fut seulement au cours de la seconde guerre mondiale que les techniciens anglais, dirigés par Watson Watt, développèrent, en installant de nombreuses stations pour la détection d'avions ennemis, la technique du radar. Et avec les radars, a commencé le traitement du signal dans sa forme actuelle.
Le traitement du signal est aussi utile pour traiter le signal biologique mesuré à l'aide d'électrode. Le cerveau lui aussi traite les signaux provenant de ses multiples capteurs : visuel, auditif, ... et ce depuis plusieurs centaines de millions d'années. La modélisation de la transmission des informations (positions, vitesses) concernant le mouvement d'une articulation a donné naissance au problème de séparation de sources, il y a une dizaine d'années. Depuis ce problème a été étudie de façon approfondi du traitement du signal pour son intérêt théorique (utilisation du concept d'indépendance et de statistique d'ordre supérieur) et ses applications nombreuses : traitement d'antenne, traitement de signal biomédical, rehaussement de parole, etc.

Dans le sujet de cette thèse, nous étudions différents aspects de la séparation aveugle de sources :

- Dans le 2ème chapitre "Séparation de sources", on exposera le problème d'une façon générale, tout en étudiant, expliquant et citant les principales idées utilisées dans ce domaine. Ce chapitre contient notement un état de l'art sur la séparation aveugle de sources.

- Le 3ème chapitre "Approche directe" est divisé en deux parties :
  - La première partie présente une approche directe de séparation, dans laquelle la séparation est obtenue en résolvant un système d'équations non linéaires.
  - La deuxième partie est focalisée sur l'étude de quelque points théoriques concernant un algorithme géométrique de séparation.
- En basculant sur des critères adaptatifs, au 4ème chapitre, on présente un critère simple pour les mélanges instantanés et convolutifs basé sur les cumulants croisées du 4ème ordre des signaux estimés, dont la validité suppose des sources de même signe de kurtosis.
- Les critères de séparation de sources supposent souvent des hypothèses sur les kurtosis des sources ou la somme des kurtosis des sources. Généralement, la pertinence de cette hypothèse est étudiée. Nous proposons dans le 5ème chapitre une étude théorique sur ce point, que nous illustrons par quelques exemples.
- Le chapitre 6 "Méthodes de type sous-espace" est consacré à l'étude de méthodes dites sous-espace qui exploitent presque uniquement les statistiques du second ordre pour séparer les sources à partir des mélanges convolutifs.
- Ce mémoire se termine par la conclusion générale et les perspectives.

Chapitre 2

Séparation de sources

2.1 Bref Historique

La séparation de sources est un problème relativement récent en traitement du signal, qui a été formulé pour la première fois en 84 par Hérault et Ans [60] et [61]. Malgré son importance en traitement de signal et ses nombreuses applications dans plusieurs domaines (notamment en traitement d'antenne et en télécommunications), son origine est liée à la modélisation d'un phénomène biologique ([4] et [60]). Bar-Ness [5] a aussi proposé une solution semblable à celle de Hérault *et al.* [61], dans le cadre du traitement d'antenne, mais de présentation confuse et sans avoir vu la portée de cette approche.
C'est maintenant, un problème très général qui se manifeste dans plusieurs domaines et dans plusieurs applications. On rencontre ce problème dans de nombreuses situations en radio-communication, (par exemple dans les systèmes radiomobiles [1]), en acoustique [95], en sismique [118]. Récemment, on a aussi utilisé des algorithmes de séparation des sources pour contrôler la dégradation de l'écran thermique d'un réacteur nucléaire [45] ou le trafic aérien d'un aéroport [23].

2.2 Modèles et problème

Le problème de **la séparation de sources** consiste à concevoir des méthodes capables de retrouver les N_s sources inconnues observées au travers de N_c mélanges inconnus des N_s sources. Ces N_c mélanges sont obtenus par un réseau de N_c capteurs.

[1] techniques de type SDMA (Spatial Division Multiple Access); système de téléphone à main-libre.

2.2.1 Modélisation du problème

Les N_s sources sont notées $s_i(t)$, avec $1 \leq i \leq N_s$ et elles sont regroupées en un vecteur $S(t)$. Soit $Y(t)$ le vecteur d'observation dont les composantes sont $y_i(t)$, $1 \leq i \leq N_c$. Compte tenu de la transmission des signaux dans le canal et des caractéristiques des capteurs, chaque observation $y_i(t)$ est une fonction inconnue des sources :

$$Y(t) = H[S(t), \ldots, S(t-M)], \tag{2.1}$$

où H est une fonction inconnue qui dépend uniquement du canal et des capteurs, M est l'ordre du canal. La structure générale de ce modèle est présenté dans la figure 2.1.

La séparation consiste à estimer un système G, telle que les composantes de

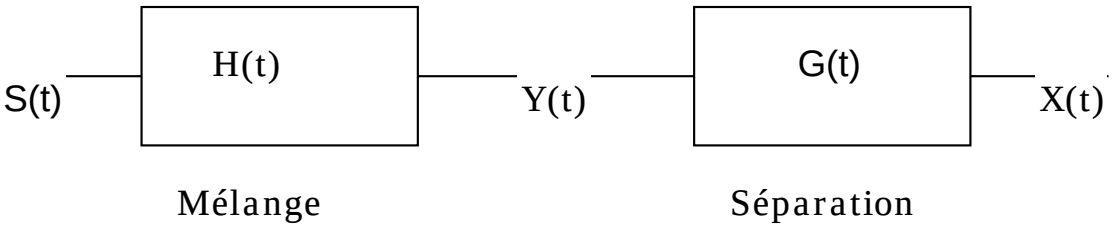


Figure 2.1: Structure générale.

$X(t) = G[H(S)]$ soient les sources inconnues. La séparation est dite aveugle si on est capable de séparer les sources sans aucune hypothèse. Mais pratiquement on ne sait pas encore résoudre le problème relatif au modèle (2.1) sans connaissances a priori sur le canal (c'est-à-dire sur H), et on choisi un modèle de mélange moins général, le plus souvent linéaire.

2.2.2 Modèle de mélanges linéaires

Dans le cas le plus général la fonction H dans l'équation (2.1) est une fonction vectorielle[2] non linéaire qui dépend aussi des instants passés de $S(t)$. Ce problème de mélanges non linéaires a été jusqu'ici peu abordé, à part l'étude de Krob [71] sur des modèles linéaires quadratiques, et les travaux récents de Pajunen *et al.* [100], Taleb et Jutten [115].

Si l'on suppose que le canal est un filtre linéaire et les capteurs aussi sont des filtres linéaires, on a alors :

$$y_j(t) = \sum_{i=1}^{N_s} h_{ji}(t) * s_i(t), \qquad 1 \leq j \leq N_c \tag{2.2}$$

[2]C'est à dire, a des vecteurs pour entrées et plusieurs sorties.

où $*$ est le produit de convolution et $h_{ji}(t)$ est un filtre linéaire qui caractérise la sensibilité de la sortie du $j$ème capteur vis à vis de la source i. Un tel mélange est dit **convolutif** et il est caractéristique d'un canal avec mémoire. Dans le cas de signaux échantillonnés, la relation (2.2), en notation vectorielle, sera notée :

$$Y(n) = [H(z)]S(n) = \sum_l H(l)S(n-l), \tag{2.3}$$

où n représente le temps discret. La représentation en z est associée naturellement aux fréquences discrètes z et l'équation (2.3) après transformation devient :

$$Y(z) = H(z)S(z), \tag{2.4}$$

dans laquelle le produit de convolution est remplacé par un produit matriciel.

Dans le cas le plus simple, $H(z)$ est une matrice de scalaires et la relation (2.4) devient :

$$Y(n) = HS(n). \tag{2.5}$$

On dit alors que le mélange est **linéaire instantané**.

2.2.3 Indétermination

On remarque facilement que la relation entrée-sortie est une application surjective, pour un mélange convolutif (2.3) ou instantané (2.5). Le même vecteur d'observation $Y(t)$ peut donc être généré à l'aide d'une infinité de vecteurs $S(t)$. En effet, l'ordre de signaux est arbitraire car toute permutation appliquée sur les sources et sur les lignes de H correspondantes donnent naissance au même vecteur $Y(t)$. Les sources ne peuvent être déterminées qu'à une permutation près.
Il existe aussi une indétermination d'échelle. En effet, posons $h_i(n)$ (respectivement $h_i(z)$) les lignes de $H(n)$ (respectivement $H(z)$), il est évident que les observations i de deux équations (2.4) et (2.5) peuvent être écrites :

$$y_i(z) = h_i(z)S(z) = \frac{h_i(z)}{\alpha_i(z)}(\alpha_i(z)S(z)) \tag{2.6}$$

$$y_i(n) = h_i(n)S(n) = \frac{h_i(n)}{\alpha_i}(\alpha_i S(n)). \tag{2.7}$$

Donc finalement, les sources sont déterminées à une permutation près et à un facteur échelle près pour les mélanges instantanés, et à une permutation près et à un filtre près dans le cas d'un mélange convolutif.
Ces indéterminations entraînent deux conséquences importantes dans le cas de mélanges instantanés :

- N_s paramètres étant indéterminés sur la matrice de mélange, on a $N_s^2 - N_s$ paramètres libres (bien sûr dans le cas où $N_s = N_c$). Et on peut imposer $h_{ii} = 1,\ 1 \leq i \leq N_s$.

- La matrice séparatrice G possède aussi $N_s^2 - N_s$ paramètres libres. Cette matrice peut être aussi contraintée (par exemple avec $g_{ii} = 1,\ 1 \leq i \leq N_s$). Quoique les contraintes sur la matrice séparatrice ne sont pas sans conséquence sur les performances [20].

2.2.4 Hypothèses

A priori les hypothèses sur le mélange et sur les sources sont faibles.

- **Hypothèse 1** : Les sources sont mutuellement indépendantes. Cette hypothèse fondamentale est commune à toutes les méthodes.

- **Hypothèse 2** : La deuxième hypothèse concerne le mélange. La plupart des auteurs discutent les cas de mélanges linéaires (donc instantanés ou convolutifs).
  Dans la suite, on considérera que $N_s = N_c$ sauf pour les méthodes sous-espaces où $N_c > N_s$, (voir chapitre 6). La matrice H est supposée de rang plein et inversible.

- **Hypothèse 3** : Les sources sont non gaussiennes sauf au plus une. On verra plus loin la nécessité de cette hypothèse.

2.3 Indépendance statistique

L'hypothèse 1 d'indépendance est essentielle pour l'estimation de la matrice séparatrice G. Par définition, deux variables aléatoires continues, u_i et u_j, sont indépendantes si leurs densités de probabilité vérifient (voir [104] et [105]) :

$$p(u_i, u_j) = p(u_i)p(u_j). \tag{2.8}$$

Pour les variables discrètes la relation (2.8) sera remplacée par une relation entre les probabilités.
Pour exploiter la notion d'indépendance, décrite dans l'équation (2.8), on peut proposer deux approches :

- Une exploitation directe fondée sur la **divergence de Kullback-Leibner** ([36], [101]).

Soient deux variables aléatoires, u_i et u_j de densités de probabilités $p_{u_i}(v)$ et $p_{u_j}(v)$, la divergence de Kullback-Leibner est la quantité :

$$\delta(p_{u_i}, p_{u_j}) \stackrel{def}{=} \int p_{u_i}(v) \log \frac{p_{u_i}(v)}{p_{u_j}(v)} dv. \tag{2.9}$$

La divergence de Kullback-Leibner est toujours positive, ou nulle si $p_{u_i}(v) = p_{u_j}(v)$.
Il faut noter aussi, que certains auteurs préfèrent utiliser la définition d'information mutuelle [62] :

$$i(p_U) = \int p_U(V) \log \frac{p_U(V)}{\Pi_{i=1}^{N} p_{u_i}(v_i)} dV \tag{2.10}$$

où U est un vecteur aléatoire et les u_i sont ces composantes. Si les u_i sont mutuellement indépendantes, alors $i(p_U) = 0$.

- **Moments et cumulants**. La plupart des algorithmes exploitent indirectement l'indépendance statistique, en utilisant des relations entre les moments ou les cumulants, cette indépendance est basée sur les propriétés de deux fonctions caractéristiques.

Par définition, on appelle première fonction caractéristique ([70], [114], [104], [105]) d'une variable aléatoire continue de dimension p :

$$U^T = (u_1, u_2, \ldots, u_p)^T, \tag{2.11}$$

l'espérance mathématique de la fonction $h(U) = \exp(jV^T U)$. Il s'agit donc d'une fonction $\Phi_U(V)$ définie par :

$$\Phi_U(V) \stackrel{\Delta}{=} E \exp(jV^T U) = \int \exp(jV^T U) dF(U), \tag{2.12}$$

où $F(U)$ est la fonction de répartition de U.
La seconde fonction caractéristique $\Psi_U(V)$ est simplement :

$$\Psi_U(V) = \ln\{\Phi(V)\}. \tag{2.13}$$

Un des principaux intérêts de ces deux fonctions est qu'elles expriment de manière très simple les moments (avec la première fonction (2.12)) et les cumulants (avec la deuxième (2.13)) de la variable aléatoire U. En effet ([70], [104], [105], [98]), le moment d'ordre q d'une variable aléatoire U est donné par :

$$\begin{aligned} Mom_q(u_1, u_2, \ldots, u_q) &\stackrel{\Delta}{=} E(u_1 u_2 \ldots u_q) \\ &= (-j)^q \frac{\partial^q \Phi_U(V)}{\partial v_1 \ \partial v_2 \ldots \partial v_q} \Big|_{V=0}. \end{aligned} \tag{2.14}$$

Pour les cumulants, on trouve que le cumulants[3] d'ordre q de U est donné par :

$$\begin{aligned} Cum_q(U) &= Cum(u_1, u_2, \ldots, u_q) \\ &= (-j)^q \frac{\partial^q \Psi_U(V)}{\partial v_1 \ \partial v_2 \ldots \ \partial v_q} \Big|_{V=0}. \end{aligned} \tag{2.15}$$

D'après la relation (2.15), on montre que les cumulants du U sont tous nuls, s'il y a au moins une composante de U indépendante des autres composantes (voir [90], [98]). En effet, supposons que parmi les q composantes de U, les r premières composantes soient indépendantes des autres, alors la première fonction caractéristique s'écrit :

$$\Phi_U(V) = E\exp(jV^TU) = E\exp(j\sum_{i=1}^{r} v_iu_i)E\exp(j\sum_{i=r+1}^{q} v_iu_i), \tag{2.16}$$

et la seconde fonction devient :

$$\Psi_U(V) = \ln E\exp(jV^TU) \tag{2.17}$$

$$= \ln(E\exp(j\sum_{i=1}^{r} v_iu_i)E\exp(j\sum_{i=r+1}^{q} v_iu_i)) \tag{2.18}$$

$$= \ln(E\exp(j\sum_{i=1}^{r} v_iu_i)) + \ln(E\exp(j\sum_{i=r+1}^{q} v_iu_i)). \tag{2.19}$$

Enfin, en ce qui concerne les cumulants, on trouve que :

$$\begin{aligned} Cum_q(U) &= (-j)^q \frac{\partial^q \ln(E\exp(j\sum_{i=1}^{r} v_iu_i))}{\partial v_1 \ \partial v_2 \ldots \ \partial v_q} \Big|_{V=0} \\ &+ (-j)^q \frac{\partial^q \ln(E\exp(j\sum_{i=r+1}^{q} v_iu_i))}{\partial v_1 \ \partial v_2 \ldots \ \partial v_q} \Big|_{V=0} = 0 \end{aligned} \tag{2.20}$$

Il est clair que la première (respectivement la deuxième) partie de (2.20) dépend uniquement de v_i où $1 \leq i \leq r < q$ (respectivement $r \leq i \leq q$), donc sa dérivée par rapport au vecteur V est nulle.

Donc l'indépendance statistique revient à annuler les cumulants de tous les ordres. Remarquons que ce critère n'est pas manipulable en pratique car on a une infinité de cumulants.

[3]Les cumulants peuvent aussi être définis par la relation suivante :

$$\begin{aligned} Cum_q(U) &= Cum(u_1, u_2, \ldots, u_q) \\ &\triangleq Mom_q(u_1, u_2, \ldots, u_q) - Mom_q(g_1, g_2, \ldots, g_q), \end{aligned}$$

où G est une variable gaussienne qui a la même espérance et la même variance que U. Cette relation met en évidence la rôle des cumulants comme une mesure d'écart à la gaussianité.

2.3.1 Exploitation de l'indépendance

On trouve beaucoup de critères basés sur les cumulants ou les moments. Mais, la construction d'un critère efficace par sélection de quelques cumulants n'est pas aisé. Une approche proposée initialement par Comon est de définir une fonction de contraste [30] et [33].
Définition une fonction de contraste J (voir [30], [31] et [93]) est une application dans $\mathbb{R}$ d'un espace de vecteurs aléatoires $X \in \mathbb{R}^n$, dépendant seulement de la loi de probabilité de x et telle que :

- $J(X)$ est symétrique par rapport aux composantes x_i de X. C.à.d. pour toute matrice de permutation P, on a $J(X) = J(PX)$.
- $J(X)$ est invariant par un changement d'échelle. Soit la matrice diagonale Δ, on trouve $J(\Delta X) = J(X)$.
- Un mélange linéaire de composantes indépendantes ne peut que dégrader le contraste : soit X un vecteur dont les composantes sont indépendantes entre elles, et une matrice quelconque H, on a $J(HX) \leq J(X)$.
- Seuls les permutations et les changements d'échelle conservent le contraste de sources indépendantes : $\forall X$ un vecteur dont les composantes sont indépendantes entre elles, et si pour une matrice H on a $J(HX) = J(X)$ alors forcément $H = \Delta P$.

Pour être réellement utile, un contraste doit également être "discriminant". Un contraste $\Gamma(X)$ est dit discriminant si $\forall x$, on a $\Gamma(AX) = \Gamma(X)$ alors $A = \Delta P$ où P est une permutation et Δ est une matrice diagonale [36].
A titre d'exemple, dans [92], Moreau a montré que la fonction :

$$J(X) = \sum_{i=1}^{n} |Cum(x_i^4)| \tag{2.21}$$

est une fonction de contraste pour les n variables aléatoires x_i, à condition que ces variables soit blanches : une étape de blanchiment spatiale est une transformation d'axes pour rendre les signaux spatialement indépendants.

2.3.2 Ordre de statistiques et limitations sur les signaux

La manipulation de statistiques de tous ordres étant impossible, on cherche alors, l'ordre suffisant pour obtenir la séparation. Par hypothèse, supposons que les sources sont centrées et commençons par les statistiques du second ordre.

Statistique du second ordre. S'il n'y a pas d'hypothèses supplémentaires (autres que les hypothèses 1, 2 et 3 de la section 2.2.4), ces statistiques sont insuffisantes pour achever la séparation. En effet, chaque matrice carrée H est décomposable sous la forme

$$H = U\Delta^{1/2}V, \tag{2.22}$$

où Δ est une matrice diagonale, U et V sont deux matrices unitaires[4]. En supposant que les sources sont à puissances unitaires, alors la matrice de covariance du vecteur observation, pour un mélange instantané, devient :

$$\begin{aligned} \Gamma = E(YY^h) &= E(U\Delta^{1/2}VSS^hV^h(\Delta^{1/2})^hU^h) && (2.23)\\ &= U\Delta^{1/2}VE(SS^h)V^h\Delta^{1/2}U^h && (2.24)\\ &= U\Delta^{1/2}VV^h\Delta^{1/2}U^h && (2.25)\\ &= U\Delta U^h && (2.26) \end{aligned}$$

Il est clair qu'on a perdu dans Γ les informations concernant la matrice V. Par conséquent, on ne peut pas trouver une inverse à gauche de la matrice H et **la séparation devient impossible, sans hypothèse supplémentaire**.

Statistique du troisième ordre. Ces statistiques sont non nulles seulement pour des signaux de densité de probabilité non symétrique. Cette restriction est sévère dans le cas général, et on exclut les statistiques d'ordre trois.

Statistique du quatrième ordre. Ces statistiques sont suffisantes pour séparer les sources ([112], [113], [17],[28] [86], ...). Nous verrons en particulier dans le chapitre 3 "Approche directe", que l'on peut montrer algébriquement l'insuffisance de l'ordre 2 et la suffisance de l'ordre 4, pour des mélanges de deux sources. L'utilisation de statistiques d'ordre supérieur à deux laisse apparaître un problème lié aux signaux gaussiens. En effet, les cumulants d'ordres supérieurs à deux d'un signal gaussien sont tous nuls. Donc la séparation des sources gaussiennes, sans hypothèse supplémentaire, est impossible, sauf dans le cas où on a une et une seule source gaussienne.
Le fait que les cumulants d'ordre élevé des signaux gaussiens soient nuls, nous permet de séparer des sources en présence de bruit gaussien additif. On pourra récupérer les sources mais bruitées par un bruit gaussien.
Les statistiques d'ordre quatre sont utilisées dans plusieurs algorithmes. On donne dans les trois équations suivantes, les relations entre les moments et les cumulants croisés du quatrième ordre pour des signaux supposés centrés [89] :

$$Cum_{13}(u_i, u_j) = Eu_iu_j^3 - 3Eu_i^2Eu_iu_j \tag{2.27}$$

[4]Une matrice unitaire est analogue à une matrice orthogonale, à coefficients complexes. Si U est une matrice unitaire, alors $U^h = U^{-1}$ où U^h est la transposée hermitienne de U.

$$Cum_{22}(u_i, u_j) = Eu_i^2 u_j^2 - Eu_i^2 Eu_j^2 - 2(Eu_i u_j)^2 \quad (2.28)$$
$$Cum_{31}(u_i, u_j) = Eu_i^3 u_j - 3Eu_j^2 Eu_i u_j. \quad (2.29)$$

2.4 État de l'art

2.4.1 Introduction

Le problème de séparation de sources a été formulé initialement par Ans, Hérault [60] et Jutten [61], dans le cas de mélange instantané. Beaucoup d'autres algorithmes sont apparus par la suite, principalement pour des mélanges instantanés à coefficients réels ou complexes ([12], [18],[29], [33], [92], etc). L'intérêt de mélange à coefficients complexes tient du fait que les mélanges convolutifs de signaux à bande étroite (courants en traitement d'antennes) peuvent être traités comme des mélanges instantanés à coefficients complexes. Ce n'est qu'à partir des années 90 que quelques auteurs se sont penchés sur le cas de mélanges convolutifs [2], [96], [85], [87]) de signaux à large bande.

Dans un cadre général, pour chaque type de mélanges (instantané ou convolutif) on peut diviser les différentes méthodes en deux, les méthodes basées sur les statistiques d'ordre élevé [92], et celles basées sur les statistiques d'ordre deux [46].
Dans des contextes particuliers, notamment avec des hypothèses sur les signaux sources, on peut élaborer des algorithmes spécifiques.
A titre d'exemple, Moreau [91], propose une fonction de contraste pour la séparation des signaux discrets (binaires). Malouche et Macchi [83] font la séparation de sources binaires à l'aide d'un réseau de neurones non linéaire bouclé. Capdevielle *et al.* [13] et [14] séparent des signaux périodiques de fréquences très proches.

2.4.2 Mélange instantané

La classification des algorithmes, suivant différentes rubriques, est très difficile. Donc j'ai essayé de classer chaque algorithme par son aspect le plus remarquable. Malgré les efforts faits, le classement donné reste un classement subjectif.

Critères basés sur les moments où les cumulants

Pour une architecture récursive, Jutten *et al.* dans [66], [64] et [63] suggèrent que la diagonale principale de la matrice C (notons que $G = (I + C)^{-1}$) est nulle, et ils ajustent par itération stochastique les autres coefficients suivant la règle :

$$c_{ij}(t+1) = c_{ij}(t) + \alpha f[\hat{x}_i(t)] g[\hat{x}_j(t)], \quad (2.30)$$

où f et g sont deux fonctions impaires si les sources sont à densité de probabilité symétrique, de sorte que les moments d'ordre impair sont nuls [35].
Pour surmonter la limitation sur la densité de probabilité des sources, Jutten *et al.* [68] ont suggéré d'ajuster les coefficients de C par annulation de cumulants croisés $Cum_{31}(x_i, x_j)$.
Jutten et Hérault [66] ont discuté le cas où on a plus de deux sources et ont présenté des résultats expérimentaux pour le cas de 3 sources et 3 capteurs.
La simplicité de l'algorithme de Jutten et Hérault a attiré l'attention de nombreux chercheurs qui ont étudié sa stabilité et ses performances ([81], [91], [47], [111], ...).
Par exemple Fort [47], a montré que si tous les signaux sont sous-gaussiens, la convergence est atteinte. Pour Sorouchyari [111], le choix de points de départ, ainsi que celui du pas d'adaptation α, ont beaucoup d'influence sur la convergence de l'algorithme. L'algorithme était aussi amélioré par une étude faite par Moreau [91] sur le choix du coefficient α.
En utilisant les cumulants, Lacoume et Ruiz [73] ont factorisé la matrice de mélange H (comme toute autre matrice) qui est décomposable en :

$$H = U\Delta^{\frac{1}{2}}V, \tag{2.31}$$

où U et V sont deux matrices unitaires et Δ est une matrice diagonale. Les auteurs montrent la possibilité de trouver U et Δ par une simple décomposition en valeurs singulières de la matrice de covariance de signaux mélangés. Par une méthode "heuristique", Lacoume et Ruiz ont estimé la valeur de V, en maximisant l'inverse de la somme de cumulants croisés de quatrième ordre :

$$F(\theta, X) = \frac{1}{(Cum_{13}(X))^2 + (Cum_{31}(X))^2 + (Cum_{22}(X))^2} \tag{2.32}$$

où θ est l'angle de rotation, si on remplace la matrice V par une rotation de Givens.

Approches Algébriques

Approche de Comon. Sachant que toute matrice carrée se factorise d'une manière unique sous la forme :

$$H = LQ\Delta, \tag{2.33}$$

où L est triangulaire inférieure avec des éléments diagonaux positifs, Q est une matrice de rotation, et Δ est une matrice signature[5], Comon [27] [29] propose

[5] C'est une matrice diagonale dont les coefficients de la diagonale principale, sont égaux à ± 1

une façon directe pour séparer le mélange de deux sources. En effet, pour séparer les sources à une permutation P et une matrice diagonale Δ près, on cherche une matrice F telle que :

$$FH = \Delta P. \tag{2.34}$$

Une solution évidente pour F est $F = Q^h L^{-1}$. Comon a montré qu'une simple factorisation de Cholesky appliquée à la matrice de covariance du mélange, nous donne une estimation de L. La détermination de Q peut être faite par sa décomposition sous la forme d'un produit de N_s^2 rotations planes (rotations de Givens). Le calcul des différents angles de Givens se fait en résolvant des équations polynômiales du second degré, dont les coefficients sont calculés en fonction de cumulants d'ordre quatre. Enfin, dans [28] Comon a appliqué son idée pour trois sources. De même, Cardoso et Comon [19] proposent des solutions algébriques basées sur un formalisme tensoriel.

Approche de Garat. Sur le même principe, Garat [52] propose une méthode algébrique qui consiste à résoudre un système d'équations basé sur les cumulants de signaux mélangés. Il prouve que les colonnes de la matrice de mélange, sont déterminées à un coefficient et une permutation près et elles sont les solutions non nulles d'une équation quadratique et homogène, dont les coefficients sont des expressions de cumulants d'ordre quatre. Dans son algorithme actuel, il se limite à trois sources. Pour un nombre de sources quelconque, il propose une version adaptative (un algorithme "ad-hoc"), de son algorithme, appliquée sur chaque paire de signaux.

Notre approche limitée [86] au cas de deux mélanges de deux sources, pour des raisons de complexité, sera détaillée au chapitre 3.

Fonctions de contraste

La notion de fonction de contraste a été introduité par Comon dans [30] et [33]. La première fonction de contraste proposée par Comon était la somme au carré des auto-cumulants d'ordre quatre des signaux estimés :

$$J(x) = \sum_i |Cum_4(x_i)|^2. \tag{2.35}$$

Cette fonction de contraste est exactement la même fonction proposée indépendamment et par une approche de maximum de vraisemblance par Gaeta [49].

Moreau et Macchi dans [92] suggère que la minimisation de la fonction de contraste proposée par Comon, par rapport à la matrice de séparation, est difficile. Pour cette raison, les auteurs proposent une autre fonction qui est la somme en valeur absolue de cumulant de quatrième ordre :

$$J(x) = \sum_i |Cum_4(x_i)|. \tag{2.36}$$

Cette fonction de contraste est valable pour des sources qui ont le même signe de kurtosis[6]. Enfin, cette fonction nécessite une étape de blanchiment.
Dans [82] et [91], Macchi et Moreau proposent une fonction de contraste qui ne nécessite pas une étape de blanchiment, en introduisant le pré-blanchiment dans leur fonction de contraste.

$$J(X) = \sum_{i}^{N_s} \frac{|Cum_4(x_i)|}{(Ex_i^2)^2} - \beta \sum_{i<j=1}^{N_s} \frac{|Cum(x_i^2 x_j^2)|}{Ex_i^2 Ex_j^2} - \gamma \sum_{i\neq j=1}^{N_s} \frac{|Cum(y_i y_j^3)}{\sqrt{Ey_i^2 (Ey_j^2)^3}} \tag{2.37}$$

Ils montrent que la fonction (2.37) [93] est un contraste pour tout vecteur X, si $\beta \geq 1,\ \gamma \geq 0$.
Dans leur approche de déflation, Loubaton et Delfosse [41] utilisent aussi un critère de contraste (voir le sous-paragraphe suivant).

Déflation

Pour les signaux de même signe de kurtosis, Delfosse et Loubaton [40] proposent une méthode de déflation (en fait, ils restituent un signal à chaque étape). Leur méthode s'inspire de l'approche développée par Shalvi et Weinstein [110].
Après une étape de blanchiment spatiale sur les signaux mélangés ils minimisent une fonction de contraste, par rapport à un vecteur séparateur g :

$$K(g) = \frac{E(gy(n))^4}{4},$$

si $g(n)$ correspond au minimum de K déterminé par un gradient stochastique, alors le signal $x(n) = g(n)y(n)$ est un estimateur de l'une des sources. Et ainsi, on sépare une source à chaque étape et on diminue le nombre des sources d'une unité à la fois. Pour avoir la séparation totale, il faut appliquer leur algorithme $N_s - 1$ fois.
Il faut noter que c'est la seule méthode pour laquelle on a pu montrer (dans le cas général de $N_s > 2$ sources) qu'il n'y a pas de solutions parasites (D'après [41], les minima locaux de K sur la sphère unité de $\mathbb{R}^p$ sont les vecteurs correspondants à l'inverse à gauche de H).
Leur approche peut être étendue au cas où les signes de kurtosis sont quelconques, voir [41].
En utilisant l'ODE (Ordinary Differential Equation, voir [10]), et dans le cas d'un

[6] Le kurtosis est l'auto-cumulant d'ordre quatre normalisé par le carré de puissance du signal, voir chapitre 5

pré-blanchiment exact, ils montrent [38] que la variance asymptotique de l'erreur au premier étage est égale :

$$\Xi = -\frac{E(x_i^6)}{2Cum_4(x_i)}I,$$

où I est la matrice identité et $Cum_4(x)$ est l'auto cumulant d'ordre 4 de x.

Séparation avec un estimateur de Maximum d'Informations Mutuelles

Comon et Lacoume ont montré [36] que l'information mutuelle est une fonction de contraste. L'information mutuelle est défini à partir de la divergence de Kullback-Leibner, voir la relation (2.9) :

$$i(p_x) = \int p_x(u) \log \frac{p_x(u)}{\Pi_{i=1}^N p_{x_i}(u_i)} du$$

on trouve un algorithme semblable dans [62].
Finalement, Pham [101] propose un algorithme d'ACI basé sur la distance de Kullback. Il a présenté dans le même article les performances de son algorithme, sa robustesse et l'absence de solutions parasites.

Méthodes en deux étapes: Blanchiment et Rotation

Plusieurs auteurs ([74], [73], [29], ...) ont proposé des algorithmes de séparation en deux étapes : un blanchiment, qui exploitent les statistiques d'ordre deux, suivi d'une rotation qui est déterminée en générale à l'aide de statistiques du quatrième ordre.
Parmi ces divers algorithmes, on remarque celui proposé par Cardoso et Laheld parcequ'il était le sujet de plusieurs papiers et ces auteurs ont proposé plusieurs versions en étudiant à chaque fois les performances de leurs algorithmes.
Pour les signaux qui ont le même signe de kurtosis Laheld *et al.* dans [75], [74] et [20] proposent deux versions d'algorithmes le PFS "Parameter Free Separation" (dénommé EASI "Equivariant Adaptive Separation via Independence" dans [20]) et le SPFS (Stabilized PFS[7]). Dans [20], et [74] on trouve quelques simulations et une étude de performances pour ces deux versions.
Leur idée principale est basée sur la décomposition de la matrice de séparation G en un produit de deux matrices $G = WU$. W est une matrice de décorrelation spatiale (blanchiment) et U est une matrice unitaire telle que la somme des kurtosis de ses sorties $z = GY$ soit extremum pour les sources de même signe de

[7] On trouve un algorithme IBSS (Iterative Blind Source Separation) semblable à ces derniers algorithmes, dans [6]

kurtosis.
Cardoso *et al.* [18] proposent une version bloc NPFS (Newton PFS) qui est une approximation de la méthode de Newton. Finalement, la matrice séparatrice est ajustée par :

$$G(t+1) = (I + \mu J(T(t)S(t)))G(t), \tag{2.38}$$

où $J(T(t)S(t))$ est le critère d'ajustement de la matrice séparatrice G. Il est lié aux sources et à la matrice globale $T(t)$.
La forme (2.38) permet d'obtenir des performances invariantes[74]. En effet, la matrice globale est ajustée de façon multiplicative sous la forme :

$$T(t+1) = \{I - \mu J(T(t)S(t))\}T(t) \tag{2.39}$$

Le résultat (2.39) est simple mais remarquable, parce qu'il implique que la loi d'évolution du système global est indépendante de la matrice de mélange H.

Séparation avec un estimateur de Maximum de Vraisemblance

Parmi les principes de séparation de sources déjà utilisé, il y a une approche très importante basée sur le maximum de vraisemblance. L'estimateur de vraisemblance est un estimateur asymptotiquement non biaisé et efficace d'où son importance.

L'approche développée par Gaeta et Lacoume [51] et [50] est divisible en deux idées essentielles. La première est la modélisation de la densité de probabilité des sources par un développement de Gram-Charlier à l'ordre quatre. La deuxième est l'utilisation de maximum de vraisemblance pour estimer les différents paramètres. Gaeta, dans [49], utilise les notations tensorielles ([11], [89]) pour montrer la validation de son approche.
Dans ([58], [59], et [72]) Harroy et Lacoume font une étude de performances sur l'algorithme de Gaeta-Lacoume et aboutissent aux bornes de Cramer - Rao, dans le cas des signaux complexes.
Indépendamment Pham *et al.* [103] proposent eux aussi une approche fondée sur le maximum de vraisemblance. En supposant que les sources sont iid avec des densités de probabilités connues, ils dérivent une procédure pour choisir le meilleur estimateur à partir des observations lorsque ces fonctions non-linéaires sont choisies dans un certain espace vectoriel. Garat [52] donne deux algorithmes relevant de cette approche : le premier est QMV-I (Quasi-Maximum de Vraisemblance) pour les signaux iid; le second est QMV-II dédiés aux signaux temporellement corrélées et basé uniquement sur les statistiques de second ordre.

Dans le cas de communications numériques les distributions des sources étant connues, Belouchrani ([6], [9]) propose un algorithme du maximum de vraisemblance pour les signaux indépendants et identiquement distribués (iid), dont les mélanges sont pollués avec un bruit additif $b(t)$ gaussien de moyenne nulle. Pour maximiser sa fonction de vraisemblance il se sert de l'algorithme EM ([43], [80]), pour lequel il donne deux algorithmes : MLS (Maximum Likelihood Séparation) et une version stochastique SMLS (Stochastic Maximum Likelihood Separation).

Statistique du second ordre

Dans de nombreux cas, les signaux ont des échantillons temporellement corrélés on peut construire des algorithmes simplifiés qui exploitent cette information et qui sont basés sur les statistiques de second ordre. L'idée principale de ces derniers algorithmes est basée sur l'utilisation de matrices de covariance, de signaux à différents instants. Toutes ces méthodes nécessitent que **les densités spectrales de sources sont distinctes**[102].
Pour des sources mutuellement indépendantes mais temporellement corrélées, il existe $\tau \neq 0$, tel que :

$$ES(t)S^T(t-\tau) = C(\tau), \tag{2.40}$$

où E est l'espérance mathématique, (S^T est le vecteur transposé de S), et la matrice $C(\tau)$ est une matrice diagonale non nulle.
Féty [46] propose de diagonaliser plusieurs matrices de covariance des signaux observés :

$$\Gamma(\tau) = HC(\tau)H^T, \tag{2.41}$$

jusqu' à la séparation souhaitable.
Cette idée a été améliorée par des travaux faits par Tong [119] puis par Comon [32]. L'algorithme de Comon (qui ressemble à l'algorithme AMUSE proposé par Tong) est détaillé dans [36], [34]. Comon suggère de construire avec un jeu de paramètres α_τ et β_τ deux matrices :

$$\Gamma_1 = \sum_\tau \alpha_\tau \Gamma(\tau) \tag{2.42}$$

$$\Gamma_2 = \sum_\tau \beta_\tau \Gamma(\tau). \tag{2.43}$$

Une fois qu'on a construit ces deux matrices, on détermine à l'aide d'un algorithme quelconque de factorisation deux matrices R et U telles que :

$$\Gamma_1 = RR^h R^{-1}\Gamma_2 R^{-h} = U\Lambda^2 U^h \tag{2.44}$$

U est la matrice de vecteurs propres et Λ est une matrice diagonale de valeurs propres.

