

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE  
À L'OBTENTION DE LA  
MAÎTRISE EN GÉNIE DE LA PRODUCTION AUTOMATISÉE  
M.Eng.

PAR  
CYRIL VIRON

ANALYSE EN COMPOSANTES INDÉPENDANTES POUR LA  
CARACTÉRISATION D'IMAGES HYPERSPECTRALES EN TÉLÉDÉTECTION

MONTRÉAL, LE 23 MAI 2006

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Jacques-André Landry, directeur de mémoire

Département de génie de la production automatisée à l'École de technologie supérieure

M. Ilian Bonev, président du jury

Département de génie de la production automatisée à l'École de technologie supérieure

M. Abdelhakim Artiba, examinateur

Département de génie de la production automatisée à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 10 MAI 2006

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

# **ANALYSE EN COMPOSANTES INDÉPENDANTES POUR LA CARACTÉRISATION D'IMAGES HYPERSPECTRALES EN TÉLÉDÉTECTION**

Cyril Viron

## **SOMMAIRE**

En réponse partielle aux problèmes écologiques actuels, l'imagerie hyperspectrale ambitionne de connaître la composition locale d'une parcelle agricole en recherchant sa signature spectrale car celle-ci caractérise de façon unique un élément. Cependant, cette signature s'avère être un mélange pondéré de celles des éléments s'y trouvant. Afin de récupérer leurs signatures à partir du mélange, l'analyse en composantes indépendantes (ACI) est légitimement utilisée ! Malgré le nombre restreint de travaux sur l'ACI en hyperspectral, devant sa popularité en traitement de signal, nous l'avons appliquée en utilisant l'algorithme FastICA, méthode la plus récente et efficace, d'abord sur des images et des signaux classiques (pour constater son efficacité), puis sur une base de signatures étalons. Le but est de comparer les composantes indépendantes à une base référencée pour former les paires les plus ressemblantes. Cependant, du fait des ambiguïtés et d'absence de critère de validation de l'ACI, il est impossible de prédire ni vérifier les paires. Pour y remédier, notre protocole expérimental est divisé entre comparaisons « théorique » et « pratique », basées sur des niveaux de confiance, permettant de former les paires considérées justes d'une part (base partielle) et expérimentales d'autre part (base totale) qui, comparées, déterminent le succès d'association. Les résultats, assujettis à deux seuils de confiance relatifs, sont excellents pour les signaux, bons pour les images mais globalement médiocres pour les signatures. La raison principale est un effet beaucoup plus visible en ce cas de la subjectivité de la prise de décision et de la décorrélation inévitable entraînant déformations et trop grande dépendance à la base. Cependant, les résultats deviennent très satisfaisants pour une sélection adéquate (cultures, arbres et minéraux). Pour tenter encore de les améliorer, des recommandations constructives ont été proposées, afin de poser le deuxième échelon de ce travail, qui se voulait novateur.

# **INDEPENDENT COMPONENT ANALYSIS FOR THE CHARACTERIZATION OF HYPERSPECTRAL IMAGES IN REMOTE SENSING**

Cyril Viron

## **ABSTRACT**

To address some current environmental problems, hyperspectral imaging is seen as a means of obtaining the local composition of an agricultural parcel. To this end, the extraction of spectral signatures is of interest as it allows the characterization of an element in a specific manner. However, the spectral signature obtained from a given parcel is in fact a weighted mixture of the various elements presents ; the individual signature of each element is then sought : independent component analysis (ICA) could be the tool of choice to accomplish this task ! In spite of limited applications of this method to this field, it was chosen for its popularity in signal processing. One of the most recent and efficient implementation, the FastICA algorithm, was applied at first to the unmixing of grayscale images, then to classic temporal signals (to verify its efficiency) and finally to a subset of the USGS spectral signature database. The approach was to compare the extracted independent components to a reference database and to form pairs based on similarity. However, due to the ambiguities and the lack of validation criterion associated with the ICA, it was impossible neither to predict nor to verify the pairs. To remedy this, our experimental protocol was divided into theoretical and practical comparisons, which are based on confidence levels and allowed to form, on one hand, the right pairs in theory (partial base) and, on the other hand, experimental pairs (entire base). These are finally compared to determine associations' success. Globally, based on two relative confidence thresholds, the results are excellent for signals, good for images but mediocre for spectral signatures. This last case is explained by a much more omnipresent effect of two general problems : decision-making's subjectivity and the unavoidable decorrelation, which involved deformations and too large a dependence on the selected base. Nevertheless, the results become very satisfactory for an adequate selection (cultures, trees and minerals). To improve the method, some constructive recommendations are proposed in order to support the second portion of this work, which wanted itself innovative.

## AVANT-PROPOS ET REMERCIEMENTS

Ce projet basé sur l'agriculture de précision m'a tenté car je le considère comme une première étape à la revalorisation de l'agriculture et à une meilleure façon d'utiliser ses ressources. En effet, essayer de trouver des méthodes scientifiques pour réduire la pollution des sols et générale de la planète a été la principale source de motivation de mon mémoire de maîtrise. Même si je n'apporte qu'une petite pierre à l'édifice, je suis heureux de pouvoir y contribuer. Dans ce travail, le but était, en utilisant l'analyse en composantes indépendantes, de trouver une méthode de télédétection pour connaître les carences dans une culture, afin de n'ajouter la quantité d'engrais requise qu'à l'endroit nécessaire ! Mon projet n'est qu'une première étape car on se doute à peine des résultats à attendre. Mais la démarche et surtout le but en tant que tel motivent le devoir de recherche de cette voie; c'est pourquoi tout au long de ce mémoire cet état d'esprit de test, de tâtonnement sera omniprésent.

Je tiens à remercier, bien évidemment, mon directeur de maîtrise, Monsieur Jacques-André Landry, Ph.D., qui m'a soutenu tout au long de cette aventure. Je souhaiterai remercier également l'étudiant en doctorat à l'École de technologie supérieure (ÉTS), Luis Da Costa, pour son code Matlab à partir duquel j'ai commencé à travailler. En outre, je tiens à remercier mon école d'ingénieur en France, l'École Nationale Supérieure des Télécommunications de Bretagne (ENST Bretagne), pour leur relation étroite avec l'ÉTS, qui m'a permis de poursuivre mon cursus étudiant de belle façon avec cette maîtrise. D'un point de vue

plus personnel, je tiens avant tout à remercier, du fond du coeur, mes parents qui, même loin, pensent et vibrent avec moi sur ce projet. Pour votre appui, merci ! Je tiens également à exprimer toute ma gratitude vis-à-vis de mes collègues et amis du LIVIA, qui, grâce à leur sympathie et bonne humeur, font de ce laboratoire un endroit d'épanouissement où il fait bon vivre et étudier. Enfin, ma chérie, je te remercie sincèrement d'avoir toujours été à mes côtés, de m'avoir soutenu, communiqué ta joie de vivre et encouragé à finir vite et bien par notamment tes mille et une idées d'endroits où partir en vacances une fois le mémoire terminé (rires).

## TABLE DES MATIÈRES

	Page
SOMMAIRE.....	i
ABSTRACT.....	ii
AVANT-PROPOS ET REMERCIEMENTS.....	iii
TABLE DES MATIÈRES .....	v
Liste des tableaux.....	x
Liste des figures .....	xii
Liste des algorithmes .....	xiv
Liste des abréviations et sigles .....	xv
INTRODUCTION .....	1
CHAPITRE 1 ANALYSE EN COMPOSANTES INDÉPENDANTES .....	9
1.1 Définition.....	9
1.2 Séparation aveugle de sources .....	10
1.2.1 Le cocktail-party .....	10
1.2.2 Hypothèse d'indépendance .....	11
1.2.3 Illustration .....	12
1.3 Applications .....	13
1.4 Modèle mathématique.....	14
1.4.1 Modèle vectoriel.....	14
1.4.1.1 $x$ , $s$ et $a$ .....	14
1.4.1.2 Modèle bruité .....	16
1.4.1.3 $y$ et $b$ .....	17
1.4.2 Modèle matriciel .....	18
1.4.2.1 $X$ , $S$ et $A$ .....	18
1.4.2.2 $Y$ et $B$ .....	19
1.4.2.3 Approches et estimées .....	21
1.5 Conditions .....	22
1.5.1 Au moins autant de mélanges que de sources .....	22
1.5.2 A matrice de rang plein .....	23

1.5.3	Indépendance statistique des composantes.....	23
1.5.4	Au plus une source gaussienne.....	23
1.6	Ambiguïtés.....	23
1.6.1	Ordre .....	23
1.6.2	Énergies .....	25
1.6.3	Matrice globale $G$ .....	26
1.6.4	Indice de performance $E_1$ .....	28
1.7	Indépendance.....	29
1.7.1	Définition mathématique .....	30
1.7.2	Mesure d'indépendance : l'information mutuelle .....	31
1.7.3	Dépendance, corrélation et non-gaussianité .....	32
1.7.3.1	Approximations.....	32
1.7.3.2	Non-gaussianité.....	33
1.7.3.3	Corrélation .....	33
1.7.3.4	Équation caractéristique .....	33
1.7.4	Indépendance et gaussianité font mauvais ménage .....	34
1.7.4.1	Sources les moins gaussiennes possibles .....	34
1.7.4.2	Au plus une source gaussienne.....	35
1.7.5	Conclusion .....	35
1.8	Fonctions de contraste .....	36
1.8.1	Statistiques d'ordres plus élevés.....	36
1.8.2	Kurtosis.....	37
1.8.3	Négentropie.....	38
1.8.3.1	Rappel sur l'entropie .....	39
1.8.3.2	Définition.....	39
1.8.3.3	Approximations.....	40
1.8.4	Autres .....	42
1.8.4.1	Information mutuelle.....	42
1.8.4.2	Maximum de vraisemblance.....	42
1.8.4.3	Infomax.....	43
CHAPITRE 2 ÉTAT DE L'ART .....		44
2.1	Choix de l'algorithme ACI.....	44
2.1.1	Problématique.....	44
2.1.1.1	Méthode d'estimation .....	45
2.1.1.2	Types d'algorithmes .....	45
2.1.2	Comparaison des différents algorithmes.....	46
2.1.2.1	Algorithmes considérés .....	46
2.1.2.2	Performances des algorithmes .....	47
2.1.3	Conséquence : choix du FastICA.....	51
2.2	Travaux sur les signatures spectrales .....	52



2.2.1	Préambule : la ségrégation hyperspectrale .....	52
2.2.2	Analogie avec l'ACI.....	54
2.2.3	Méthode des fractions d'abondance .....	55
2.2.3.1	Justification .....	55
2.2.3.2	Modèle.....	55
2.2.3.3	Travaux .....	56
2.2.4	Méthode des signatures spectrales .....	57
2.2.4.1	Modèle.....	57
2.2.4.2	Travaux .....	58
2.3	Conclusions.....	62
CHAPITRE 3 ALGORITHMES DU POINT FIXE .....		63
3.1	Prétraitement .....	63
3.1.1	Centrer $X$ .....	63
3.1.2	Blanchiment .....	64
3.2	Analogie avec les réseaux de neurones .....	68
3.3	Algorithmes du point fixe .....	70
3.3.1	Pour une composante.....	70
3.3.1.1	Convergence.....	70
3.3.1.2	Basé sur le kurtosis .....	71
3.3.1.3	Basé sur la négentropie .....	73
3.3.2	Pour toutes les composantes .....	77
3.3.2.1	Méthode déflation .....	78
3.3.2.2	Méthode symétrique.....	78
3.3.2.3	Propriétés du FastICA.....	80
3.4	Sortie du FastICA .....	81
CHAPITRE 4 PROTOCOLE EXPÉRIMENTAL .....		82
4.1	Généralités .....	82
4.1.1	Approches.....	82
4.1.2	Validité des résultats.....	84
4.1.3	Mise au point lexicale.....	86
4.1.4	Confiances .....	87
4.2	Prétraitement .....	88
4.2.1	Conformités statistiques avec le modèle.....	90
4.2.2	Formation de $A$ .....	92
4.3	Application de l'ACI .....	95
4.4	Post-traitement.....	96
4.4.1	Partie théorique .....	96
4.4.1.1	Matrice $G_{0,I}$ .....	97
4.4.1.2	Confiance théorique .....	98

4.4.2	Partie pratique.....	101
4.4.3	Pourcentage de succès .....	102
4.4.4	Mesures globales d'erreur.....	103
4.4.4.1	Erreur $E_1$ et confiances globales .....	104
4.4.4.2	Erreur $E_A$ .....	105
4.5	Variables finales interprétables.....	106
4.5.1	Nature .....	106
4.5.2	Erreur standard.....	107
4.5.3	Explications de la sélection.....	108
4.6	Algorithme.....	111
4.7	Exemple d'application du protocole expérimental .....	111
4.7.1	Choix de $S_{\mu,\Sigma}$ .....	111
4.7.2	Partie théorique .....	111
4.7.3	Partie pratique.....	114
4.7.4	Succès d'associations .....	115
4.7.5	Erreurs $E_1$ et $E_A$ .....	116
CHAPITRE 5 DISCUSSION ET INTERPRÉTATION DES RÉSULTATS .....		118
5.1	Valeurs des paramètres.....	118
5.2	Données .....	118
5.2.1	Images en niveaux de gris.....	119
5.2.2	Signaux de base.....	121
5.2.3	Signatures spectrales .....	122
5.2.4	Valeurs de $(k, \epsilon)$ en fonction de la nature des données .....	124
5.3	Commentaires de forme sur les tableaux de résultats .....	125
5.4	Remarques générales.....	127
5.4.1	Cas particulier de $n = 2$ .....	127
5.4.2	Convergence.....	127
5.4.3	Résultats sur $A$ .....	128
5.4.3.1	Uniformité en $A$ .....	128
5.4.3.2	Erreur $E_A$ .....	130
5.4.4	Dépendance à la nature de la combinaison.....	133
5.5	Interprétation des valeurs de confiances .....	134
5.5.1	Nécessité pour la distinction des cas d'associations parfaites .....	135
5.5.2	Condition non suffisante.....	137
5.5.3	Philosophie de la prise de décision.....	138
5.5.4	Valeurs des seuils empiriques locaux.....	140
5.6	Résultats particuliers .....	143
5.6.1	Images en niveaux de gris.....	143
5.6.1.1	Cas $n = 2$ .....	143
5.6.1.2	Cas $n = 3$ .....	145

5.6.1.3	Cas $n = 4$ .....	146
5.6.2	Signaux de base.....	148
5.6.2.1	5 signaux .....	148
5.6.2.2	4 signaux .....	150
5.6.3	Signatures spectrales .....	150
5.6.3.1	Cas $n = 2$ .....	151
5.6.3.2	Cas $n = 3$ .....	154
5.6.3.3	Cas $n = 4$ .....	156
5.6.4	Bilan.....	158
5.7	Note positive finale .....	159
CONCLUSION .....		161
RECOMMANDATIONS .....		164
ANNEXES		
1	: Indépendance .....	166
2	: ACI versus ACP et PP.....	174
3	: Moments et cumulants .....	181
4	: Algorithmes de l'ACI.....	186
5	: Neutralisation des deux premiers moments d'une matrice .....	193
6	: Signatures spectrales USGS .....	195
7	: Tableaux de résultats.....	201
BIBLIOGRAPHIE .....		222

## LISTE DES TABLEAUX

	Page
Tableau I	Variables de l'ACI et estimées ..... 21
Tableau II	Équivalence des variables dans les espaces normal et blanchi ..... 68
Tableau III	Arborescence de la base de données des sept signatures choisies ... 124
Tableau IV	Valeurs de $k$ , $\epsilon$ et du nombre de matrices $A$ en fonction de $n$ et des données ..... 125
Tableau V	Exemple de tableau de résultats pour des combinaisons d'images... 126
Tableau VI	Seuils empiriques de confiances locales..... 140
Tableau VII	Erreur et rejet pour $n = 2$ images ..... 144
Tableau VIII	Erreur et rejet pour $n = 3$ images ..... 145
Tableau IX	Erreur et rejet pour $n = 4$ images ..... 147
Tableau X	Erreur et rejet pour $n = 2$ signatures ..... 152
Tableau XI	Erreur et rejet pour $n = 3$ signatures ..... 154
Tableau XII	Erreur et rejet pour $n = 4$ signatures ..... 157
Tableau XIII	Répartition des signatures de la base USGS par catégorie et type ... 197
Tableau XIV	Répartition des signatures de minéraux par groupes..... 198
Tableau XV	Répartition des signatures de minéraux par familles ..... 200
Tableau XVI	Moyennes et erreurs standard pour $n = 2$ images ..... 202
Tableau XVII	Moyennes et erreurs standard pour $n = 3$ images ..... 203
Tableau XVIII	Moyennes et erreurs standard pour $n = 4$ images ..... 206
Tableau XIX	Moyennes et erreurs standard pour $n = 2$ avec $N = 5$ signaux..... 210
Tableau XX	Moyennes et erreurs standard pour $n = 3$ avec $N = 5$ signaux..... 211

Tableau XXI	Moyennes et erreurs standard pour $n = 4$ avec $N = 5$ signaux.....	212
Tableau XXII	Moyennes et erreurs standard pour $n = 2$ avec $N = 4$ signaux.....	213
Tableau XXIII	Moyennes et erreurs standard pour $n = 3$ avec $N = 4$ signaux.....	213
Tableau XXIV	Moyennes et erreurs standard pour $n = 4$ avec $N = 4$ signaux.....	214
Tableau XXV	Moyennes et erreurs standard pour $n = 2$ signatures .....	214
Tableau XXVI	Moyennes et erreurs standard pour $n = 3$ signatures .....	215
Tableau XXVII	Moyennes et erreurs standard pour $n = 4$ signatures .....	218

## LISTE DES FIGURES

	Page
Figure 1	Signature spectrale de l'élément « actinolite » ..... 3
Figure 2	Image de réflectance et pixel d'une parcelle ..... 4
Figure 3	Image hyperspectrale et signatures spectrales ..... 5
Figure 4	« Cocktail-party » ..... 10
Figure 5	Signaux sources, mélanges et composantes indépendantes ..... 12
Figure 6	Supergaussienne, gaussienne et subgaussienne ..... 37
Figure 7	Erreur $E_1$ en fonction de la précision ..... 48
Figure 8	Erreur $E_1$ en fonction de la charge informatique ..... 50
Figure 9	Erreur en fonction du nombre de sources ..... 51
Figure 10	Réseau de neurones de l'ACI ..... 69
Figure 11	Exemple de subjectivité dans les associations signaux / composantes ..... 85
Figure 12	Protocole expérimental..... 89
Figure 13	Processus de rejet des paires théoriques et pratiques en fonction des seuils de confiance locale ..... 110
Figure 14	Base de données des sept images en niveaux de gris ..... 119
Figure 15	Base de données des cinq signaux temporels classiques ..... 121
Figure 16	Base de données des sept signatures USGS choisies (réflectance de 0.2 à 3 $\mu\text{m}$ ) ..... 123
Figure 17	Meilleur, moyen et pire comportements en $A$ pour les images, les signaux et les signatures (confiance théorique en gras, pratique en normal)..... 129

Figure 18	Erreurs d'estimation de la matrice de mélange pour $n = \{2, 3, 4\}$ , sans et avec discontinuités, en fonction de la proportion $p$ du premier des $n - 1$ signaux .....	131
Figure 19	Effet de la nature de la base (décorrélation) .....	133
Figure 20	Illustration de la pertinence des confiances pour une même image présente dans trois combinaisons différentes ayant toutes un taux d'associations de 100 % (ici $S_2$ dans 2-4, 2-5 et 2-6, par colonne de gauche à droite) .....	136
Figure 21	Exemple de subjectivité pour les signatures .....	138
Figure 22	Effet de discontinuités à cause de la décorrélation pour les signaux .....	149
Figure 23	Variables décorrélées, corrélées et indépendantes .....	168
Figure 24	Géométrie de la dépendance .....	170
Figure 25	Illustration du théorème de la limite centrale .....	171
Figure 26	ACP d'un nuage de point .....	176
Figure 27	ACI versus ACP .....	177
Figure 28	PP versus ACP .....	180
Figure 29	Signatures de même catégorie ou non .....	198
Figure 30	Signatures de même groupe / famille ou non .....	199

## LISTE DES ALGORITHMES

	Page
Algorithme 1	Algorithme du point fixe du kurtosis pour une composante ..... 73
Algorithme 2	Algorithme du point fixe de la négentropie pour une composante .... 76
Algorithme 3	Algorithme de la méthode déflation ..... 79
Algorithme 4	Algorithme de la méthode symétrique ..... 80
Algorithme 5	Algorithme de base du protocole expérimental ..... 112