Le choix des paramètres α_τ et β_τ est libre à condition que les vecteurs propres de Λ sont distancés le maximum possible. **Notons que le choix des α_τ et β_τ n'était pas défini clairement**, cette raison limite en fait l'application de ce type de méthodes. Enfin, la matrice mélange H est estimée par $H = RU$ à une permutation et une matrice diagonale près.
Belouchrani ([7], [6] et [8]) a proposé un autre développement basé sur le principe de la diagonalisation conjointe (voir [113], [112]). La diagonalisation conjointe approchée d'une famille de K matrices M_l, où $l = 1, \ldots, K$, est la matrice V qui minimise le critère :

$$C(V) = -\sum_{l,i} |v_i^h M_l v_i|^2, \tag{2.45}$$

où v_i est l'ième colonne de V.
Finalement, Pham et Garat [102] et [52] montrent, en utilisant le principe de maximum de vraisemblance, l'existence de filtres optimaux pour faire la séparation. Leur fonctionnelle de vraisemblance L_τ est établie grâce à la propriété de convergence asymptotique des Transformées de Fourier Discrètes $d_S(n)$ des sources indépendantes vers des variables gaussiennes. En effet, asymptotiquement les $d_S(n)$ $(n = 0, ..., \tau/2)$ sont des vecteurs indépendants, centrés, gaussiens et de matrice de covariance diagonale $D_g(n/\tau)$ d'éléments diagonaux $g_1(n/\tau)$, ..., $g_{N_s}(n/\tau)$. Ensuite, la fonctionnelle de vraisemblance L_τ est déterminée par le logarithme de la densité de probabilité conjointe des $d_Y(n)|_{n=0}^{\tau/2}$:

$$L_\tau = -\frac{1}{2}\sum_{i=1}^{N_s}\sum_{n=0}^{\tau/2} \frac{|e_i^T H^{-1} d_S(n)|^2}{g_i(n/\tau)} - T ln|\det H|$$

où e_i^T désigne la ième ligne de la matrice identité et $d_Y(n)$ est la transformée de Fourier discrète du vecteur $Y(t)$.

$$d_Y(n) = \frac{1}{\sqrt{\tau}}\sum_{t=0} \tau - 1 e^{-j2\pi nt/\tau} Y(t). \tag{2.46}$$

Signaux non stationnaires

Matsuoka *et al.* [88] proposent un algorithme pour séparer **les signaux non-stationnaires**. En effet, ils supposent que le **rapport de puissances $Es_i^2(t)/Es_j^2(t)$ n'est pas contant en fonction du temps**. Cette hypothèse est nécessaire, lorsqu'on veut séparer les sources en n'utilisant que l'information de la matrice de covariance. En effet, on sait que la représentation (2.5) n'est pas unique, puisqu'on peut produire les signaux par des séries de sources différentes et $Y = HS = \bar{H}\bar{S}$. Avec l'hypothèse de puissance les auteurs montrent que la matrice de covariance de $\bar{S}$ est déduite de celle de S à une permutation près.
Pour des signaux centrés, ils montrent l'équivalence entre :

- Trouver une matrice de séparation à une permutation et un coefficient d'échelle près.
- Les corrélations de tous les signaux estimés sont nulles :

$$Ex_i(t)xj(t) = 0.$$

- Leur critère est :

$$Q(G, R(t)) = \frac{1}{2}\{\sum_{i=1}^{N} \log(Ex_i^2(t) - \log|EX(t)X^T(t)|\} = 0$$

  Où $R(t)$ est une matrice diagonale dont la diagonale principale est égale aux puissances de signaux estimés.

La matrice séparatrice G est obtenue par annulation de la fonction $Q(G, R(t))$ par un algorithme de gradient.

Approche Géométrique

Dans [107], [108] les auteurs proposent une idée originale de séparation basée sur la représentation géométrique des mélanges. Dans [109], on a présenté un algorithme de ce type, très simple, mais restreint au cas de deux sources, et on a discuté quelques points théoriques (voir chapitre 3).

2.4.3 Mélange convolutif

Dans le domaine de télécommunications (radio-mobiles, GSM), on ne peut pas approximer le mélange convolutif par un mélange instantané, sauf pour des signaux à bande étroite pour lesquels le mélange convolutif se réduit à un mélange instantané à coefficients complexes. Les premières études avec des signaux à large bande ont été présentées dans les années 90.

Statistique d'ordre élevée

Parmis Les premiers algorithmes, on trouve un algorithme proposé par Jutten *et al.* [67] dans le cas de deux sources et deux capteurs, pour un mélange convolutif modélisé par deux filtres à réponse impulsionnelle finie (RIF). La structure de séparation est une matrice de filtres dont les coefficients sont estimés par une généralisation du critère [64] qui était utilisé pour un mélange instantané. En effet, la séparation est assurée en annulant des moments croisés d'ordre impair des sources estimés prises à différents instants.

$$E(f(x_i(n)g(s_j(n_k))) = 0 \tag{2.47}$$

Cet algorithme peut être amélioré en minimisant la somme des trois cumulants croisés d'ordre quatre des signaux estimés [96], [95]), ou en cherchant des fonctions f et g quasi-optimales [21].

Approches fréquentielles

Il est facile de remarquer qu'un mélange convolutif ressemble au mélange instantané, après transformée de Fourier. En effet, la transformée de Fourier, de la relation (2.3), donne :

$$Y(z) = H(z)X(z), \tag{2.48}$$

expression analogue à (2.5). Apparemment, le problème semble très simple : il suffit de faire une transformation de Fourier et d'effectuer la séparation dans chaque bande, et de reconstruire chaque source.
Malheureusement, ce n'est pas le cas. En fait, la séparation est faisable à une permutation et un facteur d'échelle près dans chaque bande. Si la permutation n'est pas trop gênante dans le cas instantané (puisque on récupère toutes les sources à la sortie sans savoir leur ordre), elle est catastrophique dans le cas de signaux large-bande car la reconstruction devient délicate; la simple somme des canaux de même indice n'est pas valable. Il faut donc compléter ces traitements en bande étroite par une opération de reconstruction des sources à partir des composantes estimées à chaque canal de fréquence [14].
Pour cela, Capdevielle (dans [14] et [15]) cherche par filtrage des composantes fréquentielles estimées sur des tranches glissantes de signal, à retrouver une fonction des sources temporelles afin d'obtenir une corrélation significative des filtres issus d'une même source.

Parallèlement, Thirion *et al.* dans ([118], et [117]) proposent une méthode appliquée à la séparation de signaux sismiques. En faisant une transformée de Fourier des signaux mélanges, ils rencontrent le problème de séparation de mélangés instantanés de signaux "multi-bande étroite". Les traitements sur chaque bande de fréquence sont faits de façon indépendante, en faisant un pré-traitement qui consiste en un déroulement de phase, et en utilisant une méthode de maximum de vraisemblance [58]. Et enfin pour séparer le mélange convolutif, les auteurs proposent un algorithme qui reclasse les vecteurs d'ondes estimés en minimisant leurs fluctuations fréquentielles de phase.

Enfin Charkani et Hérault [22], en utilisant un critère basé sur des fonctions semblables aux cumulants d'ordre quatre mais dans le domaine de fréquence, ont montré la possibilité de séparer les sources dans un mélange convolutif.

Statistique du second ordre

Pour l'instant, Les algorithmes apparus pour le mélange convolutif, généralement[8], sont des méthodes de sous-espaces. Pour toutes les méthodes de sous-espaces, on pose une hypothèse commune et très importante, **le nombre de capteurs est plus grand que celui de sources** $N_s > N_c$.
On trouve principalement deux approches :

- la 1ère approche, est développée à partir de l'idée de Moulines *et al.* [94] en égalisation multivariable, on la discutera en détail dans le chapitre 6.
- Toujours, avec une méthode de type sous-espace et pour un mélange convolutif, Delfosse et Loubaton de [42] montrent la possibilité de faire une prédiction linéaire (au second ordre) suivi d'une séparation de mélange instantané (au quatrième ordre), sous l'hypothèse que :
  - le filtre de mélange $H(z)$ est causal et rationnel, et de rang plein, $\forall z$.

  En montrant que les innovations de sources sont les innovations normalisées des signaux mélangés, ils prouvent la possibilité d'extraire ces innovations à une matrice orthogonale U près, à l'ordre deux seulement. Pour identifier U, on utilise les statistiques d'ordre quatre. L'algorithme de Delfosse et Loubaton [38] est basé sur une structure de filtre de synthèse (en treillis) stable pour tout jeu des paramètres.

2.4.4 Application

Le problème de séparation de sources attire l'attention d'un grand nombre de chercheurs, notamment en raison de ses applications diverses :

1. Dans [37], il s'agit d'ACI, pour séparer dans des électrocardiogrammes, le signal correspond au battement de coeur d'un foetus de celui de sa mère.
2. Dans [97], la séparation de sources sert au rehaussement de parole.
3. Desodt *et al.* [44] l'utilisent pour séparer les signaux de radar provenant de plusieurs aires pour contrôler le trafic aérien d'un aéroport.
4. On trouve la séparation dans le domaine de radio-communication, notamment dans les systèmes radio-mobiles, techniques de type SDMA (Spatial Division Multiple Access), ou système d'un téléphone a main-libre [21].

[8]Pour les filtres RIF strictement causaux, Vangerven et Compernolle [120] ont proposé un algorithme basé sur les statistiques de second ordre, aussi bien on trouve un algorithme d'Al-Kindi et Dunlop [3].

5. En sismique, pour étudier les signaux sismiques Thirion [118] a fait de la séparation de sources.

6. Capdevielle [14], sépare les vibrations des machines tournantes, pour contrôler la fonctionnement de chaque machine.

7. Récemment, D'Urso et Cai dans [45] ont appliqué divers algorithmes de séparation de sources pour la séparation des signaux d'un réacteur nucléaire, afin de contrôler la dégradation de l'écran thermique du réacteur.

8. De plus, ces applications peuvent être facilement implantées avec des circuits discrets[65], [26] ou VLSI [121], etc

2.5 Conclusion

Nous avons vu, dans ce passage rapide sur les principes et les applications des algorithmes de séparation l'importance et la richesse de la littérature à ce sujet. Selon le cas on choisit entre tel ou tel algorithme.
Pour pouvoir séparer et récupérer les sources, les auteurs a priori supposent deux hypothèses essentielles. La première est que les sources sont statistiquement indépendantes et la deuxième consiste à poser certains modèles de mélanges (structures du canal).
On remarque que l'inconvénient majeur commun, pour les méthodes basées sur les statistiques d'ordre supérieurs, est l'estimation de ces statistiques.
Par ailleurs en général, les algorithmes bloc sont plus rapides que les algorithmes adaptatifs, mais ils sont moins robustes au bruit additif ou à l'erreur d'estimation.
Le problème de mélange convolutif paraît plus complique que celui de mélange instantané. Et dans plusieurs cas on le ramène, par une approximation ou par l'utilisation de statistiques de second ordre, à un mélange instantané. Ces algorithmes en général sont lourds de point de vu calcul et ont des faibles vitesses de convergence.
Malgré la richesse des algorithmes et des cas particuliers, on va présenter dans les chapitres qui suivent, quelques idées théoriques et quelques simulations pour divers cas et hypothèses.

Chapitre 3

Approches directes

3.1 Introduction

La plupart des méthodes de séparation de sources sont des méthodes basées sur l'estimation d'une matrice séparatrice G,

$$X(t) = GY(t), \tag{3.1}$$

telle que le produit $GH = \Delta P$ est le produit d'une matrice de permutation et d'une matrice diagonale. Les algorithmes (adaptatifs ou bloc) exploitent l'indépendance statistique des sources pour estimer G. Certaines méthodes minimisent des fonctions de coût comme [28], [16], [84] et [73]. D'autres minimisent des fonctions de contrastes ([33], [92]) ou reposent sur l'annulation de multiples critères ([66], [103]).
Dans ce chapitre, on présente deux approches de séparation directe, c'est-à-dire qui estiment la matrice de mélange H par des mesures sur le vecteur mélangé $Y(t)$. Ces approches sont valides dans le cas de mélanges instantanés de deux sources :

- La première méthode repose sur la résolution algébrique d'un système non linéaire d'équations dont les coefficients sont les statistiques des signaux mélangés.

- La seconde approche propose une identification de H à partir de considérations géométriques. Elle suppose que les sources ont des densités de probabilités à support borné.

3.2 Méthode algébrique

3.2.1 Modèle d'équations

Nous limitons notre étude au cas de deux sources indépendantes supposées centrées, non toutes les deux gaussiennes, et deux capteurs dans le contexte d'un mélange instantané. La matrice de mélange (voir fig 2.1) H est donc une matrice réelle 2×2, et l'on supposera que $h_{ii} = 1$. Dans la suite, on utilisera les notations suivantes :

- Les cumulants de signaux mélanges seront notés par

$$Cum_{kl}(y_1, y_2) = Cum(y_1^k(t) y_2^l(t)) = C_{kl}, \tag{3.2}$$

- Les statistiques d'ordre deux et quatre des sources supposées centrées seront notées par :

$$\begin{aligned} p_i &= Es_i^2(t), && (3.3)\\ \gamma_i &= Es_i^4(t), && (3.4)\\ \beta_i &= Cum(s_i^4) = \gamma_i - 3p_i. && (3.5) \end{aligned}$$

3.2.2 Estimation de la matrice de mélange

En écrivant les statistiques des mélanges jusqu'à l'ordre deux, on obtient un système de 3 équations à quatre inconnues : h_{12}, h_{21}, p_1 et p_2. La résolution est donc impossible.
Les statistiques à l'ordre trois s'annulent pour des sources centrées à densité de probabilité paire, on propose d'utiliser directement les statistiques jusqu'à l'ordre quatre.
Les relations fondées sur les moments d'ordre deux et quatre, donnent huit équations non linéaires très complexes mal adaptées à une résolution algébrique (voir annexe A.1). En revanche, le calcul des divers cumulants d'ordre deux et d'ordre quatre conduit à un système de huit équations non linéaires à six inconnues $h_{12}, h_{21}, p_1, p_2, \beta_1$ et β_2 :

$$\begin{aligned} C_{11} = Cum_{11}(y_1, y_2) &= h_{21}p_1 + h_{12}p_2, && (3.6)\\ C_{20} = Cum_{20}(y_1, y_2) &= p_1 + h_{12}^2 p_2, && (3.7)\\ C_{02} = Cum_{02}(y_1, y_2) &= h_{21}^2 p_1 + p_2, && (3.8)\\ C_{31} = Cum_{31}(y_1, y_2) &= h_{21}\beta_1 + h_{12}^3 \beta_2, && (3.9)\\ C_{13} = Cum_{13}(y_1, y_2) &= h_{21}^3 \beta_1 + h_{12}\beta_2, && (3.10) \end{aligned}$$

$$C_{22} = Cum_{22}(y_1, y_2) = h_{21}^2\beta_1 + h_{12}^2\beta_2, \quad (3.11)$$
$$C_{40} = Cum_{40}(y_1, y_2) = \beta_1 + h_{12}^4\beta_2, \quad (3.12)$$
$$C_{04} = Cum_{04}(y_1, y_2) = h_{21}^4\beta_1 + \beta_2, \quad (3.13)$$

Malgré l'apparence compliquée de ce système d'équations non linéaires, on peut le résoudre assez simplement et discuter les solutions en traitant tous les cas possibles. Dans le cas particulier où on a un signal gaussien, on a montré que la séparation est possible. Dans le cas où les deux signaux sont gaussiens, la séparation est impossible. Ces résultats sont bien connus dans le domaine de la séparation de sources, ont été expliqués de façon directe et évidente dans [86].

3.2.3 Sources gaussiennes

Si les deux sources sont gaussiennes, alors leurs cumulants d'ordre supérieurs à deux sont nuls, les équations ((3.9), ...(3.13)) disparaissent. Et donc on n'a pas suffisamment d'équations pour identifier les inconnues.
Si l'une des sources est gaussienne, par exemple la première source ($\beta_1 = 0$), la séparation devient possible. En effet, les équations (3.6) à (3.13) deviennent :

$$C_{11} = h_{21}p_1 + h_{12}p_2 \quad (3.14)$$
$$C_{20} = p_1 + h_{12}^2 p_2 \quad (3.15)$$
$$C_{02} = h_{21}^2 p_1 + p_2 \quad (3.16)$$
$$C_{31} = h_{12}^3\beta_2 \quad (3.17)$$
$$C_{13} = h_{12}\beta_2 \quad (3.18)$$
$$C_{22} = h_{12}^2\beta_2 \quad (3.19)$$
$$C_{40} = h_{12}^4\beta_2 \quad (3.20)$$
$$C_{04} = \beta_2. \quad (3.21)$$

Les cinq dernières équations (3.17) jusqu'à (3.21) contiennent trois variables. Il est facile de montrer que :

$$\beta_2 = C_{04} \neq 0 \quad (3.22)$$
$$h_{12} = \frac{C_{13}}{C_{04}}. \quad (3.23)$$

Les trois relations (3.17), (3.19) et (3.20) donnent :

$$\frac{C_{31}^4}{C_{04}^4} = \frac{C_{22}^6}{C_{04}^6} = \frac{C_{40}^3}{C_{04}^3}. \quad (3.24)$$

cette relation entre les cumulants, permet de vérifier si l'une des deux sources est gaussienne. On peut trouver les autres variables à partir des relations (3.14), ..., (3.16) :

$$h_{21} = \frac{C_{11} - C_{02}h_{12}}{C_{20} - C_{11}h_{12}}, \tag{3.25}$$

$$p_2 = \frac{C_{02} - h_{21}^2 C_{20}}{1 - h_{12}^2 h_{21}^2}, \tag{3.26}$$

$$p_1 = C_{20} - h_{12}^2 p_2. \tag{3.27}$$

L'identification est donc complète.

3.2.4 Sources Non gaussiennes

Solutions exploitant les cumulants du quatrième ordre

Dans cette section, on considère seulement les relations basées sur les cumulants du quatrième ordre (3.9), ..., (3.13). Des deux dernières relations (3.12) et (3.13), nous tirons β_i :

$$\beta_1 = \frac{C_{40} - C_{04}h_{12}^4}{1 - h_{21}^4 h_{12}^4}, \tag{3.28}$$

$$\beta_2 = \frac{C_{04} - C_{40}h_{21}^4}{1 - h_{21}^4 h_{12}^4}. \tag{3.29}$$

Il est clair que les deux dernières relations sont valables si $1 - h_{21}^4 h_{12}^4 \neq 0$, c'est à dire si :

$$1 - h_{21}^2 h_{12}^2 \neq 0. \tag{3.30}$$

Cette dernière équation se divise en deux autres conditions :

- $1 - h_{21}h_{12} \neq 0$, ce qui est équivalent à dire que la matrice mélange H est une matrice inversible. Cette condition est satisfaite par hypothèse.
- $1 + h_{21}h_{12} \neq 0$ est un cas particulier de calcul que l'on discutera dans la section suivante.

En remplaçant β_i par leurs valeurs (3.28) et (3.29) dans les autres équations, on trouve :

$$C_{31}(1 + h_{21}^2 h_{12}^2)(1 + h_{21}h_{12}) = C_{40}h_{21}(1 + h_{21}h_{12} + h_{21}^2 h_{12}^2) + C_{04}h_{12}^3 \tag{3.31}$$

$$C_{13}(1 + h_{21}^2 h_{12}^2)(1 + h_{21}h_{12}) = C_{40}h_{21}^3 + C_{04}h_{12}(1 + h_{21}h_{12} + h_{21}^2 h_{12}^2) \tag{3.32}$$

$$C_{22}(1 + h_{21}^2 h_{12}^2) = C_{40}h_{21}^2 + C_{04}h_{12}^2. \tag{3.33}$$

Les relations (3.32) et (3.33) nous permettent de calculer le coefficient h_{12} :

$$h_{12} = \frac{C_{13} - C_{22}h_{21}}{C_{04} - C_{13}h_{21}}. \tag{3.34}$$

Il est facile de vérifier que :

$$C_{04} - C_{13}h_{21} = \beta_2(1 - h_{12}h_{21}). \tag{3.35}$$

Le dénominateur de (3.34) $C_{04} - C_{13}h_{21}$ est différent de zéro si la matrice H est inversible.
Finalement, en utilisant (3.34) dans (3.33), nous déduisons :

$$\begin{aligned}(C_{40}C_{13}^2 - C_{22}^3)h_{21}^4 + 2C_{13}(C_{22}^2 - C_{40}C_{04})h_{21}^3 \\ +(C_{40}C_{04}^2 + C_{04}C_{22}^2 - 2C_{22}C_{13}^2)h_{21}^2 + C_{04}(C_{13}^2 - C_{22}C_{04}) = 0.\end{aligned} \tag{3.36}$$

Cette dernière équation est une équation algébrique du quatrième degré en h_{21}. La résolution d'une telle équation est faisable par l'intermédiaire de plusieurs changements de variables. Les racines de cette équation ont une forme analytique complexe à écrire, mais qui ne pose pas de problèmes numériques, son intégration dans notre algorithme est facile à faire.

Cas particulier

Nous étudions ici le cas particulier $1 + h_{21}h_{12} = 0$. Il correspond à une matrice de mélange particulière :

$$H = \begin{pmatrix} 1 & -\frac{1}{h_{21}} \\ h_{21} & 1 \end{pmatrix}. \tag{3.37}$$

Ce cas peut être détecté par les relations suivantes :

$$\frac{C_{20}}{C_{02}} = \frac{C_{31}}{C_{13}} = \frac{C_{40}}{C_{22}} = \frac{C_{22}}{C_{04}}. \tag{3.38}$$

En tenant compte de la forme particulière de la matrice de mélange, on peut calculer l'inconnu h_{21} :

$$h_{21} = \pm\sqrt{\frac{C_{02}}{C_{20}}}. \tag{3.39}$$

Solutions plus simples

L'utilisation des cumulants de second ordre simplifie la résolution du système (3.6), ..., (3.13). En utilisant les trois cumulants de second ordre, on montre que le lieu des solutions possibles est une hyperbole :

$$h_{12} = \frac{C_{11} - C_{20}h_{21}}{C_{02} - C_{11}h_{21}}. \tag{3.40}$$

La satisfaction de deux relations (3.40) et (3.34), nous permet de déduire que :

$$(C_{13}C_{20} - C_{22}C_{11})h_{21}^2 + (C_{02}C_{22} - C_{04}C_{20})h_{21} + C_{11}C_{04} - C_{13}C_{02} = 0. \tag{3.41}$$

Finalement, on obtient une équation du second degré. Cette solution est semblable à celle trouvée par Comon [29] en utilisant une parametrisation de la matrice mélange. Pour identifier les coefficients de la matrice de mélange, Comon a proposé un algorithme basé sur deux étapes. La première étape est un blanchiment spatial fait par une décomposition de Cholesky de la matrice de covariance de signaux mélangés. La seconde étape il cherche à trouver, par des équations algébriques l'angle d'une rotation de Givens.

3.2.5 Performances et limitations

Rappelons la définition de la diaphonie résiduelle sur une sortie donnée, c'est le rapport de puissance d'erreur d'estimation de cette sortie sur la puissance de la source elle même :

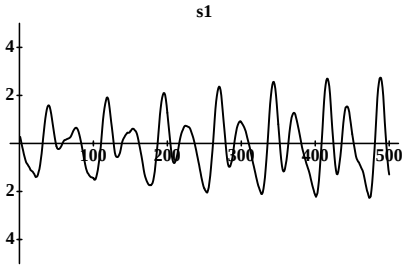
$$\text{Diaphonie} = \frac{E(\hat{s}_i - s_i)^2}{Es_i^2}. \tag{3.42}$$

Dans le cas des signaux stationnaires, et en estimant les statistiques sur 500 échantillons, on sépare les sources avec une diaphonie résiduelle de l'ordre de -24 dB. En appliquant l'algorithme sur des signaux non stationnaires comme des signaux de paroles, on trouve une diaphonie résiduelle de l'ordre de -20 dB, voir la figure 3.1.
Même si les performances de cette méthode ne sont pas excellentes, son importance se manifeste sur tout par :

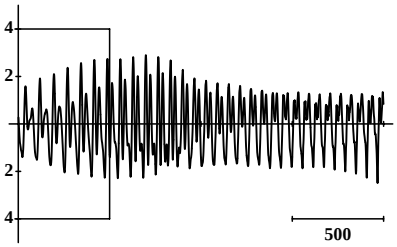
- le fait, qu'on peut l'utiliser comme une méthode simple pour expliquer le problème de la séparation et la limitation vis à vis de la nature des signaux (gaussiens ou pas).
- elle montre d'une façon simple l'insuffisance des statistiques de second ordre pour faire la séparation dans le contexte où on n'a pas d'hypothèses supplémentaires.

- elle peut fournir une bonne initialisation pour un algorithme adaptatif.

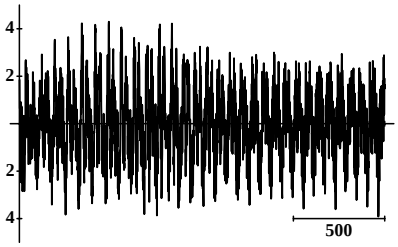
En fin, on remarque bien qu'il y a deux grandes limitations de cet algorithme. En effet, le nombre de sources est égal à deux, et d'autre part les performances de cette méthode dépendent beaucoup de la précision faite sur l'estimation de différents cumulants utilisés.



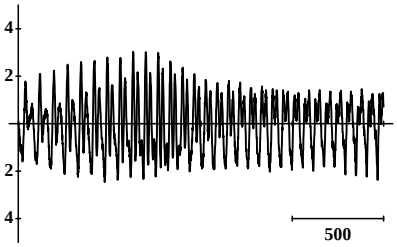
(a) Les 500 échantillons de la source S_1.



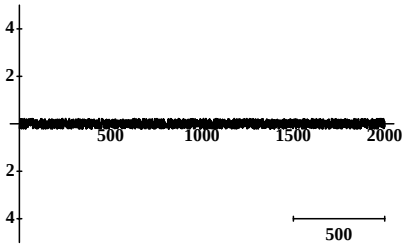
(b) La source S_1.



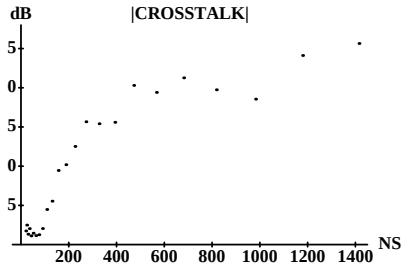
(c) Le signal mélange Y_1.



(d) La source estimée X_1.



(e) Erreur d'estimation.



(f) La valeur absolue de la diaphonie en fonction de nombre d'échantillons utilisés, chaque point est la moyenne de dix manipulations.

Figure 3.1: Performances de la méthode directe.

3.3 Approches géométriques

Dans cette section, nous proposons une méthode géométrique simple pour la séparation aveugle de sources, développée en coopération avec C. Puntonet ([106], [108] et [107]), dans le cadre d'une action intégrée.

3.3.1 Représentation géométrique des solutions

Pour N_c observations, les signaux mélangés $y_j(t)$ $(1 \leq j \leq N_c)$ sont liés aux sources $s_i(t)$ $(1 \leq i \leq N_s)$ par :

$$y_j(t) = \sum_{i=1}^{N_s} h_{ij} s_j(t). \tag{3.43}$$

Pour simplifier la représentation géométrique de la méthode, on discutera seulement le cas où $N_s = N_c = 2$. La relation 3.43 devient :

$$\begin{cases} y_1(t) = s_1(t) + a s_2(t) \\ y_2(t) = b s_1(t) + s_2(t). \end{cases} \tag{3.44}$$

Si les sources sont statistiquement indépendantes et bornées, les points $(y_1(t), y_2(t))$ sont inscrits dans un parallélogramme dans le plan (y_1, y_2) (voir 3.2). En utilisant la relation (3.44), on remarque :

- Pour $s_1(t) = s_1 = cte$, on a $y_2(t) = \frac{1}{a} y_1(t) + (b - \frac{1}{a}) cte$.
- Pour $s_2(t) = s_2 = cte$, on a $y_2(t) = b y_1(t) + (1 - ab) cte$.

On en déduit que les pentes du parallélogramme sont égales à b et $1/a$. L'estimation de ces pentes nous permet d'estimer directement la matrice de mélange.

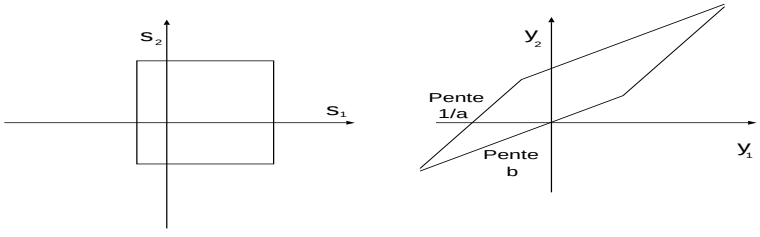


Figure 3.2: Les espaces des sources et du mélange, pour des sources bornées.

Il est facile de voir que si les sources sont simplement "semi-bornées" (par exemple $s_i(t) \in [0, \infty[$), la séparation par la méthode géométrique est encore possible. Le mélange de sources, au lieu d'être inscrit dans un parallélogramme, est représenté dans un secteur angulaire (voir Fig. 3.3).

A titre d'exemple, on a tracé les distributions correspondant à divers signaux mé-

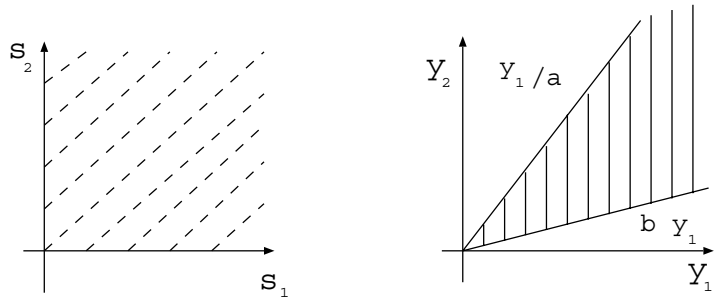


Figure 3.3: L'espace source et celui de mélange, pour des sources positives.

langés $(y_1(t), y_2(t))$ $(1 \leq t \leq 5000)$ dans le plan (y_1, y_2). La figure 3.4 correspond à un mélange de deux sources de deux densités de probabilités (ddp) uniformes.

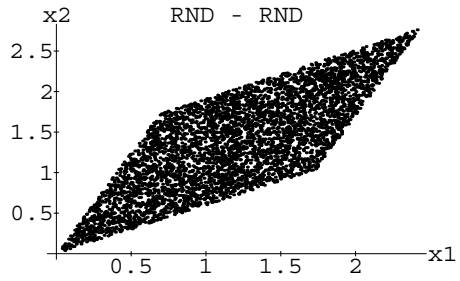


Figure 3.4: Distributions de deux sources uniformes.

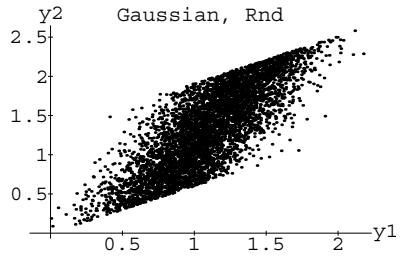


Figure 3.5: Distributions des mélanges d'une source uniforme et d'une source gaussienne.

Dans la figure 3.5, les deux sources ont des ddp uniforme et gaussienne.
La méthode géométrique peut aussi être appliquée pour des sources déterministes, la figure 3.6 montre les distributions pour les deux sources $s_1(t) = sin(2\pi f_0 t)$ et $s_2(t) = sin(2\pi f_1 t + \phi)$).
A la figure 3.7, on a tracé les distributions des mélanges de deux sources gaus-

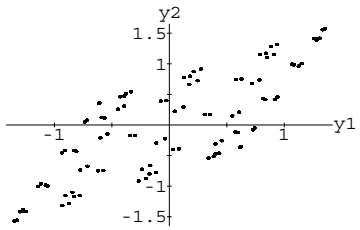


Figure 3.6: Distributions des mélanges de deux sources sinusoïdales.

siennes blanches. Ce mélange ne peut pas être résolu ni par la méthode géométrique, ni par des méthodes statistiques.

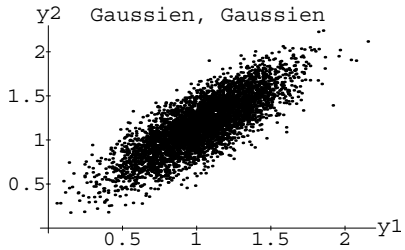


Figure 3.7: Distributions des mélanges de deux sources gaussiennes.

3.3.2 Algorithme

D'après la figure 3.2, il est évident qu'on peut calculer les deux pentes par une simple recherche du minimum et du maximum du rapport $r(t) = \frac{y_2(t)}{y_1(t)}$. Dans le cas de signaux bornés $s_i(t) \in [0, M_i]$, on peut calculer les coefficients de la matrice mélange soit par $r_{min} = \hat{b}$ et $r_{max} = 1/\hat{a}$, soit par $r_{min} = 1/\hat{a}$ et $r_{max} = \hat{b}$. Le choix arbitraire entre la première ou la deuxième solution correspond à l'indétermination de permutation. En effet, les deux solutions correspondent à :

$$\hat{H}_1 = \begin{pmatrix} 1 & \hat{a} \\ \hat{b} & 1 \end{pmatrix}, \tag{3.45}$$

ou

$$\hat{H}_2 = \begin{pmatrix} 1 & 1/\hat{b} \\ 1/\hat{a} & 1 \end{pmatrix}. \tag{3.46}$$

On obtient donc deux solutions pour la matrice globale, $T_i = \hat{H}_i^{-1} H$:

$$T_1 = \frac{1}{1-\hat{a}\hat{b}} \begin{pmatrix} 1-\hat{a}b & a-\hat{a} \\ b-\hat{b} & 1-a\hat{b} \end{pmatrix}, \tag{3.47}$$

ou

$$T_2 = \frac{1}{1-\hat{a}\hat{b}} \begin{pmatrix} \hat{a}(b-\hat{b}) & \hat{a}(1-a\hat{b}) \\ \hat{b}(1-\hat{a}b) & \hat{b}(a-\hat{a}) \end{pmatrix}. \tag{3.48}$$

Il est clair que si $\hat{a} \to a$ et $\hat{b} \to b$, les deux solutions sont les solutions séparatrices à un facteur d'échelle et à une permutation près.
On propose un algorithme décomposable en deux étapes :

- une étape de translation pour placer l'origine sur un sommet.

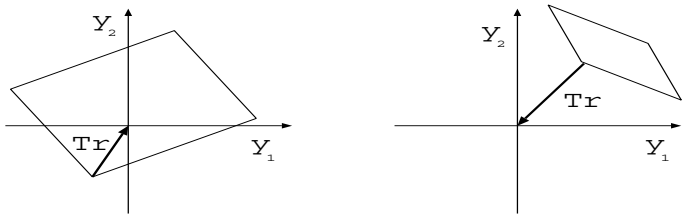


Figure 3.8: Translation d'un sommet.

- La deuxième étape consiste à chercher le minimum (respectivement le maximum) de $r(t)$ (voir [109]).

L'algorithme en code MATLAB est

Algorithme 1 *Algorithme géométrique*
for $t = 1 : N$
$Tr(t) = (y(1,t))^2 + (y(2,t))^2$;
end;
$[T, to] = max(Tr)$;
$e = zeros(2, N)$;
for $t = 1 : N$,
$y(:,t) = y(:,t) - y(:,to)$;
end
for $t = 1 : N$,
if $((y(1,t) == 0)|(y(2,t) == 0)) == 1$,
else
$rm(t) = y(2,t)/y(1,t)$;
end;
end;
$rmin = min(rm)$;
$rmax = max(rm)$;
$b = rmin$;
$a = inv(rmax)$;
$M1 = [1 \quad a; b \quad 1]$;

3.3.3 Performances théoriques de la méthode

Supposons qu'on ait une erreur d'estimation ϵ, le paramètre estimé est alors : $\hat{b} = b + \epsilon$. Donc au lieu de mesurer la vraie pente (correspond à $y_2 = by_1$), on mesure celle de droite D, qui est définie par $y_2 = (b + \epsilon)y_1$. Soit α le secteur de confiance angulaire, associé à cette erreur (voir figure 3.9).
La droite D est la transformation de la droite Δ dans le plan de sources, par le mélange linéaire. D'après la relation :

$$\begin{aligned} \begin{pmatrix} s_2 \\ s_1 \end{pmatrix} &= \frac{1}{1-ab} \begin{pmatrix} 1 & -a \\ -b & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \\ &= \frac{1}{1-ab} \begin{pmatrix} 1 & -a \\ -b & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ (b+\epsilon)y_1 \end{pmatrix}, \end{aligned}$$

on déduit que la droite D est la transformé de la droite Δ d'équation :

$$s_2 = \frac{\epsilon}{1 - ab - a\epsilon} s_1 \tag{3.49}$$

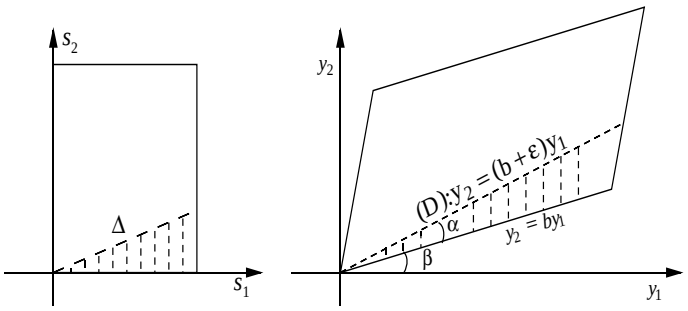


Figure 3.9: Secteur de confiance.