## LISTE DES ABRÉVIATIONS ET SIGLES

ACI	Analyse en Composantes Indépendantes
ACP	Analyse en Composantes Principales
ACY	Algorithme d'Amari-Cichocki-Yang
AVIRIS	Airborne Visible InfraRed Imaging Spectrometer (spectromètre visible infrarouge portable)
BS	Algorithme de Bell et Sejnowski
BSS	Blind Source Separation (séparation aveugle de sources)
CI	Composante Indépendante
cICA	contextual ICA (ACI contextuelle)
DKL	Divergence de Kullback-Leibler
DVP	Décomposition en Valeurs et Vecteurs Propres
DVS	Décomposition en Valeurs Singulières
EASI	Equivariant Adaptive Separation via Independence (séparation adaptative équivariante de sources indépendantes)
ExtBS	Algorithme étendu de Bell et Sejnowski
FastICA	Fast Independent Component Analysis
FP	Fixed-Point Algorithm (algorithme du point fixe)
FPsymth	Point-fixe à non-linéarité tanh et méthode symétrique
FPth	Point-fixe à non-linéarité tanh et méthode déflation
HOS	Higher-Order Statistics (statistiques d'ordres plus élevés)
HUT	Helsinki University of Technology (Université de Technologie d'Helsinki)

ICA	Independent Component Analysis (analyse en composantes indépendantes)
LSMA	Linear Spectral Mixture Analysis (analyse de mélanges spectraux linéaires)
MAP	Maximum a posteriori
NDVI	Normalized Difference Vegetation Index (indice de végétation par différence normalisée)
NPCA-RLS	Nonlinear PCA by Recursive Least Squares (ACP non-linéaire par la technique récursive des moindres carrés)
PCA	Principal Component Analysis (analyse en composantes principales)
PP	Projection Pursuite
USGS	United States Geological Survey
WACY	White ACY (ACY blanchi)
$\alpha_i$	Fraction d'abondance
$\delta$	Symbole de Kronecker
$\delta(\cdot \cdot)$	Divergence de Kullback-Leibler
$\epsilon$	Pas correctif de $A$
$\gamma_4$	Kurtosis normalisé
$\kappa_n$	Cumulant d'ordre $n$
$\kappa_4, \text{kurt}(\cdot)$	Kurtosis ou cumulant d'ordre 4
$\sigma_n$	Erreur standard
$\Theta(\cdot)$	Fonction échelon
$k$	Pas de pourcentage du premier signal dans $A$ ( $k \in [0, 1]$ )
$m$	Nombre de mélanges
$n$	Nombre de signaux sources

$p$	Proportion du premier signal dans $A$
$(s_i)_{i \in [1, n]}$	signal source (vecteur)
$(v_i)_{i \in [1, n]}$	mélange blanchi (vecteur)
$w_i$	colonne de $W$ , calculée par les algorithmes du point fixe
$(x_j)_{j \in [1, m]}$	mélange (vecteur)
$(y_i)_{i \in [1, n]}$	composante indépendante (vecteur)
$A$	Matrice de mélange
$\tilde{A}$	Estimée de la matrice de mélange $A$
$B$	Matrice de séparation
$C$	Matrice de covariance de $X$
$C_G$	Confiance globale théorique
$C_G^{\text{base}}$	Confiance globale pratique
$C_L$	Confiance locale théorique
$C_L^{\text{base}}$	Confiance locale pratique
$D$	Matrice diagonale scalaire ( $\pm 1$ )
$D_C$	Matrice des valeurs propres de $C$
$E$	Matrice des vecteurs propres de $C$
$E_1$	Erreur globale, indice de performance
$E_A$	Matrice d'erreur entre $A$ et $\tilde{A}$
$E(S_j/Y_i)$	Erreur du signal $S_j$ , connaissant $Y_i$
$E_{SY}$	Matrice de terme général $E(S_j/Y_i)$
$E(Y_i/S_j)$	Erreur de la composante $Y_i$ , connaissant $S_j$
$E_{YS}$	Matrice de terme général $E(Y_i/S_j)$
$E(Y_i, S_j)$	Erreur du couple $(Y_i, S_j)$

$G$	Matrice globale
$G_P$	Matrice globale parfaite ( $G_P = PD$ )
$G_{0,I}$	Matrice globale théorique
$G_{0,I}^{\text{base}}$	Matrice globale pratique
$H(\cdot)$	Entropie
$I(\cdot)$	Information mutuelle
$J(\cdot)$	Négentropie
$M_{\mu,\Sigma}$	Matrice de vecteur moyenne $\mu$ et de matrice de covariance $\Sigma$
$M_{0,\Sigma}$	Matrice centrée et de matrice de covariance $\Sigma$
$M_{0,I}$	Matrice centrée et décorrélée
$N$	Nombre de signaux sources de la base totale (protocole)
$P$	Matrice de permutation
$R(M_i)$	Signature spectrale du matériau $M_i$
$R(x, y, \lambda)$	Réflectance $R$ du pixel $(x, y)$ à la longueur d'onde $\lambda$
$S$	Matrice des $n$ signaux sources de la base partielle
$S^{\text{base}}$	Matrice des $N$ signaux sources de la base totale
$S_L$	Seuil local théorique
$S_L^{\text{base}}$	Seuil local pratique
$U$	Matrice de blanchiment
$V$	Matrice des mélanges blanchis
$W$	Matrice de mélange blanchie
$W^T$	Matrice de séparation blanchie
$X$	Matrice des signaux mélangés (ou matrice des mélanges)
$Y$	Matrice des composantes indépendantes

## INTRODUCTION

### **Caractérisation d'images hyperspectrales**

Pour débiter, nous allons évoquer le contexte de ce mémoire de maîtrise, qu'est l'agriculture de précision; puis, nous définirons les termes de « signature spectrale » et « image hyperspectrale ».

### **Contexte : l'agriculture de précision**

L'agriculture traditionnelle consiste, généralement, à utiliser des engrais sur l'ensemble d'un champ à cultiver pour permettre une meilleure production, sans distinction des réels besoins nutritifs des différentes parties du champ. Ainsi, il se peut fort bien que seulement un tiers du champ ait besoin d'engrais; mais par commodité et/ou habitude, tout le champ sera « aspergé ». S'en suit alors une pollution des sols plus importante. Dans le contexte actuel de notre société qui se veut et se doit d'être de plus en plus écologique et environnementaliste, une nouvelle manière de concevoir l'agriculture a été pensée : l'agriculture de précision.

Le leitmotiv de cette nouvelle « philosophie » agricole est d'ajouter la bonne dose (d'engrais, de dés herbant ou autres produits) seulement au bon endroit et seulement au bon moment. En d'autres termes, cela revient, d'une part, à non plus penser un champ dans sa globalité mais par ses subdivisions ou parcelles; et, d'autre part, à combler les carences de chacune de ces parcelles de terre de manière plus individuelle et privilégiée. Les intérêts de cette nouvelle façon de penser sont évidemment d'ordre écologique (diminution de la pollution des sols par infiltration) et économique (réduction de coûts pour l'agriculteur).

Pour identifier les carences locales et les combler, il est primordial de connaître la constitution du sol. Une fois celle-ci connue et à partir de l'expertise des biologistes et agriculteurs qui connaissent les éléments nécessaires à une bonne pousse, il devient évident de savoir

ce qu'il faut modifier : les éléments (engrais, . . .) à rajouter et leurs quantités. C'est une adaptation personnalisée à chaque parcelle.

L'objectif est donc de savoir ce que contient le sol, localement. Pour ce faire, on va, tout d'abord, mentalement subdiviser le champ en parcelles pour ainsi satisfaire à la problématique de la localité (il sera montré plus tard que ce n'est pas si mental que ça, mais dépendant de la résolution de la caméra). Dans un deuxième temps, le but est de rechercher ce que l'on appelle les *signatures spectrales* de ces parcelles, pour satisfaire à la problématique de leur contenu.

### **Signature spectrale**

La signature spectrale d'un élément, qu'il soit humain, animal, végétal ou minéral. . . , représente son comportement, sa réponse unique à un rayonnement lumineux; c'est-à-dire la façon dont un corps réagit, émet sous l'incidence d'un rayonnement électromagnétique, qui est constitué d'énergie transportée sur une plage étendue de longueurs d'ondes. On appelle *signature spectrale* la réponse sur cette plage et *réflectance* la réponse pour une longueur d'onde donnée. Ainsi, une signature spectrale (voir l'exemple de l'actinolite, signature issue de « United States Geological Survey » (USGS), figure 1)<sup>1</sup> est un signal unidimensionnel qui représente la réflectance en fonction des longueurs d'ondes des bandes spectrales utilisées (typiquement ici de 0,2 à 3  $\mu\text{m}$ , soit du proche ultra-violet à l'infrarouge). À noter que, puisque la réflectance est le rapport de ce qui est émis par ce qui est reçu, sa valeur est normalisée et donc comprise entre 0 et 1. Il en va de même par conséquent pour la signature spectrale.

Ce qu'il est primordial de comprendre est le fait que, pour une condition d'éclairage donnée, *à un élément correspond une et une seule signature spectrale et réciproquement !* Par exemple, les pics et bosses de la figure 1 sont typiques de l'actinolite; et seule l'actinolite présente cette signature. Le terme « signature » prend alors tout son sens. De même,

<sup>1</sup><http://speclab.cr.usgs.gov/spectral.lib04/PLOTS.lib04a/Actinolite.HS22.3B.tgif.gif>

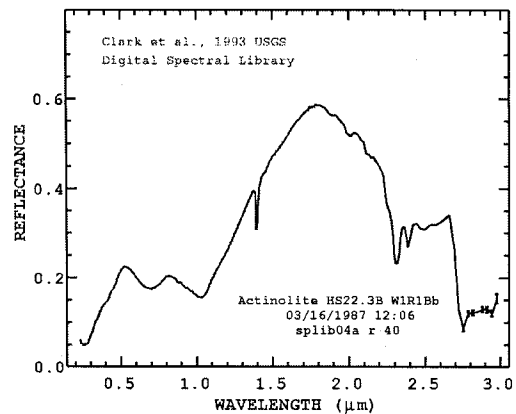


Figure 1 Signature spectrale de l'élément « actinolite »

les os sont visibles aux rayons X, alors que les êtres humains ne le sont pas, mais le sont dans le visible... Ainsi, connaître les signatures spectrales d'une parcelle nous renseigne de façon unique et parfaite sur la nature des éléments qui y sont présents. Cette phrase est la phrase-clé du mémoire de maîtrise car grâce à elle, on transforme le problème agricole, sur le terrain, en un problème théorique, mathématique, de traitement du signal, comme il le sera expliqué plus tard.