Supposons maintenant que nos sources aient des ddp uniformes et que $s_1 \in [0, M_1]$ et $s_2 \in [0, M_2]$, alors la probabilité P_ϵ d'avoir un point dans le secteur de confiance est donnée par :

$$\begin{aligned} P_\epsilon &= \int_0^{M_1} \left(\int_0^{\frac{\epsilon s_1}{1-ab-a\epsilon}} \frac{1}{M_1 M_2} ds_2 \right) ds_1 \\ &= \frac{\epsilon M_1}{2M_2(1-ab-\epsilon a)}. \end{aligned} \tag{3.50}$$

Pour un nombre total d'échantillons égale à N, le nombre d'échantillons dans le secteur de confiance N_s, est asymptotiquement tel que :

$$\frac{N_s}{N} = P_\epsilon,$$

pour N suffisamment grand. Pour obtenir une mesure avec une précision ϵ et une probabilité P_ϵ, il faut mesurer au moins un point dans la zone de confiance, c'est à dire $N_s \geq 1$, d'où $N \geq 1/P_\epsilon$. Enfin le nombre N doit vérifier l'inégalité suivante :

$$N >> \frac{2M_2(1-ab-a\epsilon)}{\epsilon M_1}. \tag{3.51}$$

En supposant que les sources ont la même puissance et si on suppose que les deux paramètres de la matrice mélange sont estimés par $\hat{a} = a + \epsilon_a$ et $\hat{b} = b + \epsilon_b$, alors la diaphonie résiduelle est égale à :

$$C_i = \frac{\epsilon_i}{1 - ab - a\epsilon_b - b\epsilon_a}, \text{avec } i \in \{a, b\}. \tag{3.52}$$

Pour une diaphonie résiduelle $C_a = C_b$ donnée, on en déduit, ϵ_a et ϵ_b puis P_{ϵ_a} et P_{ϵ_b}, ce qui impose un nombre minimum de point :

$$N > \max\left(\frac{1}{P_{\epsilon_a}}, \frac{1}{P_{\epsilon_b}} \right) \tag{3.53}$$

Ainsi, on peut calculer N en fonction de la diaphonie résiduelle désirée. Les performances expérimentales, ont été étudiées dans voir [109] et [108].

3.3.4 Estimateur Maximum de vraisemblance

Dans cette sous-section, on montre que la méthode géométrique correspond en fait à un estimateur du maximum de vraisemblance. Pour simplifier les équations et la compréhension, on suppose que les deux sources (s_1, s_2) sont statistiquement indépendantes et de ddp uniformes :

$$p(s_i) = \Pi_{[0,\theta]} = \begin{cases} 1 & \text{si} \quad 0 < s_i < \theta \\ 0 & \text{ailleurs} \end{cases} \tag{3.54}$$

La ddp du vecteur (s_1, s_2) est donnée par :

$$p_\theta(s_1, s_2) = (\frac{1}{\theta}\Pi_{[0,\theta]}(s_1))(\frac{1}{\theta}\Pi_{[0,\theta]}(s_2)) \tag{3.55}$$

D'après cette équation, on déduit que :

$$p_\theta(s_1, s_2) = \frac{1}{\theta^2} \quad \text{si} \begin{cases} 0 < s_1 < s_2 < \theta \\ \text{ou} \\ 0 < s_2 < s_1 < \theta \end{cases} \tag{3.56}$$

$$p_\theta(s_1, s_2) = 0 \quad \text{ailleurs}, \tag{3.57}$$

autrement, on peut écrire les relations (3.56) et (3.57) par :

$$p_\theta(s_1, s_2) = \frac{1}{\theta^2} \quad \text{si} \begin{cases} 0 < \min(s_1, s_2) < \max(s_1, s_2) \\ \text{et} \\ \min(s_1, s_2) < \max(s_1, s_2) < \theta \end{cases} \tag{3.58}$$

$$p_\theta(s_1, s_2) = 0 \quad \text{ailleurs} \tag{3.59}$$

En notant par $\Pi_{[a,b]}(t)$ la fonction définie par :

$$\Pi_{[a,b]}(t) = \begin{cases} 1 & \text{si} \quad a \leq t \leq b \\ 0 & \text{ailleurs} \end{cases} \tag{3.60}$$

Cette dernière équation nous permet d'écrire que :

$$p_\theta(s_1, s_2) = \frac{1}{\theta^2}\Pi_{[0,\max(s_1,s_2)]}(\min(s_1, s_2)\Pi_{[\min(s_1,s_2),\theta]}(\max(s_1, s_2)) \tag{3.61}$$

Regardons les choses d'un autre point de vue : si on a mesuré un couple (s_1, s_2), alors la valeur de θ qui maximise la ddp $p_\theta(s_1, s_2)$, et qui est déduite du principe du maximum de vraisemblance, est donnée par :

$$p_\theta(s_1, s_2) = \frac{1}{\theta^2}\Pi_{[\min(s_1,s_2),\theta]}(\max(s_1, s_2)) \tag{3.62}$$

$$= \frac{1}{\theta^2}\Pi_{[\max(s_1,s_2),\infty]}(\theta) \tag{3.63}$$

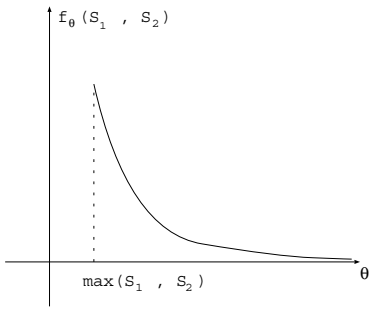


Figure 3.10: La fonction de vraisemblance

En regardant la fonction de vraisemblance tracée dans la figure 3.10, on peut voir que cette fonction est nulle pour toute valeur de $\theta < \max(s_1, s_2)$. Pour une valeur de $\theta > \max(s_1, s_2)$, la fonction de vraisemblance prend la valeur $\frac{1}{\theta^2}$. Par conséquent, le maximum de cette fonction est atteint pour le valeur de $\theta = \max(s_1, s_2)$, et cette valeur nous permet de déterminer les bornes exactes de ddp de sources. D'une autre façon, en maximisant la fonction de vraisemblance, on pourra estimer par un rectangle de paramètres variables (longueur, largeur) la distribution de deux sources, ce qui revient à estimer aussi le parallélogramme correspondant dans le plan de mélange.
Cette démonstration rapide montre qu'on peut considérer la méthode géométrique, en quelque sorte, comme l'estimateur de maximum de vraisemblance.

3.3.5 Performances et limitations

On résume l'algorithme géométrique par :

- C'est un algorithme simple et qui tend vers les solutions exactes si les sources sont à ddp de supports bornées

- Convergence rapide : on a une convergence avec 1000 ou 1500 échantillons.

- Par contre il est sensible au bruit, surtout un bruit gaussien additif. En effet, on remarque d'après la figure 3.11 que les bornes du parallélogramme ne sont plus nettes avec un bruit gaussien additif.
  Par contre, si les sources sont bruitées par un bruit uniforme alors la séparation est faisable. Puisque Cela revient à translater les bornes de notre parallélogramme.
  Dans le cas où le rapport signal sur bruit est suffisant alors on arrive à estimer les coefficients de la matrice mélange.

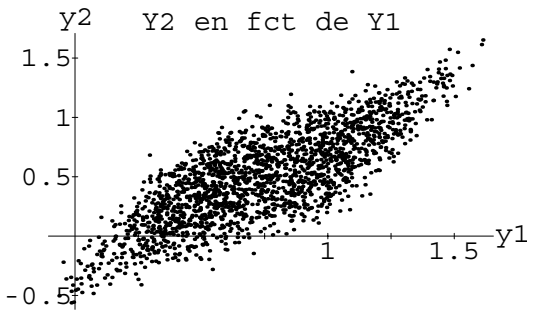


Figure 3.11: Sources polluées par un bruit gaussien.

- Pour deux sources de ddp bornés, on a mesuré une diaphonie résiduelle de -20 dB, en appliquant l'algorithme sur des signaux de 500 échantillons.
- La méthode est généralisable à plus de deux sources, mais l'algorithme dans son cas actuel n'est pas bien adapté à toutes les situations possibles (matrice de mélange et nature de signaux). On travaille actuellement sur un algorithme qui sera valable pour un nombre $n > 2$ de sources.

3.4 Conclusion

Dans ce chapitre, nous avons proposé deux approches directes, l'une algébrique, l'autre géométrique. Les méthodes restent simples dans le cas de deux mélanges de deux sources, mais leur généralisation à un nombre quelconque de sources est délicate.
Pour l'approche algébrique, les calculs deviennent trop complexes; Pour l'approche géométrique, l'algorithme, très simple pour deux sources, devient aussi très complexe [116] et il semble que les méthodes statistiques restent plus simples, au moins par leur caractère général.

Chapitre 4

Un critère simple pour les mélanges instantanés et convolutifs

4.1 Introduction

La plupart des algorithmes de séparation de sources sont basés sur des statistiques d'ordre élevé de signaux estimés, généralement d'ordre quatre, sauf si on a plus d'hypothèses comme on l'a vu dans le chapitre 2. Par exemple, dans [73], les auteurs estiment les paramètres de mélange par la maximisation de :

$$d = \frac{1}{Cum_{13}^2(x_i, x_j) + Cum_{22}^2(x_i, x_j) + Cum_{31}^2(x_i, x_j)}, \tag{4.1}$$

et dans [16], Cardoso a proposé deux méthodes, l'une basée sur les moments d'ordre quatre et l'autre sur les cumulants d'ordre quatre. Enfin Comon, dans [28], trouve la séparation par une résolution d'un système d'équations utilisant les cumulants des sources estimées.
Pour des mélanges instantané et convolutif, Nguyen Thi *et al.* [97] proposent une solution qui annule la somme quadratique des cumulants croisés Cum_{13} et Cum_{31}. Cependant expérimentalement, Oliva [99] a montré que cet algorithme présente, pour certains types des signaux, des solutions parasites, qui peuvent être éliminées par l'annulation d'un cumulant croisé supplémentaire Cum_{22}.
Ces résultats nous ont poussés à chercher le coût le plus simple possible, et à nous interroger en particulier sur les solutions générées par l'annulation du seul cumulant Cum_{22}.
Initialement, nous avons considéré le cas de deux mélanges instantanés de deux sources vues par deux capteurs (les résultats ont été publiés dans l'annexe [84]).

Par la suite, nous avons pu généraliser ce critère dans le cas de N_s sources et de N_c capteurs pour des mélanges instantanés et convolutifs [85]. Dans ce chapitre nous nous focalisons sur ces derniers résultats plus généraux.

4.2 Mélange instantané de N_s sources

Le cas où le nombre de sources est différent du nombre de capteurs est discuté dans [63], et de la même manière que [49] on peut considérer que le nombre de sources et de capteurs sont identiques.
Pour des mélanges instantanés, la matrice de mélange H est alors une matrice carrée $N_s \times N_s$ à coefficients réels, et la matrice séparante G est aussi une matrice carrée $N_s \times N_s$ à coefficients réels (voir fig 2.1).
La matrice globale correspondante à l'application entre $s(t)$ et $x(t)$ est notée :

$$T = HG, \tag{4.2}$$

on a donc :

$$X(t) = TS(t). \tag{4.3}$$

4.2.1 Cumulant croisé d'ordre 4

Pour des mélanges non bruités, en se basant sur les propriétés de multilinéarité des cumulants [98] et la relation entre les sources et les sorties, on peut montrer que, pour $m \in [1, N_s]$ et $n \in [1, N_s]$:

$$\begin{aligned} Cum_{22}(x_m, x_n) &= Cum(x_m, x_m, x_n, x_n) \\ &= Cum(\sum_{a=1}^{N_s} t_{ma} s_a, \sum_{b=1}^{N_s} t_{mb} s_b, \sum_{c=1}^{N_s} t_{nc} s_c, \sum_{d=1}^{N_s} t_{nd} s_d) \\ &= \sum_{a=1}^{N_s} \sum_{b=1}^{N_s} \sum_{c=1}^{N_s} \sum_{d=1}^{N_s} Cum(t_{ma} s_a, t_{mb} s_b, t_{nc} s_c, t_{nd} s_d) \\ &= \sum_{a,b,c,d} t_{ma} t_{mb} t_{nc} t_{nd} Cum(s_a, s_b, s_c, s_d). \end{aligned} \tag{4.4}$$

Les sources étant supposées mutuellement indépendantes, les cumulants d'ordre quatre de la forme $Cum(s_a, s_b, s_c, s_d)$ sont nuls si deux sources au moins sont indépendantes, et l'on peut écrire :

$$\begin{aligned} Cum(s_a, s_b, s_c, s_d) &= Cum(s_a, s_a, s_a, s_a) \delta_{a,b}^{c,d} \\ &= \beta_a \delta_{a,b}^{c,d}, \end{aligned} \tag{4.5}$$

où $\beta_a = Cum(s_a, s_a, s_a, s_a)$ et $\delta_{a,b}^{c,d}$ est la généralisation du symbole de Kronecker :

$$\delta_{a,b}^{c,d} = \begin{cases} 1 & \text{si} \quad a = b = c = d \\ 0 & \text{sinon} \end{cases} \tag{4.6}$$

L'équation (4.4) devient :

$$Cum_{22}(x_m, x_n) = \sum_{a=1}^{N_s} t_{ma}^2 t_{na}^2 \beta_a. \tag{4.7}$$

Si les sources ont le même signe de kurtosis, on peut facilement montrer [84] que :

$$Cum_{22}(x_m, x_n) = 0 \Longleftrightarrow t_{ma} t_{na} = 0 \quad \forall a, \text{et } m \neq n. \tag{4.8}$$

A partir de ces dernières relations, on peut déduire que chaque colonne de la matrice globale T contient au maximum un coefficient non nul.

4.2.2 Mélange non bruité

De la relation (4.8), on a déduit la propriété :
Propriété 1 : les colonnes de la matrice globale ont au maximum un coefficient non nul.
Cette propriété implique que les signaux de sorties $x_i(t)$ sont nuls ou bien spatialement indépendants.
En effet, si le coefficient $t_{ma} \neq 0$, alors la relation (4.8) montre que $t_{na} = 0$, $\forall n \neq m$. Une étude plus détaillée sur le nombre et la nature des solutions, est donnée dans l'annexe B.5.
Si le nombre de capteurs est égal[1] à celui des sources, $N_s = N_c$, alors les sorties sont les sources à une permutation et un facteur d'échelle près.
Les solutions correspondantes à des sorties nulles peuvent être éliminées d'une manière simple, en fixant les coefficients de la diagonale principale de la matrice poids (G) à un :

$$g_{ii} = 1. \tag{4.9}$$

En effet, notons par H_j les colonnes de la matrice mélange et G_i la ligne i de la matrice séparante. La condition (4.9) implique que $G_i \neq 0$. Dans ce cas, toutes les lignes de la matrice globale sont non nulles. Car en supposant que la ligne i de cette matrice est nulle :

$$t_{ij} = G_i H_j = 0, \quad \forall j, \tag{4.10}$$

[1]Si on a plus de capteurs que de sources, on peut se ramener au cas d'égalité voir [79].

le vecteur G_i sera orthogonal à tous les vecteurs H_j. Or cette proposition est impossible puisque : la matrice H étant une matrice régulière, ses lignes sont linéairement indépendantes; et les vecteurs $G_i \neq 0$. Il est donc impossible de trouver un vecteur non nul $G_i \neq 0$, dans l'espace $\mathbb{R}^{N_s}$, qui soit perpendiculaire à N_s vecteurs linéairement indépendants. Il existe donc j tel que $t_{ij} \neq 0$. La contrainte (4.9) entraîne la seconde propriété :

Propriété 2 : les lignes de la matrice globale ont au moins un coefficient différent de zéro.

Les deux propriétés, Propriété 1 et Propriété 2, prouvent que la matrice globale est le produit d'une matrice diagonale et d'une matrice de permutation.
L'annulation des cumulants 22 nous permet de faire la séparation à une matrice diagonale et une matrice de permutation près. Donc l'annulation de notre critère est équivalent à :

$$G = \Lambda P H^{-1}, \tag{4.11}$$

où Λ est une matrice diagonale et P est une matrice permutation. Si on suppose que les sources sont pollués par un bruit gaussien[2], alors les sorties estimées sont alors :

$$\begin{aligned} X &= GY = \Lambda P H^{-1} H S + \Lambda P H^{-1} N \\ &= \Lambda P S + \Lambda P H^{-1} N. \end{aligned} \tag{4.12}$$

L'estimation de G n'est pas influencée par le bruit gaussien additif. Cependant, les sources estimées sont pollués par un bruit résiduel correspondant au terme $\Lambda P H^{-1} N$ de (4.12). Enfin on a ramené le problème à un problème de détection d'un signal bruité par un bruit gaussien.

4.2.3 Notations tensorielles

Dans la suite de ce chapitre, on utilisera les notations tensorielles pour alléger l'écriture des relations et des équations. Pour cela, on va rappeler dans cette sous-section les notations tensorielles qui seront utilisées dans la suite.
Rappelons les définitions utiles dans la notation tensorielle. Soit $\mathcal{H}$ l'espace d'Hilbert et $\mathcal{H}^D$ l'espace dual. Notons par $S^1 \in \mathcal{H}$ un tenseur covariant, et dans ce cas $S_1 = {S^1}^T \in \mathcal{H}^D$ est un tenseur contravariant et il est le tenseur dual de S^1. Dans les deux cas, le nombre 1 ne répresente pas un indice ni une puissance. il correspond à la variance d'un tenseur.
Par définition, la nature et les dimensions du tenseur sont caractérisées par sa

[2]Les relations développées dans cette section restent invariant si on ajoute aux signaux mélangés un bruit gaussien, voir après.

variance. Par exemple S_3^2 est un tenseur 2 fois covariant et 3 fois contravariant, le nombre de composantes de ce tenseur est $2+3$ fois la dimension de l'espace $\mathcal{H}$, finalement son terme général sera noté par s_{klm}^{ij} avec i, j, k, l et m prenent des valeurs entre 1 et la dimension de l'espace $\mathcal{H}$.
Dans la suite on notera par :

1. $\bullet$ est le produit de contraction : le produit de contraction tensoriel est l'équivalent de produit matriciel. Pour un tenseur $A^1 = B_1^1 \bullet C^1$, le terme général est calculé par $a^i = \sum_j b_j^i c^j$. Il nous reste à remarquer que B_1^1 est un tenseur une fois contravariant et une fois covariant.

2. $\otimes$ est le produit tensoriel : le produit tensoriel de deux tenseurs B_1^1 et C_2^1 est un tenseur $A_3^2 = B_1^1 \otimes C_2^1$ de terme général $a_{ijk}^{lm} = b_i^l c_{jk}^m$. Il est clair que le tenseur A contient $(N_c)^5$ coefficients différents[3], et que la notation tensorielle nous permet de l'écrire de la façon la plus compacte possible.

3. Enfin, l'addition et la soustraction entre tenseurs ressemblent à celles utilisées dans le cas matriciel, par consequent elles s'appliquent sur les tenseurs de même variance.

Lorsque l'on présisera les composantes des tenseurs, la notation d'Einstein sera adoptée (voir [11]). Un exemple sur la notation d'Einstein :
Soit la somme $a_l = \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} b_{li} c_{ij} d_j$, selon la notation d'Einstein cette équation s'écrit sous la forme $a_l = b_{li} c_{ij} d_j$.

4.2.4 Cas des mélanges bruités

Si les observations sont polluées par un bruit additif $N(t)$ on aura la structure donnée par la figure 4.1.
On suppose que les composantes du vecteur $N(t)$ sont indépendantes des sources

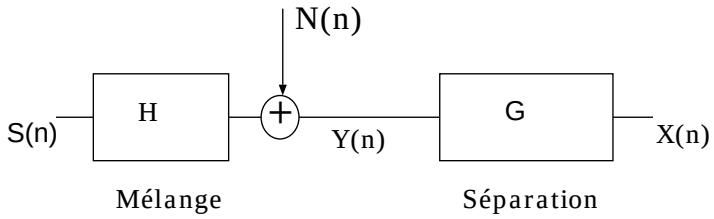


Figure 4.1: Structure générale avec bruit.

$s_i(t)$, avec $(i = 1, \ldots, N_s)$ et non corrélées entre elles. En plus, ses composantes

[3]En supposant que la dimension de l'espace $\mathcal{H}$ est égale au nombre des sources.

sont supposées gaussiennes. Cette hypothèse est forte mais elle est justifiable à partir du moment où on sait que, physiquement, on a du bruit gaussien.
Dans la suite, les notations utilisées sont les notations tensorielles :

1. S^1 le tenseur source.
2. N^1 le tenseur bruit.
3. T_1^1 le tenseur global.
4. H_1^1 le tenseur de mélange.
5. G_1^1 le tenseur séparant.

La relation entre les signaux mélangés et les sorties est :

$$\begin{aligned} X^1 &= G_1^1 \bullet Y^1 \\ &= G_1^1 \bullet (H_1^1 \bullet S^1 + N^1) \\ &= T_1^1 \bullet S^1 + G_1^1 \bullet N^1 \end{aligned} \tag{4.13}$$

Dans le cas général, le cumulant Cum_{22} est un tenseur deux fois contravariant et deux fois covariant (voir [89] et [49]), il sera donc noté par Cum_2^2. Les moments d'ordre N seront notés par :

$$Mom_p^{N-p} = ES^{1\,\otimes(N-p)} \otimes S_1^{\otimes p}, \tag{4.14}$$

où $A^{\otimes p} = \underbrace{A \otimes \ldots \otimes A}_{p}$.
La valeur du tenseur Cum_2^2 peut-être exprimée, selon (voir [11] et [50]) pour des signaux centrés :

$$Cum_2^2 = Mom_2^2 - 2Mom_1^1 \otimes Mom_1^1 - Mom_2^0 \otimes Mom_0^2 \tag{4.15}$$

En utilisant les propriétés de multilinéarités des cumulants, l'indépendance spatiale entre les sources, l'indépendance entre le bruit et les sources, et en notant en plus que les bruits sont des signaux indépendants gaussiens, on montre facilement que :

$$\begin{aligned} Cum_{X_2^2} &= T_1^1 \otimes T_1^{1\,T} \otimes T_1^1 \otimes T_1^{1\,T} \bullet Cum_{S_2^2} \\ &\quad + G_1^1 \otimes G_1^{1\,T} \otimes G_1^1 \otimes G_1^{1\,T} \bullet Cum_{N_2^2}, \end{aligned} \tag{4.16}$$

sachant que $Cum_{X_2^2} = Cum(X^1, X_1, X^1, X_1)$ est le cumulant d'ordre 2×2. Le cumulant d'ordre quatre d'un bruit gaussien est nul, et la relation (4.16) devient :

$$Cum_{X_2^2} = T_1^1 \otimes T_1^{1^T} \otimes T_1^1 \otimes T_1^{1^T} \bullet Cum_{S_2^2} \tag{4.17}$$

Si on note le terme général de S^1 par s^i , alors le terme $ijkl$ du tenseur cumulant [89] est :

$$\begin{aligned} Cum_{S_{2\,jl}^{2\,ik}} &= Es_i s_j s_k s_l - Es_i s_j Es_k s_l \\ &\quad - Es_i s_l Es_j s_k - Es_i s_k Es_j s_l, \end{aligned} \tag{4.18}$$

Avec les propriétés du bruit, la relation est identique à celle obtenue dans le cas non bruité. En tenant compte de l'indépendance des sources, on pourra montrer que :

$$\begin{aligned} Cum_{X_{2\,jl}^{2\,ik}} &= T_m^i T_j^{nT} T_o^k T_l^{pT} C_{S_{2\,mm}^{2\,mm}} \delta_{mo}^{np} \\ &= T_m^i T_m^j T_m^k T_m^l C_{S_{2\,mm}^{2\,mm}} \end{aligned} \tag{4.19}$$

Et on déduit que :

$$Cum_{X_{2\,ik}^{2\,ik}} = (T_m^i)^2 (T_m^k)^2 \beta_m \tag{4.20}$$

Pour avoir la relation (4.19), on a développé le calcul dans l'annexe B.4. Il faut finalement noter que la relation (4.20), en notations tensorielles, est identique à la relation (4.7).

4.3 Généralisation au mélange convolutif

Nous allons montrer dans cette section que la séparation est obtenue à une permutation P et une matrice polynômiale diagonale Δ près, en utilisant un critère basé sur les cumulants 22.
Si le mélange est convolutif alors la matrice H est une matrice de filtres (ses coefficients sont de filtres). Dans le domaine de fréquence z, cette matrice devient une matrice polynômiale en z. La relation entre les sources et les sources estimés est :

$$X(n) = \sum_p T(p) S(n-p), \tag{4.21}$$

que l'on notera :

$$X(z) = T(z) S(z). \tag{4.22}$$

Avec la notation tensorielle [89], on peut écrire :

$$X^1(n) = \sum_p T_1^1(p) \bullet S^1(n-p), \tag{4.23}$$

ou encore :

$$X^1(z) = T_1^1(z) \bullet S^1(z), \tag{4.24}$$

Par abus de la notation d'Einstein (où comme généralisation de cette notation), on écrit la relation (4.23) comme étant :

$$X^1(n) = T_1^1(p) \bullet S^1(n-p), \tag{4.25}$$

où la sommation sur les retards p est implicite. Si on considère les tenseurs de sorties à différents instants $X^1(n-p_i)$, en utilisant la relation (4.17), on obtient :

$$\begin{aligned} &Cum(X^1(n-p_1), X_1(n-p_2), X^1(n-p_3), X_1(n-p_4)) = \\ &\quad T_1^1(q_1) \otimes T_1^{1T}(q_2) \otimes T_1^1(q_3) \otimes T_1^{1T}(q_4) \\ &\quad \bullet Cum(S^1(n-p_1-q_1), S_1(n-p_2-q_2), S^1(n-p_3-q_3), S_1(n-p_4-q_4)). \end{aligned}$$

dont le terme général est :

$$\begin{aligned} Cum_{X_{ik}^{jl}}(p_1,p_2,p_3,p_4) &= t_i^a(q_1) t_b^{jT}(q_2) t_k^c(q_3) t_d^{lT}(q_4) \\ &\quad Cum(s_a(n-p_1-q_1), s_b(n-p_2-q_2), \\ &\qquad s_c(n-p_3-q_3), s_d(n-p_4-q_4)). \end{aligned} \tag{4.26}$$

Les $t_i^a(q_1)$ (respectivement $t_i^{aT}(q_1)$) correspondent aux coefficients $t_{ai}(q_1)$ (respectivement $t_{ia}(q_1)$) de la matrice $T(q_1)$.
L'indépendance des sources nous permet d'écrire que :

$$\begin{aligned} &Cum(s_a(n-p_1-q_1), s_b(n-p_2-q_2), s_c(n-p_3-q_3), s_d(n-p_4-q_4)) = \\ &\quad \delta_{ab}^{cd} Cum(s_a(n-p_1-q_1), s_a(n-p_2-q_2), s_a(n-p_3-q_3), s_a(n-p_4-q_4)), \end{aligned}$$

et de simplifier la relation (4.26), qui devient :

$$\begin{aligned} Cum_{X_{ik}^{jl}}(p_1,p_1,p_3,p_4) &= T_i^a(q_1) T_a^{jT}(q_2) T_k^a(q_3) T_a^{lT}(q_4) \\ &\quad Cum_{s_a}(n-q_1-p_1, n-q_2-p_2, n-q_3-p_3, n-q_4-p_4). \end{aligned} \tag{4.27}$$

Enfin, la stationarité des sources et l'utilisation de $T_i^a(q_1) = t_{ia}(q_1)$ et la stationarité des signaux, nous permet de réécrire la relation (4.27) :

$$\begin{aligned} Cum_{X_{ik}^{jl}}(p_1,p_2,p_3,p_4) &= t_{ia}(q_1) t_{ja}(q_2) t_{ka}(q_3) t_{la}(q_4) \\ &\quad Cum_{s_a}(p_1+q_1, p_2+q_2, p_3+q_3, p_4+q_4). \end{aligned} \tag{4.28}$$

Si de plus, on suppose que les signaux s_a sont des signaux temporellement iid, alors :

$$Cum_{s_a}(p_1 - q_1, p_2 - q_2, p_3 - q_3, p_4 - q_4) = \delta_{(q_1+p_1)(q_2+p_2)}^{(q_3+p_3)(q_4+p_4)} Cum_{s_a}(p, p, p, p), \quad (4.29)$$

où $p = p_1 + q_1$, pour alléger l'écriture, on note $Cum_{s_a}(p, p, p, p)$ par β_a.
Si les sources ont des kurtosis (β_a) de même signe, on peut déduire de la relation (4.28) que :

$$Cum_{X_{ij}^{ij}}(p_1, p_2, p_3, p_4) = 0 \Longleftrightarrow$$
$$t_{ia}(q_1)t_{ia}(q_1 + p_1 - p_2)t_{ja}(q_1 + p_1 - p_3)t_{ja}(q_1 + p_1 - p_4)\beta_a = 0. \quad (4.30)$$

En supposant que les filtres $t_{ia}(z)$ sont des filtres à réponse impulsionnelle finie, causaux d'ordre M, on a :

$$t_{ia}(p) = 0, \quad (4.31)$$

pour $p < 0$ et $p > M$. Donc le paramètre q_1 est limité par :

$$\max(0, p_2 - p_1, p_3 - p_1, p_4 - p_1) \leq q_1$$
et
$$q_1 \leq \min(M, M + p_2 - p_1, M + p_3 - p_1, M + p_4 - p_1).$$

Enfin, en choisissant $p_1 = p_2 = 0$ et $p_3 = p_4 = q_1 - q_2 \neq 0$, alors que q_1 et q_2 sont deux nombres entiers, l'équivalence (4.30) devient :

$$Cum_{X_{ik}^{jl}}(0, 0, q_1 - q_2, q_1 - q_2) = 0 \Longleftrightarrow t_{ia}^2(q_1)t_{ja}^2(q_2)\beta_a = 0. \quad (4.32)$$

Comme nous l'avons déjà dit, cette condition nécessite que les sources aient le même signe de kurtosis. La dernière relation montre que :

$$t_{ia}(q_1)t_{ja}(q_2) = 0, \quad (4.33)$$

$\forall a, q_1, q_2$, et $i \neq j$. Ce qui revient à écrire que :

$$q_1 = q_2 \Rightarrow T(q_1) = PD, \quad (4.34)$$

où P est une matrice de permutation, D est une matrice diagonale quelconque [84].
Ces résultats sont généralisables pour $q_1 \neq q_2$, et on a :

$$T(q_2) = PD'. \quad (4.35)$$

Il est clair que D' est en général différente de D, ce qui conduit à la matrice globale T donnée par :

$$T(z) = P \sum_i D(i) z^{-i}. \quad (4.36)$$

La séparation est donc obtenue à une permutation et un filtre près.

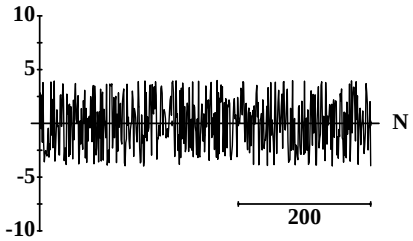
4.4 Conclusion et résultats expérimentaux

Dans ce chapitre, on a présenté un nouveau critère de séparation de sources basé sur le cumulant (2x2) de signaux estimés. Et on a montré que :

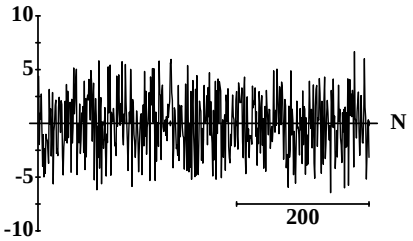
- L'annulation du Cum_{22} est suffisant pour faire la séparation dans le cas d'un mélange instantané. Ce résultat est valable même si le mélange est bruité par un bruit gaussien.
- Si les sources sont iid, ce critère peut être généralisé pour la séparation de sources dans des mélanges convolutifs. Dans ce cas le critère devient :

$$J(n) = \sum_{i,j,p} Cum^2(x_i(n), x_i(n), x_j(n-p), x_j(n-p)). \tag{4.37}$$

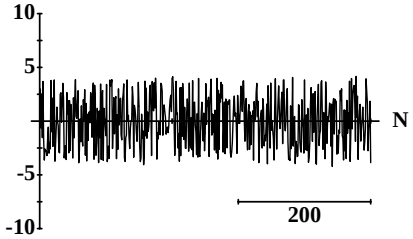
Pour les résultats expérimentaux, on a réussi à séparer un réseau de deux sources et deux capteurs, par notre algorithme, avec une diaphonie résiduelle de -21dB. On a testé l'influence du nombre d'échantillons utilisés dans l'estimation du cumulant, sur les performances, on a trouvé qu'à partir de 50 échantillons, les performances sont acceptables.
Les figures 4.4 (a), (b), (c) et (d) montrent les différents signaux et l'erreur d'estimation. Les deux dernières figures 4.4 (e) et (f) montrent la convergence et la diaphonie en fonction de nombre d'itérations (voir [84]).



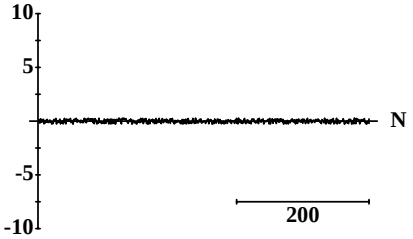
(a) La source $s_1(t)$



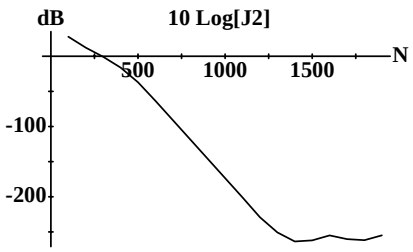
(b) Le premier signal mélange $y_1(t)$



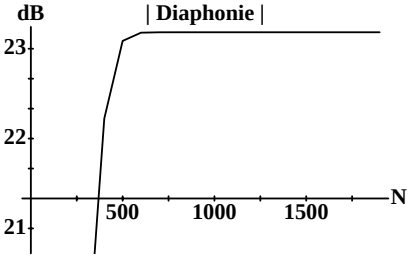
(c) La source estimée $x_1(t)$



(d) L'erreur d'estimation $s_1(t) - x_1(t)$



(e) Convergence du critère



(f) La diaphonie résiduelle en valeur absolue

Chapitre 5

Kurtosis

5.1 Introduction

L'importance du kurtosis (cumulant normalisé) apparaît clairement dans la manipulation de signaux non-gaussiens. Il intervient aussi explicitement dans plusieurs algorithmes de séparation, dont la validité impose certaines contraintes sur le kurtosis de sources ([92], [84], [76] et [96]). Généralement, ces conditions portent sur le signe des kurtosis ou le signe de la somme des kurtosis, et les auteurs considèrent que les hypothèses sont vérifiées pourvu que le signal soit simplement sur-gaussien (si le kurtosis est supposé positif) ou sous-gaussien (si le kurtosis est supposé négatif).
Dans le chapitre 4, nous avons nous aussi proposé un critère (Cum_{22}, [84]) qui suppose que les sources ont le même signe du kurtosis. Nous nous sommes penchés sur les propriétés des signaux satisfaisant cette contrainte, et ces résultats originaux, que nous avons obtenus et qui vont à l'encontre d'idées reçues, sont exposés dans ce chapitre.

5.2 Définition et propriétés des kurtosis

5.2.1 Définition

Pour un signal aléatoire centré $x(t)$ de densité de probabilité (ddp) $p(x)$, on définit le kurtosis $K[p(x)]$ comme étant son auto-cumulant d'ordre quatre normalisé par le carré de sa puissance :

$$K[p(x)] = \frac{Cum_4(x)}{(Ex^2)^2} = \frac{Ex^4 - 3(Ex^2)^2}{(Ex^2)^2} = \frac{Ex^4}{(Ex^2)^2} - 3. \tag{5.1}$$

Si $x(t)$ n'est pas centré, cette dernière relation devient plus compliquée [89] :

$$\begin{aligned} K[p(x)] &= \frac{Cum_4(x)}{(Ex^2)^2} \\ &= \frac{Ex^4 - ExEx^3 - 3(Ex^2)^2 + 12(Ex)^2Ex^2 - 6(Ex)^4}{(Ex^2)^2}. \end{aligned} \tag{5.2}$$

5.2.2 Propriétés et exemples

Dans la plupart des cas, les contraintes portent sur le signe du kurtosis mais pas sur sa valeur. Dans ce chapitre, nous nous intéressons, donc au signe du kurtosis noté $ks(x)$, qui est le même que celui de l'auto-cumulant d'ordre quatre. On peut déduire immédiatement quelques propriétés de $ks(x)$:

1. $ks(x)$ est invariant par une transformation linéaire. On peut remarquer d'après la relation (5.2) que :

$$Cum_4(ax + b) = a^4 Cum_4(x), \tag{5.3}$$

   donc $ks(ax+b) = ks(x)$. Dans la suite, on supposera sans aucune restriction que les signaux sont centrés et de variance unité.

2. Comme n'importe quelle fonction, la ddp $p(x)$ est décomposable en une somme de deux fonctions, l'une paire $p_p(x)$ et l'autre impaire $p_i(x)$:

$$p(x) = p_p(x) + p_i(x). \tag{5.4}$$

   Alors, on constate que :

   - $ks(x)$ dépend seulement de $p_p(x)$, parce que (5.1) ne dépend que des moments d'ordre pair de $x(t)$.
   - La partie paire $p_p(x)$ a les propriétés d'une ddp :

     (a) $p_p(x) \geq 0$

     (b) $\forall x$ et $\int_{\mathbb{R}} p(x)dx = \int_{\mathbb{R}} p_p(x)dx = 1$.

     Dans ce qui suit, on suppose donc que la ddp $p(x)$ en question est paire.

3. Le kurtosis d'un signal gaussien est nul.

4. Notons aussi que **le signe de kurtosis ne change pas si le signal est pollué par un bruit gaussien additif** : en effet, (voir chapitre 4), les cumulants d'un signal bruité, par un bruit additif indépendant du signal,

sont égaux aux cumulants du signal plus ceux du bruit (4.16). Donc un bruit additif gaussien ne change pas les cumulants d'ordre supérieur à deux du signal, puisque les cumulants d'ordre supérieur à deux de signaux gaussiens sont nuls.

Généralement de façon hâtive, un signal dont le signe du kurtosis est positif (resp. négatif) est dit sur-gaussien (resp. sous-gaussien) ([47] et [10]).
Par définition, une variable est sur-gaussienne si on a :

$$\exists A \in \mathbb{R} / \forall x \geq A, \quad p(x) > g(x), \tag{5.5}$$

où $g(x) = \frac{1}{\sqrt{2\Pi}} \exp(-\frac{1}{2}x^2)$ est la ddp d'une variable gaussienne centrée réduite. Dans le cas où l'inégalité (5.5) est inversé, la variable aléatoire x est dite sous-gaussienne. Il est évident que pour les ddp de la forme exponentielle :

$$p_\alpha(x) \simeq c_1 exp(-c_2|x|^\alpha), \tag{5.6}$$

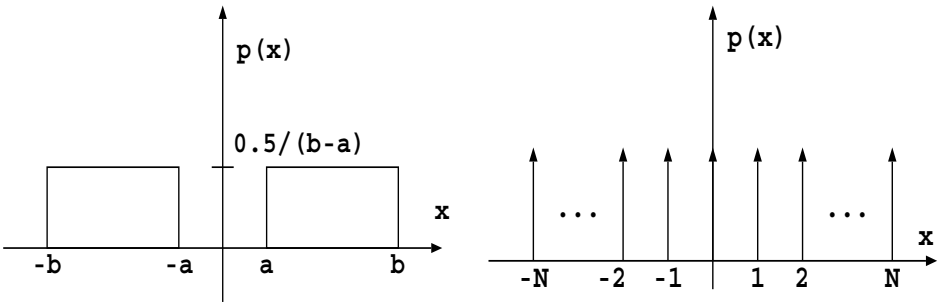
où c_i sont deux constantes de normalisation, la nature sur-gaussienne ou sous-gaussienne dépend directement de la valeur de α. On vérifie également que le signe du kurtosis dépend également de α :

- pour $\alpha = 2$, la variable est gaussienne et $k_s(x) = 0$.
- sinon, on a :

$$ks(x) = \text{signe}(2 - \alpha). \tag{5.7}$$

Ce résultat sera déduit directement du lemme de la section 5.3.1.
A titre d'exemple, on illustre dans le tableau suivant les signes de kurtosis $ks(x)$, pour certaines ddp bien connues :

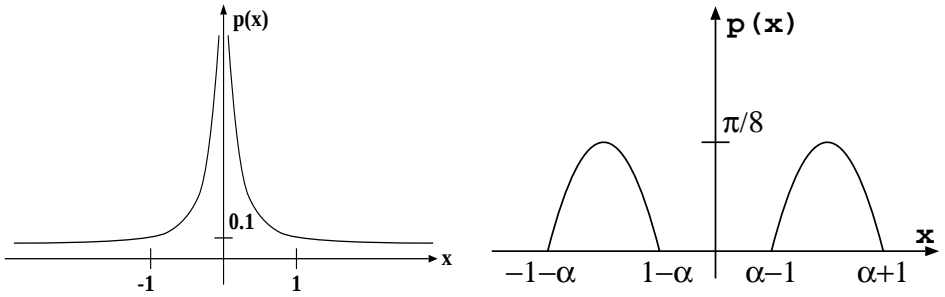
Signal	$Cum_4(x)$	$ks(x)$	ddp	figure
Uniforme	$-\frac{2a^4+2b^4 7a^3b+7b^3a+12a^2b^2}{15}$	-	$p(x) = \begin{cases} \frac{1}{2(b-a)} & a < \lvert x\rvert < b \\ 0 & \text{ailleurs} \end{cases}$	(5.1, a)
Peigne	$\frac{-N(N+1)(2N^2+2N+1)}{15}$	-	$p(x) = \begin{cases} \frac{1}{N} & \text{si } x \in \{-N, \ldots, N\} \\ 0 & \text{ailleurs} \end{cases}$	(5.1, b)
Gamma	$\frac{26}{3}$, si $\sigma = 1$	+	$p(x) = \frac{\sqrt[4]{3}}{2\sqrt{2\pi\sigma\lvert x\rvert}} \exp(\frac{-\sqrt{3}\lvert x\rvert}{2\sigma})$	(5.2, c)
Cosinus	$\frac{192}{\pi^4} - 2 - 2\alpha^4$	-	$p(x) = \begin{cases} \frac{\pi}{8}\cos(\frac{\pi}{2}(t-\alpha)) & \text{si } \lvert x \pm \alpha\rvert < 1 \\ 0 & \text{ailleurs} \end{cases}$	(5.2, d)



(a) Uniforme bimodale. (b) Peigne.

Figure 5.1: Différentes ddp.

A première vue, il semble que $ks(x)$ **sont positive pour une variable sur-gaussienne et négative pour une sous-gaussienne** (donc en particulier pour des variables aléatoires à support borné).



(c) Loi Gamma (d) Loi bimodale en Cosinus

Figure 5.2: Différentes ddp.

5.3 Résultats théoriques

Dans cette section, nous allons étudier les relations entre le signe de kurtosis et la nature (sur- ou sous-gaussienne) des signaux.

5.3.1 Lemme

Soient $p(x)$ une ddp continue centrée paire, de variance unité et $g(x)$ celle d'une variable gaussienne réduite centrée. Si $p(x)$ et $g(x)$ ont deux points d'intersection, alors $ks(x)$ est positif (respectivement négatif) si $p(x)$ est sur-gaussienne (respectivement sous-gaussienne).

Démonstration
Compte tenu de la parité de $p(x)$ et $g(x)$, on restreint l'étude à $\mathbb{R}^+$. Supposons que $p(x)$ et $g(x)$ ont un seul point d'intersection dans $\mathbb{R}^+$, $\rho > 0$. Le cumulant d'ordre quatre d'une variable gaussienne étant nul, on a :

$$\int_{\mathbb{R}} x^4 g(x)dx = 3(\int_{\mathbb{R}} x^2 g(x)dx)^2 = 3, \tag{5.8}$$

puisque $g(x)$ est supposée réduite. La variance de $p(x)$ étant égale à 1, la définition (5.1) avec (5.8) permettent s'écrire :

$$K[p(x)] = \int_{\mathbb{R}} x^4 (p(x) - g(x))dx. \tag{5.9}$$

La fonction $p(x)$ étant paire, on a :

$$\begin{aligned} K[p(x)] &= 2\int_0^{\infty} x^4 (p(x) - g(x))dx \\ &= 2\int_0^{\rho} x^4 (p(x) - g(x))dx + 2\int_{\rho}^{\infty} x^4 (p(x) - g(x))dx. \end{aligned} \tag{5.10}$$

Si on considère que $p(x)$ est sur-gaussienne ($p(x) > g(x)$, quand $x \to \infty$), alors le signe de $p(x) - g(x)$ est constant sur les deux l'intervalles [0,ρ] et [ρ, ∞]. D'après le théorème de la moyenne généralisé[1], on a :

$$\begin{aligned} K[p(x)] &= 2[\xi^4 \int_0^\rho (p(x) - g(x))dx + \lambda^4 \int_\rho^\infty (p(x) - g(x))dx] \\ &= 2[\lambda^4 \int_\rho^\infty (p(x) - g(x))dx - \xi^4 \int_0^\rho (g(x) - p(x))dx], \end{aligned} \tag{5.13}$$

avec :

$$0 < \xi < \rho < \lambda. \tag{5.14}$$

Il ne faut pas oublier que $p(x)$ et $g(x)$ sont deux ddp donc :

$$\int_0^\infty (p(x) - g(x))dx = \int_0^\rho (p(x) - g(x))dx + \int_\rho^\infty (p(x) - g(x))dx = 0. \tag{5.15}$$

D'après la relation (5.15), et le fait que $p(x)$ est sur-gaussienne, on déduit que :

$$\int_\rho^\infty (p(x) - g(x))dx = \int_0^\rho (g(x) - p(x))dx > 0. \tag{5.16}$$

Enfin de (5.16), (5.13) et (5.14), on peut remarquer que :

$$K[p(x)] = 2(\lambda^4 - \xi^4) \int_\rho^\infty (p(x) - g(x))dx > 0. \tag{5.17}$$

Si $p(x)$ est sur-gaussienne, (et s'il y a un seul point d'intersection positif entre $p(x)$ et $g(x)$), alors le kurtosis est positif.
De la même façon, on démontre que le signe du kurtosis d'une ddp sous-gaussienne est négatif.
Ce lemme confirme la proposition connue et adoptée dans la littérature. Une question se pose : est-ce que cette proposition tient toujours debout dans le cas général, où $p(x)$ et $g(x)$ ont plus d'un point d'intersection ?

---

[1] Si $f(x)$ et $g(x)$ sont **continues** sur $[a, b]$ et si $g(x)$ est une fonction *positive décroissante*, alors il existe un point $\xi \in [a, b]$ tel que :

$$\int_a^b f(x)g(x)dx = g(a) \int_a^\xi f(x)dx \tag{5.11}$$

Or si $g(x)$ est une fonction *positive croissante*, alors il existe un point $\xi \in [a, b]$ tel que :

$$\int_a^b f(x)g(x)dx = g(b) \int_\xi^b f(x)dx \tag{5.12}$$

5.3.2 Cas général

S'il y a plus de deux points d'intersections dans l'intervalle $\mathbb{R}^+$, entre $p(x)$ et $g(x)$, on peut montrer avec des exemples qu'il n'y a pas de règle générale : un signal sur-gaussien, selon la définition (5.5), peut avoir indifféremment un kurtosis positif ou négatif.
Considérons par exemple la ddp $p(x)$, somme de deux exponentielles :

$$p(x) = \frac{b}{4}(\exp(-b|x-a|) + \exp(-b|x+a|)) \tag{5.18}$$

On illustre $p(x)$ dans la figure 5.3 pour $x > 0$.
Dans les deux figures 5.4(a) et 5.4(b), on donne deux cas particulier pour $a = 5$,

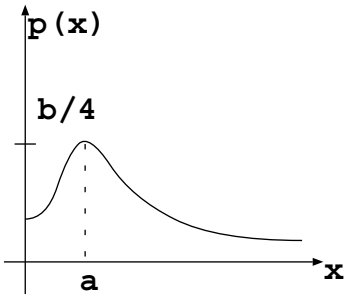
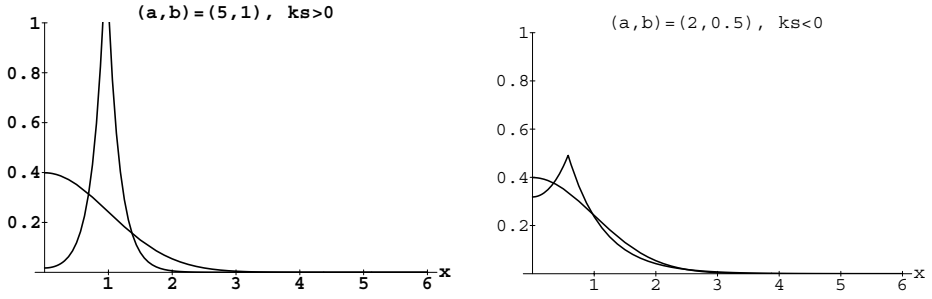


Figure 5.3: Exemple d'une densité, somme de deux fonctions exponentielles

$b = 1$ (figure 5.4(a)) où le kurtosis est positif et pour $a = 2$, $b = 0.5$ (figure 5.4(a)), le kurtosis devient négatif :



(a) (b)

Figure 5.4: L'intersection de $p(x)$ avec $g(x)$ pour deux valeurs de paramètres différentes.

Les moments d'ordre 4 et 2 de $p(x)$ sont faciles à calculer, et on trouve :

$$Ex^4 = a^4 + \frac{12}{b^2}a^2 + \frac{24}{b^4}, \tag{5.19}$$

$$Ex^2 = a^2 + \frac{2}{b^2} = \sigma^2 \tag{5.20}$$

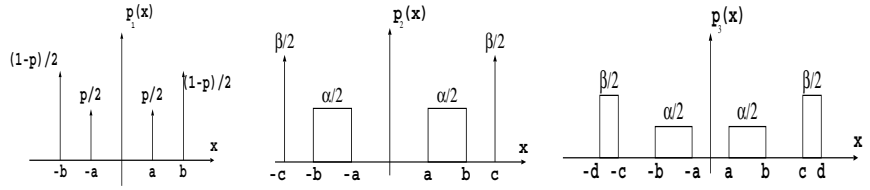
Alors, son kurtosis est :

$$K[p(x)] = \frac{12 - 2a^4b^4}{b^4\sigma^4}. \tag{5.21}$$

Suivant les valeurs de a et b, le kurtosis de ce signal change son signe. En effet, $K[p(x)] \geq 0$ si $0 < ab \leq \sqrt[4]{6}$. Suivant la définition (5.5), $p(x)$ est un signal sur-gaussien, mais son signe $ks(x)$ change suivant les valeurs du couple (a, b).

5.3.3 Ddp bornées

En pratique, on peut considérer que les signaux artificiels physiquement réalisables sont bornés. Et on peut objecter que le signal continu (5.18) est un exemple purement mathématique. Considérons donc ici le cas de quelques signaux bornés (figure 5.5).
Le cumulant d'ordre quatre d'un signal quaternaire (figure 5.5,(a)) est alors :



(a) (b) (c)

Figure 5.5: Trois exemples sur des ddp bornées

$$Cum_4(x) = a^4p(1-3p) - 6a^2b^2p(1-p) - b^4(1-p)(2-3p). \tag{5.22}$$

Il est évident que le signe de (5.22) change avec les valeurs des paramètres a, b et p. Par exemple, si $a = 0$ alors $Cum_4(x)$ a le même signe que $p - 2/3$.
Passons à des choses plus pratiques et plus raisonnables, on sait que la plupart des signaux dans le domaine de la télécommunication sont de types M-aire ou uniformes. Pour cela on a proposé deux ddp qui sont des mélanges de ddp Prenons les deux ddp dans les figures 5.5 (b et c). Le kurtosis du premier signal est :

$$\begin{aligned} K(p_1(x)) = {} & \frac{\alpha}{5}(b^5 - a^5) + (1 - \alpha(b-a))c^5 - \frac{\alpha^2}{3}(b^3 - a_3)^2 \\ & -3(1 - \alpha(b-a))_2c^6 - 2\alpha(1 - \alpha(b-a))c^3(b^3 - a^3). \end{aligned} \tag{5.23}$$

Cette équation est égale à zéro pour $a = 2, b = 9, c = 1$, et $\alpha = 0.06345$, et c'est clair que le signe (5.23) passe du négatif (pour $b = 0$, $\alpha > 0$ et $a \longrightarrow \infty$) au positif (pour $a = 0$, $\alpha > 0$ et $b \longrightarrow \infty$). Donc le kurtosis peut changer de signe. Le kurtosis du deuxième signal (figure 5.5. c) est :

$$\begin{aligned} K(p_2(x)) &= \frac{\alpha}{5}(b^5 - a^5) + \frac{\beta}{5}(d^5 - c^5) - \frac{\alpha^2}{3}(b^3 - a^3)^2 \\ &\quad -\frac{\beta^2}{3}(d^3 - c^3)^2 + |2\frac{\alpha\beta}{3}(b^3 - a^3)(d^3 - c^3), \end{aligned} \tag{5.24}$$

avec la condition :

$$\alpha(b - a) + \beta(d - c) = 1, \tag{5.25}$$

afin que $p(x)$ soit une ddp Pour bien illustrer le changement de signe de ce kurtosis, on trace la fonction :

$$K^\star(p(x)) = \frac{1}{2}[K(p(x)) + \mid K(p(x)) \mid]. \tag{5.26}$$

Dans ce cas, si $K(p(x)) > 0$, alors $K^\star(p(x)) = K(p(x))$. Sinon $K^\star(p(x)) = 0$. On voit bien le changement de signe du kurtosis illustré dans la figure 5.6.

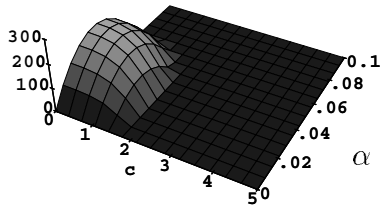
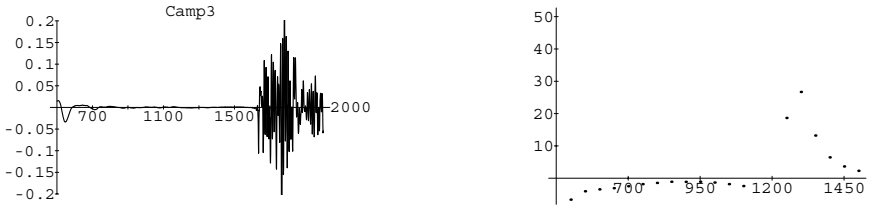


Figure 5.6: La représentation de $K^\star(p(x))$ en fonction de c et α

5.4 Résultats expérimentaux

Dans le cas des signaux réels, l'estimation du kurtosis est faite dans une fenêtre (glissante ou non). Si le signal est stationnaire, les estimations seront d'autant plus précises que la fenêtre d'estimations large. Dans le cas où le signal est non stationnaire, comme dans le cas des signaux de paroles, la largeur de la fenêtre doit être suffisamment petite, afin d'être adaptée à la durée de quasi-stationarité du signal.

Par exemple, un signal de parole est caractérisé par des périodes voisées et non voisées interrompus par des zones de silence. Si on considère le silence comme une partie du signal, la ddp présente un pic autour du zéro et le signal lui-même est mieux représenté par une ddp de type Gamma, voir tableau 5.2.2.
Expérimentalement, on remarque que le signe du kurtosis dépend de la position et de la largeur de fenêtre d'estimation. La figure 5.7 montre la variation du kurtosis en fonction de la position de la fenêtre. On a estimé le kurtosis avec des fenêtres de 500 échantillons décalés chaque fois de 50 échantillons. On remarque clairement le changement de signe du kurtosis (qui était négatif dans la zone de silence).



(a) Le signal (b) kurtosis estimé sur des fenêtres courtes

Figure 5.7: Résultats expérimentaux

Le signal lui-même est donnée par la figure (figure 5.8).

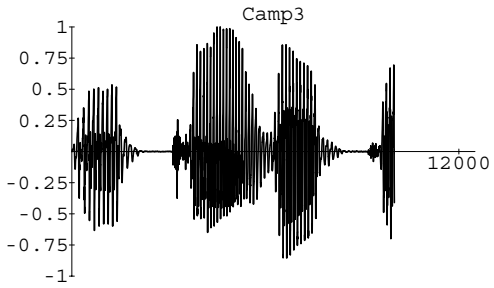


Figure 5.8: Camp3

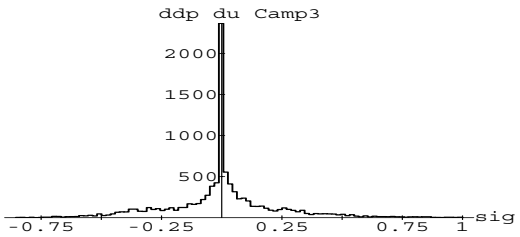


Figure 5.9: L'histogramme estimé avec 100 portions

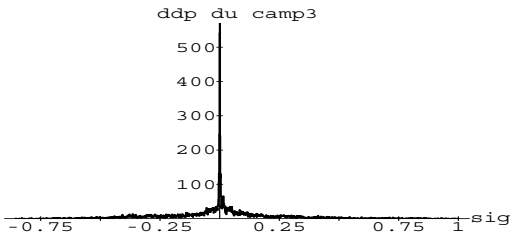


Figure 5.10: L'histogramme estimé avec 1500 portions

Son histogramme est représenté dans les figures 5.9 et 5.10, ce qui donne une estimation de sa ddp.

5.5 Conclusions

Les résultats de ce chapitre nous permettent de conclure que le signe du kurtosis est invariant par une transformation linéaire. En plus si la ddp a seulement deux points d'intersections avec une ddp gaussienne alors on peut déduire le signe du kurtosis à partir de la nature (sur- ou sous-gaussienne) du signal.
Dans le cas général, **ni la forme de la ddp et ni la nature de signal ne peuvent nous informer sur le signe du kurtosis**.
Finalement, pour des signaux non stationnaires, l'estimation du kurtosis (ou d'autres statistiques du signal) doit être effectuée sur des fenêtres étroites d'estimation. On observe alors des changements de signe du kurtosis, que l'on peut éliminer en excluant les périodes de silence (voir figures 5.11).

(a) Le signal Camp3, en éliminant le zone de silence (b) L'histogramme estimé avec 100 portions

Figure 5.11: Résultats expérimentaux

En estimant son kurtosis avec une fenêtre de 500 échantillons décalés chaque fois de 50 échantillons, on remarque que le signe de son kurtosis reste négatif (figure 5.12), ce qui n'était pas le cas à la figure 5.7 sur le signal complet.

Cette dernière remarque donne une explication théorique sur la nécessité de

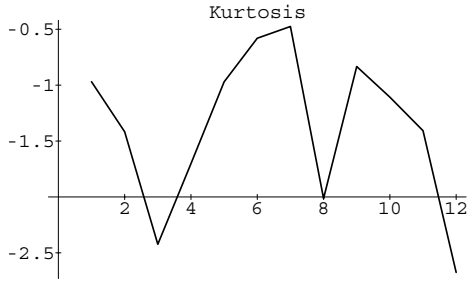


Figure 5.12: Kurtosis de Camp3 privé de ces zones de silence

l'adaptation utilisée dans [96] pour séparer des sources non-stationnaires.

Chapitre 6

Méthodes de type sous-espace

6.1 Introduction

La plupart des algorithmes de séparation mettent en oeuvre des statistiques d'ordres supérieur à deux (sauf en présence d'hypothèses supplémentaires sur les sources, voir Belouchrani [7], Pham [102]), et traitent le problème dans le cas d'un mélange linéaire instantané [64], [16], [30], [50], [75], [92], [84], [41], etc.
Le problème du mélange convolutif est très important pour l'application en télécommunications. Les premières solutions proposées ont été basées sur les statistiques d'ordre supérieur à deux, [95], [85] et [123]. Cependant, dans le cas de mélanges strictement causaux, les techniques d'ordre deux sont suffisantes [53].

L'utilisation des méthodes de type sous espace est récemment apparue, en particulier pour le problème d'identification aveugle [94], [1]. Dans ce chapitre, nous proposons d'étudier des solutions pour le problème de séparation de sources fondées sur cette approche de type de sous-espace, qui ne nécessite pratiquement que les statistiques d'ordre deux.
Ce chapitre commence par un exposé sur le modèle de mélange et sur les hypothèses de base. Les deux parties suivantes étudient, d'un point de vue théorique et algorithmique, des méthodes associées à deux paramétrisations différentes. La dernière partie fournit les résultats expérimentaux obtenus en utilisant la première paramétrisation.

6.2 Modèle et critère

6.2.1 Modèle

Soient $S(n)$ le vecteur source à N_s composantes et $Y(n)$ le vecteur des signaux mélangés à N_c composantes qui sont les sorties des capteurs. Notons $H(z)$ la matrice polynômiale de dimensions $N_c \times N_s$ qui représente le filtre de mélange :

$$Y(n) = [H(z)]S(n) \tag{6.1}$$

Nous avons vu dans l'introduction qu'une infinité de couple, $(H(z), S(n))$ peuvent produire la même observation $y(n)$, ce qui induit une indétermination à un filtre près sur la séparation de sources. Pour rendre sa représentation unique, on peut imposer une hypothèse supplémentaires.

- les signaux $S(n)$ sont indépendants et identiquement distribués (iid). Cette hypothèse est inacceptable dans la séparation aveugle de sources, puisqu'elle modifierait profondément les sources. Par contre, C'est une hypothèse usuelle dans le problème d'identification aveugle.
- le filtre $H(z)$ est un filtre à réponse impulsionnelle finie (RIF).

Dans la suite, on retiendra la seconde hypothèse. Le filtre $H(z)$ est un filtre RIF, causal et d'ordre M, c'est-à-dire que le maximum degré de polynômes $h_{ij}(z)$ est M, où $h_{ij}(z)$ est le coefficient d'indice ij de $H(z)$, $1 \leq i \leq N_c$, et $1 \leq j \leq N_s$.

6.2.2 Paramétrisation par matrice de Sylvester

Par concaténation des vecteurs de mélanges $Y(n), \ldots, Y(n-N)$ aux $(N+1)$ instants d'observation $\{n, \ldots, n-N\}$, on construit un grand vecteur $Y_N(n)$ de dimension $N_c(N+1)$. De même, par concaténation des vecteurs sources $S(n)$, $S\,(\,n-1\,)$,..., $S(n-M-N)$, on construit un grand vecteur $S_{(M+N)}(n)$ de dimension $N_s(N+M+1)$. La relation (6.1) peut être alors :

$$Y_N(n) = \begin{pmatrix} Y(n) \\ \ldots \\ Y(n-N) \end{pmatrix} \tag{6.2}$$

$$= T_N(H)S_{N+M}(n) = T_N(H) \begin{pmatrix} S(n) \\ \ldots \\ S(n-N-M) \end{pmatrix}, \tag{6.3}$$

où $T_N(H)$ est une matrice de Sylvester associée au filtre $H(z) = \sum_{i=0}^{M} H(i)z^{-i}$,de dimensions $N_c(N+1) \times (M+N+1)N_s$, et N est le nombre d'observations. Cette

matrice a la forme particulière suivante :

$$T_N(H) = \begin{bmatrix} H(0) & H(1) & H(2) & \dots & H(M) & 0 & 0 & \dots & 0 \\ 0 & H(0) & H(1) & \dots & H(M-1) & H(M) & 0 & \dots & 0 \\ \vdots & \vdots & & & & & & & \\ 0 & \dots & \dots & & 0 & H(0) & H(1) & \dots & H(M) \end{bmatrix},$$

et s'appelle matrice de Sylvester.
Pour séparer les sources, l'idée de base est de trouver une matrice G qui soit l'inverse à gauche de $T_N(H)$. Pour que l'inverse à gauche de $T_N(H)$ existe, cette matrice doit avoir plus de lignes que de colonnes, c'est-à-dire :

$$N_c(N+1) \geq N_s(M+N+1). \tag{6.4}$$

Pour vérifier cette dernière condition, on suppose que le nombre des capteurs est plus grand que celui des sources (voir hypothèse H4 dans la section suivante). Si $N_c > N_s$, on aura :

$$N > \frac{N_s M}{N_c - N_s} - 1 \tag{6.5}$$

Pour satisfaire (6.5), il suffit donc de prendre suffisamment d'instants d'observations : $N \geq N_s M$, pour assurer que le $T_N(H)$ soit rectangulaire dans la bonne dimension.

6.3 Hypothèses

Récapitulons et justifions les différentes hypothèses sur le filtre et les signaux :

- H1 : les signaux sources sont indépendants et non gaussiens.

  Comme nous l'avons vu, l'hypothèse d'indépendance est liée au problème de séparation de sources, en général. En revanche, l'hypothèse de signaux non-gaussiens est liée à l'utilisation des statistiques d'ordre supérieur, en particulier.
  Or, sous certaines conditions, la séparation peut être obtenue en utilisant les statistiques du second ordre. Dans ce cas, on pourra séparer même les signaux gaussiens (voir dans la suite).

- H2 : le filtre $H(z)$ est un filtre RIF.

  Cette hypothèse H2 est nécessaire pour définir le filtre d'une façon unique, et nous permet ensuite de séparer les sources à des coefficients et à une

permutation près, c'est-à-dire avec les mêmes indéterminations que pour des mélanges linéaires instantanées.

Si le filtre est un filtre à réponse impulsionnelle infinie (RII)[1], on peut montrer ([39], [38]) que la séparation ne peut être obtenue qu'à un filtre près.

- H3 : $H(z)$ est un filtre causal de dimension $N_c \times N_s$ et d'ordre M.

Les solutions développées dans la suite supposent en effet une connaissance exacte de l'ordre M. Cette connaissance pouvant être inexacte. Dans la première partie, on suppose que l'estimation de M est correcte [1], on s'interrogera sur la robustesse de la méthode dans la partie expérimentale.

- H4 : le nombre des capteurs est supposé strictement supérieur au nombre des sources[2] : $N_c > N_s$.

Cette hypothèse est fondamentale et très importante pour satisfaire l'hypothèse suivante.

- H5 : $H(z)$ est de rang plein (ou irréductible) et à colonnes réduites.
  - Si $H(z)$ est de rang plein, alors $\text{Rang}(H(z)) = N_s$:

$$H(z) \neq 0, \qquad \forall z. \tag{6.6}$$

    Cette condition implique que les composantes $h_{ij}(z)$ n'aient pas de zéro commun. En général, cette hypothèse n'est satisfaite que si $N_c > N_s$. En pratique, cette hypothèse est satisfaite lorsque les canaux de transmission entre les sources et les divers capteurs seront suffisamment différents.
    D'après cette hypothèse, $H(z)$ devient un filtre à phase minimale, admettant un filtre inverse à gauche causal $G(z)$.
  - Si $H(z)$ est à colonnes réduites, alors les colonnes de $H(z)$ forment une base polynômiale minimale de l'espace engendré par l'image de la matrice de $T_N(H)$, (voir [69]). Donc, $T_N(H)$ caractérise complètement

[1]Si $H(z)$ est un filtre RII, on peut le mettre sous la forme d'un produit de deux filtres : $H(z) = A(z)B(z)^{-1}$, où $A(z)$ et $B(z)$ sont deux filtres RIF. Le problème revient à traiter la partie correspondant à $A(z)$ et à l'effet de $B(z)$. Dans ce document, on suppose que le filtre est de type RIF.

[2]C'est une hypothèse courante dans toutes les méthodes de type sous-espace.

$H(z)$. Ces résultats s'énoncent avec le lemme suivant :
Lemme Soit $H'(z) = (H'_1(z), \ldots, H'_q(z))$ un $N_c \times N_s$ filtre. Supposons que $\deg(H'_i(z)) \geq M_i, \quad \forall i$, alors :

$$Image(T_N(H')) \subset Image(T_N(H)), \tag{6.7}$$

ssi $H'(z) = H(z)R(z)$, où $R(z)$ est une $n_s \times N_s$ matrice de rang plein, et dans ce cas on a une égalité dans (6.7). Si $\deg(H'(z)) = M_i, \quad \forall i$, alors R devient une matrice scalaire.
Pour la démonstration de ce lemme voir [1].

- H6 : les colonnes de $H(z)$ doivent avoir même degré, i.e. $M_i = M$.

Pour que $T_N(H)$ soit inversible, son nombre de lignes doit être égal à son rang :

$$Rang(T_N(H)) = (M + N + 1)N_s.$$

Or, on peut montrer [69] que le rang de la matrice de Sylvester est égale à $N_s(N+1) + \sum_{i=1}^{N_s} M_i$ (voir appendice C.1). Pour $T_N(H)$ soit inversible, il faut que $M_i = M$.
Nous verrons plus loin que cette hypothèse est liée à la *paramétrisation* du mélange et qu'elle peut être relâchée en adaptant une autre paramétrisation. Si les degrés des colonnes de $H(z)$ sont différents, tout en supposant que ces degrés sont parfaitement connus, on montre dans la suite que la séparation est faisable en utilisant seulement les statistiques du second ordre.

6.4 Cas d'un filtre dont les colonnes ont le même degré

Soit G l'inverse à gauche de $T_N(H)$. On peut écrire que : $G.T_N(H) = I_{(M+N+1)N_s}$, où $I_{(M+N+1)N_s}$ est la matrice identité de dimension $(M+N+1)N_s \times (M+N+1)N_s$. Il est facile de remarquer, d'après (6.3), que les $(M+N)N_s$ premières lignes de $G.Y_N(n)$ coïncident avec les $(M+N)N_s$ dernières lignes de $G.Y_N(n+1)$.
Pour estimer G, on peut proposer comme [54] de minimiser l'erreur quadratique, par rapport à G :

$$C_1(G) = E\|[I_{(M+N)N_s} \quad 0_{N_s}]GY_N(n) - [0_{N_s} \quad I_{(M+N)N_s}]GY_N(n+1)\|^2, \tag{6.8}$$

où E est l'espérance mathématique et 0_{N_s} est la matrice rectangulaire nulle de dimension $(M+N)N_s \times N_s$.

La solution triviale de la minimisation de ce critère est la matrice $G = 0$. Pour cette raison, la minimisation sans contrainte n'est pas suffisante pour chercher une inverse à gauche de $T_N(H)$. On discutera des contraintes possibles dans la section suivante.
Dans le cas général, on peut montrer que la minimisation de (6.8) est une matrice G_{min} qui vérifie $G_{min}T_N(H) = \mathcal{A}$, où $\mathcal{A}$ est une matrice diagonale par bloc de dimension $(M+N+1)N_s \times (M+N+1)N_s$. Donc on a :

$$G_{min}T_N(H) = \mathcal{A} = \begin{pmatrix} A & 0 & 0 & \dots \\ 0 & A & 0 & \dots \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & A \end{pmatrix} = \text{diag}(A, A, \dots, A), \qquad (6.9)$$

et A est une matrice quelconque de dimension $N_s \times N_s$ (voir l'appendice C.2).
Il est clair que la minimisation de $C(G)$ n'est pas suffisante pour achever la séparation de sources puisque les signaux de sorties $Z(n)$ sont les signaux sources à une matrice et à un coefficient scalaire près correspondant à un mélange instantané. En effet, en notant $Z_{(M+N)}(n) = (Z(n)^T, Z(n-1)^T, ..., Z(n-M-N)^T)$, on a :

$$\begin{aligned} Z_{(M+N)}(n) &= GY_N(n) \\ &= \mathcal{A}S_{(M+N)}(n). \end{aligned}$$

On peut donc déduire que :

$$Z(n) = AS(n). \qquad (6.10)$$

En particulier, on a :

$$A = G_{11}H(0),$$

où G_{11} est le premier élément-bloc de G, de dimension $N_s \times N_c$, et $H(0)$ est le premier coefficient du filtre $H(z)$. On peut donc conclure que **l'ordre deux est suffisant** pour transformer le mélange convolutif en un mélange instantané.
Finalement, pour pouvoir séparer le mélange instantané, en absence d'hypothèses sur les sources, on doit se servir des statistiques d'ordre supérieur. De plus, la matrice A doit être inversible, donc la minimisation doit être réalisée avec une contrainte qui assure l'inversibilité de la matrice A.

6.4.1 Minimisation adaptative ou bloc ?

Contrainte pour une minimisation adaptative

Deux idées simples ont été rejetées rapidement :

- la première consisterait à imposer le produit $A = G_{11}H(0)$ inversible.
  Cependant, en raison de la nature rectangulaire de G_{11} et de $H(0)$, il est impossible de trouver une contrainte générale sur G_{11} qui garantisse l'inversibilité de A.

- la seconde consisterait à minimiser le critère avec $G_{11} = H(0)^{-1}$.
  Or $H(0)$ est inconnue et nous n'avons pas trouvé de méthode pour l'estimer complètement à l'ordre deux (nous avons vu que c'était possible à l'ordre 4, au chapitre 4).