### Image hyperspectrale

Pratiquement, pour obtenir les signatures, on utilise des images hyperspectrales du champ en question. Elles sont obtenues en pratique, dans notre cas, par photographies aéroportées à l'aide d'une caméra hyperspectrale; d'où la notion de « télédétection ».

En d'autres termes, un avion survole le champ à analyser et mesure la réflectance de chaque parcelle à un ensemble de longueurs d'onde données. On obtient ainsi, pour chaque longueur d'onde, une image de niveaux de gris, dite de réflectance (voir figure 2)<sup>2</sup>, dont la valeur de chaque pixel, représentatif d'une parcelle, est sa valeur de réflectance, à cette longueur d'onde. La taille de chaque parcelle est définie par la résolution de la caméra

<sup>2</sup>Données de PCI Geomatica et schéma de Clément Chion, M.Ing., doctorant au LIVIA

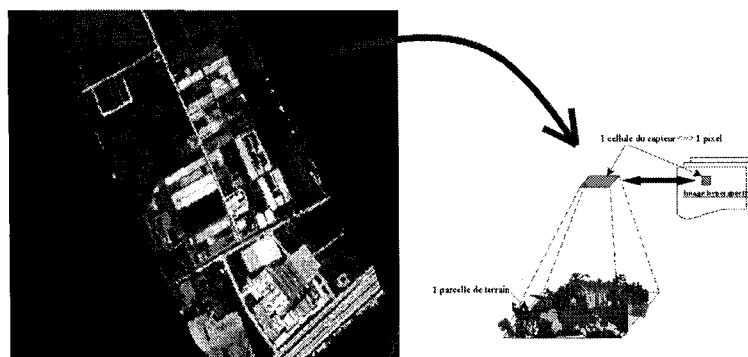


Figure 2 Image de réflectance et pixel d'une parcelle

hyperspectrale, qui détermine le nombre de pixels utilisés pour représenter l'image de réflectance. On en déduit ensuite la superficie réelle de chaque parcelle, qui est généralement de l'ordre de quelques mètres carrés, avec les caméras standard actuelles. La source d'énergie est l'éclairage ambiant, dans notre cas l'ensoleillement, et la caméra capte cette énergie aux longueurs d'ondes désirées. À chaque longueur d'onde correspond une image de réflectance. Au final, soit pour toutes les longueurs d'onde du spectre à étudier, on se retrouve avec un « volume », formé d'une superposition de couches d'images de réflectance. Ce volume est appelée *image hyperspectrale* (voir figure 3)<sup>3</sup>.

En considérant les valeurs de réflectance de toutes les couches (ou longueurs d'onde) pour un même pixel, on obtient la signature spectrale de ce pixel, donc de la parcelle correspondante; et en conséquence, sa constitution en éléments ! On parle alors de *caractérisation d'images hyperspectrales* : dire ce que contient un sol à partir de ses signatures, obtenues à partir des images hyperspectrales !

### **Problématique : mélange versus signature**

Cependant, comme il a été mentionné plus haut, de par la résolution de la caméra, un pixel peut représenter une zone de quelques mètres carrés (ou dizaines voire centaines), il est

<sup>3</sup>Source : NEMO Project Office, United States Navy



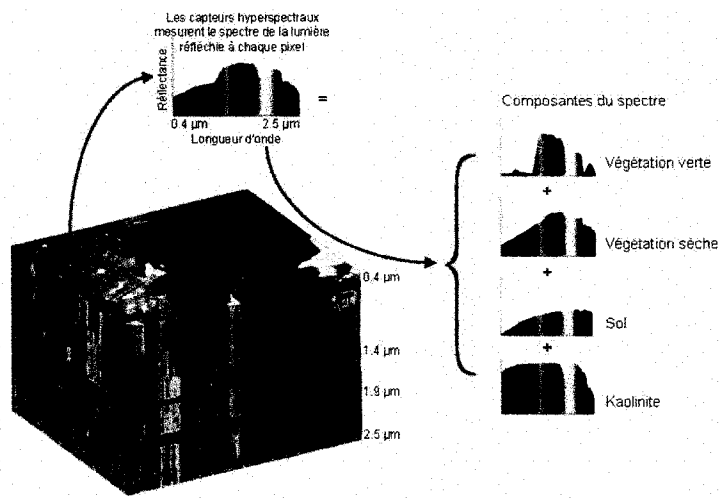


Figure 3 Image hyperspectrale et signatures spectrales

donc évident que plusieurs éléments y cohabitent et que la signature spectrale issue de ce pixel sera celle du mélange pondéré (la pondération étant a priori inconnue !) de tous les éléments présents et non les signatures de chacun de ces éléments. En effet, à titre d'exemple, sur la figure 3, la signature issue du pixel (celle du haut) est un mélange des éléments de droite, à savoir : la végétation verte, la végétation sèche, le sol et la kaolinite, qui ont chacun leur propre signature. Or, ce sont les signatures de ces éléments qui nous intéressent; pas celle du mélange, qui intrinsèquement n'offre *aucune* information interprétable ! L'objectif est donc bien de « récupérer » les signatures individuelles des éléments à partir de celle du mélange pondéré.

Correctement reformulée, la problématique de ce projet est donc d'investiguer s'il est possible de dériver les signatures individuelles, et leur pondération, à partir de la signature composée qu'elles forment et que l'on observe.

### **Solution proposée : l'ACI**

En l'occurrence, il existe une méthode statistique qui permet effectivement d'extraire des signaux indépendants à partir de mélanges. Il s'agit de l'« *Analyse en Composantes indé-*

*pendantes* » (ACI) ou "*Independent Component Analysis*" (ICA) en anglais. Cette analyse sera expliquée plus en détail par la suite. Pour le moment, il suffit de savoir que le rôle de cette méthode est de récupérer des signaux à partir de mélanges; et ce, sans connaissance a priori des proportions du mélange (ou pondération). La seule condition sous-jacente à l'ACI est qu'elle doit disposer d'au moins autant de mélanges que de sources. En d'autres termes, par rapport à notre problème pratique, il faut recueillir au moins autant de signatures de mélanges que de signatures des éléments qui les composent. Cela va signifier concrètement qu'au lieu de considérer une seule parcelle pour, par exemple, quatre éléments, on devra en considérer au moins quatre; d'où une multiplication par quatre de la superficie réelle d'étude. Cela entraîne également le fait que l'agriculteur et le scientifique ont a priori connaissance de la nature globale du champ, considéré lui-même comme a priori homogène dans cette zone de quatre parcelles. Mais cette idée ne pose pas de problème particulier car elle reste conforme au protocole d'études. En effet, rappelons que le but est de connaître les concentrations en éléments et qu'a priori un agriculteur sait ce qu'il fait pousser dans son champ. Ainsi, agrandir la zone d'études ne présente pas de problème théorique particulier et l'ACI pourra être utilisée.

### **Hypothèses simplificatrices**

Cependant, il est intéressant de noter ici (et ce sera redit dans l'état de l'art) qu'a priori, nous n'avons que peu d'indices sur la teneur des résultats; c'est-à-dire sur la capacité de l'ACI à correctement démixer les signatures composant ces mélanges. En effet, peu de recherches ont été menées et l'un des problèmes majeurs est que, comme l'indique son « I », l'ACI ne traite que des signaux *indépendants*. Or, il apparaît quasi-évident que les éléments nutritifs (et autres) dans le sol interagissent ensemble. Ainsi, nous posons l'une de nos hypothèses en affirmant que l'on considère les éléments comme indépendants les uns des autres; on parle donc de *signatures pures*. Malgré tout, cette hypothèse n'est pas spécialement limitative pour nous car de par le but même de l'ACI, qui est de séparer des signaux, les signatures spectrales vont être interprétées par l'ACI comme de simples

signaux unidimensionnels. À vrai dire pour l'ACI que ces signaux représentent de la voix, du temps, des signatures, . . . , le traitement qu'elle en fera sera le même. Les données seront interprétées de façon identique.

### **Redéfinition du mandat**

Le contexte agricole environnant mon mémoire de maîtrise n'apparaît finalement que comme un « prétexte » ; car comme il vient d'être décrit, mon travail de séparation des signatures, va, en définitive et sans amoindrir la teneur de mes travaux, se transformer en un problème mathématique de traitement de signal ! Cette introduction n'était que la mise en contexte du problème qui m'intéresse vraiment : la capacité de l'ACI à démêler les signatures spectrales de parcelles. L'agriculture de précision en est l'un de ses champs d'applications possibles. Pour bien souligner la distance théorique prise avec le terrain, je travaillerai par la suite sur des signatures pures de laboratoire, soit extrêmement précises, sur lesquelles n'apparaissent pas, entre autres, les effets de la dispersion par l'air pendant les mesures. Le problème est une nouvelle fois simplifié.

Ainsi, pour résumer, à partir de maintenant, mon travail se détache du problème agricole pour ne considérer les signatures spectrales « que » comme des signaux à une dimension. Le but est donc d'extraire, à partir de mélanges, les signaux (signatures) originaux et indépendants, constitutifs de ces mélanges.

### **Démarche**

En outre, vu que le but ici est de savoir si l'utilisation de l'ACI est pertinente à la résolution de notre problème de séparation de signatures spectrales, on veut juger de sa robustesse et de sa capacité à correctement démêler. Ainsi, contrairement au cas théorique où les composantes indépendantes sont récupérées en sortie de l'ACI, uniquement à partir des mélanges, sans aucune connaissance des signaux originaux et des proportions du mélange, la démarche ici sera d'ordre pratique, basée sur la comparaison de signatures originales

et des composantes indépendantes. En effet, pratiquement, des signatures originales de la base USGS sont choisies, puis mélangées dans des proportions connues. Ces mélanges sont introduits à l'ACI, qui rend en sortie les composantes indépendantes. On compare celles-ci avec les signatures originales, en se basant sur une mesure de confiance. On forme les paires les plus ressemblantes selon ce critère de confiance et on vérifie la validité des associations, en relevant les taux d'erreurs. Cette procédure sera détaillée dans le chapitre 4 dédié au protocole expérimental.