- la troisième idée est fondée sur le fait qu'après convergence de l'algorithme, on doit avoir :

  $$G_0 Y_N(n) = AS(n),$$

  où G_0 est la première ligne bloc de G de dimension $N_s \times N_c(N+1)$. De plus, on doit avoir :

  $$E[(G_0 Y_N(n))(G_0 Y_N(n))^T] = G_0 R_{Y_N}(0) G_0^T = E[AS(n)(AS(n))^T] = AR_s A^T.$$

  On sait que la matrice de covariance des sources R_S est diagonale et de rang plein à cause de l'indépendance des sources, et pour alléger les notation on notera la matrice de covariance du vecteur $Y_N(n)$ par R_{Y_N}. En imposant :

  $$G_0 R_{Y_N}(0) G_0^T = I_{N_s}, \tag{6.11}$$

  on peut donc assurer l'inversibilité de A.
  Il faut noter que l'inverse à gauche de $T_N(H)$ est définie à une matrice appartenant au noyau de $T_N(H)$ près, c'est-à-dire si on a

  $$\mathcal{X} \in \mathcal{N}(T_N(H)) \Rightarrow \mathcal{X} T_N(H) = 0.$$

  Si G est une inverse à gauche de $T_N(H)$ alors $G + \mathcal{X}$ l'est aussi. Dans tous les cas, on trouvera la même matrice A,

  $$(G + \mathcal{X}) T_N(H) = \mathrm{diag}(A, A, \ldots, A). \tag{6.12}$$

Minimisation par Bloc

La minimisation peut être obtenue par une méthode de Lagrange généralisée. Cependant, cette approche est une approche par bloc, alors que nous avons choisi de nous restreindre à des méthodes adaptatives.

Cette méthode n'a donc pas été poursuivie au-delà d'un développement du critère sous forme de relations matricielles. La démonstration est basée sur une généralisation d'une minimisation de Lagrange sous contrainte (voir appendice C.4).

6.4.2 Minimisation adaptative : algorithme LMS

Le problème revient à minimiser la fonction (6.8) : $C_1(G) = E\|[I_{(M+N)N_s}\ 0_{N_s}]GY_N(n) - [0_{N_s}\ I_{(M+N)N_s}]GY_N(n+1)\|^2$, sous la contrainte (6.11) : $G_0 R_{Y_N}(0) G_0^T = I_{N_s}$.

Dans la littérature, on trouve des méthodes pour résoudre des problèmes analogues, par exemple :

$$\text{Minimiser } h^T S h \text{ avec la Contrainte } \quad \|h\| = 1, \tag{6.13}$$

$$\text{Minimiser } h^T S h \text{ avec la Contrainte } \quad h^T R h = 1, \tag{6.14}$$

où h est un vecteur, et R et S sont des matrices. Le passage du système (6.14) vers celui (6.13) est presque immédiat : il suffit de décomposer la matrice $R = R^{\frac{1}{2}} R^{\frac{1}{2}}$ et de poser $g^T = h^T R^{\frac{1}{2}}$.
Les deux principales différences, entre les deux approches (6.13) et (6.14), et notre problème, sont les suivantes :

- Dans notre cas, G est une matrice de dimension $(M+N+1)N_s \times N_c(N+1)$, et non pas un simple vecteur;
- La contrainte (6.11) concerne seulement la première ligne bloc de la matrice G, et non pas la matrice toute entière.

Donc, on ne peut pas appliquer directement les approches utilisées dans la littérature. En utilisant des stratégies semblables[3] à celles utilisées par les deux approches précédentes, on a essayé de résoudre notre problème.

Pour minimiser la fonction coût sans contrainte, on a choisi un simple algorithme de type LMS dont les détails de calcul sont développés dans l'appendice C.3.

Pour satisfaire la contrainte, on décompose à chaque itération, avec une décomposition de Cholesky [55], la matrice $G_0 R_{Y_N}(0) G_0^T = K K^T$, où la matrice K est une matrice symétrique définie positive. A chaque itération, on normalise la matrice G_0 par l'inverse de K.

$$G_0 \leftarrow K^{-1} G_0. \tag{6.15}$$

Les performances obtenues par cette méthode seront bonnes si les signaux sont stationnaires, mais elles seront très médiocres si les sources sont des signaux de paroles (non-stationnaires), voir paragraphe 6.6. Cette raison nous a amené à améliorer la vitesse de convergence par un algorithme de gradient conjugué.

[3] Normalement les auteurs cherchent à minimiser le critère du 6.13 ou celui 6.14 en faisant une projection sur la contrainte à chaque étape de l'algorithme

6.4.3 Algorithme du Gradient conjugué

Dans cette partie, on utilise l'algorithme du gradient conjugué généralisé, qui diffère de l'algorithme du Gradient Conjugué "classique" (Annexe C.5).
En effet, l'algorithme du Gradient conjugué généralisé (voir [24], [48] et [122]) cherche à minimiser le rapport généralisé de Rayleigh (voir [55]) :

$$f(X) = \frac{X^h A X}{X^h B X}, \tag{6.16}$$

où X est un vecteur complexe, et X^h est son vecteur transposé conjugué. A et B sont deux matrices symétriques définies positives (le développement de l'algorithme gradient conjugué généralisé est mis en annexe C.6).

Pour minimiser une fonction de coût, sous une contrainte donnée, par un algorithme de Gradient Conjugué, les auteurs ([24], [48], etc) cherchent à écrire la fonction de coût sous la forme de la nominateur de l'équation (6.16) et la contrainte sous la forme de déneminateur de (6.16). Finalement la contrainte sera satisfait par une projection à chaque itération (voir annexe C.6).

Pour appliquer l'algorithme, il faut d'abord réécrire la fonction coût (6.8) sous la forme (6.16).
Notons $\nu = \mathrm{Col}(G)$, de dimension $(M+N+1)N_sN_c(N+1)$ le vecteur obtenu par concaténation des colonnes de G. En utilisant le produit de Kronecker noté $\bigotimes$, la relation (6.8) devient :

$$\begin{aligned} C_1(\nu) &= \nu^T \{ R_{Y_N}(0) \bigotimes \begin{pmatrix} I_{N_s} & 0 & 0 \\ 0 & 2I_{(M+N-1)N_s} & 0 \\ 0 & 0 & I_{N_s} \end{pmatrix} \\ &\quad - R_{Y_N}(1) \bigotimes \begin{pmatrix} 0 & I_{(M+N)N_s} \\ 0 & 0 \end{pmatrix} \\ &\quad - R^T_{Y_N}(1) \bigotimes \begin{pmatrix} 0 & 0 \\ I_{(M+N)N_s} & 0 \end{pmatrix} \} \nu, \end{aligned} \tag{6.17}$$

où $R_{Y_N}(0) = \mathrm{E}[Y_N(n)Y_N(n)^T]$ et $R_{Y_N}(1) = \mathrm{E}[Y_N(n)Y_N(n+1)^T]$. Les termes 0 de l'équation (6.17) correspondent à des matrices rectangulaires (ou carrés) nulles, dont les dimensions sont déduites directement à partir de la position de cette matrice dans l'équation. Le détail de la démonstration est donné dans l'annexe C.7, est utilisé l'annexe C.8 (produit de Kronecker et ses propriétés)
Le critère de départ (6.8) a été transféré dans l'équation (6.17) sous une forme quadratique de ν.

Pour appliquer l'algorithme du gradient conjugué correspondant à la minimisation de (6.16), il nous reste à voir s'il est possible d'écrire la contrainte (6.11) sous la forme :

$$\nu^T B \nu = 1. \tag{6.18}$$

La contrainte (6.11) peut être écrite sous différentes formes :

$$E G_0 Y_N(n) Y_N^T(n) G_0^T = G_0 R_{Y_N}(0) G_0^T = I_{N_s} \tag{6.19}$$

$$(I_{N_s} \;\; 0) G R_{Y_N}(0) G^T \begin{pmatrix} I_{N_s}, \\ 0 \end{pmatrix} = I_{N_s}. \tag{6.20}$$

Une autre forme de la contrainte est obtenue à partir de :

$$\begin{aligned} G_0 Y_N(n) &= [I_{N_s} \;\; 0] G Y_N(n) \\ &= \{Y_N^T \bigotimes [I_{N_s} \;\; 0]\} \nu, \end{aligned} \tag{6.21}$$

ce qui permet d'écrire la contrainte :

$$I_{N_s} = E\{Y_N^T(n) \bigotimes [I_{N_s} \;\; 0]\} \nu \nu^T \{Y_N(n) \bigotimes \begin{pmatrix} I_{N_s} \\ 0 \end{pmatrix}\}. \tag{6.22}$$

ou bien de façon équivalente :

$$\| G_0 R_{Y_N}(0) G_0^T - I_{N_s} \|^2 = 0. \tag{6.23}$$

Les trois formes de contraintes ((6.19), ..., (6.23)) ne correspondant pas à la forme souhaitable (6.16) : $\nu^T B \nu$, on abordera le problème autrement.

Soit $K1$ la racine carrée de $R_{Y_N}(0)$ (obtenue par une décomposition de Cholesky de $R_{Y_N}(0)$) :

$$R_{Y_N}(0) = K1 K1^T. \tag{6.24}$$

En introduisant une matrice C, on peut écrire la contrainte (6.11) sous la forme

$$C.C^T = I_{N_s} \tag{6.25}$$

En effet, à partir de (6.24), on a :

$$G_0 K1 = [I_{N_s} \;\; 0_{N_s}] G K1, \tag{6.26}$$

et en notant col(C), l'operateur colonne sr la matrice C, on peut vérifier que :

$$\mathrm{Col}(C) = \{K1^T \bigotimes [I_{N_s} \;\; 0]\} \nu. \tag{6.27}$$

Cependant, la relation (6.25) correspond à N_s^2 contraintes simultanéesque l'on ne peut pas mettre sous la forme de (6.18). En remarquant que la matrice C est une matrice orthogonale, on montre que :

$$\|C\|^2 = N_s. \tag{6.28}$$

on peut déduire une contrainté nécessaire (de la forme (6.18)) mais non suffisante :

$$\mathrm{Col}(\frac{1}{\sqrt{N_s}}C)^T \ \mathrm{Col}(\frac{1}{\sqrt{N_s}}C) = 1. \tag{6.29}$$

Le facteur N_s ne change rien à la solution puisqu'on minimise une fonction coût (le produit de la fonction coût par un nombre constant ne modifie pas la valeur atteinte au minimum de cette fonction coût).

On peut donc appliquer l'algorithme du gradient conjugué avec une condition supplémentaire, la normalisation à chaque itération de la matrice G_0, de façon similaire à ce qui a été proposé dans la section 6.4.2.

En pratique, On a remarqué sur des exemples simplifiés que l'algorithme du gradient conjugué est plus efficace que celui du gradient simple. Malheureusement, il met en jeu des matrices de grandes dimensions : Par exemple, la matrice obtenue dans (6.17) par le produit de Kroneker est de très grande dimension : pour deux sources, un filtre d'ordre un, on arrive à une matrice carrée de 240×240.
Ce point qui nous a poussé à étudier et à proposer un autre critère.

6.4.4 Un second critère

Principe et algorithme

A cause de sa faible vitesse de convergence[4], l'algorithme LMS précédent est peu efficace pour les signaux fortement non stationnaires (comme les signaux de paroles par exemple). Ces raisons nous ont poussés à chercher un autre critère C2.

Partons toujours de l'idée que l'on cherche une matrice G telle que $T_N(H) = \mathrm{diag}\{A, A, \ldots, A\}$. On peut écrire :

$$GY_N(n) = \begin{pmatrix} AS(n) \\ AS(n-1) \\ \vdots \\ AS(n-M-N) \end{pmatrix}, \tag{6.30}$$

[4]On a remarqué qu'en pratique le critère C1 converge très lentement, voir la section de performance 6.6.

et :

$$GY_N(n+1) = \begin{pmatrix} AS(n+1) \\ AS(n) \\ \vdots \\ AS(n-M-N+1) \end{pmatrix}. \tag{6.31}$$

Notons par G_i $(0 \leq i \leq M+N)$ la ième ligne bloc de dimension $N_s \times N_c(N+1)$ de la matrice G. En utilisant les relations (6.30 et 6.31), on a alors :

$$G_i Y_N(n) = AS(n-i), \tag{6.32}$$

et :

$$G_i Y_N(n+1) = AS(n+1-i). \tag{6.33}$$

On en déduit que :

$$G_i Y_N(n) = G_{(i+1)} Y_N(n+1), \tag{6.34}$$

avec $(0 \leq i \leq M+N)$. En notant par $\mathcal{G}$ la matrice de dimension $N_s \times N_c(N+1)(M+N+1)$ telle que : $\mathcal{G} = (G_0, G_1, \ldots, G_{(M+N)})$ et par $\mathcal{Y}$ la matrice de dimension $N_c(N+1)(M+N+1) \times (N+M)$:

$$\mathcal{Y} = \begin{pmatrix} Y_N(n) & 0 & \ldots & 0 \\ -Y_N(n+1) & Y_N(n) & 0 & \ldots \\ 0 & -Y_N(n+1) & \ldots & \vdots \\ \vdots & \ldots & \ldots & \vdots \\ 0 & \ldots & -Y_N(n+1) & Y_N(n) \\ 0 & \ldots & 0 & -Y_N(n+1) \end{pmatrix}, \tag{6.35}$$

on trouve :

$$\begin{aligned} \mathcal{G}\mathcal{Y} &= (G_0, G_1, \ldots, G_{(M+N)}) \begin{pmatrix} Y_N(n) & 0 & \ldots & 0 \\ -Y_N(n+1) & Y_N(n) & 0 & \ldots \\ 0 & -Y_N(n+1) & \ldots & \vdots \\ \vdots & \ldots & \ldots & \vdots \\ 0 & \ldots & -Y_N(n+1) & Y_N(n) \\ 0 & \ldots & 0 & -Y_N(n+1) \end{pmatrix} \\ &= 0. \end{aligned}$$

On peut trouver les coefficients de la matrice $\mathcal{Y}$ directement à partir des signaux des capteurs. En effet, en notant $y_i(j)$ le signal sur l'ième capteur à l'instant j,

$0 \leq i \leq N_c(N+1)-1,\ 0 \leq j < N+M$, on peut montrer facilement à partir des définitions de $Y_N(n)$ et $\mathcal{Y}$ que :

$$\begin{cases} \mathcal{Y}_{(i+jN_c(N+1))j} & = & y_{i[N_c]}(n - \ \mathrm{ent}(\frac{i}{N_c})) \\ & \text{et} & \\ \mathcal{Y}_{(i+(j+1)N_c(N+1))j} & = & -y_{i[N_c]}(n+1 - \ \mathrm{ent}(\frac{i}{N_c})) \end{cases}$$

où $\mathrm{ent}(\frac{i}{N_c})$ est le quotient entier de i par N_c, et $i[N_c]$ est i modulo N_c. Notre critère devient La minimisation de $\mathcal{G}\mathcal{Y}\mathcal{Y}^T\mathcal{G}^T$ sous la même contrainte que précédemment : $G_0 R_{Y_N} G_0^T = I$.

Cas de deux sources

Pour fixer les idées, on présente le développement de ce critère pour deux sources, la généralisation est évidente. En notant G_{0i} la ième ligne de G_0,
$1 \leq i \leq N_s = 2$, la contrainte devient :

$$\begin{pmatrix} G_{01} \\ G_{02} \end{pmatrix} R_{Y_N} (G_{01}^T G_{02}^T) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.36}$$

Notons par $\mathcal{G}_i$, $1 \leq i \leq 2$, la ième ligne de $\mathcal{G}$, le critère sera donc appliqué en N_s étapes $(N_s = 2)$, donc ici succesivement sur $\mathcal{G}_1$ et $\mathcal{G}_2$:

- 1ère étape

$$\text{Minimiser} \quad \mathcal{G}_1 \sum_n \mathcal{Y}(n)\mathcal{Y}^T(n)\mathcal{G}_1^T. \tag{6.37}$$

  Le fait de minimiser la somme sur n de $\mathcal{Y}(n)\mathcal{Y}^T(n)$ donne une plus grande robustesse à la minimisation. Le critère doit être minimisé sous la contrainte

$$G_{01} R_{Y_N} G_{01}^T = 1. \tag{6.38}$$

  En remplaçant G_{01} par sa valeur en fonction de $\mathcal{G}_1$, on trouve facilement que :

$$\mathcal{G}_1 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_1^T = 1. \tag{6.39}$$

- 2ème étape :

$$\text{Minimiser } \mathcal{G}_2 \sum_n \mathcal{Y}(n)\mathcal{Y}^T(n)\mathcal{G}_2^T, \tag{6.40}$$

  avec :

$$G_{02} R_{Y_N} G_{02}^T = 1. \tag{6.41}$$

On peut écrire cette dernière équation sous la forme :

$$\mathcal{G}_2 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_2^T = 1. \tag{6.42}$$

Il faut remarquer que pour satisfaire exactement la contrainte (6.36), on doit ajouter à la deuxième étape la contrainte supplémentaire :

$$G_{01} R_{Y_N} G_{02}^T = 0, \tag{6.43}$$

que l'on peut encore écrire :

$$\mathcal{G}_1 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_2^T = 0. \tag{6.44}$$

Finalement, en associant une des deux contraintes à la fonction à minimiser, on peut réécrire la 2ème étape sous la forme suivante :

$$\text{Minimiser } \mathcal{G}_2 \sum_n \mathcal{Y}(n)\mathcal{Y}^T(n)\mathcal{G}_2^T + (\mathcal{G}_1 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_2^T)^2, \tag{6.45}$$

sous la contrainte de normalisation :

$$\mathcal{G}_2 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_2^T = 1. \tag{6.46}$$

Le deuxième terme dans (6.45) est le carré d'un scalaire. On peut donc le remplacer par son produit avec son transposé, ce qui donne finalement le critère :

$$\text{Minimiser } \mathcal{G}_2[\sum \mathcal{Y}\mathcal{Y}^T + \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_1^T \mathcal{G}_1 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix}]\mathcal{G}_2^T \tag{6.47}$$

sous la contrainte :

$$\mathcal{G}_2 \begin{pmatrix} R_{Y_N} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{G}_2^T = 1. \tag{6.48}$$

L'idée, d'extraire un signal et un seul à chaque étape, est une idée de déflation [41].

Validité de ce critère : cas de deux sources

En observant ce critère, on remarque que sa première étape est toujours faisable : en effet, c'est la minimisation d'un critère quadratique avec une certaine normalisation. La deuxième étape ($N_s = 2$) est plus complexe (le terme à minimiser est

la somme de deux termes quadratiques, et l'on peut se demander s'il y a toujours une solution.

Par exemple, supposons que l'on ait calculé $\mathcal{G}_1$ tel que $\mathcal{G}_1\mathcal{Y}(n) = 0$ et qui vérifie $\mathcal{G}_{10}R_{Y_N}\mathcal{G}_{10}^T = 1$ (où $\mathcal{G}_{i0}$ est la ième ligne de la matrice G_0).
Il reste à trouver $\mathcal{G}_2$ qui vérifie :

$$\mathcal{G}_2\mathcal{Y}(n) = 0, \tag{6.49}$$

$$\text{et } < \mathcal{G}_{20}R_{Y_N}^{1/2}, \mathcal{G}_{10}R_{Y_N}^{1/2} >= 0, \tag{6.50}$$

où $< \mathcal{G}_{20}R_{Y_N}^{1/2}, \mathcal{G}_{10}R_{Y_N}^{1/2} >$ est le produit scalaire de ces deux vecteurs, et $R_{Y_N}^{1/2}$ est la racine carrée de la matrice $R_{Y_N}(0)$.
Soit u la différence entre le vecteur $\mathcal{G}_{20}R_{Y_N}^{1/2}$ et sa projection sur $\mathcal{G}_{10}R_{Y_N}^{1/2}$. Alors u est perpendiculaire à $\mathcal{G}_{10}R_{Y_N}^{1/2}$. En notant par A/B la projection du vecteur A sur B, on trouve :

$$\begin{aligned} u &= \mathcal{G}_{20}R_{Y_N}^{1/2} - (\mathcal{G}_{20}R_{Y_N}^{1/2})/(\mathcal{G}_{10}R_{Y_N}^{1/2}) && (6.51)\\ &= \mathcal{G}_{20}R_{Y_N}^{1/2} - \frac{< \mathcal{G}_{20}R_{Y_N}^{1/2}, \mathcal{G}_{10}R_{Y_N}^{1/2} >}{\mathcal{G}_{10}R_{Y_N}\mathcal{G}_{10}^T}\mathcal{G}_{10}R_{Y_N}^{1/2} && (6.52)\\ &= (\mathcal{G}_{20} - < \mathcal{G}_{20}R_{Y_N}^{1/2}, \mathcal{G}_{10}R_{Y_N}^{1/2} > \mathcal{G}_{10})R_{Y_N}^{1/2} && (6.53) \end{aligned}$$

En posant $\alpha =< \mathcal{G}_{20}R_{Y_N}^{1/2}, \mathcal{G}_{10}R_{Y_N}^{1/2} >$ et $\delta G_2 = \mathcal{G}_2 - \alpha\mathcal{G}_1$, il est facile de vérifier que :

- $\delta G_2\mathcal{Y}(n) = 0$ puisque $\mathcal{G}_2\mathcal{Y}(n) = 0$ par hypothèse et $\mathcal{G}_1\mathcal{Y}(n) = 0$ résultat de la première étape.
- $< \delta G_{20}R_{Y_N}^{1/2}, \mathcal{G}_{10}R_{Y_N}^{1/2} >= 0$ par la construction du vecteur.

Enfin, il reste à satisfaire la condition de normalisation (6.41). Pour cela, on normalise à chaque étape le vecteur :

$$\delta G_2 \leftarrow \frac{\delta G_2}{\sqrt{\delta G_{20}R_{Y_N}\delta G_{20}^T}}, \tag{6.54}$$

où δG_{20} est la pémiere partie du vecteur δG_2 correspondante à 2ème ligne de G.

6.5 Cas où les degrés des colonnes sont différents

Dans cette section, on discute le cas où les colonnes de $H(z)$ sont de degrés différents. La suite du calcul restera dans le domaine temporel. Mais les résultats et les

démonstrations développés dans cette section peuvent être aussi présentés avec un autre formalisme fréquentiel. Le développement dans le domaine fréquentiel n'apportant pas de résultats différents, sera développé dans l'annexe C.9.

Cas de deux sources

Pour simplifier les notations, on supposera au début que $N_s = 2$ et l'ordre de deux colonnes de $H(z)$ satisfait $M_2 > M_1$. Cette hypothèse n'altère en rien la généralité de la méthode. Le résultat pour $N_s > 2$ peut être déduit facilement.

Le rang de $T_N(H)$ n'est pas plein puisque les degrés de $H(z)$ sont différents. On propose alors de construire une matrice de taille $N_c(N+1) \times [\sum_{i=1}^{N_s} M_i + N_s(N+1)]$, notée $U_N(H)$, dont laquelle $T_N(H_1)$ et $T_N(H_2)$ sont les matrices de Sylvester appliquées sur H_1 et sur H_2 respectivement :

$$U_N(H) = (T_N(H_1), T_N(H_2)) \tag{6.55}$$

Le vecteur d'observation est alors :

$$Y_N(n) = U_N(H)(S^T_{1,M_1+N}(n), S^T_{2,M_2+N}(n))^T. \tag{6.56}$$

où $S_{i,M_i+N}(n) = (s_i(n), s_i(n-1), \ldots, s_i(n-M_i-N))$ et $s_i(n-j)$ est la ième source à l'instant $(n-j)$.
Notons par G une inverse à gauche de $U_N(H)$. G est une matrice de dimension $\sum_{i=1}^{2} M_i + N_s(N+1) \times N_c(N+1)$ que l'on écrit sous la forme :

$$G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}, \tag{6.57}$$

où G_i est de dimension $M_i + N_s(N+1) \times N_c(N+1)$. On doit avoir :

$$GY_N(n) = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} Y_N(n) \tag{6.58}$$

$$= \begin{pmatrix} S_{1,M_1+N}(n) \\ S_{2,M_2+N}(n) \end{pmatrix}. \tag{6.59}$$

On remarque que les premières $(M_i + N)$ lignes de $G_i Y_N(n)$ sont identiques aux $(M_i + N)$ dernières lignes de $G_i Y_N(n+1)$. On peut écrire :

$$(I_{M_i+N}\ 0) G_i Y_N(n) = (0\ I_{M_i+N}) G_i Y_N(n+1), \tag{6.60}$$

avec $i = 1, 2$. Posons $GU_N(H) = B$:

$$\begin{aligned} GU_N(H) &= \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} (T_N(H_1), T_N(H_2)) \\ &= \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \end{aligned} \tag{6.61}$$

où les B_{ij} sont des matrices de $(M_i + N + 1) \times (M_j + N + 1)$ ($i = 1, 2$ et $j = 1, 2$). On peut récrire (6.60) en tenant compte de la relation (6.61) :

$$(I_{M_i+N}\ 0)(B_{i1}\ B_{i2}) \begin{pmatrix} S_{1,M_1+N}(n) \\ S_{2,M_2+N}(n) \end{pmatrix} = (0\ I_{M_i+N})(B_{i1}\ B_{i2}) \begin{pmatrix} S_{1,M_1+N}(n+1) \\ S_{2,M_2+N}(n+1) \end{pmatrix}.$$

Donc on résume ces dernières relations par :

$$\sum_{j=1}^{2} ((I_{(M_i+N)}\ \ 0)(0\ \ B_{ij}) - (0\ \ I_{(M_i+N)})(B_{ij}\ \ 0))S_{j,M_j+N+1}(n) = 0 \tag{6.62}$$

Si les $S_{j,M_j+N+1}(n)$ sont suffisamment excités (une hypothèse classique en automatique qui correspond à des signaux de bande suffisamment large) et indépendants (l'hypothèse habituelle dans le problème de séparation de sources), alors :

$$(I_{(M_i+N)}\ \ 0)(0\ \ B_{ij}) = (0\ \ I_{(M_i+N)})(B_{ij}\ \ 0) \tag{6.63}$$

Les B_{ij}, solutions de cette équation, sont des matrices très particulières et vérifient le lemme suivant :

Lemma 6.5.1 *Soit A une matrice $p \times q$ qui vérifie la condition suivante :*

$$(I_{(p-1)}\ \ 0)(0\ \ A) = (0\ \ I_{(p-1)})(A\ \ 0), \tag{6.64}$$

- *si $p = q$ alors $A = aI_p$, où a est un scalaire quelconque.*
- *si $p > q$ alors $A = 0$.*
- *si $q > p$ alors*

$$A = \begin{pmatrix} a_0 & a_1 & \dots & a_{q-p} & 0 & \dots & 0 \\ 0 & a_0 & \dots & \dots & a_{q-p} & 0 & \dots \\ \vdots & \dots & \vdots & \dots & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & a_0 & \dots & a_{q-p} \end{pmatrix}. \tag{6.65}$$

Ce lemme est montré dans l'annexe C.2. En l'appliquant à la matrice B, on trouve :

$$B_{11} = a_1 I_{M_1+N+1}, \tag{6.66}$$

$$B_{22} = a_2 I_{M_2+N+1}, \tag{6.67}$$

$$B_{21} = 0, \tag{6.68}$$

$$B_{12} = \begin{pmatrix} b_0 & b_1 & \dots & b_{M_2-M_1} & 0 & \dots & 0 \\ 0 & b_0 & \dots & 0 & b_{M_2-M_1} & 0 & \dots \\ \vdots & & & & & & \\ 0 & \dots & 0 & b_0 & b_1 & \dots & b_{M_2-M1} \end{pmatrix}. \tag{6.69}$$

En notant par b le vecteur $b = (b_0, b_1, \dots, b_{M_2-M_1})^T$, on trouve :

$$G_2 Y_N(n) = a_2 S_{2,M_2+N}(n) \tag{6.70}$$

$$G_1 Y_N(n) = a_1 S_{1,M_1+N}(n) + \begin{pmatrix} b^T S_{2,M_2-M_1}(n) \\ b^T S_{2,M_2-M_1}(n-1) \\ \dots \\ b^T S_{2,M_2-M_1}(n-M_1-N) \end{pmatrix}. \tag{6.71}$$

L'équation (6.70) montre bien qu'on a extrait la deuxième source à un coefficient près. Pour avoir la première source, on doit appliquer un algorithme d'extraction classique sur l'équation (6.71), en utilisant le signal de référence donné par (6.70). Ceci peut être obtenu par décorrélation, à partir de statistiques au second ordre.

Cas de plus de deux sources

Dans le cas où on a plus de deux sources $N_S > 2$, on peut rencontrer deux situations différentes : soit les degrés des colonnes du filtre $H(z)$ sont strictement décroissants, soit certaines des colonnes ont des degrés identiques.

1. Si les degrés des colonnes de $H(z)$ sont toutes différents, c'est-à-dire $M_1 < M_2 < \dots < M_{N_s}$, alors la généralisation est presque évidente. En effet, selon la même paramétrisation que dans la section précédente, on écrit :

   $$G.U_N(H) = B \tag{6.72}$$

   où B est une matrice de taille $(N_s(N+1) + \sum_{i=1}^{N_s} M_i) \times (N_s(N+1) + \sum_{i=1}^{N_s} M_i)$ formée par $N_s \times N_s$ blocs B_{ij}, chaque bloc B_{ij} est de dimension $(M_i + N_s(N+1)) \times (M_j + N_s(N+1))$. D'après le lemme de la section précédente, on trouve que :

   $$B_{N_s j} = \begin{cases} 0 & \text{si} \quad j < N_s \\ aI & \text{si} \quad j = N_s \end{cases} \tag{6.73}$$

La relation (6.73) montre la possibilité d'extraire la dernière source. De cette façon, on peut retrouver les sources l'une après l'autre.

2. Si M_{N_s} est strictement supérieur aux autres degrés, alors la séparation de la source N_s est possible. Supposons qu'il existe un entier p tel que $M_{N_s} = M_{(N_s-1)} = \ldots = M_{(N_s-p)} > M_{(N_s-p-1)} \ldots$. Alors, d'après le même lemme, on montre que :

$$B_{ij} = \begin{cases} 0 & \text{si} \quad j < N_s - p < i \\ a_{ij} I & \text{si} \quad i \text{ et } j \in \{N_s - p, \ldots, N_s\} \end{cases} \tag{6.74}$$

D'où finalement, dans le cas général, on trouve que B est une matrice triangulaire par bloc suivant :

$$B = \begin{pmatrix} a_1 I & B_{12} & \ldots & \ldots & B_{1N_s} \\ 0 & a_2 I & B_{22} & \ldots & B_{2N_s} \\ 0 & \ldots & \vdots & \ldots & B_{iN_s} \\ 0 & \ldots & \boxed{\text{A1}} & \vdots & B_{jN_s} \\ 0 & \ldots & \vdots & \ldots & B_{kN_s} \\ 0 & \vdots & \ldots & \boxed{\text{A2}} & B_{jN_s} \\ 0 & \ldots & \vdots & \ldots & B_{lN_s} \\ 0 & \ldots & 0 & \ldots & a_l I \end{pmatrix},$$

où $A1$ et $A2$ sont des blocs carrés tel que leurs matrices coefficients sont de la forme $a_t I$. Si les degrés sont tous différents, alors on a la séparation complète au second ordre. Mais si on a p degrés qui sont identiques, alors on doit ensuite faire une séparation instantané. Pour mieux comprendre prenons un exemple.

Exemple : On suppose que l'on dispose de quatre sources $N_s = 4$ et que $M_4 = M_3 > M_2 > M_1$, alors B a la forme suivante :

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ B_{31} & B_{32} & B_{33} & B_{34} \\ B_{41} & B_{42} & B_{43} & B_{44} \end{pmatrix}. \tag{6.75}$$

Dans ce cas, et d'après la relation (6.74), on trouve que les blocs $B_{21} = B_{31} = B_{41} = B_{32} = B_{42} = 0$, tandis que les blocs $B_{11}, B_{22}, B_{33}, B_{34}, B_{43}, et B_{44}$ sont des

matrices identités à un scalaire près. La matrice B devient :

$$B = \begin{pmatrix} a_1 I & B_{12} & B_{13} & B_{14} \\ 0 & a_2 I & B_{23} & B_{24} \\ 0 & 0 & a_3 I & a_4 I \\ 0 & 0 & a_5 I & a_6 I \end{pmatrix}. \tag{6.76}$$

Avec la paramétrisation choisie, le produit de G par $Y_N(n)$ nous donne :

$$GY_N(n) = GU_N(H) \begin{pmatrix} S_{1,M_1+N}(n) \\ S_{2,M_2+N}(n) \\ S_{3,M_3+N}(n) \\ S_{4,M_3+N}(n) \end{pmatrix}. \tag{6.77}$$

Or $GU_N(H) = B$, et d'après la relation (6.76), on a :

$$GY_N(n) = \begin{pmatrix} a_1 I & B_{12} & B_{13} & B_{14} \\ 0 & a_2 I & B_{23} & B_{24} \\ 0 & 0 & a_3 I & a_4 I \\ 0 & 0 & a_5 I & a_6 I \end{pmatrix} \begin{pmatrix} S_{1,M_1+N}(n) \\ S_{2,M_2+N}(n) \\ S_{3,M_3+N}(n) \\ S_{4,M_3+N}(n) \end{pmatrix} =$$

$$\begin{pmatrix} a_1 S_{1,M_1+N}(n) + B_{12} S_{2,M_2+N}(n) + B_{13} S_{3,M_3+N}(n) + B_{14} S_{4,M_3+N}(n) \\ a_2 S_{2,M_2+N}(n) + B_{23} S_{3,M_3+N}(n) + B_{24} S_{4,M_3+N}(n) \\ a_3 S_{3,M_3+N}(n) + a_4 S_{4,M_3+N}(n) \\ a_5 S_{3,M_3+N}(n) + a_6 S_{4,M_3+N}(n) \end{pmatrix}$$

Il est clair que les deux dernières lignes nous permettent de faire une séparation instantanée pour séparer les deux dernières sources. Une fois qu'on a séparé les sources $s_3(n)$ et $s_4(n)$, on extrait les deux autres sources en appliquant un algorithme d'extraction deux fois de suite pour avoir $s_2(n)$ puis $s_1(n)$.

Notons bien que cette paramétrisation est plus générale que celle utilisée dans la section 6.4.1. En effet, si les colonnes du filtre $H(z)$ ont le même degré, on aboutira à une matrice de séparation B carré par bloc et tel que chaque coefficient bloc est une matrice $a_{ij} I$, ce qui nous oblige à terminer par une étape de séparation instantanée. On retrouve ainsi le résultat de la section précédente.
Ces résultats sont équivalents à ceux obtenus par Gorokhov [57], [56].

6.6 Résultats expérimentaux

Les expériences sont faites sur l'algorithme de type LMS de minimisation du premier critère C1 correspondant à la premier paramétrisation. D'un point de vue théorique, cette méthode suppose que l'ordre du filtre M est connu. Cette hypothèse est assez forte, car M est généralement inconnu et son estimation, possible par de nombreuses méthodes, risque d'être erronée. La robustesse de l'algorithme à cette hypothèse nous a conduit à effectuer deux autres ensembles d'expérience correspondant à la surestimation de l'ordre M ou à la sous-estimation de cet ordre.

6.6.1 Signaux, mélanges

On a pris un mélange de 2 sources ($N_s = 2$), observés par 3 ou 4 capteurs ($N_c = 3$ ou 4). Le filtre $H(z)$ est composée de filtres $h_{ij}(z)$ d'ordre 0 à 5, passe -bas, passe-haut ou passe-bande. Dans la suite, l'ordre et la nature des filtres seront précisés. La minimisation de la fonction coût est très lente. En général, l'algorithme converge au bout de 7000 à 10000 itérations. Expérimentalement, on a remarqué de bons résultats dans le cas des signaux stationnaires, et de très mauvais résultats pour les signaux non stationnaires, comme pour des signaux de parole par exemple.
Pour faire la séparation instantanée, on a utilisé une version modifiée et améliorée de l'algorithme de Jutten (voir [66], [64]), proposé par Moreau [91].
Il faut noter que en premier temps, on a utilisé un filtre RIF d'ordre 3, qui a le même degré sur toutes ses colonnes.

Signaux stationnaires

Les signaux stationnaires, utilisés pour faire ces manipulations, sont des signaux indépendants et identiquement distribués (iid), avec des densités de probabilité uniformes.
Si on trace la matrice produit $G.T_N(H)$, on remarque que cette matrice est presque une matrice diagonale par bloc, voir les figure 6.1. Ce résultat est prévu par les études théoriques : la séparation à l'ordre 2, laisse une indétermination à une matrice réelle A près.

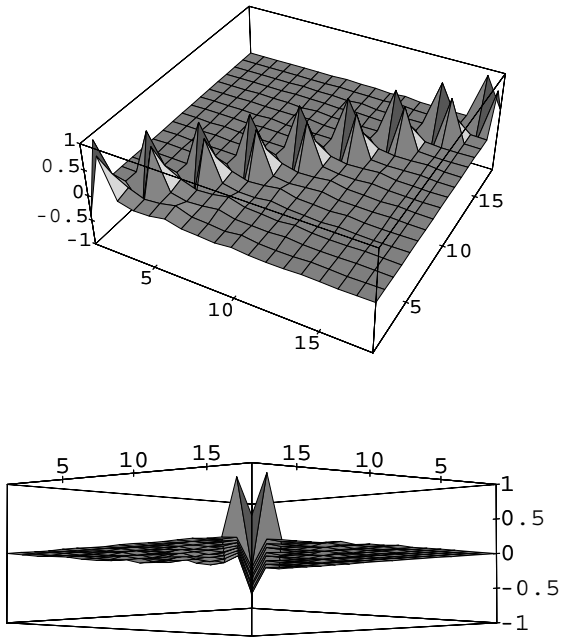


Figure 6.1: $GT_N(H)$

Après la séparation instantanée à l'ordre 4, on mesure une diaphonie résiduelle de l'ordre de -21 dB. Les figures 6.2 et 6.3 qui correspondent aux différences entre la première source et le premier signal de mélange, la première source et la première sortie, respectivement, donnent une idée de la qualité de la séparation.