Cependant, l'ACI sera appliquée dans un premier temps sur des images en niveaux de gris. En effet, sans rentrer dans les détails, notre équipe de recherche a déjà réalisé de nombreux tests et nous connaissons l'efficacité de la méthode sur les images. L'intérêt est ici de pouvoir estimer l'effet de certains paramètres sur les résultats, tels que les proportions du mélange et l'effet du bruit, avant de l'appliquer, dans un second temps, aux signatures spectrales.

## **Plan**

Finalement, le plan suivant a été adopté pour le mémoire. Dans le premier chapitre, le modèle mathématique de l'ACI est décrit, avant d'aborder, dans le deuxième, son état de l'art. Les algorithmes de l'ACI, et plus spécifiquement le FastICA, sont ensuite présentés au chapitre 3. Le chapitre 4 explicite, quant à lui, le protocole expérimental de la démarche expliquée ci-haut. Les résultats et leurs interprétations sont examinés au chapitre 5. Enfin, une conclusion et les recommandations futures terminent ce mémoire.

Je vous souhaite une agréable lecture.

## CHAPITRE 1

### ANALYSE EN COMPOSANTES INDÉPENDANTES

Dans ce chapitre, nous allons présenter et définir l'« Analyse en Composantes Indépendantes ». Par choix, nous avons décidé de présenter l'ACI, avant de faire l'état de l'art de la question. En effet, celui-ci nécessite des connaissances théoriques, qui vont être introduites dans ce chapitre. Débutons par les origines de l'ACI.

#### 1.1 Définition

L'appellation « Analyse en Composantes Indépendantes » est apparue progressivement au début des années 90, en même temps que s'affinait la notion de « séparation aveugle de sources » (« Blind Source Separation » (BSS) en anglais). En effet, l'ACI tire son origine de la BSS, technique de traitement de signal, que nous présentons dans la partie suivante. D'ailleurs, les premiers à avoir expliqué, ou du moins intuité, la notion d'ACI sont, en 1991, Jutten et Héroult (Jutten et Héroult, 1991), sans jamais avoir employé l'expression d'ACI; il était question de BSS. Le premier à avoir nommé l'ACI comme telle est Pierre Comon dont l'article (Comon, 1994) fait figure de référence. Dans ce dernier, il définit l'ACI ainsi : « *L'analyse en composantes indépendantes (ACI) d'un vecteur aléatoire consiste en la recherche d'une transformation linéaire qui minimise la dépendance statistique entre ses composantes.* » En termes plus simplifiés, Jean-François Cardoso (Cardoso, 2002, 2004a), autre référence dans le domaine du traitement de signal, affirme que : « *L'analyse en composantes indépendantes est la décomposition d'un vecteur aléatoire en composantes linéaires « les plus indépendantes possible »* ».

Ces deux définitions mettent en lumière la notion essentielle de l'ACI : *l'indépendance*. On y reviendra dans la partie 1.7. L'ACI apparaît ainsi comme l'« extraction » de signaux (composantes) à partir d'un mélange (vecteur aléatoire) linéaire de ceux-ci. Le vocabulaire employé est celui de la BSS; le lien avec l'ACI est alors flagrant.

## 1.2 Séparation aveugle de sources

En effet, la séparation aveugle de sources (Cardoso, 2004b; Jutten et Herault, 1991; Bell et Sejnowski, 1995a,b; Amari et al., 1996; Hyvärinen et al., 2001; Cardoso, 2002) est une technique de traitement de signal qui consiste grosso modo à récupérer des signaux sources à partir d'enregistrements de capteurs, qui sont des mélanges de ces signaux sources. Pour mieux comprendre ce qu'est la BSS, nous présentons le problème dit du « cocktail-party », qui est l'exemple typique dans la littérature.

### 1.2.1 Le cocktail-party

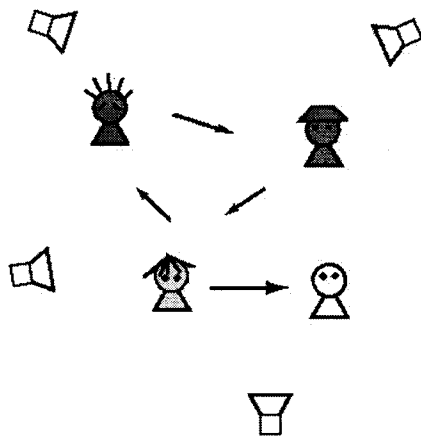


Figure 4 « Cocktail-party »

Considérons une soirée avec  $n$  invités qui discutent normalement les uns avec les autres (voir Figure 4). La voix de chacune des  $n$  personnes est représentée par un signal temporel  $s_i(t)_{i \in [1, n]}$ , appelé *signal source*. On note  $S(t) = (s_1(t), s_2(t), \dots, s_n(t))^T$  le vecteur colonne des  $n$  signaux sources.  $m$  microphones (*capteurs*), disposés dans la salle (avec  $m \geq n$ ), enregistrent l'ensemble de tous ces signaux vocaux. De par leur position différente dans la salle, les microphones génèrent individuellement des enregistrements distincts, mais sur lesquels l'ensemble de toutes les voix est présent. Ainsi, chaque mi-

crophone enregistre un mélange *linéaire* pondéré de toutes les voix ou signaux sources. Suivant le même modèle mathématique que  $s_i$ , chacun des  $m$  mélanges est défini comme une suite temporelle  $x_j(t)_{j \in [1, m]}$ . On note  $X(t) = (x_1(t), x_2(t), \dots, x_m(t))^T$  le vecteur colonne des  $m$  mélanges. La pondération linéaire se modélise mathématiquement par la matrice de mélange  $A = (a_{ji})_{j \in [1, m] \times i \in [1, n]}$ , de taille  $[m, n]$ , telle que :

$$x_j(t) = \sum_{i=1}^n a_{ji} s_i(t), \forall j \in [1, m] \quad (1.1)$$

$$X(t) = AS(t) \quad (1.2)$$

Le but du cocktail-party (et de la BSS) est de calculer les signaux sources  $s_i$  à partir des mélanges  $x_j$ ; et ce, sans aucune connaissance a priori des  $s_i$ , ni de  $A$ ; uniquement des  $x_j$ .

### 1.2.2 Hypothèse d'indépendance

Pour ce faire (Cardoso, 2004b), on ne considère pas la physique du problème car de nombreux phénomènes physiques (tels que la propagation des sources aux capteurs, ou encore les caractéristiques des instruments de mesure, ...) sont non maîtrisables. En revanche, la solution consiste à poser l'hypothèse forte mais acceptable physiquement : l'*indépendance statistique* des signaux sources ( $s_i$ ) ! La physique s'est transformée en statistique : on parle alors de méthode *aveugle* car toute l'information provient uniquement des données elles-mêmes et non de connaissance a priori. Cette hypothèse d'indépendance constitue le fondement de l'ACI. Elle sera exhaustivement explicitée dans la partie 1.7.

Par analogie avec les définitions précédentes de Comon et Cardoso de la partie 1.1, le « vecteur aléatoire » en question est  $x_j$ , la « transformation linéaire » est assurée par la matrice  $A$  et les « composantes indépendantes » sont les estimées des signaux sources inconnus  $s_i$ . Le but de l'ACI est d'extraire ces composantes indépendantes à partir des mélanges, sous l'unique condition (certes forte) de leur indépendance statistique.

Il est important de remarquer que bien que liées, la BSS et l'ACI ne sont pas identiques (Hyvärinen, 1999e). En effet, l'ACI est un modèle de données théorique à plusieurs applications, dont l'une d'elles est la BSS. Les autres applications possibles sont présentées en 1.3. Réciproquement, la BSS peut être résolue par d'autres méthodes que l'ACI.

### 1.2.3 Illustration

Nous pouvons illustrer cette séparation de sources sur un exemple tout simple (Hyvärinen et Oja, 2000). Soit deux signaux classiques  $s_1$  et  $s_2$ , respectivement une sinusoïde et un signal en dents de scie (figure 5(a)), considérés comme inconnus, que l'on mélange (figure 5(b)) de façon inconnue. Ces mélanges sont en revanche connus et on obtient finalement les estimées des signaux sources initiaux : les composantes indépendantes (figure 5(c)). On remarque qu'elles sont égales aux sources, au signe près : il s'agit d'une des ambiguïtés de l'ACI, qui seront abordées en 1.6.3.

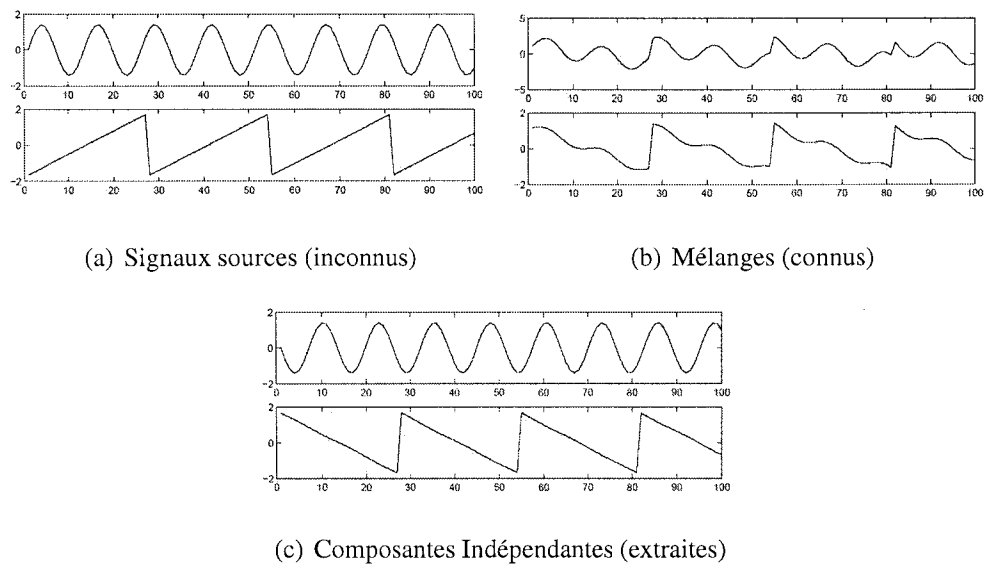


Figure 5 Signaux sources, mélanges et composantes indépendantes



### 1.3 Applications

Il existe quatre types d'applications générales de l'ACI (Hyvärinen, 1999e; Hyvärinen et Oja, 1998b) :

- a. la séparation de sources aveugles;
- b. la déconvolution aveugle (Bell et Sejnowski, 1995a,b; Comon, 1994), identique à la BSS, excepté que les mélanges ne sont pas des combinaisons linéaires des signaux sources; mais des convolutions de signaux, à déconvoluer;
- c. l'extraction de caractéristiques, appliquée notamment aux images (Hyvärinen et Oja, 1998a; Bell et Sejnowski, 1997), où l'on se rend compte que les composantes extraites sont les caractéristiques fondamentales de l'image, typiquement des morceaux de contours;
- d. la réduction de la redondance (Field, 1994).

Nous ne développons succinctement que la séparation aveugle de sources.

La BSS est l'application principale dont l'ACI est solution; car, à partir du moment où les données à extraire ou démélanger sont interprétables comme des signaux, l'ACI est appropriée. À ce titre, il existe de nombreuses applications appartenant au problème de la BSS (Hyvärinen, 1999e; Hyvärinen et Oja, 2000; Hyvärinen et al., 2001; Bell et Sejnowski, 1995a,b; Amari et al., 1996) :

- a. détection d'artefacts dans les électro- (EEG) et magnéto-encéphalogrammes (MEG) (Vigario et al., 1998; Vigario, 1997; Makeig et al., 1996);
- b. étude des séries économiques temporelles (Kiviluto et Oja, 1998; Back et Weigend, 1997);
- c. traitement de l'image (Hurri et al., 1996; Hyvärinen et al., 2003), en milieu bruité (Hyvärinen, 1999c,d, 1998a);
- d. télécommunications (Ristaniemi et Joutsensalo, 1999);

- e. imagerie hyperspectrale, application de notre mémoire qui sera exhaustivement expliquée dans le chapitre suivant sur l'état de l'art, dans la partie 2.2; nous en faisons seulement mention ici.

Plus généralement, Cardoso (Cardoso, 2002) résume l'utilité de l'ACI et ses applications possibles : « cette idée (ACI) est très fructueuse, de manière générale, dans tous les cas où un système de plusieurs capteurs, fournissant des signaux cohérents, est à l'écoute d'un ensemble discret de sources de signal que l'on cherche à extraire des observations et que l'on peut, *pour des raisons physiques, supposer mutuellement statistiquement indépendantes* ».

Cela signifie qu'à partir du moment où les données sont interprétables comme des signaux et hypothétiquement supposables indépendantes, l'ACI est utilisable; d'où son application aux voix, signaux nerveux, signatures spectrales, ...

## 1.4 Modèle mathématique

En reprenant le modèle de la séparation aveugle de sources (voir 1.2.1), nous allons présenter le modèle mathématique de l'ACI, d'un point de vue vectoriel et matriciel. On remarquera que les variables ne sont plus considérées comme temporelles mais comme des variables aléatoires.

### 1.4.1 Modèle vectoriel

#### 1.4.1.1 $x$ , $s$ et $a$

Soit un ensemble connu de  $m$  mélanges  $(x_j)_{j \in [1, m]}$  de  $k$  observations chacun (généralement  $k \gg m$ ). Par analogie avec le cocktail-party,  $k$  peut représenter, par exemple, le nombre d'échantillons temporels discrets du signal. On pose chaque mélange comme combinaison linéaire de  $n$  signaux sources  $(s_i)_{i \in [1, n]}$  inconnus de  $k$  observations également

(de même  $k \gg n$  et  $m \geq n$  (voir 1.5.1)); soit :

$$\forall j \in [1, m], x_j = (x_{j1}, x_{j2}, \dots, x_{jk}) \quad (1.3)$$

$$\forall i \in [1, n], s_i = (s_{i1}, s_{i2}, \dots, s_{ik}) \quad (1.4)$$

Ces signaux sources  $s_i$  vérifient les *propriétés mathématiques fondamentales* suivantes du modèle de l'ACI (conséquence de leur indépendance, voir partie 1.7.1) :

- a. leur nombre  $n$  est connu;
- b. ils sont centrés;
- c. ils sont statistiquement mutuellement indépendants.

Mathématiquement, les deux dernières propriétés signifient que chaque signal source  $s_i$  est de moyenne nulle et que chaque couple  $(s_i, s_j)_{i \neq j}$  a une covariance nulle; soit :

$$\left\{ \begin{array}{ll} E[s_i] = 0 & , \forall i \in [1, n] \\ \text{cov}(s_i, s_j) = E[s_i s_j] - E[s_i]E[s_j] = 0 & , \forall j \neq i, j \in [1, n] \\ \text{var}(s_i) = \sigma_{s_i}^2 = E[s_i^2] - E[s_i]^2 = E[s_i^2] & \end{array} \right. \quad (1.5)$$

À titre de précision rigoureuse, les deux dernières égalités sont les égalités caractéristiques de variables décorréelées et non indépendantes. Cependant, des variables indépendantes sont décorréelées; donc ces égalités s'appliquent tout de même aux  $s_i$ . On étudiera plus en détail en 1.7 ces deux notions.

En outre, étant donné que les  $s_i$  sont des vecteurs lignes, le carré  $s_i^2$  est un abus de langage et représente en fait un produit scalaire. Pour y remédier, on pose l'égalité suivante, qui sera admise par la suite, pour tout vecteur ligne :

$$s_i^2 = s_i s_i^T \quad (1.6)$$

En introduisant les  $m \times n$  coefficients réels  $(a_{ji})_{(j,i) \in [1,m] \times [1,n]}$ , par définition inconnus eux aussi, on obtient les expressions vectorielles de l'ACI :

$$x_{jl} = \sum_{i=1}^n a_{ji} s_{il}, \forall (j, l) \in [1, m] \times [1, k] \quad (1.7)$$

$$x_j = \sum_{i=1}^n a_{ji} s_i \quad (1.8)$$

Ainsi, à partir de (1.8) et (1.5), il apparaît, par linéarité, que chaque mélange est de moyenne nulle :

$$E[x_j] = 0, \forall j \in [1, m] \quad (1.9)$$

#### 1.4.1.2 Modèle bruité

Nous pouvons ouvrir ici une parenthèse, à titre de précision, en remarquant que si l'on voulait vraiment être rigoureux et convenablement représenter le phénomène physique de mélanges de signaux, il faudrait considérer la présence du bruit et donc rajouter une composante de bruit, typiquement gaussien, inhérente à chaque transformation. En posant  $(n_j)_{j \in [1,m]}$  les  $m$  composantes de bruit gaussien de  $k$  observations chacune, l'équation (1.8) se réécrit :

$$\forall j \in [1, m], x_j = \sum_{i=1}^n a_{ji} s_i + n_j \quad (1.10)$$

Dans la suite, nous ne traiterons pas de l'effet du bruit sous cette forme théorique. Mais son étude pourrait être un excellent travail d'investigation future; c'est pourquoi nous la mentionnons en fin de mémoire comme une de nos recommandations. Ainsi, dans toute la suite, le bruit est considéré comme absent et l'équation vectorielle de l'ACI est (1.8). Pour plus de détails, il est possible de se référer aux travaux suivants (Hyvärinen, 1999c,d, 1998a). Refermons la parenthèse.

### 1.4.1.3 $y$ et $b$

D'après la partie 1.2, on sait que le but de l'ACI est de calculer les estimées des signaux sources, appelées « composantes indépendantes » (on utilisera l'abréviation « CI »), à partir des mélanges.

Pour ce faire, on note, tout d'abord,  $(y_i)_{i \in [1, n]}$  la composante indépendante, estimée de  $(s_i)_{i \in [1, n]}$ , telle que, suivant (1.4) :

$$\begin{cases} y_i = \tilde{s}_i \\ y_i = (y_{i1}, y_{i2}, \dots, y_{ik}) = (\tilde{s}_{i1}, \tilde{s}_{i2}, \dots, \tilde{s}_{ik}), \forall i \in [1, n] \end{cases} \quad (1.11)$$

Il est important de noter que l'utilisation du même indice  $i$  ici est un abus d'écriture car comme nous le verrons dans la partie 1.6.1, l'ordre des CIs n'est pas forcément respecté. Mais par souci de clarté, nous utilisons cette convention pour le moment.

Par définition, les  $y_i$  présentent les mêmes propriétés mathématiques (1.5) que les  $s_i$ ; soit en adaptant la notation :

$$\begin{cases} E[y_i] = 0 & , \forall i \in [1, n] \\ \text{cov}(y_i, y_j) = 0 & , \forall j \neq i, j \in [1, n] \\ \text{var}(y_i) = \sigma_{y_i}^2 = E[y_i^2] \end{cases} \quad (1.12)$$

D'autres propriétés seront énoncées en 1.6.2, après développement de certaines notions.

Ensuite, si l'on introduit, de même qu'en 1.4.1.1,  $n \times m$  coefficients réels  $(b_{ij})$ , on peut exprimer chacune des CIs comme combinaison linéaire des mélanges, sous la forme :

$$y_{il} = \sum_{j=1}^m b_{ij} x_{jl}, \forall (i, l) \in [1, n] \times [1, k] \quad (1.13)$$

$$y_i = \sum_{j=1}^m b_{ij} x_j \quad (1.14)$$

Le but est donc de calculer les  $b_{ij}$ , pour, à partir des  $x_j$ , obtenir les  $y_i$ . On reviendra sur ce calcul en 1.4.2.2.

## 1.4.2 Modèle matriciel

### 1.4.2.1 $X, S$ et $A$

En notation matricielle, on introduit les *matrices des signaux mélangés* (ou matrice des mélanges)  $X$  et *des signaux sources*  $S$ , de tailles respectives  $[m, k]$  et  $[n, k]$ , définies par :

$$X = (x_{jl})_{(j,l) \in [1,m] \times [1,k]} = (x_1, x_2, \dots, x_m)^T \quad (1.15)$$

$$S = (s_{il})_{(i,l) \in [1,n] \times [1,k]} = (s_1, s_2, \dots, s_n)^T \quad (1.16)$$

où le vecteur ligne  $(x_j)_{j \in [1,m]}$  (respectivement  $(s_i)_{i \in [1,n]}$ ) est le  $j^{\text{ième}}$  mélange de  $X$  (resp. le  $i^{\text{ième}}$  signal source de  $S$ ), tel que défini en 1.4.1.1.