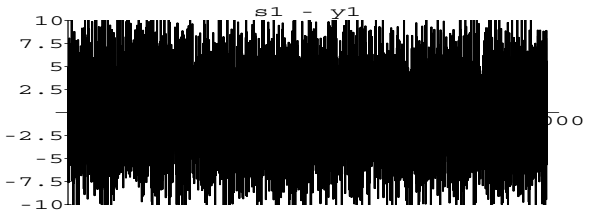


Figure 6.2: Différence entre la source s_1 et le mélange y_1, avec 10000 échantillons

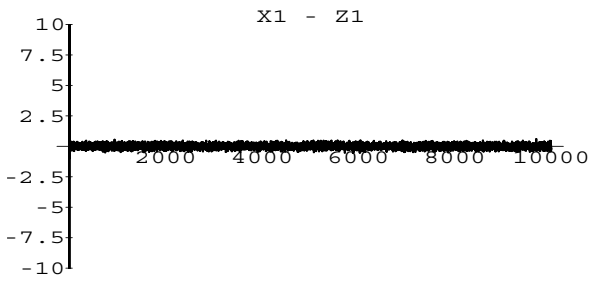


Figure 6.3: Différence entre la source s_1 et la source estimée x_1

Signaux de Paroles

Dans le cas des signaux de parole, on a eu de mauvais résultats. En effet, au bout de 10000 itérations l'algorithme donne encore une matrice G, tel que, le produit de cette matrice avec la matrice de Sylvester n'est pas une matrice diagonale par bloc, voir figure 6.4.

Donc le mélange est toujours convolutif, ce qui explique les mauvais résultats

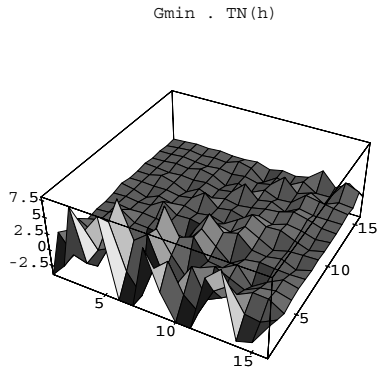
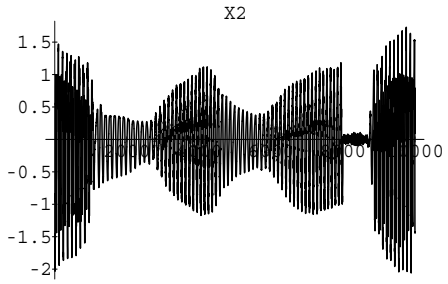
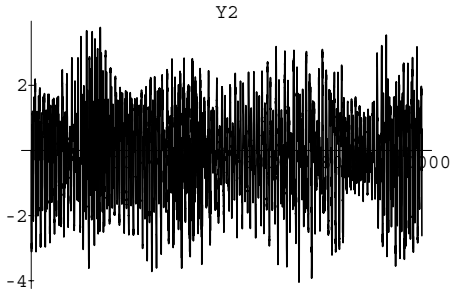


Figure 6.4: Performance.

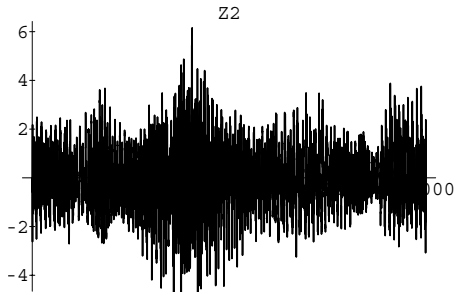
lors de la séparation instantanée. Le figure 6.5 montre que la séparation n'est pas réalisée.



Seconde source, avec 10000 échantillons



Mélange 2, avec 10000 échantillons.



Sortie 2, avec 10000 échantillons

Figure 6.5: Différents signaux

Mélange de parole et de bruit

En mélangeant un signal de parole avec un bruit blanc, on a utilisé les mêmes conditions d'expérience des deux dernières sections (même nombre d'échantillons et même filtre) pour faire la séparation. Le résultat global était mauvais. En effet, on a trouvé que :

$$G_0 T_N(H) = \begin{pmatrix} -4.05914 & 0.8059 \\ -2.88508 & -0.0635 \end{pmatrix} \tag{6.78}$$

Il est clair d'après cette dernière relation, que les deux sorties sont presque proportionnelles à la première source (le signal de parole). D'où, la première source a été trouvé sur les voies de sorties.

La figure 6.6 montre que la séparation des mélanges convolutifs n'est toujours

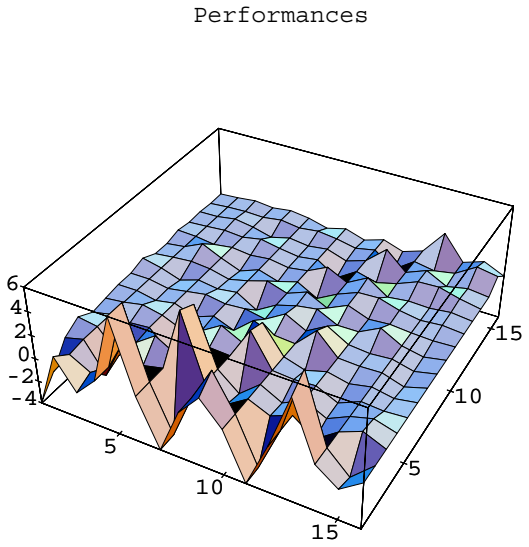


Figure 6.6: Performances pour un signal de parole avec un iid.

pas réalisée.

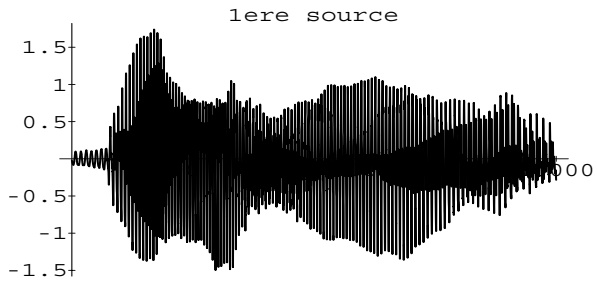


Figure 6.7: La première source.

Les différents signaux sont donnés par les figures 6.7, ..., 6.10.

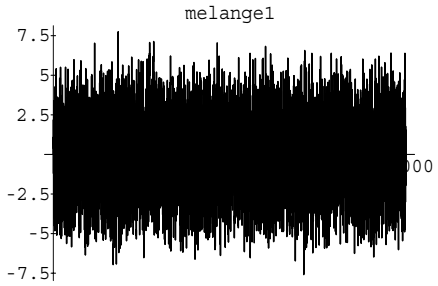


Figure 6.8: Le premier signal mélange.

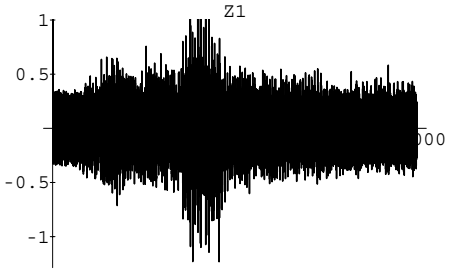


Figure 6.9: La première sortie.

Figure 6.10: La deuxième sortie.

On se demande : est ce qu'on peut améliorer les performances en ne choisissant que des filtres passe-bas? Dans ce cas, est ce qu'on est capable de séparer des signaux de parole avec le même algorithme?

Performances de C1 : avec des filtres passe bas

Pour un filtre RIF $H(z)$ du 3ème ordre, dont les composantes sont toutes de type passe-bas, on a remarqué de meilleures performances par rapport à celles trouvés dans les sections précédentes, utilisant toujours les mêmes paramètres de manipulation, voir figure 6.11.

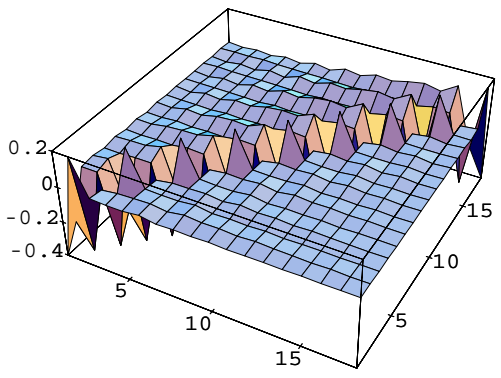


Figure 6.11: $GT_N(H)$

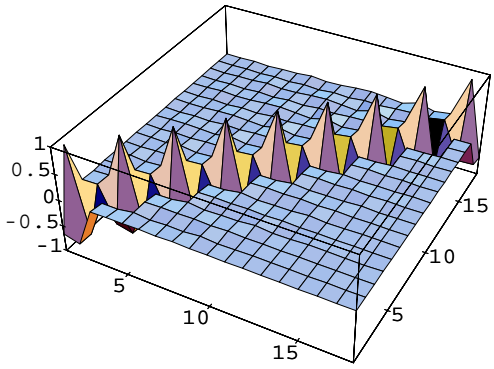


Figure 6.12: $GT_N(H)$ avec d'autres filtres passe bas

D'après les figures 6.11 et 6.12, on remarque bien l'amélioration de la diaphonie résiduelle par rapport à celle trouvée dans la figure 6.1.

Malheureusement, pour les signaux de paroles, ça n'a pas modifié grande chose (figure 6.13). On sait que l'estimation des matrices de covariance a une influence

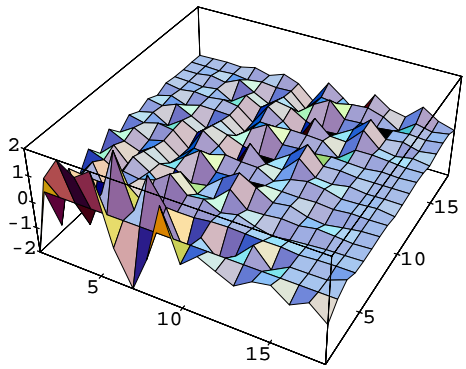


Figure 6.13: $GT_N(H)$

directe sur les performances de l'algorithme, pour cela on a essayé deux types d'estimateurs.

Estimation de la matrice de covariance

La matrice de covariance $R_{Y_N} = EY_N(n)Y_N(n)^T$ a une influence directe sur la convergence du critère C1. Pour diminuer l'influence d'estimation de cette matrice, on a fait quelques essais pour bien choisir l'estimateur et surtout pour bien choisir les paramètres d'estimation.

Pour estimer la matrice R_{Y_N}, on a choisi les deux estimateurs suivants :

- l'estimateur classique :

$$\hat{R}_N(k) = \frac{(k-1)\hat{R}_N(k-1) + Y_N(k)Y_N(k)^T}{k}. \tag{6.79}$$

- l'estimateur par filtrage passe bas :

$$\hat{R}_N(k) = (1-\alpha)\hat{R}_N(k-1) + \alpha Y_N(k)Y_N(k)^T, \tag{6.80}$$

avec k le numéro d'itération. Les signaux utilisés sont des bruits blancs filtrés avec des filtres RIF de différents ordres. Les courbes suivantes illustrent $\|R_{Y_N} - \hat{R}_N\|$, dans plusieurs manipulations :

- Estimateur classique :

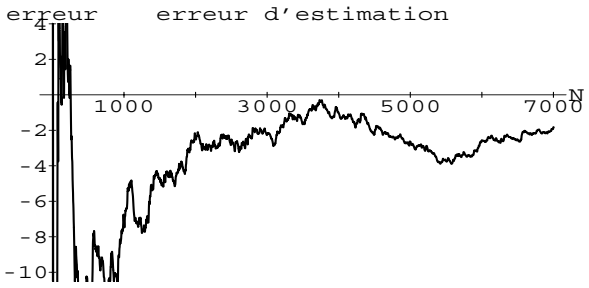
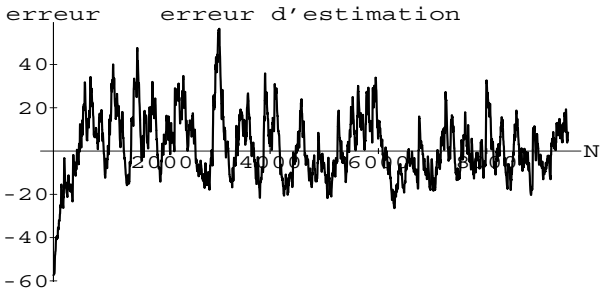
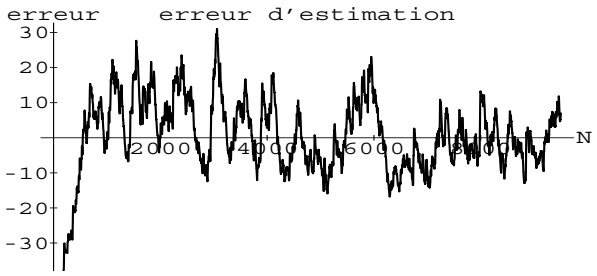


Figure 6.14: Estimateur classique

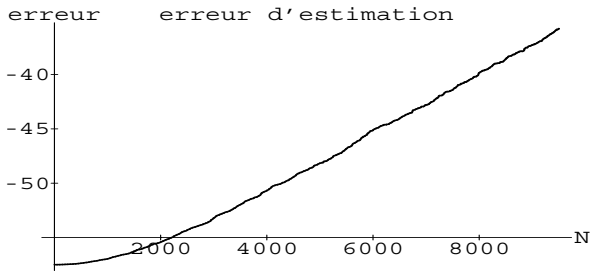
- Estimateur par un filtre passe-bas :



$\alpha = 0.1$



$\alpha = 0.005$



$\alpha = 0.0001$

Figure 6.15: Estimateur par un filtre passe-bas

Expérimentalement, on trouve clairement que l'estimateur classique est plus biaisé que celui d'un filtre passe-bas. Par contre la variance de l'erreur d'estimation de ce dernier est beaucoup plus grande.
Finalement, la valeur de α a une grande influence sur notre deuxième estimateur. Si α augmente, l'estimation devient plus rapide, mais la variance d'erreur d'estimation augmente aussi.

Donc, si on a suffisamment d'échantillons, il est préférable d'estimer R_{Y_N} par un estimateur classique ou à l'aide d'un filtre passe-bas avec une α petite. Si on n'a pas suffisamment d'échantillons, ou bien si le signal est non stationnaire (signal de parole par exemple), il vaut mieux estimer R_{Y_N} avec un filtre passe-bas.

6.6.2 Mauvaise estimation du degré M du filtre

Du point de vue théorique, l'utilisation de C1 comme critère de séparation nécessite (l'hypothèse H6[5], section 6.3) l'égalité entre les degrés des différentes colonnes de $H(z)$. Nous nous sommes interrogés sur la tolérance de l'algorithme à cette hypothèse.

Or, expérimentalement , on a observé qu'une mauvaise estimation (une sous ou sur estimation) de M nous permet dans certains cas de faire la séparation. A titre d'exemple, pour le même filtre décrit dans la section précédente, et sous-estimant la valeur de $M = 3$ à $M = 2$, on a trouvé de bonnes performances, voir figure 6.16.

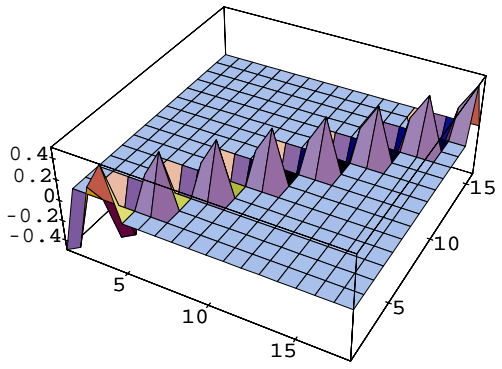


Figure 6.16: $GT_N(H)$ avec une sous estimation de M

[5]L'hypothèse H6 assure l'existence à gauche d'une inverse de la matrice de Sylvester $T_N(H)$.

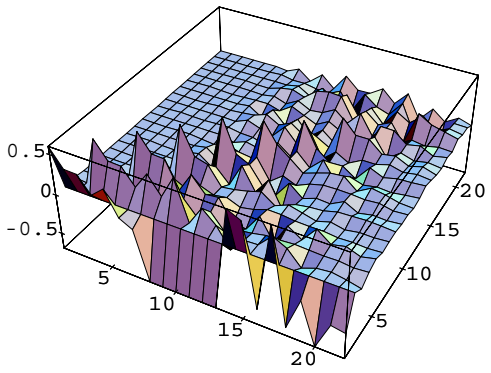


Figure 6.17: $GT_N(H)$ une sous-estimation

Par contre, une sous-estimation d'un filtre typique d'ordre cinq par un filtre de degré deux ne nous permet pas de faire la séparation, voir fig 6.20.

En effet, ce résultat était prévu puisque notre filtre était un filtre possédant plusieurs résonances, donc c'est un filtre typique d'ordre cinq, dont on illustre le module dans la figure 6.18 et la phase dans la figure 6.19.

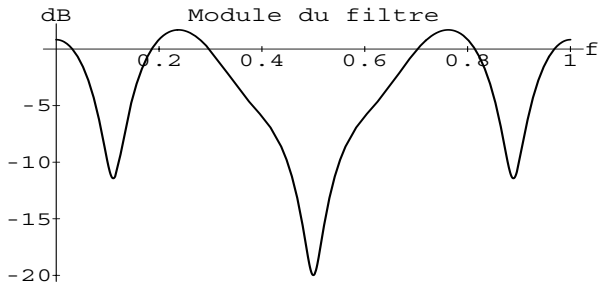


Figure 6.18: Le module du filtre d'ordre cinq.

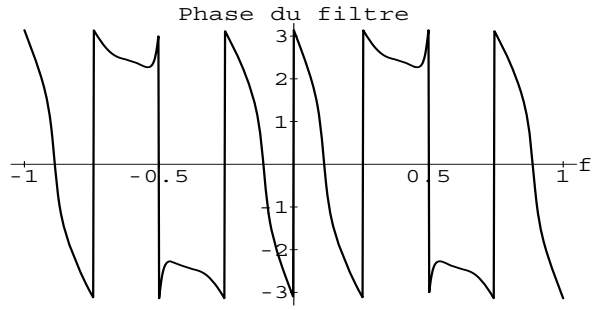


Figure 6.19: La phase du filtre d'ordre cinq.

Une sous-estimation de M revient à faire une troncature du filtre. Si les coefficients des termes de degrés élevés sont petits, alors la séparation par le critère C1 reste possible.

Une surestimation de M est toujours inadmissible. Théoriquement, l'image de $T_N(H)$, dans le cas d'une surestimation de M, nous permet d'identifier le filtre $H(z)$ à un filtre scalaire prés, de degré égale à la différence entre M et la valeur surestimé de M [1]. D'où l'explication des mauvaises performances obtenues. On illustre dans la figure 6.20, la matrice $GT_N(H)$ d'une manipulation où l'on surestime le degré $M = 3$ à $M = 4$.
Il faut noter, qu'on a toujours utilisé les mêmes paramètres pour faire les diffé-

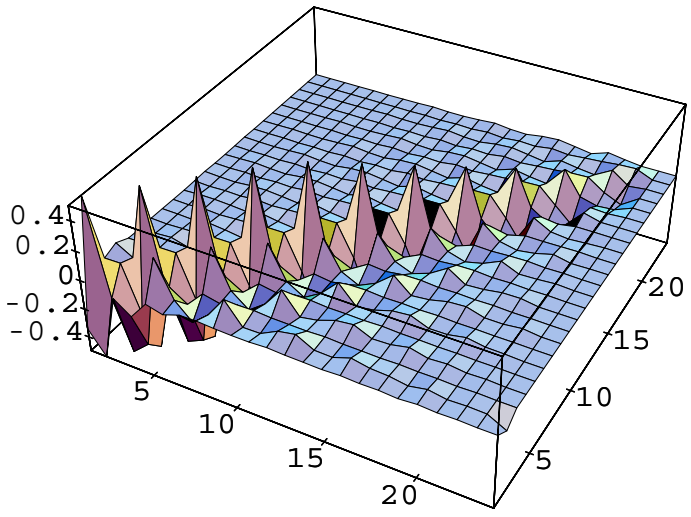


Figure 6.20: $GT_N(H)$ avec une surestimation

rentes manipulation de cette section.

Notons aussi une augmentation remarquable, au delà de la valeur nécessaire (voir début de ce chapitre), du champ d'observation N. Cela va introduire deux problèmes : le premier se résume par une augmentation du temps de calcul et de l'espace mémoire utilisé, tandis que le deuxième problème est un problème d'erreurs numériques qui va dégrader plus ou moins les résultats.

Finalement, un filtre qui n'a pas les mêmes degrés sur ces colonnes, conduit la convergence de C1 vers des résultats mauvais. On a préféré de ne pas illustre les résultats pour ne pas alourdir ce chapitre et en tout cas les courbes ne portent pas trop d'intérêt dans ce cas.

6.7 Conclusion

Les méthodes de type sous-espace sont des méthodes très élégantes en théorie. En pratique, on doit manipuler des grandes matrices, et les algorithmes sont très lourds en temps de calcul.

Ces méthodes montrent néanmoins que l'on peut séparer les sources presque entièrement avec des critères de second ordre statistiques, sous la condition d'avoir plus de capteurs que de sources et de connaître (au moins savoir estimer) les degrés des colonnes de $H(z)$.

Ce résultat, est à rapprocher des travaux de Van Gerven [53] qui montre que les résolutions des équations de corrélation entre les échantillons des signaux à différents instants, conduit à une solution unique dans le cas de deux sources dans un mélange convolutif, réalisé avec des filtres RIF strictement causaux.

Chapitre 7

Conclusion générale

La richesse de la littérature de séparation de sources est remarquable pour un problème relativement récent, apparu en 1984 [60].

Dans cette thèse, nous avons développé et exploré plusieurs aspects de ce problème.
Comme on l'a vu dans l'état de l'art, la plupart des algorithmes ont été conçus pour des mélanges linéaires instantanés. Dès l'année 90, sont apparus quelques algorithmes de séparation pour les mélanges convolutifs [68], [95]. Tout ces algorithmes reposant sur deux hypothèses fondamentales :

- L'indépendance statistique de sources.
- La nature du modèle utilisé.

L'exploitation de la première hypothèse est très variée selon les divers algorithmes (maximum de vraisemblance, fonction de contraste, cumulants, etc).
Actuellement, la plupart des auteurs ont traite majoritairement le cas de mélanges linéaires instantanés, quelques uns ont abordé les mélanges convolutifs.

En utilisant le modèle d'un mélange instantané, on a développé deux approches directes (chapitre III), l'une est basé sur la résolution d'un système d'équations, l'autre approche basée sur la structure géométrique de la distribution de signaux mélangés. Les deux approches sont simples et convergent rapidement, mais en pratique ne permettent que de traiter des mélanges de deux sources et deux capteurs. Leurs performances sont aussi moins bonnes qu'un algorithme adaptatif. Ces deux approches peuvent cependant constituer une bonne initialisation pour accélérer un algorithme adaptatif.

Le chapitre IV est focalisé sur une approche adaptative qui utilise des cumulants d'ordre quatre, et qui est valable pour les mélanges instantanés comme convolutifs. Le critère proposé dans ce chapitre est très simple. Les algorithmes dérivées de ce critère convergent assez rapidement, 500 - 1000 itérations.
Mais les performances dépendent de la nature de sources. Expérimentalement, on a ainsi trouvé de performances de l'ordre de -25 dB de diaphonie résiduelle pour les sources stationnaires et -20 dB pour les signaux de parole. Le problème de ce critère se manifeste surtout dans l'allure des cumulants qui sont trop plats au voisinage des solutions. Les algorithmes de minimisation des critères sont aussi peu précis.

L'hypothèse sur le signe de kurtosis de sources est utilisé très souvent dans les algorithmes de séparation. Au chapitre V, nous avons montré que dans le cas général de sources de ddp multimodales, ni la forme de la ddp et ni la nature asymptotique des sources (sous ou sur-gaussiennes) nous permettent de conclure sur le signe de kurtosis.

Finalement, dans le dernier chapitre VI, on a abordé le problème de mélange convolutif, selon une approche sous-espace. Ces algorithmes nécessitent un nombre de capteurs plus grand que celui de sources. On a montré que ces méthodes sous-espace permettent de séparer des sources, ou de ramener le mélange convolutif en un mélange instantané, en utilisant seulement les statistiques du second ordre.
Dans ce chapitre, on a proposé trois critères différents. Le premier critère est une adaptation d'un algorithme d'identification aveugle [54], et le deuxième critère est une approche de déflation. Ces deux critères exploitent une paramétrisation basée sur la matrice de Sylvester. Cette paramétrisation nécessite une hypothèse assez forte, concernant les degrés de colonnes de la matrice de mélange. Enfin, à l'aide d'une autre paramétrisation, on a proposé le troisième critère.
Malheureusement, ces méthodes souffrent, en pratique, d'une faible vitesse de convergence et mettent en jeu de très grandes matrices.

Annexes

Appendix A

Méthode directe

A.1 Moments de signaux mélangés

A partir du modèle de séparation, on peut calculer les moments croisées des signaux mélangés $y_1(t)$ et $y_2(t)$ en fonction des coefficients de la matrice mélange h_{ij} et des moments de sources $s_1(t)$ et $s_2(t)$. Tous calculs faits, on trouve un système de huit équations nonlinéaires fonction de huit variables h_{ij}, p_i, γ_i, $(i,j \in [1,2])$:

$$
\begin{aligned}
Mom_{11}(y_1,y_2) &= h_{11}h_{21}p_1 + h_{12}h_{22}p_2, && \text{(A.1)}\\
Mom_{20}(y_1,y_2) &= h_{11}^2 p_1 + h_{12}^2 p_2, && \text{(A.2)}\\
Mom_{02}(y_1,y_2) &= h_{21}^2 p_1 + h_{22}^2 p_2, && \text{(A.3)}\\
Mom_{31}(y_1,y_2) &= 3h_{11}h_{12}(h_{11}h_{22} + h_{21}h_{12})p_1p_2 \\
&\quad + h_{11}^3 h_{21}\gamma_1 + h_{12}^3 h_{22}\gamma_2, && \text{(A.4)}\\
Mom_{13}(y_1,y_2) &= 3h_{21}h_{22}(h_{11}h_{22} + h_{21}h_{12})p_1p_2 \\
&\quad + h_{11}h_{21}^3\gamma_1 + h_{12}h_{22}^3\gamma_2, && \text{(A.5)}\\
Mom_{22}(y_1,y_2) &= (h_{11}^2h_{22}^2 + 4h_{11}h_{21}h_{12}h_{22} + h_{12}^2h_{21}^2)p_1p_2 \\
&\quad + h_{11}^2h_{21}^2\gamma_1 + \ h_{12}^2h_{22}^2\gamma_2, && \text{(A.6)}\\
Mom_{40}(y_1,y_2) &= h_{11}^4\gamma_1 + 6h_{11}^2h_{12}^2p_1p_2 + h_{12}^4\gamma_2, && \text{(A.7)}\\
Mom_{04}(y_1,y_2) &= h_{21}^4\gamma_1 + 6h_{21}^2h_{22}^2p_1p_2 + h_{22}^4\gamma_2. && \text{(A.8)}
\end{aligned}
$$

Ce système est de forme très complexe par rapport à celui correspondant aux cumulants (voir paragraphe 3.2.2); les cumulants ont déjà un intérêt purement formel par rapport aux moments.

Appendix B

Méthodes adaptatives

B.1 Annulation ou minimisation

On montre dans cette section que l'annulation de $Cum_{22}(X)$ est équivalent à sa minimisation. La structure d'un mélange instantané, pour deux capteurs et deux sources (voir fig 2.1), et la relation (4.4) nous permettent de calculer la dérivée de Cum_{22} par rapport aux coefficients de la matrice séparatrice G :

$$\frac{\partial Cum_{22}(g_{12}, g_{21})}{\partial g_{12}} \geq 0.$$

En développant cette relation et en notant β_i le kurtosis de la source i, on trouve :

$$t_{11}\beta_1 t_{21}^2 h_{21} + t_{22}^2 \beta_2 t_{12} h_{22} \geq 0.$$

Et ce sera facile de prouve que :

$$(\beta_1 t_{21}^2 h_{21}^2 + t_{22}^2 \beta_2 h_{22}^2) g_{12} \geq -(\beta_1 t_{21}^2 h_{21} h_{11} + t_{22}^2 \beta_2 h_{22} h_{12}).$$

Si on suppose que les signaux ont de kurtosis positifs, alors le coefficient g_{12min} correspondant au minimum de critère est :

$$g_{12min} \geq -\frac{\beta_1 t_{21}^2 h_{21} h_{11} + \beta_2 t_{22}^2 h_{12} h_{22}}{\beta_1 t_{21}^2 h_{12}^2 + \beta_2 t_{22}^2 h_{22}^2}. \tag{B.1}$$

Cette relation nous permet de calculer les valeurs correspondant aux coefficients de la matrice totale $T = G.H$:

$$t_{11} = \frac{t_{22}^2 \beta_2 h_{22}(h_{11} h_{22} - h_{21} h_{12})}{\beta_1 t_{21}^2 h_{21}^2 + \beta_2 t_{22}^2 h_{22}^2} \tag{B.2}$$

$$t_{12} = \frac{t_{21}^2 \beta_1 h_{21}(h_{12} h_{21} - h_{22} h_{11})}{\beta_1 t_{21}^2 h_{21}^2 + \beta_2 t_{22}^2 h_{22}^2}. \tag{B.3}$$

Aussi bien, la valeur de fonction de coût pour cette valeur minimum :

$$Cum_{22}(g_{12min}, g_{21}) = \frac{\beta_1\beta_2(h_{11}h_{22} - h_{12}h_{21})^2 t_{22}^2 t_{21}^2}{\beta_1 t_{21}^2 h_{21}^2 + \beta_2 t_{22}^2 h_{22}^2}. \tag{B.4}$$

Les extremums sont données par la relation :

$$\frac{\partial Cum_{22}(g_{12min}, g_{21})}{\partial g_{21}} = \frac{\beta_1\beta_2 t_{21} t_{22}(h_{11}h_{22} - h_{12}h_{21})^2(\beta_1 t_{21}^3 h_{12} h_{21}^2 + \beta_2 t_{22}^3 h_{11} h_{22}^3)}{(\beta_1 t_{21}^2 h_{21}^2 + \beta_2 t_{22}^2 h_{22}^2)^2} = 0. \tag{B.5}$$

Finalement, on trouve :

$$t_{22} = 0 \tag{B.6}$$

$$t_{21} = 0 \tag{B.7}$$

$$t_{21} = -t_{22}\sqrt[3]{\frac{\beta_2 h_{11} h_{22}^2}{\beta_1 h_{12} h_{21}^2}} \tag{B.8}$$

Les solutions (B.6) et (B.7) correspondent aux solutions théoriques : $g_{21} = -h_{22}/h_{12}$ et $g_{12} = -h_{11}/h_{21}$, $g_{21} = -h_{21}/h_{11}$ et $g_{12} = -h_{12}/h_{22}$ respectivement. La solution (B.8) est un maximum. **La minimisation de ce critère est équivalent à son annulation**.

B.2 Existence des solutions parasites

Par un simple calcul, on peut montrer que ce critère n'admet pas des solutions parasites. En effet, le système :

$$\begin{aligned} \frac{\partial Cum_{22}(g_{12}, g_{21})}{\partial g_{12}} &= 0 \\ \frac{\partial Cum_{22}(g_{12}, g_{21})}{\partial g_{21}} &= 0 \end{aligned} \tag{B.9}$$

est équivalent à

$$\begin{cases} t_{11}\beta_1 t_{21}^2 h_{21} + t_{22}^2\beta_2 h_{22} t_{12} = 0 \\ t_{21}\beta_1 t_{11}^2 h_{11} + t_{22}\beta_2 h_{12} t_{12}^2 = 0, \end{cases} \tag{B.10}$$

Les solutions sont donc :

$$\begin{aligned} t_{11} &= t_{12} = 0 \\ t_{21} &= t_{22} = 0 \end{aligned} \tag{B.11}$$

$$\begin{aligned} t_{11} &= t_{22} = 0 \\ t_{12} &= t_{21} = 0 \end{aligned} \tag{B.12}$$

$$t_{21} = -t_{22}\sqrt[3]{\frac{\beta_2 h_{11} h_{22}^2}{\beta_1 h_{12} h_{21}^2}} \tag{B.13}$$

Les solutions (B.11) et (B.12) sont exactement les mêmes dans le cas d'annulation.

B.3 Dérivée des coûts par rapport aux statistiques de signaux

En supposant que les éléments de la diagonale de la matrice G sont égaux à un, les relations entre les sources estimées et les signaux mélanges sont :

$$x_i = y_i + g_{ij} y_j \tag{B.14}$$

$$y_i = \frac{x_i - g_{ij} x_j}{1 - g_{ij} g_{ji}}. \tag{B.15}$$

C'est qui nous donne sans aucune difficultés :

$$\begin{aligned} \frac{\partial x_i}{\partial g_{ij}} &= y_j = \frac{x_j - w_{ji} x_i}{1 - g_{ij} g_{ji}} \\ \frac{\partial x_i}{\partial g_{ji}} &= 0. \end{aligned} \tag{B.16}$$

Enfin, la dérivée est :

$$\begin{aligned} \frac{\partial Cum_{22}(x_i, xj)}{\partial g_{ij}} = \frac{2}{1 - g_{ij} g_{ji}} &[E x_i x_j^3 - g_{ji} E x_i^2 x_j^2 - 3 E x_j^2 E x_i x_j \\ &+ g_{ji} E x_i^2 E x_j^2 + 2 g_{ji} (E x_i x_j)^2]. \end{aligned} \tag{B.17}$$

Cette valeur sera utilisée la valeur pour ajuster les coefficient a chaque étape par une méthode de gradient.

B.4 Calcul direct du coût

Le $Cum_{22}(x_m, x_n)$ des deux signaux centrés est [89].

$$Cum_{22}(x_m, x_n) = Ex_m^2 x_n^2 - Ex_m^2\ Ex_n^2 - 2(Ex_m x_n)^2. \tag{B.18}$$

Il est facile d'exprimer chaque terme dans l'équation (B.18) en fonction des signaux sources. Pour cela, développons le moment $Ex_m^2 x_n^2$:

$$\begin{aligned} Ex_m^2 x_n^2 &= E(\sum_i t_{mi} s_i)^2 (\sum_i t_{ni} s_i)^2 \\ &= E \sum_{a,b,c,d} t_{ma} t_{mb} t_{nc} t_{nd} s_a s_b s_c s_d \\ &= \sum_{a,b,c,d} t_{ma} t_{mb} t_{nc} t_{nd} E s_a s_b s_c s_d. \end{aligned} \tag{B.19}$$

En considérant que les sources sont spatialement indépendantes et centrées, on trouve :

$$Es_a s_b s_c s_d = \begin{cases} p_a p_c & \begin{cases} \text{si} \quad a = b \neq c = d \\ \text{et} \quad a = d \neq b = c \end{cases} \\ p_a p_b \quad \text{si} \quad a = c \neq b = d \\ \gamma_a \quad \text{si} \quad a = b = c = d \\ 0 \quad \text{sinon,} \end{cases} \tag{B.20}$$

où $p_a = Es_a^2$, et $\gamma_a = Es_a^4$. Les relations (B.19) et (B.20) nous donnent :

$$Ex_m^2 x_n^2 = \sum_a t_{ma}^2 t_{na}^2 \gamma_a + \sum_{a \neq b} p_a p_b t_{ma} t_{nb} (t_{ma} t_{nb} + 2 t_{na} t_{mb}). \tag{B.21}$$

Par un calcul similaire, on développe la puissance des signaux de sorties :

$$\begin{aligned} Ex_m^2 &= (\sum_i t_{mi} s_i)^2 \\ &= E \sum_{a,b} t_{ma} t_{mb} s_a s_b \\ &= \sum_{a,b} t_{ma} t_{mb} E s_a s_b \\ &= \sum_a t_{ma}^2 p_a \end{aligned} \tag{B.22}$$

De la même façon, on montre que :

$$Ex_m x_n = \sum_a t_{ma} t_{na} p_a \tag{B.23}$$

En utilisant les relations (B.21), (B.22) et (B.23) on déduit :

$$\begin{aligned} Cum_{22}(x_m, x_n) &= \sum_a t_{ma}^2 t_{na}^2 (\gamma_a - 3P_a^2) \\ &= \sum_a t_{ma}^2 t_{na}^2 \beta_a. \end{aligned} \tag{B.24}$$

Où $\beta_a = Cum(s_a, s_a, s_a, s_a)$.

B.5 Les solutions possibles

On a choisi une démonstration basée sur les propositions logiques. Pour ça, on note l'équivalence entre la condition $t_{ma} = 0$ et la proposition logique p_m^a (si $p_m^a = 1$ donc p_m^a est vraie, alors $t_{ma} = 0$).
On peut reécrire la relation 4.8 par :

$$p_m^a + p_n^a \quad \forall a, \text{ et } m \neq n. \tag{B.25}$$

Pour une valeur quelconque de a, on trouve l'équation logique suivante :

$$\begin{gathered} (p_1^a + p_2^a)(p_1^a + p_3^a) \cdots (p_1^a + p_{N_c}^a). \\ (p_2^a + p_3^a) \ldots (p_2^a + p_{N_c}^a). \\ \vdots \\ (p_{(N-1)}^a + p_{N_c}^a) \end{gathered} \tag{B.26}$$

Le système (B.26) contient la multiplication de $\frac{N_c(N_c-1)}{2}$ terms $(p_i^a + p_j^a)$. En utilisant les tables de Karnaugh et l'algebre de boole, on peut montre que le système B.26 est équivalent à :

$$\begin{gathered} (p_1^a + p_2^a p_3^a \cdots p_{N_c}^a). \\ (p_2^a + p_3^a p_4^a \cdots p_{N_c}^a). \\ \vdots \\ (p_{(N_c-1)}^a + p_{N_c}^a) \end{gathered} \tag{B.27}$$

On peut demontre par récurrence que ce dernier système est équivalent à une équation composée par la somme de N_c termes, et chaque terme est le produit de $N_c - 1$ propositions logiques p_i^a :

$$\underbrace{p_1^a p_2^a \cdots p_{(N_c-1)}^a + \cdots + \underbrace{p_1^a p_i^a \cdots p_{N_c}^a}_{(N_c-1)\ coef} + \cdots + p_2^a p_3^a \cdots p_{N_c}^a}_{N_c\ terms} \tag{B.28}$$

Pour simplifier la notation, on note le produit des $(N_c - 1)$ coefficients de p_i^a par q_i^a. Et tout calcul fait la relation (B.28) devient :

$$(\sum_{i=1}^{N_c} q_i^1)(\sum_{i=1}^{N_c} q_i^2)\cdots(\sum_{i=1}^{N_c} q_i^{N_c}) = \sum_{i=1}^{N_c}\sum_{j=1}^{N_c}\cdots\sum_{k=1}^{N_c} q_i^1 q_j^2 \cdots q_k^{N_c} \tag{B.29}$$

Il est facile de démontrer ces trois résultats :

1. Il y a $N_c^{N_c}$ solutions au total.
2. Parmis ces solutions, $N_c^{N_c} - N_c!$ sont des solutions parasites correspondant à une de sorties au moins est nulle.
3. Pour chaque sortie y_i, il y a au total $(N_c - 1)^{N_c}$ solutions parasites.

B.6 Séparation dans le domaine z

Dans cette section on développe une approche intéressante pour étudier un critère de séparation de sources. Si cette approche n'a aucun sens du point de vue signal, elle restera un moyen puissant pour simplifier le calcul. D'après le paragraphe précédent, le calcul d'un critère basé sur les cumulants et pour un mélange convolutif est relativement compliqué, et il nécessite l'utilisation de la notation tensorielle. Pour cette raison on a essayé de développer un moyen de faire le calcul dans le domaine de z.

B.6.1 Fonction de coût

Cette approche consiste à appliquer les définitions de moments et de cumulants directement sur les signaux décrits dans le domaine de z (bien sûr sans l'utilisation habituelle de ployspectre). On définit normalement la relation entre les différents signaux (sources, signaux mélangés et sources estimés), dans les domaines temporel et fréquentiel par :

$$\begin{aligned} Y(n) &= \sum_{i=0}^{N} H(i)S(n-i), \\ Y(z) &= H(z)S(z). \end{aligned} \tag{B.30}$$

$$\begin{aligned} X(n) &= \sum_{i=0}^{N} G(i)Y(n-i), \\ X(z) &= G(z)Y(z). \end{aligned} \tag{B.31}$$

La matrice globale $T(z)$ (de terme général $t_{ij}(z)$), s'écrit :

$$T(z) = G(z)H(z) = \sum_{0}^{M} T(i)z^{-i}, \tag{B.32}$$

où M est l'ordre du filtre, alors :

$$X(n) = \sum_{i=0}^{N} T(i)S(n-i), \tag{B.33}$$

$$X(z) = T(z)S(z). \tag{B.34}$$

Avec les notations tensorielles [11], les relations (B.33) et (B.34) deviennent :

$$X^1(n) = \sum_{\alpha} T_1^1(\alpha) \bullet S^1(n-\alpha), \tag{B.35}$$

$$X^1(z) = T_1^1(z) \bullet S^1(z). \tag{B.36}$$

La transformation en z d'un signal aléatoire n'a aucun sens, du moment que le polynôme obtenu est un polynôme à coefficients aléatoires. Donc on ne peut pas l'utiliser, mais si on applique l'espérance mathématique sur ce polynôme, on obtiendra un polynôme avec des coefficients réels et fixés, pour un signal stationnaire réel.
En appliquant la définition de cumulant ([89] et [49]) sur le tenseur $X^1(z)$, on obtient :

$$\begin{aligned} Cum(X^1(z), X_1(z), X^1(z), X_1(z)) &= T_1^1(z) \otimes T_1^{1T}(z) \\ \otimes T_1^1(z) \otimes T_1^{1T}(z) &\bullet Cum(S^1(z), S_1(z), S^1(z), S_1(z)), \end{aligned} \tag{B.37}$$

On a vu dans le domaine temporel que l'indépendance de sources impose $Cum(S^1(z), S_1(z), S^1(z), S_1(z))$ à être diagonale. En rappelant que δ_{ab}^{cd} le symbole de Kronecker et par $\beta_a(z)$ l'auto-cumulant d'ordre quatre de la source s_a, le terme $ijkl$ de (B.37) s'écrit :

$$\begin{aligned} Cum_{ik}^{jl}(X)(z) &= h_i^a(z)h_b^{jT}(z)h_k^c(z)h_d^{lT}(z)Cum(s_a(z), s_b(z), s_c(z), s_d(z))\delta_{ab}^{cd} \\ &= t_{ia}(z)t_{ja}(z)t_{ka}(z)t_{la}(z)\beta_a(z) \end{aligned} \tag{B.38}$$

L'annulation des cumulants de la forme de (B.38) nous donne :

$$Cum_{ij}^{ij}X(z) = 0 \Longleftrightarrow t_{ia}^2(z)t_{ja}^2(z)\beta_a(z) = 0 \tag{B.39}$$

La relation (B.39) est semblable à la relation obtenue dans le domaine temporel (voir [84] et [85]).

Pour un mélange instantané, on a aussi trouvé que l'annulation des cumulants 22 suffit pour obtenir la séparation de sources [84], avec la seule condition que les sources ont le même signe de kurtosis.
Cependant, dans le domaine z, cette hypothèse sur le signe de kurtosis n'est pas réaliste. En effet, les sources ont le même signe de kurtosis en z, ce qui revient à dire que leur cumulant d'ordre quatre dans le domaine z a le même signe et pour tout z. Donc leurs transformées en z sont des polynômes réels,$\forall z$. Autrement dit, ces polynômes sont des constantes, ce qui donne que les sources sont des impulsions de Dirac.

B.6.2 Le cumulant d'ordre quatre d'un signal iid

Une première hypothèse simple [96] est de supposer que les sources sont des signaux iid.
La transformation en z d'un signal $s(t)$ est $S(z) = \sum_i s(i)z^{-i}$, où $s(i)$ est un échantillon de $s(t)$ à l'instant t. Par généralisation de la définition, le cumulant d'ordre quatre de $S(z)$ est :

$$Cum_4(S(z)) = ES^4(z) - 3(ES^2(z))^2. \tag{B.40}$$

Il est facile de montrer que :

$$ES^4(z) = \sum_{i,j,k,l} Es(i)s(j)s(k)s(l)z^{-(i+j+k+l)} \tag{B.41}$$

Les signaux étant centrés et iid, il reste :

$$ES^4(z) = \sum_i Es(i)^4 z^{-4i} + 6\sum_{i>j} Es(i)^2 Es(j)^2 z^{-2(i+j)} \tag{B.42}$$

De plus en utilisant l'hypothèse de la stationarité des signaux sources, on montre que :

$$ES^4(z) = Es^4 \sum_i z^{-4i} + 6(Es^2)^2 \sum_{i>j} z^{-2(i+j)} \tag{B.43}$$

De même on a :

$$ES^2(z) = \sum_{i,j} Es(i)s(j)z^{-i-j} = Es^2 \sum_i z^{-2i} \tag{B.44}$$

Notons par $P_s = Es^2$ la puissance du signal s, et par $Cum_4(s) = Es^4 - 3P_s^2$ le cumulant d'ordre quatre de ce signal. Il est évident de voir que (B.40) devient en utilisant (B.43) et (B.44) :

$$\begin{aligned} Cum_4(S(z)) &= Es^4 \sum_i z^{-4i} + 6(Es^2)^2 \sum_{i>j} z^{-2(i+j)} - 3P_s^2 \sum_{i,j} z^{-2(i+j)} \\ &= Cum_4(s) \sum_i z^{-4i} \end{aligned} \tag{B.45}$$

En reportant (B.45) dans la relation (B.39), que $\forall i, j$:

$$\sum_{a,k} t_{ia}^2(z) t_{ja}^2(z) Cum_4(s_a) z^{-4k} = 0 \tag{B.46}$$

Pour alléger l'écriture, on note $R_a(z) = t_{ia}^2(z) t_{ja}^2(z)$, et le degré de $R_a(z)$ par d_{R_a}. Le coefficient de $R_a(z)$ du terme z^{-q} est :

$$r_a(q) = \sum_{k,l,m,n} t_{ia}(k) t_{ia}(l) t_{ja}(m) t_{ja}(n), \tag{B.47}$$

avec $k + l + m + n = q$, $t_{ia}(k)$ est le coefficient de z^{-k} dans $t_{ia}(z)$.
Notons $Cum_4(s_a) = \beta_a$. Alors (B.46) s'écrit :

$$\sum_{a,k,q} r_a(q) \beta_a z^{-4k-q} = 0, \tag{B.48}$$

où $1 \leq a \leq N_s$, $0 \leq q \leq d_{R_a}$, et $k \geq 0$.
Si on suppose dans la relation (B.48) que k est un nombre fini, c'est à dire que le développement en z du signal est tronqué : $0 \leq k \leq N$.
Alors la relation (B.48) devient :

$$\begin{aligned} &\sum_{a} r_a(d_{R_a}) \beta_a z^{-4N-d_{R_a}} + \sum_{a,k<N} r_a(d_{R_a}) \beta_a z^{-4k-d_{R_a}} \\ &+ \sum_{a,q<d_{R_a}} r_a(q) \beta_a z^{-4N-q} + \sum_{a,k<N,q<d_{R_a}} r_a(q) \beta_a z^{-4k-q} = 0 \end{aligned} \tag{B.49}$$

Dans (B.49), le coefficient de chaque terme z^{-p} doit être nul. Pour $z^{-4N-d_{R_a}}$, on a alors :

$$\sum_{a} r_a(d_{R_a}) \beta_a = 0, \tag{B.50}$$

où $r_a(d_{R_a})$ est le coefficient de plus haut degré de $R_a(z)$, c'est à dire :

$$r_a(d_{R_a}) = t_{ia}^2(max_i) t_{ja}^2(max_j), \tag{B.51}$$

et $t_{ia}(max_i)$ est le coefficient de plus haut degré de $t_{ia}(z)$. Finalement à partir de (B.51), tout les coefficients de $r_a(d_{R_a})$ sont positifs. En supposant que toutes les sources ont le même signe de kurtosis alors tous les β_a ont le même signe.
Maintenant, la relation (B.50) nous assure que :

$$\sum_{a} r_a(d_{R_a}) \beta_a = 0 \Longleftrightarrow t_{ia}^2(max_i) t_{ja}^2(max_j) = 0 \quad \forall a. \tag{B.52}$$

Cette dernière équation nous aide à simplifier la relation (B.49) qui devient :

$$\sum_{a,k,q<d_{R_a}} r_a(q) \beta_a z^{-4k-q} = 0. \tag{B.53}$$

Cette relation est identique à la relation (B.48) avec la seule différence que $q < d_{R_a}$. Donc par récurrence on montre que $R_a(z) = 0$ et :

$$t_{ia}^2(z)t_{ja}^2(z) = 0 \tag{B.54}$$

Et cette dernière relation montre que :

$$t_{ia}(z)t_{ja}(z) = 0. \tag{B.55}$$

Enfin, la séparation à un filtre près est assuré par l'équation (B.55), [85]. Si on ne prend pas des hypothèses supplémentaires sur la matrice filtre alors on peut avoir des solutions parasites correspondant à des sorties nulles.

B.6.3 Hypothèses sur les filtres

La relation (B.55) est équivalent à $T(z)$ est le produit d'une matrice diagonale et une matrice de permutation près (voir la sous-section précédente et [85]). Mais il faut s'assurer que toutes les lignes de $T(z)$ ne sont pas nulles (dans le cas contraire, on aura des sorties nulles).
Le filtre global $T(z)$ est calculé comme étant le produit de $T(z) = G(z)H(z)$. Notons $T_i(z)$ (respectivement $G_i(z)$) la ième ligne de $T(z)$ (respectivement de $G(z)$). Alors :

$$T_i(z) = G_i(z)H(z). \tag{B.56}$$

On suppose que $H(z)$ est à colonne réduite. Selon [69], une matrice est dite à colonne réduite si et seulement si le degré du déterminant de cette matrice est égal à la somme des degrés de ses colonnes. Or par définition, le degré d'une colonne est le degré le plus élevé des coefficients (polynômes en z) de cette colonne.
Pour une matrice à colonne réduite, il est connu d'après [69] que le degré du vecteur $G_i(z)H(z)$ est égal à la somme de degrés de $G_i(z)$ et de $H(z)$. On définit le degré d'une matrice polynômiale comme étant le degré le plus élevé de ses coefficients (qui sont des polynômes).
Donc si $G_i(z)$ n'est pas un vecteur nul alors $T_i(z)$ n'est pas un vecteur nul. Et la séparation est achevée à une permutation et une matrice diagonale polynômiale près. Pour que $G_i(z)$ soit différente de zéro, il suffit de poser $g_{ii}(z) = 1$.

B.6.4 Séparation temporelle

On peut déduire du critère précédent, dans le domaine fréquentiel, un critère temporel. En effet, le cumulant fréquentiel utilisé dans le paragraphe précédent

est équivalent à :

$$\begin{aligned} Cum_2^2(X(z)) &= Cum(x_m(z), x_m(z), x_n(z), x_n(z)) \\ &= Cum(\sum_i x_m(i)z^{-i}, \sum_j x_m(j)z^{-j}, \sum_k x_n(k)z^{-k}, \sum_l x_n(l)z^{-l}) \\ &= \sum_{i,j,k,l} z^{-(i+j+k+l)} Cum(y_1(i), y_1(j), y_2(k), y_2(l)), \end{aligned} \tag{B.57}$$

$\forall m < n$. Donc l'annulation de (B.57) est équivalent à :

$$Cum_2^2(X(z)) = 0 \Longleftrightarrow \sum_{i+j+k+l=\mathrm{c}} Cum(x_m(i), x_m(j), x_n(k), x_n(l)) = 0, \tag{B.58}$$

$\forall m < n$, $0 \leq \mathrm{c} \leq 4N + 1$ et N est la borne du développement du signal dans le domaine de z, voir l'équation (B.49).
La séparation est finie une fois qu'on a estimé une matrice polynômiale $G(z)$ de taille $N_s \times N_s$. Si le degré de cette matrice est d_G, alors on a $(d_G+1)N_s^2$ coefficients à déterminer. La séparation nécessite que le degré de $G(z)$ soit plus grand que celui de $H(z)$, car le degré de liberté du modèle dépend directement du degré de $G(z)$, tandis que le nombre des paramètres inconnues est lié au degré de $H(z)$.

Appendix C

Méthodes sous-espace

C.1 Rang de la matrice de Sylvester

Soit $T_N(H)$ une matrice de Sylvester correspondant à une fonction de transfert RIF $H(z)$ $N_c \times N_s$, et dont les colonnes constituent une base minimale polynomiale de l'espace rationnel de l'espace qu'elles engendrent (i.e. le rang de $H(z)$ est égal à N_s pour tout z et $H(z)$ est à colonnes réduites).
La matrice $T_N(H)$ a pour dimensions $N_c(N+1) \times (M+N+1)N_s$ où M est le plus haut degré dans $H(z)$. Pour calculer son rang, on évalue la dimension de son noyau à gauche, i.e. l'ensemble des vecteurs lignes g de dimension $N_c(N+1)$ pour lesquels $gT_N(H) = 0$. En passant aux transformées en z, ceci équivaut à $g(z)H(z) = 0$. On se ramène donc au calcul de la dimension de l'ensemble des polynômes $1 \times N_c$ de degré au plus N pour lesquels $g(z)H(z) = 0$.

Pour faire ce calcul, on introduit une base minimale polynomiale $G(z)$ de l'orthogonal de l'espace engendré par les colonnes de $H(z)$. $G(z)$ est donc une matrice polynomiale $(N_c - N_s) \times N_c$ qui est de rang $N_c - N_s$ en tout point du plan complexe, et qui est à ligne réduite (propriété duale d'être à colonnes réduites). Un vecteur $g(z)$ est tel que $g(z)H(z) = 0$ ssi $g(z) = r(z)G(z)$ pour un certain polynôme $1 \times (N_c - N_s)$ $r(z)$. Il est alors clair que la dimension de l'espace des $g(z)$ de degrés inférieurs à N et qui vérifient $g(z)H(z) = 0$ est égale à celle des $r(z)$ pour lesquels le degré de $r(z)G(z)$ est au plus N. Pour évaluer la dimension de cet espace, on se sert d'une propriété importante des matrices polynomiale à lignes réduites. Cette propriété nous permet d'affirmer que si les $r_i(z)$ sont les composantes de $r(z)$ et si les $G_i(z)$ sont les lignes de $G(z)$, alors le degré de $r(z)G(z)$ coïncide avec le max des degrés des $r_i(z)G_i(z)$ (si l'un des $r_i(z)$ est nul, on compte le degré de $r_i(z)G_i(z)$ à 0. Notons $(M_i^{\perp})_{i=1,N_c-N_s}$ les degrés des lignes de $G(z)$. Alors, le degré de $r(z)G(z)$ est inférieur à N ssi les

degrés des $r_i(z)G_i(z)$ sont inférieurs à N pour tout i. Mais, le degré de $r_i(z)G_i(z)$ est égal à $\deg(r_i) + \deg(G_i) = \deg(r_i) + M_i^\perp$. De ceci, on déduit immédiatement la dimension de l'espace des $r(z)$ tels que degré de $r(z)G(z)$ est inférieur à N. Faisons le calcul dans le cas où N est supérieur ou égal au sup des $M_i^\perp$. Dans ce cas, aucun des r_i ne doit être nul, pour chaque i, degré de $r_i(z)G_i(z)$ inférieur à N équivaut à

$$\deg(r_i) \leq (N - M_i^\perp)$$

L'espace des polynômes correspondant est de dimension $(N - M_i^\perp + 1)$, et en faisant la somme sur i, on trouve que le noyau à gauche de $T_N(H)$ a pour dimension :

$$\sum_{i=1}^{N_c-N_s} (N - M_i^\perp + 1) = (N_c - N_s)(N+1) - \sum_{i=1}^{N_c-N_s} M_i^\perp$$

Mais, si on désigne par $(M_i)_{i=1,N_s}$ les degrés des colonnes de $H(z)$, on sait que

$$\sum_{i=1}^{N_c-N_s} M_i^\perp = \sum_{i=1}^{N_s} M_i$$

Puisque le rang de $T_N(H)$ est égal $(N+1)N_c$ – (dimension du noyau à gauche), on en déduit que si N est plus grand que le sup des $M_i^\perp$, alors,

$$\text{Rang}(T_N(H)) = p(N+1) + \sum_{i=1}^{N_s} M_i$$

De cette formule, on peut trouver un critère d'existence d'une inverse à gauche de $T_N(H)$. En effet, la matrice sera inversible à gauche ssi son rang est égal à son nombre de colonnes. Puisque le nombre de colonnes vaut $N_s(M+N+1)$, on voit qu'une condition nécessaire et suffisante d'existence d'une inverse à gauche est que $N_sM = \sum_{i=1}^{N_s} M_i$. Mais, par définition, M est égal au sup des M_i. Donc, la condition équivaut à $M_i = M$ pour tout $i = 1, N_s$. Pour que $T_N(H)$ soit inversible à gauche pour N supérieur au sup des $M_i^\perp$, il faut et il suffit que les degrés des colonnes de $H(z)$ soient tous égaux.

C.2 Type de la solution

En posant $\mathcal{A} = GT_N(H)$, on aura $GY_N(n) = \mathcal{A}S_{(M+N)}(n)$. La fonction coût est un critère quadratique et son minimum est zéro. Donc la matrice G, qui minimise la fonction coût, à la propriété suivante, (voir 6.8) :

$$\begin{aligned} [I_{(M+N)N_s} \;\; 0_{N_s}]\mathcal{A}S_{(M+N)}(n) &= [0_{N_s} \;\; I_{(M+N)N_s}]\mathcal{A}S_{(M+N)}(n+1) \qquad \text{(C.1)} \\ \Longleftrightarrow [0_{N_s} \;\; (I_{(M+N)N_s} \;\; 0_{N_s})\mathcal{A}]S_{(M+N+1)}(n+1) &= [(0_{N_s} \;\; I_{(M+N)N_s})\mathcal{A} \;\; 0_{N_s}]S_{(M+N+1)}(n+1). \end{aligned}$$

Et ceci est vrai pour tout n. Si le vecteur $S_{(M+N+1)}(n+1)$ a ses composantes linéairement indépendantes, alors :

$$[0_{N_s} \quad (I_{(M+N)N_s} \quad 0_{N_s})\mathcal{A}] = [(0_{N_s} \quad I_{(M+N)})\mathcal{A} \quad 0_{N_s}].$$

Supposons que :

$$\mathcal{A} = (A_{ij}) = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}.$$

où A_{ij} est une matrice de dimension $N_s \times N_s$, a et d sont de dimension $N_s \times (M+N+1)N_s$, et b et c sont de dimension $(M+N)N_s \times (M+N+1)N_s$. D'après l'équation C.1, on peut déduire que :

$$[b \quad 0_{N_s}] = [0_{N_s} \quad c]. \tag{C.2}$$

Cette dernière équation nous montre que :

- La première colonne bloc (c.à.d les p premières colonnes) de b est égale à zéro. Donc on trouve :

$$A_{i1} = 0, \quad \forall\ 1 < i \leq M+N+1. \tag{C.3}$$

- La dernière colonne bloc de c est égale à zéro.

$$A_{i(M+N+1)} = 0, \quad \forall\ 1 \leq i < m+N+1. \tag{C.4}$$

- $A_{ij} = A_{(i-1)(j-1)}$

Donc la matrice $\mathcal{A}$ est une matrice diagonale par bloc.

C.3 Dérivation du critère

Le critère cherche à minimiser la fonction coût :

$$\text{Min } C(G) = E\|[I_{(M+N)N_s} \quad 0_{N_s}]GY_N(n) - [0_{N_s} \quad I_{(M+N)N_s}]GY_N(n+1)\|^2 \tag{C.5}$$

En développant cette fonction, on trouve :

$$\begin{aligned} C(G) &= \text{E Trace}\{[(I_{(M+N)N_s} \quad 0_{N_s})GY_N(n) - (0_{N_s} \quad I_{(M+N)N_s})GY_N(n+1)] \\ &\quad [(I_{(M+N)N_s} \quad 0)GY_N(n) - (0_{N_s} \quad I_{(M+N)N_s})GY_N(n+1)]^T\} \\ &= \text{Trace}\{(I_{(M+N)N_s} \quad 0_{N_s})GEY_N(n)Y_N^T(n)G^T \begin{pmatrix} I_{(M+N)N_s} \\ 0_{N_s}^T \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
- \quad & (I_{(M+N)N_s} \quad 0_{N_s}) G E Y_N(n) Y_N^T(n+1) G^T \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix} \\
- \quad & (0_{N_s} \quad I_{(M+N)N_s}) G E Y_N(n+1) Y_N^T(n) G^T \begin{pmatrix} I_{(M+N)N_s} \\ 0_{N_s}^T \end{pmatrix} \\
+ \quad & (0_{N_s} \quad I_{(M+N)N_s}) G E Y_N(n+1) Y_N^T(n+1) G^T \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix} \} \qquad \text{(C.6)}
\end{aligned}
$$

Notons par $R_{Y_N}(0)$ la matrice de covariance de $Y_N(n)$, tel que $R_{Y_N}(0) = E Y_N(n) Y_N^T(n)$, et $R_{Y_N}(1) = E Y_N(n) Y_N^T(n+1)$. Tout en supposant que les signaux sont stationnaires, on démontre que :

$$
\begin{aligned}
C(G) \quad = \quad & \text{Trace}\{(I_{(M+N)N_s} \quad 0_{N_s}) G R_{Y_N}(0) G^T \begin{pmatrix} I_{(M+N)N_s} \\ 0_{N_s}^T \end{pmatrix} \} \\
- \quad & 2 \ \text{Trace}\{(I_{(M+N)N_s} \quad 0_{N_s}) G R_{Y_N}(1) G^T \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix} \} \\
+ \quad & (0_{N_s} \quad I_{(M+N)N_s}) G E Y_N(n+1) Y_N^T(n+1) G^T \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix} \} \qquad \text{(C.7)}
\end{aligned}
$$

On peut montrer que si $f(G) =$Trace $AGBG^TC$, alors la dérivé [1] de cette fonction par rapport à G est $\frac{df(G)}{dG} = A^TC^TGB^T + CAGB$. En effet, en utilisant la notation d'Einstein :

$$
\begin{aligned}
\frac{\partial \text{Trace}(AGBG^TA^T)}{\partial G} \quad &= \quad \frac{\partial (a_{il} g_{lm} b_{mn} g_{nr}^T a_{ri}^T)}{\partial G} \\
&= \quad \frac{\partial (a_{il} g_{lm} b_{mn} g_{rn} a_{ir})}{\partial G} \qquad \text{(C.8)}
\end{aligned}
$$

Le coefficient général de cette matrice est égale :

$$
\frac{\partial \text{Trace}}{\partial g_{pq}} = a_{ip} b_{qn} g_{rn} a_{ir} + a_{il} g_{lm} b_{mq} a_{ip} \qquad \text{(C.9)}
$$

Enfin, $\frac{df(G)}{dG} = A^TC^TGB^T + CAGB$.
D' où la dérivé de $C(G)$ par rapport à G est :

$$
\begin{aligned}
\frac{1}{2} \frac{\partial C(G)}{\partial G} \quad = \quad & \begin{pmatrix} I_{(M+N)N_s} & 0 \\ 0 & 0 \end{pmatrix} G R_{Y_N}(0) - \begin{pmatrix} 0 & I_{(M+N)N_s} \\ 0 & 0 \end{pmatrix} G R_{Y_N}^T(n+1) \\
- \quad & \begin{pmatrix} 0 & 0 \\ I_{(M+N)N_s} & 0 \end{pmatrix} G R_{Y_N}(1) + \begin{pmatrix} 0 & 0 \\ 0 & I_{(M+N)N_s} \end{pmatrix} G R_{Y_N}(0)
\end{aligned}
$$

[1] La dérivé d'une fonction $f(A)$ par rapport à la matrice A est une matrice D de même dimensions que A. Ses coefficients sont calculés par : $D = (d_{ij}) = (\frac{\partial f(A)}{a_{ij}})$.

$$
\begin{aligned}
&= \begin{pmatrix} I_{N_s} & 0 & \\ 0 & 2I_{(M+N-1)p} & 0 \\ 0 & 0 & I_{N_s} \end{pmatrix} GR_{Y_N}(0) \\
&- \begin{pmatrix} 0 & I_{(M+N)N_s} \\ 0 & 0 \end{pmatrix} GR^T_{Y_N}(n+1) - \begin{pmatrix} 0 & 0 \\ I_{(M+N)N_s} & 0 \end{pmatrix} GR_{Y_N}(1)
\end{aligned}
$$

Finalement, on pourra minimiser la fonction coût par une simple règle d'adaptation :

$$G(k+1) = G(k) - \mu \frac{dC(G)}{dG} \tag{C.10}$$

C.4 Minimisation par une méthode de Lagrange généralisée

Le système qu'on cherche à résoudre est :

$$\left\{ \begin{array}{c} \text{Min } C(G) = E\|[I_{(M+N)N_s} \;\; 0_{N_s}]GY_N(n) - [0_{N_s} \;\; I_{(M+N)N_s}]GY_N(n+1)\|^2 \\ \text{avec } G_0 R_{Y_N}(0) G_0^T = I_{N_s} \end{array} \right. \tag{C.11}$$

La contrainte a une forme matricielle symétrique (car $R_{Y_N}(0)$ est symétrique). Donc, on a $N_s(N_s+1)/2$ contraintes différentes. On a adopté une méthode semblable à celle de Lagrange (pour la minimisation avec contrainte). Le principe de cette méthode est de classer les contraintes dans un vecteur $Cont$ de N_s^2 coefficients. Soit $\Lambda = (\Lambda_I)$ le vecteur ligne de coefficient de Lagrange de dimension $N_s^2 \times 1$ dont on fait correspondre la matrice $\Gamma = (\Gamma_{ij})$. Le critère généralisé de Lagrange devient alors la minimisation de la fonction :

$$f(G) = C(G) - \Lambda Cont \tag{C.12}$$

La minimisation sous contrainte, revient à minimiser la fonction $f(G)$. La minimisation de $f(G)$ par rapport à Λ assure la contrainte en question, ce qui n'est pas le cas de la minimisation par rapport à G. En effet, la dérivé de $f(G)$ par rapport à G est égale à :

$$\frac{\partial f(G)}{\partial G} = \frac{\partial C(G)}{\partial G} - \frac{\partial(\Lambda Cont)}{\partial G} \tag{C.13}$$

La première partie, de cette égalité a déjà été calculé. Il nous reste à calculer la deuxième partie. Alors, on montre que la relation s'écrit, en utilisant toujours la notation d'Einstein, que :

$$\Lambda Cont = \Lambda_I Cont_I = \Lambda_{i+N_s(j-1)} Cont_{i+N_s(j-1)} \tag{C.14}$$

$$= \Lambda_{i+N_s(j-1)}(g_{il} R_{lk} g_{jk} - {\delta_{ij}}^2) \tag{C.15}$$

La contrainte dépend uniquement de G_0, alors :

$$\frac{\partial \Lambda Cont}{\partial G_0} = (\Gamma_{pj} g_{jk} R_{kq} + \Gamma_{ip} g_{il} R_{lq}) \tag{C.16}$$

Car la contrainte étant symétrique par définition, la matrice Γ l'est aussi. On trouve finalement, que :

$$\frac{\partial \Lambda Cont}{\partial G} = \begin{pmatrix} 2\Gamma G_0 R \\ 0_{N_s} \\ \vdots \\ 0_{N_s} \end{pmatrix}. \tag{C.17}$$

Le système final revient à résoudre, (voir l'équation (C.3)) :

$$\begin{aligned} \begin{pmatrix} G_0 \\ 2G_1 \\ \vdots \\ 2G_{(M+N-1)} \\ G_{(M+N)} \end{pmatrix} R_{Y_N}(0) &= \begin{pmatrix} G_1 \\ G_1 \\ \vdots \\ G_{(M+N)} \\ 0_{N_s} \end{pmatrix} R_{Y_N}^T(1) \\ &+ \begin{pmatrix} 0_{N_s} \\ G_0 \\ \vdots \\ G_{(M+N-2)} \\ G_{(M+N-1)} \end{pmatrix} R_{Y_N}(1) + \begin{pmatrix} \mu G_0 R_{Y_N}(0) \\ 0_{N_s} \\ \vdots \\ 0_{N_s} \\ 0_{N_s} \end{pmatrix} \end{aligned}$$

Notons $R_{Y_N}(0) = R$ et $R_{Y_N}(1) = Q$, alors le système devient :

$$\begin{aligned} G_0 R &= G_1 Q^T + \mu G_0 R \\ 2G_i R &= G_{(i+1)} Q^T + G_{(i-1)} Q \\ G_{(M+N)} R &= G_{(M+N-1)} Q \end{aligned} \tag{C.18}$$

Si les deux matrices R et Q sont inversibles alors le système admet une solution évidente.
Dans notre cas, ces deux matrices ne sont pas inversibles, car le nombre de capteurs est plus grand que celui des sources. Cette approche est une approche par bloc. Notre but est d'appliquer la méthode des sous-espaces avec des algorithmes adaptatifs, on s'arrête ici.

[2] δ_{ij} est le symbole de kronecker, donc il est égale à un si seulement si $i = j$, sinon il est nul

C.5 Principe et l'algorithme du Gradient Conjugué

C.5.1 Méthode de descente

A chaque itération d'une méthode de descente ([25], [77], [78], ...), on détermine un vecteur p_k et un scalaire α_k, ce qui permet de calculer le vecteur x_{k+1}, qu'on cherche à estimer :

$$x_{k+1} = x_k + \alpha_k p_k. \tag{C.19}$$

Soit A une matrice symétrique. Alors, on trouve une équivalence entre la solution de $Ax = b$ et :

$$\min_x J(x) = (Ax|x) - 2(b|x), \tag{C.20}$$

où $(Ax|b)$ est le produit scalaire $b^T Ax$, et $J(x)$ est une fonctionnelle quadratique définie positive.
Posons $e(x)$ l'erreur d'estimation $e(x) = x - \bar{x}$. Il est encore équivalent de minimiser $J(x)$ ou de minimiser $E(x)$, avec :

$$E(x) = (A(x - \bar{x})|(x - \bar{x}) = (Ae(x)|e(x)). \tag{C.21}$$

En notant par $r(x)$ le résidu, on aura :

$$\begin{aligned} r(x) &= -\frac{1}{2}\vec{\nabla}J(x) \\ &= A\bar{x} - Ax = Ae(x), \end{aligned}$$

où $\vec{\nabla}J(x)$ est le vecteur gradient de $J(x)$. On peut exprimer la relation (C.21) par :

$$E(x) = (r(x)|A^{-1}r(x)). \tag{C.22}$$

Une fois qu'on a défini notre fonctionnelle (C.22), on cherche à minimiser cette fonctionnelle pour bien déterminer α_k et p_k (voir la relation (C.19)).

C.5.2 Choix de α_k

Tout d'abord, on choisit α_k telle que :

$$E(x_k + \alpha_k p_k) = \min_{\alpha \in \mathbb{R}} E(x_k + \alpha p_k) \tag{C.23}$$

Tout calcul fait, on trouve :

$$\alpha_k = \frac{(r_k|p_k)}{(Ap_k|p_k)}, \tag{C.24}$$

$\forall p_k \neq 0$ et pour α_k optimal, on a les deux relations suivantes, $\forall k \geq 0$:

$$\begin{aligned} r_{(k+1)} &= r_k - \alpha_k A p_k & \text{(C.25)} \\ (p_k | r_{(k+1)}) &= 0. & \text{(C.26)} \end{aligned}$$

Une fois, on a choisi notre α_k optimal, il nous reste à choisir la valeur de p_k (voir le relation (C.19)).

C.5.3 Choix de p_k

Pour choisir la valeur de p_k, il faut remarquer que la relation (C.22) peut s'écrire sous la forme :

$$E(x_{k+1}) = E(x_k)(1 - \gamma_k) \qquad \text{(C.27)}$$

avec :

$$\begin{aligned} \gamma_k &= \frac{(r_k | p_k)^2}{(A p_k | p_k)(A^{-1} r_k | r_k)} \\ &\geq \frac{1}{K(A)} (\frac{r_k}{\|r_k\|} | \frac{p_k}{\|p_k\|})^2, & \text{(C.28)} \end{aligned}$$

où $K(A) = \frac{\lambda_1}{\lambda_N}$ avec λ_1 (respectivement λ_N) est la plus grande (respectivement la plus petite) valeur propre de A.
Pour bien converger, il faut choisir p_k de façon que $\gamma_k \geq \mu > 0$, avec μ une valeur qui ne dépend pas de k.

C.5.4 Gradient conjugué

Dans une méthode de gradient conjugué [25], on va chercher p_k dans le plan formé par les deux directions orthogonales r_k et $p_{(k-1)}$ (voir la relation (C.26)). Pour cela, on suppose que :

$$p_k = r_k + \beta_k p_{(k-1)} \qquad \text{(C.29)}$$

β_k est choisi telle que γ_k (voir la relation (C.28) soit maximale :

$$\max_{\beta_k} \rightarrow \beta_k = -\frac{(A p_{(k+1)} | r_k)}{(A p_{(k-1)} | p_{(k-1)})}. \qquad \text{(C.30)}$$

Alors, on montre que, $\forall k \geq 0$:

$$\begin{aligned} (r_{(k+1)} | r_k) &= 0 & \text{(C.31)} \\ \beta_k &= \frac{\|r_k\|^2}{\|r_{(k-1)}\|^2} & \text{(C.32)} \end{aligned}$$

C.5.5 L'algorithme du gradient conjugué

En choisissant les deux valeurs initiales x_0 et $p_0 = r_0 = b - Ax_0$ (voir la relation (C.20), on peut résumer l'algorithme [78] par :

$$\left\{\begin{array}{rcl} \alpha_k & = & \frac{\|r_k\|^2}{(Ap_k|p_k)} \\ x_{k+1} & = & x_k + \alpha_k p_k \\ r_{(k+1)} & = & r_k - \alpha_k A p_k \\ \beta_{(k+1)} & = & \frac{\|r_{(k+1)}\|^2}{\|r_k\|^2} \\ p_{(k+1)} & = & r_{(k+1)} + \beta_{(k+1)} p_k \end{array}\right.$$

Cette méthode est alors l'une des mieux adaptées à la relation des systèmes linéaires dont la matrice est symétrique définie positive.