En outre, à partir de (1.5), on peut déduire de telles propriétés pour  $S$ , en passant des données vectorielles à matricielles; à savoir que :

$$\begin{cases} E[S] = 0 \\ C_S = E[SS^T] = \text{diag}(\sigma_{s_1}^2, \sigma_{s_2}^2, \dots, \sigma_{s_n}^2) \end{cases} \quad (1.17)$$

où  $C_S = E[SS^T]$  est la matrice de covariance de  $S$ , diagonale car  $S$  est décorrélée (covariance nulle des  $(s_i, s_j)$ ).

On introduit également la *matrice  $A$  de mélange*, de dimension  $[m, n]$  définie, suivant le même principe que (1.15) et (1.16), par :

$$A = (a_{ji})_{(j,i) \in [1,m] \times [1,n]} = (a_1, a_2, \dots, a_m)^T \quad (1.18)$$

À partir de ces trois matrices et des équations (1.7), (1.8) et (1.17), on obtient, d'une part, l'égalité matricielle caractéristique de l'ACI :

$$X = A S \quad (1.19)$$

et, d'autre part, la preuve de la nullité théorique de la moyenne de  $X$  :

$$E[X] = 0 \quad (1.20)$$

Cette équation explique le fait que les données à traiter par l'ACI, dans le cas pratique, subissent comme premier prétraitement, d'être centrées (annulation de leur moyenne); comme nous le verrons en 3.1.1.

En outre, si l'on considère les lignes  $a_j$  de  $A$ , l'équation (1.8) peut se réécrire :

$$\forall j \in [1, m], x_j = a_j S \quad (1.21)$$

et chaque  $a_j$  apparaît alors comme la contribution, le « poids » de chacun des signaux sources au mélange  $j$ .

#### 1.4.2.2 $Y$ et $B$

À partir de (1.11), on peut introduire la *matrice  $Y$  des composantes indépendantes*, matrice estimée de la matrice  $S$  des signaux sources. Suivant le même principe que (1.16), on a :

$$Y = \tilde{S} = (y_{il})_{(i,l) \in [1,n] \times [1,k]} = (y_1, y_2, \dots, y_m)^T \quad (1.22)$$

où chaque vecteur ligne  $y_i$  est une CI, similaire à (1.11).

De plus,  $Y$  satisfait aux mêmes propriétés que (1.17); à savoir :

$$\begin{cases} E[Y] = 0 \\ C_Y = E[YY^T] = \text{diag}(\sigma_{y_1}^2, \sigma_{y_2}^2, \dots, \sigma_{y_n}^2) \end{cases} \quad (1.23)$$

Finalement, de même qu'en (1.18), on peut regrouper les  $b_{ij}$  en une matrice  $B$ , de taille  $[n, m]$ , telle que :

$$B = (b_{ij})_{(i,j) \in [1,n] \times [1,m]} = (b_1, b_2, \dots, b_n)^T \quad (1.24)$$

On peut alors exprimer l'équation (1.14) sous forme matricielle :

$$Y = BX \quad (1.25)$$

Cette équation représente la *deuxième équation caractéristique de l'ACI*, après (1.19).

L'objectif devient alors, d'un point de vue matriciel, le calcul de  $Y$  à partir de  $X$ . Il est donc évident que la matrice  $B$  joue un rôle essentiel dans le modèle de l'ACI : elle est nommée la *matrice de séparation*. Elle apparaît, de plus, par (1.25), comme l'estimée de l'inverse de  $A$ , soit :

$$B = \widetilde{A}^{-1} \quad (1.26)$$

En comparant (1.25) à (1.19), on a :

$$Y = BX = (BA)S = GS \quad (1.27)$$

La matrice  $G = BA$  est appelée *matrice globale* (Amari et al., 1996; Cichocki et al., 2003) ou *matrice quasi-identité* (Cardoso et Laheld, 1996). Vu qu'idéalement, on veut que  $Y = S$ , on peut donc intuitivement aux premiers abords (on verra en pratique en 1.6.3 qu'il en est autrement) que  $G = I$ .

Suivant ce raisonnement, avec  $\widetilde{A}$ , l'estimée de  $A$ , qui se retrouve être l'inverse de  $B$ , on obtient la version *estimée* de (1.19) :

$$X = \widetilde{A}\widetilde{S} = \widetilde{A}Y \quad (1.28)$$

Sans rentrer dans les détails de 1.6.3 sur la nature exacte de  $G$ , à partir des premières variables de l'ACI qui viennent d'être présentées (d'autres seront présentées dans le cha-



pitre 3), nous pouvons faire un résumé synthétique de leurs relations, dépendamment de l'approche utilisée.

### 1.4.2.3 Approches et estimées

Comme nous le verrons dans le chapitre 4, il existe deux approches expérimentales possibles : l'approche théorique et l'approche pratique.

En théorie,  $S$  et  $A$  sont inconnues. Cela signifie que les variables  $Y$ ,  $B$  et  $G$  n'ont pas lieu d'être; dans le sens où ce ne sont plus des estimées mais directement les variables que l'on recherche, car on ne peut estimer des inconnues. Ainsi, le modèle se simplifie en  $Y = S$ ,  $B = A^{-1}$  et  $G = I$ .

Dans le cas pratique, le but est de comparer des données réelles ( $S$  et  $A$ ), dont on dispose initialement, avec leurs estimées calculées par l'ACI  $(Y, \tilde{A})$ .  $S$  et  $A$  sont alors connues et les trois variables ( $Y$ ,  $B$  et  $G$ ) sont bien des données pertinentes. Le tableau I récapitule la nomenclature et les liens existant entre les variables initiales ( $S$ ,  $A$ ,  $A^{-1}$ ) et leurs estimées  $(\tilde{S}, \tilde{A}, \tilde{A}^{-1})$ ; sachant que la relation entre  $A^{-1}$  et  $B$  s'obtient directement à partir de la définition de  $G$ , et que celle entre  $A$  et  $\tilde{A}$  se démontre à partir de (1.28) :  $X = \tilde{A}Y = (\tilde{A}G)S = AS$ .

Tableau I

Variables de l'ACI et estimées

Intitulé	Variable initiale	Estimée	Relation
Matrice des sources	$S$	$Y$	$Y = GS$
Matrice de mélange	$A$	$\tilde{A}$	$A = \tilde{A}G$
Matrice de séparation	$A^{-1}$	$B$	$B = GA^{-1}$

À présent, nous allons nous intéresser aux conditions et hypothèses globales qui régissent l'analyse en composantes indépendantes.

## 1.5 Conditions

Il existe quatre conditions principales à l'ACI (Hyvärinen, 1999e) :

- a. il doit y avoir au moins autant de mélanges que de sources;
- b.  $A$  se doit d'être une matrice de rang plein;
- c. les sources doivent être statistiquement indépendantes;
- d. au plus une source peut être gaussienne.

Nous les détaillons les unes après les autres.

### 1.5.1 Au moins autant de mélanges que de sources

Tout d'abord, l'une des conditions importantes de l'ACI est que le nombre  $m$  de mélanges soit supérieur ou égal au nombre  $n$  de sources, soit :

$$m \geq n \tag{1.29}$$

Cela vient du fait que, si  $m < n$ , alors on se retrouve avec moins d'équations (nombre de mélanges, de combinaisons linéaires) que d'inconnues (signaux sources); d'où si des solutions existent, il y en a une infinité ! En outre (Hyvärinen, 1999e), en s'assurant que  $m \geq n$ , on se retrouve dans le cas inverse : plus d'équations que d'inconnues et s'il existe des solutions, il existe des méthodes pour réduire ce nombre d'équations; l'analyse en composantes principales (ACP) en est une (on en discute en annexe 2). On se retrouve alors typiquement avec  $m = n$ ; c'est pourquoi le choix est généralement de considérer :  $m = n$ . On rappelle que  $n$  est supposé connu.

### 1.5.2 A matrice de rang plein

Une matrice de rang plein est une matrice qui est inversible ou au moins pseudo-inversible. C'est une condition évidente pour pouvoir calculer  $B$ .

### 1.5.3 Indépendance statistique des composantes

Il s'agit ici d'un simple rappel de l'hypothèse fondamentale de l'ACI. Cette condition sera plus explicitée dans la partie 1.7, traitant de la notion d'indépendance.

### 1.5.4 Au plus une source gaussienne

La dernière condition est qu'au plus une source ou composante soit de distribution gaussienne. On en dit pas plus pour le moment car de même, cette condition sera expliquée en 1.7.

## 1.6 Ambiguïtés

Il existe deux ambiguïtés au modèle de l'ACI (Hyvärinen et Oja, 2000, 1997). En effet, il est impossible de déterminer l'ordre des composantes indépendantes, ni leurs variances ou énergies. Le terme « ambiguïtés » signifie que changer un des ces deux paramètres sur les signaux sources n'entraîne aucune modification sur les mélanges et a fortiori aucune non plus sur les composantes indépendantes résultantes. Après avoir explicité ces ambiguïtés, des hypothèses seront posées afin d'annuler leurs effets.

### 1.6.1 Ordre

La première ambiguïté à noter est qu'il est impossible de connaître a priori l'ordre des composantes. Considérons pour le prouver trois signaux initiaux ( $s_1$ ,  $s_2$  et  $s_3$ ) que l'on mélange et envoie à l'ACI dans cet ordre. Une idée a priori pertinente serait de penser

que les trois composantes indépendantes ( $y_1$ ,  $y_2$  et  $y_3$ ) rendues en sortie de l'ACI sont les estimées respectives des trois signaux dans le même ordre. Or, il n'en est rien.

En effet, soit  $X = (x_1, x_2, x_3)^T$  et  $S = (s_1, s_2, s_3)^T$ . On décide de modifier l'ordre des signaux  $s_i$  en introduisant une matrice de permutation  $P$ , dont c'est le but.

On rappelle qu'une matrice de permutation est la matrice nulle où se trouve un et seul 1 par ligne et colonne. En d'autres termes, c'est la matrice identité (qui est un cas particulier de matrice de permutation), dont on a interverti les lignes et/ou les colonnes.