C.6 Gradient conjugué généralisé

C.6.1 Description

Cet algorithme est une version modifié de celui du gradient conjugué classique. Les points majeurs de différence entre cet algorithme et la version classique sont :

- on cherche à minimiser le quotient de **Rayleigh généralisé**

$$f(X) = \frac{(AX|X)}{(BX|X)} = \frac{X^T AX}{X^T BX}, \tag{C.33}$$

  au lieu de $f(X) = (AX|X) - 2(b|X)$ dans la version classique.

- le vecteur X_k sera ajusté par

$$X_{(k+1)} = X_k + \alpha_k D_k, \tag{C.34}$$

  et le vecteur D_k est donné par :

  - Si k est un multiple de la dimension du vecteur X, $D_k = R_k$ comme un algorithme de gradient classique.
  - Sinon $D_{(k+1)} = R_{(k+1)} + \beta_k D_k$ et $\beta_k = -\frac{(AD_k|R_{(k+1)})}{(AD_k|D_k)} = -\frac{D_k^T A R_{(k+1)}}{D_k^T A D_k}$

- Puisque la minimisation est faite avec une contrainte $X^T BX = 1$, alors on normalise, à chaque itération X_k par :

$$X_{(k+1)} = \frac{X_{(k+1)}}{\sqrt{X_{(k+1)}^T B X_{(k+1)}}} \tag{C.35}$$

C.6.2 Calcul du pas optimal

Pour éviter de refaire le calcul dans le cadre d'une fonction déduite du quotient de Rayleigh (C.33) et qui sera utile dans notre application, on propose de faire le calcul de α optimal pour la fonction suivante :

$$f(X) = \frac{(AX|X) - 2(C|X)}{(BX|X)} \tag{C.36}$$

Le vecteur gradient de cette fonction est :

$$\vec{\nabla} f(X) = \frac{2(AX - C - f(X)BX)}{(BX|X)} \tag{C.37}$$

La valeur de α, voir (C.34) est calculée par :

$$f(X_{(k+1)}) = f(X_k + \alpha_k D_k) = \min_{\alpha \in \mathbb{R}} f(X_k + \alpha D_k) \tag{C.38}$$

Dans la suite et pour alléger l'écriture, on notera X_k par X et D_k par D. D'après (C.36), on trouve que :

$$f(X + \alpha D) = \frac{(AX|X) - 2(C|X) + 2\alpha[(AX|D) - (C|D)] + \alpha^2(AD|D)}{(BX|X) + 2(C|X) + 2\alpha(BX|D) + \alpha^2(BD|D)} \tag{C.39}$$

En annulant la dérivée première par rapport à α dans (C.39) , on trouve :

$$\begin{aligned}&[(Ax|D) - (C|D) + \alpha(AD|D)][(BX|X) + 2\alpha(BX|D) + \alpha^2(BD|D)] \\ &\quad -[(BX|D) + \alpha(BD|D)][(AX|X) - 2(C|X) + 2\alpha[(AX|D) - (C|D)] + \alpha^2(AD|D)] = 0\end{aligned}$$

Enfin, on trouve que la dernière équation revient à :

$$\begin{aligned}&\alpha^2\{(AD|D)(BX|D) - (BD|D)[(AX|D) - (C|D)]\} \\ &\quad +\alpha\{(AD|D)(BX|X) - (BD|D)[(AX|X) - 2(C|X)]\} \\ &\quad +(AX|D)(BX|X) - (C|D)(BX|X) - (BX|D)[(AX|X) - 2(C|X)] = 0\end{aligned}$$

En posant $m_1 = (AD|D)$, $m_2 = (BX|D)$, $m_3 = (BD|D)$, $m_4 = (AX|D) - (C|D)$ et en supposant que $(BX|X) = 1$ (normalisation du vecteur X), alors on trouve que :

$$\alpha^2(m_1 m_2 - m_3 m_4) + \alpha(m_1 - m_3 f(X)) + m_4 - m_2 f(X) = 0 \tag{C.40}$$

Finalement, on note $a = m_1 m_2 - m_3 m_4$, $b = m_1 - m_3 f(X)$, $c = m_4 - m_2 f(X)$ et l la valeur optimale de α, alors :

$$\alpha = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{C.41}$$

C.6.3 Algorithme

L'algorithme du gradient conjugué, donné par [24], est :

1. Initialisation
   - $X(0) = X0$, une valeur initiale du vecteur X de dimension $d \times 1$.
   - $X(0) = \frac{X(0)}{\sqrt{X(0)^+ BX(0)}}$, normalisation
   - $D(0) = f(X(0))BX(0) - AX(0)$, le résiduel ou le négatif du gradient
2. La partie recursive $t > 0$
   - $f(X(t)) = X(t)^+ AX(t)$
   - $m_1 = D(t)^+ AD(t)$
   - $m_2 = X(t)^+ BD(t)$
   - $m_3 = D(t)^+ BD(t)$
   - $m_4 = X(t)^+ AD(t)$
   - $a = m_1 m_2 - m_3 m_4$
   - $b = m_1 - f(X(t))m_3 - 2j\text{Imag}\{m_4 m_2^*\}$
   - $c = m_4^* - f(X(t))m_2^*$
   - $\alpha(t) = \frac{-b+\sqrt{b^2-4ac}}{2a}$
   - $X(t+1) = X(t) + \alpha(t)D(t)$
   - $X(t+1) = \frac{X(t+1)}{\sqrt{X(t+1)^+ BX(t+1)}}$
   - $\beta(t) = -\frac{D(t)^+ AR(t+1)}{D(t)^+ AD(t)}$
   - $R(t) = -AX(t+1) + BX(t+1)f(X(t+1))$

$$\begin{cases} \text{si } t \equiv 0(\text{mod } d) & D(t+1) = R(t) \\ \text{sinon} & D(t+1) = R(t) + \beta(t)D(t) \end{cases}$$

C.7 Transformation de la fonction coût

D'après la relation (C.53) et l'annexe C.8 (propriétés du produit de Kronecker), la fonction $C(G)$, voir (6.8), devient :

$$C(G) = E\|[Y_N^T(n) \bigotimes (I_{(M+N)N_s} \;\; 0_{N_s}) - Y_N^T(n+1) \bigotimes (0_{N_s} \;\; I_{(M+N)N_s})]\ \text{Col}(G)\|^2 \tag{C.42}$$

Notons $\mathrm{Col}(G) = \nu$, on aura :

$$\begin{aligned} C(\nu) &= \nu^T E[Y_N^T(n) \bigotimes (I_{(M+N)N_s} \;\; 0_{N_s}) - Y_N^T(n+1) \bigotimes (0_{N_s} \;\; I_{(M+N)N_s})]^T \\ & \quad [Y_N^T(n) \bigotimes (I_{(M+N)N_s} \;\; 0_{N_s}) - Y_N^T(n+1) \bigotimes (0_{N_s} \;\; I_{(M+N)N_s})]\nu. \quad \text{(C.43)} \end{aligned}$$

D'après les propriétés (C.50) et (C.48), on déduit :

$$\begin{aligned} C(\nu) &= \nu^T E[Y_N(n) \bigotimes \begin{pmatrix} I_{(M+N)N_s} \\ 0_{N_s}^T \end{pmatrix} - Y_N(n+1) \bigotimes \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix}] \\ & \quad [Y_N^T(n) \bigotimes (I_{(M+N)N_s} \;\; 0_{N_s}) - Y_N^T(n+1) \bigotimes (0_{N_s} \;\; I_{(M+N)N_s})]\nu. \\ C(\nu) &= \nu^T E[Y_N(n) Y_N^T(n) \bigotimes \begin{pmatrix} I_{(M+N)N_s} \\ 0_{N_s}^T \end{pmatrix} (I_{(M+N)N_s} \;\; 0_{N_s}) \\ & \quad - Y_N(n) Y_N^T(n+1) \bigotimes \begin{pmatrix} I_{(M+N)N_s} \\ 0_{N_s}^T \end{pmatrix} (0_{N_s} \;\; I_{(M+N)N_s}) \\ & \quad - Y_N(n+1) Y_N^T(n) \bigotimes \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix} (I_{(M+N)N_s} \;\; 0_{N_s}) \\ & \quad + Y_N(n+1) Y_N^T(n+1) \bigotimes \begin{pmatrix} 0_{N_s}^T \\ I_{(M+N)N_s} \end{pmatrix} (0_{N_s} \;\; I_{(M+N)N_s})]\nu \quad \text{(C.44)} \end{aligned}$$

En utilisant les notations de $R_{Y_N}(0)$ et $R_{Y_N}(1)$, tout en supposant que les signaux soient stationnaires, on trouve que :

$$\begin{aligned} C(\nu) &= \nu^T \{R_{Y_N}(0) \bigotimes \begin{pmatrix} I_{(M+N)N_s} & 0 \\ 0_{N_s}^T & 0 \end{pmatrix} - R_{Y_N}(1) \bigotimes \begin{pmatrix} 0 & I_{(M+N)N_s} \\ 0 & 0 \end{pmatrix} \\ & \quad - R_{Y_N}^T(1) \bigotimes \begin{pmatrix} 0 & 0 \\ I_{(M+N)N_s} & 0 \end{pmatrix} + R_{Y_N}(0) \bigotimes \begin{pmatrix} 0 & 0 \\ 0 & I_{(M+N)N_s} \end{pmatrix}\} \nu \\ C(\nu) &= \nu^T \{R_{Y_N}(0) \bigotimes \begin{pmatrix} I_{N_s} & 0 & 0 \\ 0 & 2I_{(M+N-1)N_s} & 0 \\ 0 & 0 & I_{N_s} \end{pmatrix} - R_{Y_N}(1) \bigotimes \begin{pmatrix} 0 & I_{(M+N)N_s} \\ 0 & 0 \end{pmatrix} \\ & \quad - R_{Y_N}^T(1) \bigotimes \begin{pmatrix} 0 & 0 \\ I_{(M+N)N_s} & 0 \end{pmatrix}\} \nu \quad \text{(C.45)} \end{aligned}$$

C.8 Produit de Kronecker

Le produit de Kronecker de deux matrices A de dimension $m \times n$ et B de dimension $p \times q$ est une matrice C de dimension $mp \times nq$, noter $A \bigotimes B$. C est formé par mn matrices C_{ij} de dimension $N_s \times N_c$, tel que :

$$C_{ij} = a_{ij} B \quad \forall\, 1 \leq i \leq m \;;\; 1 \leq j \leq n \qquad \text{(C.46)}$$

le produit de Kronecker a certaines propriétés :

$$(A+B)\bigotimes(C+D) = A\bigotimes C + A\bigotimes D + B\bigotimes C + B\bigotimes D \quad \text{(C.47)}$$

$$(A\bigotimes b)(C\bigotimes D) = AC\bigotimes BD \quad \text{(C.48)}$$

$$(A\bigotimes B)C = A\bigotimes(b\bigotimes C) = A\bigotimes B\bigotimes C \quad \text{(C.49)}$$

$$(A\bigotimes B)^T = A^T\bigotimes B^T \quad \text{(C.50)}$$

$$(A\bigotimes B)^{-1} = A^{-1}\bigotimes B^{-1} \text{ si } A^{-1} \text{ et} B^{-1}\text{existes} \quad \text{(C.51)}$$

$$\text{rang}(A\bigotimes B) = \text{ rang } (A).\ \text{rang } (B). \quad \text{(C.52)}$$

Enfin, il a aussi une propriété très importante :

$$AVB = C \Longleftrightarrow (B^T\bigotimes A)\ \text{Col}(V) = \ \text{Col}(C). \quad \text{(C.53)}$$

où Col(V) est un grand vecteur formé par les colonnes de V, mis bout à bout. Si V est de dimension $n \times p$ alors Col(V) est de dimension $np \times 1$.

C.9 Critère dans le domaine fréquentiel

Pour une simplification de notation, on développe ce calcul dans le cas de deux sources $N_s = 2$. La généralisation pour $N_s > 2$ est évidente.
Si $H(z)$ a de différents dégrées de colonnes, alors :

$$H(z) = (H_1(z), H_2(z)), \quad \text{(C.54)}$$

avec deg$H_1(z) = M_1$ et deg$H_2(z) = M_2$. On sait que le vecteur d'observation est noté par :

$$Y_N(n) = U_N(H)(S^T_{1,M_1+N}(n), s^T_{2,M_2+N}(n))^T. \quad \text{(C.55)}$$

Notons par G une inverse à gauche de $U_N(H)$, G est une matrice de dimension $\sum_{i=1}^{p} M_i + p(N+1) \times N_c(N+1)$, tel que :

$$G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}. \quad \text{(C.56)}$$

où G_i est la matrice qui correspond à l'inverse de $T_N(H_i)$ (voir section 6.5). Notons par $G_{i,k}$ la $k$ième ligne de G_i.
Comme dans $C1$, les $(M_i + N)$ premières lignes de $G_iY_N(n)$ sont identiques aux $(M_i + N)$ dernières lignes de $G_iY_N(n+1)$. Donc on a :

$$G_{i,k}Y_N(n) = G_{i,k+1}Y_N(n+1). \quad \text{(C.57)}$$

Dans la section C.1, on montre l'existence d'un filtre $G(z)$, égal à l'inverse à gauche de $H(z)$ (bien sûr $H(z)$ satisfait les hypothèses nécessaires), avec $\deg G(z) \leq N$. Posons $G_{i,k}(z)$ le polynôme correspondant à la matrice $G_{i,k}$. D'après la relation (C.57), on trouve que :

$$[G_{i,k}(z)]Y(n) = [G_{i,k+1}(z)]Y(n+1) \quad \text{(C.58)}$$

$$G_{i,k+1}(z)Y(z) = z^{-1}G_{i,k}(z)Y(z) \quad \text{(C.59)}$$

$$G_{i,k+1}(z)H(z)S(z) = z^{-1}G_{i,k}(z)H(z)S(z) \quad \text{(C.60)}$$

Pour les mêmes raisons, déjà utilisées dans la section 6.5 sur les sources, on montre que la dernière relation revient à :

$$G_{i,k+1}(z)H(z) = z^{-1}G_{i,k}(z)H(z) \quad \text{(C.61)}$$

donc la dernière ligne est :

$$G_{i,M_i+N}(z)H(z) = z^{-M_i-N}G_{i,0}(z)H(z) \quad \text{(C.62)}$$

D'après la relation (C.54), on trouve que :

$$G_{i,M_i+N}(z)H_j(z) = z^{-M_i-N}G_{i,0}(z)H_j(z) \quad \text{(C.63)}$$

Dans cette dernière équation, le degré du premier terme est :

$$\deg(G_{i,M_i+N}(z)H_j(z)) \leq N + M_j \quad \text{(C.64)}$$

$$\deg(z^{-M_i-N}G_{i,0}(z)H_j(z)) \geq M_i + N \quad \text{(C.65)}$$

Les trois relations (C.63), (C.64) et (C.65) nous montrent que :

- si $M_j = M_i$ alors $G_{i,0}(z)H_j(z)$ est une constante α_{ij}.
- si $M_j < M_i$ alors $G_{i,0}(z)H_j(z)$ est un polynôme identiquement nul.
- si $M_j > M_i$ alors $G_{i,0}(z)H_j(z)$ est un polynôme $r_{ij}(z)$ de degré $M_j - M_i$.

Finallement, on trouve que le filtre resultant du produit de $G(z)$ par $H(z)$ et relatif aux sources sera :

$$\begin{pmatrix} G_{1,0}(z)H_1(z) & G_{1,0}(z)H_2(z) \\ G_{2,0}(z)H_1(z) & G_{2,0}(z)H_2(z) \end{pmatrix} = \begin{pmatrix} \alpha_{11} & r_{12}(z) \\ 0 & \alpha_{22} \end{pmatrix} \quad \text{(C.66)}$$

Cette équation montre bien que la séparation sur la deuxième voie est immédiate, tandis que sur la première voie on a un résidu de la deuxième source filtré par $r_{12}(z)$. La généralisation est immédiate à partir des relations (C.63), (C.64) et (C.65).

Bibliographie

[1] K. Abed Meraim, P. Loubaton, and E. Moulines. A subspace algorithm for certain blind identification problems. *IEEE on IT*, January 97.

[2] D. Achavar. *Séparation de sources: généralisation à un modèle convolutif.* PhD thesis, Université de Montpellier II, décembre 1993.

[3] M.J. Al-Kindi and J. J. Dunlop. Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel. *Signal Processing*, 17:241 – 250.

[4] B. Ans, J. C. Gilhodes, and J. Hérault. Simulation de réseaux neuronaux (sirene). II. hypothèse de décodage du message de mouvement porté par les afférences fusoriales IA et II par un mécanisme de plasticité synaptique. *C. R. Acad; Sci. Paris*, série III:419 – 422, 1983.

[5] Y. Barness, J. Carlin, and M. Steinberger. Bootstrapping adaptive cross pol cancelers for satellite communications. In *International Conference on Communication*, Philadelphia, Pennsylvania, June 1982.

[6] A. Belouchrani. *Séparation autodidacte de sources: Algoritmes, performances et application à des signaux expérimentaux.* PhD thesis, ENST Paris, 1995.

[7] A. Belouchrani and K. Abed-Meraim. Séparation aveugle au second ordre de sources corrélées. In *GRETSI*, pages 309–312, Juan-Les-Pins, France, September 1993.

[8] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines. Second-order blind separation of correlated sources. In *Int. Conf. on Digital Sig.*, pages 346–351, Nicosia, Cyprus, july 1993.

[9] A. Belouchrani and J. F. Cardoso. Maximum likelihood source separation for discrete sources. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications*, pages 768 – 771, Edinburgh, Scotland, September 1994. Elsevier.

[10] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations.* Spring-Verlag, 1990.

[11] L. Brillouin. *Les tenseurs en mécanique et en élasticité.* Jacques Gabay, 1987.

[12] G. Burel. Blind separation of sources: a nonlinear neural algorithme. *Neural Networks*, 5:937–947, 1992.

[13] V. Capdevielle. Séparation de sources sinusoidales de frequences proches. In *GRETSI*, pages 337–340, Juan-Les-Pins, France, Septembre 1993.

[14] V. Capdevielle. *Séparation de sources large bande a l'aide des moments d'ordre supérieur.* PhD thesis, INP Grenoble, 1995.

[15] V. Capdevielle, C. Seviére, and J. L. Lacoume. Separation of wide band sources. In *HOS 95*, pages 66–70, Girona - Spain, 12-14 June 1995.

[16] J. F. Cardoso. Blind identification of independent signals. In *Workshop on Higher-Order Spectral Analysis*, Vail (CO), USA, June 1989.

[17] J. F. Cardoso. Source separation using higher order moments. In *Proceeding of ICASSP*, pages 2109–2212, Glasgow, Scotland, May 1989.

[18] J. F. Cardoso, A. Belouchrani, and B. Laheld. A new composite criterion for adaptive and iterative blind source separation. In *Proc. Icassp*, Australia, 1994.

[19] J. F. Cardoso and P. Comon. Tensor-based independent component analysis. In L. Torres, E. Masgrau, and M. A. Lagunas, editors, *Signal Processing V, Theories and Applications*, pages 673–676, Barcelona, Espain, 1990. Elsevier.

[20] J. F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Trans on Signal Processing*, 44(12), December 1996.

[21] N. Charkani. *Séparation auto-adaptative de sources pour les mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures.* PhD thesis, INP Grenoble, Novembre 1996.

[22] N. Charkani and J. Hérault. On the performances of the fourth-order cross cumulants in blind separation of sources. In *HOS 95*, pages 86–90, Girona - Spain, 12-14 June 1995.

[23] E. Chaumette, P. Common, and D. Muller. Application of ica to airport surveillance. In *HOS 93*, pages 210–214, South Lake Tahoe - California, 7-9 June 1993.

[24] H. Chen, T. K. Sarkar, S. A. Dianat, and J. D. Brule. Adaptive spectral estimation by the conjugate gradient method. *IEEE Trans. on Acoustics. Speech and Signal Processing*, ASSP - 34(2):272 – 284, April 1986.

[25] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation.* Masson, 1990.

[26] M. H. Cohen and A. G. Andreou. Current-mode subthreshold mos implementation of the Hérault-Jutten autoadaptative network. *IEEE Signal Processing Magazine*, 27(5), May 1992.

[27] P. Comon. Séparation de mélanges de signaux. In *GRETSI*, pages 137–140, Juan-Les-Pins, France, June 1989.

[28] P. Comon. Separation of sources using higher-order cumulants. In *SPIE Vol. 1152 Advanced Algorithms and Architectures for Signal Processing IV*, San Diego (CA), USA, August 8-10, 1989.

[29] P. Comon. Separation of stochastic processes whose a linear mixture is observed. In *Workshop on Higher-Order Spectral Analysis*, pages 174–179, Vail (CO), USA, June 1989.

[30] P. Comon. Analyse en composantes indépendantes et identification aveugle. *Traitement du signal*, 7(5):435–450, Dec 1990.

[31] P. Comon. Blind identification in presence of noise. In *EUSIPCO*, Brussels, Belgium, August 1992.

[32] P. Comon. Remarque sur la diagonilisation tensorielle par la méthode de jacobi. In *GRETSI*, pages 125–128, Juan-Les-Pins, France, Septembre 1993.

[33] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287 – 314, April 1994.

[34] P. Comon. Quelques developpements recents en traitement du signal. Technical report, Univ. Nice Sophia-Antipolis, 18 september 1995. Habilitation de diriger des recherches.

[35] P. Comon, C. Jutten, and J. Hérault. Blind separation of sources, Part II: Statement problem. *Signal Processing*, 24(1):11–20, November 1991.

[36] P. Comon and Lacoume J. L. Statistiques d'ordres supérieurs pour le traitement du signal. In *Traitement du signal - Developpements Recents*, Les Houches, France, Août - Septembre 1993.

[37] L. De Lathauwer, D. Callaerts, B. De Moor, and J. Vandewalle. Separation of wide band sources. In *HOS 95*, pages 134–138, Girona - Spain, 12-14 June 1995.

[38] N. Delfosse. *Séparation aveugle des sources : méthodes de type sous - espace.* PhD thesis, ENST Paris, Décembre 1995.

[39] N. Delfosse. Séparation aveugle des sources : méthodes de type sous-espace. Technical report, Telecom paris, June 1995.

[40] N. Delfosse and P. Loubaton. Séparation adaptative de sources indépendantes par une approche de déflation. In *GRETSI*, pages 309–312, Juan-Les-Pins, France, Septembre 1993.

[41] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 45(1):59–83, July 1995.

[42] N. Delfosse and P. Loubaton. Séparation aveugle adaptative de mélanges convolutifs. In *Actes du XVeme colloque GRETSI*, pages 281 – 284, Juan-Les-Pins, France, 18 - 21 september 1995.

[43] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat; Soc. B*, 39:1–38, 1977.

[44] G. Desodt and D. Muller. Complex independent components analysis applied to the separation of radar signals. In L. Torres, E. Masgrau, and M. A. Lagunas, editors, *Signal Processing V, Theories and Applications*, pages 665–668, Barcelona, Espain, 1994. Elsevier.

[45] G. D'urso and L. Cai. Sources separation method applied to reactor monitoring. In *Proc. Workshop Athos working group*, Girona, Spain, June 1995.

[46] L. Féty. *Méthodes de traitement d'antenne adaptées aux radiocommunications.* PhD thesis, ENST Paris, 1988.

[47] J. C. Fort. Stabilité de l'algorithme de séparation de sources de Jutten et Hérault. *Traitement du signal*, 8(1):35–42, 1991.

[48] Z. Fu and E. M. Dowling. Conjugate gradient eigenstructure tracking for adaptive spectral estimation. *IEEE Trans. on Signal Processing*, 43(5):1151 – 1160, May 1995.

[49] M. Gaeta. *Les statistiques d'ordre supérieur appliquées à la séparation de sources.* PhD thesis, CEPHAG - ENSIEG Grenoble, Juillet 1991.

[50] M. Gaeta and J. L. Lacoume. Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes. *Traitement du Signal*, 7(5):419–434, 1990.

[51] M. Gaeta and J. L. Lacoume. Sources separation without a priori knowledge: the maximum likelihood solution. In L. Torres, E. Masgrau, and M. A. Lagunas, editors, *Signal Processing V, Theories and Applications*, pages 621–624, Barcelona, Espain, 1994. Elsevier.

[52] P. Garat. *Approche statistique pour la Séparation aveugle de sources.* PhD thesis, INP Grenoble, Décembre 1994.

[53] S. Van Gerven. *Adaptive Noise Cancellation and Signal Separation with Applications to Speech Enhancement.* PhD thesis, K.U.Leuven, E.S.A.T., Belgium, March 1996.

[54] D. Gesbert, P. Duhamel, and S. Mayrargue. Subspace-based adaptive algorithms for the blind equalization of multichannel fir filters. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications*, pages 712–715, Edinburgh, Scotland, September 1994. Elsevier.

[55] G. H. Golub and C. F. Van Loan. *Matrix computations.* The johns hopkins press- London, 1984.

[56] A. Gorokhov and P. Loubaton. Second order blind identification of convolutive mixtures with temporally correlated sources: A subspace based approch. In *Signal Processing VIII, Theories and Applications*, pages 2093 – 2096, Triest, Italy, September 1996. Elsevier.

[57] A. Gorokhov and P. Loubaton. Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources. *Submitted to IEEE Trans. on Circuits and Systems*, 95.

[58] F. Harroy. *Séparation de sources non gaussiennes: performance et robustesse statistique.* PhD thesis, INP Grenoble, septembre 1994.

[59] F. Harroy, J. L. Lacoume, and M. A. Lagunas. A general adaptive algorithm for non-gaussian source separation without any constraint. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications*, pages 1161–1164, Edinburgh, Scotland, September 1994. Elsevier.

[60] J. Hérault and B. Ans. Réseaux de neurones à synapses modifiables : décodage de messages sensoriels composites par une apprentissage non supervisé et permanent. *C. R. Acad; Sci. Paris*, série III:525–528, 1984.

[61] 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*, pages 1017–1022, Nice, France, 20-24, Mai 1985.

[62] Bell A. J. and Sejnowski T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, November 1995.

[63] C. Jutten. *Calcul neuro-mimétique et traitement du signal: Analyse en composantes indépendantes.* PhD thesis, UJF-INP Grenoble, 1987.

[64] C. Jutten and J. Hérault. Une solution neuromimétique du problème de séparation de sources. *Traitement du signal*, 5(6):389–403, 1988.

[65] C. Jutten and J. Hérault. Analog implementation of a neuromimetic auto-adaptive algorithm. In J. Hérault and F. Fogelman-Soulié, editors, *Neuro-Computing: Algorithms, architectures and applications*, pages 145–152, Les Arcs, France, Feb. 27-March 3 1989. Springer Verlag, NATO ASI Series, Series F, Vol 68.

[66] C. Jutten and J. Hérault. Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[67] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, and Caelen J. Blind separation of sources: An algorithm for separation of convolutive mixtures. In *International Signal Processing Workshop on Higher Order Statistics*, pages 273–276, Chamrousse, France, July 1991.

[68] C. Jutten, L. Nguyen Thi, and J. Hérault. Nouveaux algorithmes de séparation de sources. In *Con. satellite du cong. Europée de mathématiques: Aspects théoriques des reseaux de neurones*, Paris, France, July 1992.

[69] T. Kailath. *Linear systems.* Prentice Hall, 1980.

[70] M. Kendall and A. Stuart. *The advanced theory of statistics.* Charles Griffin & Company Limited, 1991. volume 1.

[71] M. Krob and M. Benidir. Fonction de contraste pour l'identification aveugle d'un modéle linéaire quadratique. In *GRETSI*, pages 101–104, Juan-Les-Pins, France, Septembre 1993.

[72] J. L. Lacoume and F. Harroy. Performances in blind source separation. In *HOS 95*, pages 25–29, Girona - Spain, 12-14 June 1995.

[73] J. L. Lacoume and P. Ruiz. Sources identification: A solution based on cumulants. In *IEEE ASSP WorkshopV*, Mineapolis, USA, August 1988.

[74] B. Laheld. *Séparation auto-adaptative de sources. Implantations et performances.* PhD thesis, ENST Paris, 1994.

[75] B. Laheld and J. F. Cardoso. Séparation adaptative de sources en aveugle. implantation complexe sans contraintes. In *Actes du XIVeme colloque GRETSI*, pages 329–332, Juan-Les-Pins, France, 13-16 september 1993.

[76] B. Laheld and J. F. Cardoso. Adaptive source separation without pre-whitening. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications*, pages 183–186, Edinburgh, Scotland, September 1994. Elsevier.

[77] P. Lascaux and R. Théodor. *Analyse numérique matricielle appliquée à l'art de l'ingénieur.* Masson, 1994. Tome 1.

[78] P. Lascaux and R. Théodor. *Analyse numérique matricielle appliquée à l'art de l'ingénieur.* Masson, 1994. Tome 2.

[79] C. Latombe. *Détection et caractérisation des signaux à plusieurs composantes à partir de la matrice interspectrale.* PhD thesis, INPG Grenoble, 1983.

[80] M. Lavielle, E. Moulines, and J. F. Cardoso. On a stochastic approximation version of the EM algorithm. In *HOS 95*, pages 61–63, Girona - Spain, 12-14 June 1995.

[81] O. Macchi and E. Moreau. Self-adaptive source separation, part I: convergence analysis of a direct linear neural network controled by the Hérault-jutten. To appear in IEEE Trans. on Siganl Processing.

[82] O. Macchi and E. Moreau. Self-adaptive source separation using correlated signals and cross-cumulants. In *Proc. Workshop Athos working group*, Girona, Spain, June 1995.

[83] Z. Malouche, O. Macchi, and E. Moreau. Séparation adaptative de sources binaires à l'aide d'un réseau de neurones non linéaires bouclés. In *Actes du XVeme colloque GRETSI*, pages 305–309, Juan-Les-Pins, France, 18 - 21 september 1995.

[84] A. Mansour and C. Jutten. Fourth order criteria for blind separation of sources. *IEEE Trans on SP*, 43(8):2022 – 2025, August 1995.

[85] A. Mansour and C. Jutten. A simple cost function for instantaneous and convolutive sources separation. In *Actes du XVeme colloque GRETSI*, pages 301–304, Juan-Les-Pins, France, 18-21 septembre 1995.

[86] A. Mansour and C. Jutten. A direct solution for blind separation of sources. *IEEE Trans on SP*, 44(3):746 – 748, March 96.

[87] A. Mansour, C. Jutten, and P. Loubaton. Subspace method for blind separation of sources and for a convolutive mixture model. In *Signal Processing VIII, Theories and Applications*, pages 2081 – 2084, Triest, Italy, September 1996. Elsevier.

[88] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.

[89] P. Mccullagh. *Tensor methods in statistics*. Chapman and Hall, 1987.

[90] J. M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *IEEE Proc*, 79:277–305, 1991.

[91] E. Moreau. *Apprentissage et adaptativité. Séparation auto-adaptative de sources indépendantes*. PhD thesis, Université de Paris-Sud, 1995.

[92] E. Moreau and O. Macchi. New self-adaptive algorithms for source separation based on contrast functions. In *IEEE Signal Processing Workshop on Higher-Order Statistics*, pages 215–219, South Lac Tahoe, USA (CA), June 1993.

[93] E. Moreau and O. Macchi. Séparation auto-adaptative de sources sans blanchiment prealable. In *Actes du XIVème colloque GRETSI*, pages 325–328, Juan-Les-Pins, France, 13-16 september 1993.

[94] E. Moulines, Duhamel P., J. F. Cardoso, and Mayrargue S. Subspace methods for the blind identification of multichannel FIR filters. *IEEE Trans on SP*, 43(2):516–525, February 1995.

[95] L. Nguyen Thi. *Séparation aveugle de sources à large bande dans un mélange convolutif*. PhD thesis, INP Grenoble, Janvier 1993.

[96] L. Nguyen Thi and C. Jutten. Blind sources separation for convolutive mixtures. *Signal Processing*, 45(2):209–229, 95.

[97] L. Nguyen Thi, C. Jutten, and J. Caelen. Speech enhancement: Analysis and comparison of methods in various real situations. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI, Theories and Applications*, pages 303–306, Brussels, Belgium, August 1992. Elsevier.

[98] C. L. Nikias and J. M. Mendel. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, 10(3):10–37, 1993.

[99] X. Oliva Galvan. Blind separation of sources: Some adaptative algorithms. Technical report, Lab CEPHAG in Grenoble, April 1993.

[100] P. Pajunen, A. Hyvarinen, and J. Karhunen. Nonlinear blind source separation by self-organizing maps. In *ICONIP 96*, volume 2, pages 1207–1210, Hong-Kong, September 1996.

[101] D. T. Pham. Séparation aveugle de sources via une analyse en composantes indépendantes. In *Actes du XVeme colloque GRETSI*, pages 289–292, Juan-Les-Pins, France, 18 - 21 september 1995.

[102] D. T. Pham and P. Garat. Séparation aveugle de sources temporellement corrélées. In *XIV Colloque GRETSI*, pages 317–320, Juan-Les-Pins, France, September 1993.

[103] D. T. Pham, P. Garat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI, Theories and Applications*, pages 771–774, Brussels, Belgium, August 1992. Elsevier.

[104] B. Picinbono. *Signaux aléatoires, Tome 1.* Dunod, 1993.

[105] B. Picinbono. *Signaux aléatoires, Tome 2.* Dunod, 1994.

[106] C. G. Puntonet. *Nuevos algoritmos de separación de fuentes en medios lineales.* PhD thesis, University of Granada, SPAIN, September 1994.

[107] C. G. Puntonet, A. Prieto, C. jutten, M. Rodriguez-Alvarez, and J. Ortega. Separation of sources: A geometry-based procedure for reconstruction of n-valued signals. *Signal Processing*, 46(3):267–284, 1995.

[108] C. G. Puntonet, M. Rodriguez, and A. Prieto. A geometrical based procedure for source separation mapped to a neural network. In *From natural to artificial neural computation*, pages 898–905, Torremolimos, Malaga , Spain, 7 - 9 June 1995.

[109] G. Puntonet, C., A. Mansour, and C. Jutten. Geometrical algorithm for blind separation of sources. In *Actes du XVeme colloque GRETSI*, pages 273 – 276, Juan-Les-Pins, France, 18 - 21 september 1995.

[110] O. Shalvi and E. Weistein. New criteria for blind deconvolution of non-minimum-phase systems (channels). *IEEE trans. on Information Theory*, 36(2):312–321, March 1990.

[111] E. Sorouchyari. Blind separation of sources, Part III: Stability analysis. *Signal Processing*, 24(1):21–29, 1991.

[112] A. Souloumiac. *Utilisation des statistiques d'ordre supérieur pour le filtrage et la séparation de sources en traitement d'antenne.* PhD thesis, ENST Paris, 1993.

[113] A. Souloumiac and J. F. Cardoso. Comparaison de méthodes de séparation de sources. In *GRETSI*, pages 661–664, Juan-Les-Pins, France, Septembre 1991.

[114] A. Stuart and J. Keith Ord. *Kendall's advanced theory of statistics.* Oxford University Press, 1991. volume 2.

[115] A. Taleb. Introduction à la séparation de mélanges non-linéaires. Technical report, LTIRF - INPG, Juillet- Août 1996. Rapport de fin d'étude.

[116] A. Taleb. Méthodes géométriques dans la séparation de sources. Technical report, LTIRF - INPG, Juin 1996. Rapport de DEA.

[117] N. Thirion. *Séparation d'ondes en prospection sismique*. PhD thesis, CEPHAG, INPG Grenoble, 1995.

[118] N. Thirion, J. Mars, and J. L. Lacoume. Séparation aveugle de signaux large bande: un nouveau challenge en prospection sismique. In *Actes du XVeme colloque GRETSI*, pages 1335–1338, Juan-Les-Pins, France, 18 - 21 september 1995.

[119] L. Tong, G. Xu, and T. Kailath. A new approach to blind identification and equalization of multipath channels. In *The 25th Asilomar Conference*, pages 856–860, Pacific Grove, 1991.

[120] S. Van Gerven and D. Van Compernolle. Feedforward and feedback in a symmetric adaptive noise canceller: Stability analysis in a simplified case. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI, Theories and Applications*, pages 1081–1084, Brussels, Belgium, August 1992. Elsevier.

[121] E. Vittoz and X. Arreguit. Cmos integration of Hérault-Jutten cells for separation of sources. Kluwer Academic Publishers, May 1989.

[122] J. F. Yang and M. Kaveh. Adaptive eigensubspace algorithms for direction or frequency estimation and tracking. *IEEE Trans. on Acoustics. Speech and Signal Processing*, ASSP - 36(2):241 – 251, February 1988.

[123] D. Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Trans. on Signal Processing*, 42(8):2158 – 2167, August 1994.