Soit, par exemple,

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (1.30)$$

Après permutation, on obtient les nouveaux signaux permutés  $S_P$ , définis par :

$$S_P = PS = (s_3, s_1, s_2)^T \quad (1.31)$$

L'ordre est donc passé de  $\{1,2,3\}$  à  $\{3,1,2\}$  et de plus, cette matrice  $P$  a modifié la nature des mélanges  $X$ , ce que nous ne souhaitons pas. Pour remédier au problème, il suffit d'introduire l'inverse de la matrice de permutation  $P^{-1}$  (toujours inversible car  $\det P = \pm 1$ ) telle que :

$$X = AP^{-1}PS = A_P S_P \quad (1.32)$$

Ainsi,  $A_P = AP^{-1}$  devient la nouvelle matrice de proportion des mélanges (permutée !), toute aussi inconnue que  $A$  en théorie et  $S_P$  est la matrice des signaux permutés. L'ordre a été modifié mais sans aucune modification sur les mélanges, ni sur les composantes indépendantes, puisqu'elles ne dépendent uniquement que des mélanges et du nombre de sources, qui sont restés inchangés. Ainsi, en sortie de l'ACI, on obtient bien  $y_1$ ,  $y_2$  et  $y_3$  mais il est impossible de déterminer si leur ordre de sortie correspond bien à l'ordre d'entrée sans permutation  $\{1,2,3\}$ , avec permutation  $\{3,1,2\}$  ou bien un autre ordre !

Par conséquent, l'ordre ne peut être considéré comme un critère de repérage des composantes, dans le cas pratique; ce qui est bien dommage, comme nous pourrions le constater par la suite, dans la partie 4.1.2.

### 1.6.2 Énergies

La seconde ambiguïté de l'ACI est la valeur non identifiable des énergies des composantes indépendantes (Hyvärinen et Oja, 1997). Le terme « énergie » désigne la variance pour une variable aléatoire et la matrice de covariance pour une matrice; soit (1.5), (1.17) pour  $S$  et (1.12), (1.23) pour  $Y$  :

$$E_{s_i} = E[s_i^2] = \sigma_{s_i}^2 \quad (1.33)$$

$$E_S = E[SS^T] \quad (1.34)$$

Il en va de même pour  $y$  et  $Y$ .

Suivant le même principe que la partie précédente, on décide de modifier les énergies des signaux sources, en utilisant un facteur d'échelle. Ainsi, au lieu de considérer  $S$ , considérons  $S_\alpha = \alpha S$ , avec  $\alpha$  un réel non nul. L'énergie de ces signaux (toujours centrés) devient, après calcul,  $E_{S_\alpha} = \alpha^2 E_S$ .

À noter que  $\alpha$  peut être une variable aléatoire, pas seulement une constante. En effet, en posant  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ , avec  $n$  le nombre de signaux, chaque  $\alpha_i$  est le facteur de normalisation propre au signal  $i$ .

L'énergie des mélanges  $E_X$  est alors également modifiée. Pour résoudre ce problème, il suffit de multiplier la matrice de mélange  $A$  par le facteur  $1/\alpha$ . Ainsi, la nouvelle matrice de proportion des mélanges est  $A_\alpha = \frac{1}{\alpha} A$ , toujours inconnue en théorie et l'énergie de  $X$  est conservée (il suffit de calculer  $E_{A_\alpha S_\alpha}$ ).

Il apparaît donc qu'il est possible de modifier les énergies des signaux sources, sans modifier celle des mélanges, ni celle des composantes indépendantes; d'où une ambiguïté d'énergie. Un cas particulier de celle-ci est l'ambiguïté de signe; car, si  $\alpha$  ou des  $\alpha_i$  sont négatifs, le signe des sources correspondantes est modifié, sans altérer celui des composantes ou des mélanges.

Pour annuler cette ambiguïté d'énergie, on impose *les énergies de tous les signaux  $s_i$  égales à 1*. D'après (1.17), la matrice de covariance de  $S$  est ainsi réduite à la matrice identité; soit :

$$E_{s_i} = E[s_i^2] = 1 \quad (1.35)$$

$$E_S = E[SS^T] = I \quad (1.36)$$

Par analogie, les deux équations précédentes s'appliquent également aux  $y_i$  et à  $Y$ .

On peut noter, comme simple remarque pour le moment, que l'une des applications de cette hypothèse sera le « blanchiment » des mélanges (voir 3.1.2).

Les propriétés statistiques théoriques de  $S$  et  $Y$  se résument donc à :

$$\begin{cases} E[S] = E[Y] = 0 \\ E[SS^T] = E[YY^T] = I \end{cases} \quad (1.37)$$

### 1.6.3 Matrice globale $G$

Ces deux ambiguïtés peuvent se résumer en affirmant (Giannakopoulos, 1998), qu'en sortie de l'ACI, contrairement à ce qui était affirmé de façon intuitive en 1.4.2.2, les composantes indépendantes ne sont pas identiquement égales aux signaux sources; mais le sont à une permutation près (ordre) et/ou à un facteur d'échelle près (énergie et possiblement signe) :  $Y$  et  $S$  sont dits (Cardoso, 2002) *essentiellement égaux*.

Mathématiquement (Bayliss et al., 1997; Amari et al., 1996; Giannakopoulos, 1998), à partir de (1.27), cela se traduit par l'égalité :

$$G = BA = PD \quad (1.38)$$

où :

- a.  $P$  est une matrice de permutation, de taille  $[n, n]$ , du même type général que celle en 1.6.1, symbolisant l'ambiguïté de l'ordre;
- b.  $D$  est une matrice diagonale scalaire, de taille  $[n, n]$  :  $D = \text{diag}(d_1, d_2, \dots, d_n)$ , avec les  $(d_i)_{i \in [1, n]}$  réels non nécessairement identiques, qui symbolise les ambiguïtés de l'énergie.

Plus précisément, l'hypothèse d'unicité des énergies de  $S$  et  $Y$  entraîne que les termes diagonaux de  $D$  sont  $\pm 1$  (ou encore  $|D| = I$ ); ce qui équivaut à dire que *les sources et les composantes indépendantes sont égales au signe près*; et que l'ambiguïté d'énergie, qui se traduisait par la présence de scalaires dans  $D$ , est réduite à une ambiguïté de signe.

La preuve est la suivante. À partir de (1.27) et (1.37), l'énergie de  $Y$  devient :

$$E[YY^T] = GE[SS^T]G^T = GG^T = I \quad (1.39)$$

On en déduit anecdotiquement l'orthogonalité de  $G$  et d'après (1.38), on obtient alors :

$$GG^T = I = PDD^T P^T = PD^2 P^T \quad (1.40)$$

Or,  $P$ , étant une matrice de permutation, est orthogonale et l'équation précédente se réduit à  $D^2 = I$ ; d'où le résultat et par la même occasion, l'explication de l'illustration de la partie 1.2.3.

Mathématiquement, en introduisant la fonction  $\sigma$  de permutation, qui traduit l'ambiguïté d'ordre, la conclusion précédente s'exprime sous la forme :

$$y_{\sigma(i)} = \pm s_i \quad (1.41)$$

Ce qui signifie que le signal source  $s_i$  est estimé, au signe près, par la composante indépendante  $y_j$ , où  $j = \sigma(i)$  est l'image de  $i$  par la permutation  $\sigma$ .

Ainsi, en résumé, dans le cadre théorique de l'ACI :

- a.  $Y$  et  $S$  sont *essentiellement égaux* : c'est-à-dire égaux ou à l'ordre près et/ou au signe près;
- b.  $G$  est l'identité ou une matrice de permutation ( $P$ ) ou une matrice diagonale de termes  $\pm 1$  ( $D$ ) ou le produit des deux ( $PD$ );
- c. d'après le tableau I,  $\tilde{A}$  et  $B$  sont donc respectivement les estimées de  $A$  et  $A^{-1}$ , à une permutation et/ou au signe près.

#### 1.6.4 Indice de performance $E_1$

L'équation (1.38) montre, en outre, que, dans le cas où l'ACI a correctement extrait les composantes indépendantes, la matrice globale  $G = BA$  a une forme bien précise : produit d'une matrice de permutation  $P$  par une matrice diagonale scalaire  $D$ . Un outil mathématique utile pour juger de la bonne performance d'un algorithme de l'ACI serait donc le calcul de la différence entre  $G$  et ce produit car alors, plus celle-ci serait faible et meilleure serait la décomposition de l'ACI. Pour ce faire, il a été introduit un *indice de performance* (Amari et al., 1996; Giannakopoulos, 1998), qui mesure justement cet écart. Cet indice  $E_1$  est défini par l'expression suivante :

$$E_1 = \sum_{i=1}^n \left( \sum_{j=1}^n \frac{|g_{ij}|}{\max_k |g_{ik}|} - 1 \right) + \sum_{j=1}^n \left( \sum_{i=1}^n \frac{|g_{ij}|}{\max_k |g_{kj}|} - 1 \right) \quad (1.42)$$

où  $G = (g_{ij}) = BA$ .

Son intervalle de valeurs possibles est  $[0, 2n(n-1)]$ ; le maximum étant obtenu pour le cas (très) particulier d'une matrice  $G$  unitaire (que des 1). Il est nul (minimum) quand  $G$  est bien de la forme  $PD$  et a été construit ainsi pour satisfaire cette propriété.



La preuve est simple. On suppose pour la démonstration que  $G = PD$ . Ainsi, la normalisation en valeur absolue par le maximum de chaque ligne (ou colonne) englobe les valeurs  $\frac{|g_{ij}|}{\max_k |g_{ik}|}$  (ou  $\frac{|g_{ij}|}{\max_k |g_{kj}|}$ ) de la nouvelle matrice  $G$  entre 0 et 1 (1 compris). Ainsi,  $D = I$  et  $G = P$ , une matrice de permutation; c'est-à-dire que des zéros, excepté un et un seul 1 par ligne et colonne. La soustraction par 1 entraîne alors la nullité de  $E_1$  dans le cas parfait !

En d'autres termes,  $E_1 = 0$  indique une *extraction parfaite des signaux sources par l'ACI* (plutôt essentiellement parfaite, si l'on veut rester rigoureux); et représente bien un indice de mesure adéquat car plus il est faible, meilleure est cette extraction : sa minimalité est donc recherchée !

À noter que, d'une part, il existe un second indice de performance  $E_2$  (Giannakopoulos, 1998), qui est l'équivalent quadratique de  $E_1$ . En effet, au lieu de valeurs absolues, on considère des carrés; soit :

$$E_2 = \sum_{i=1}^n \left( \sum_{j=1}^n \frac{g_{ij}^2}{\max_k g_{ik}^2} - 1 \right) + \sum_{j=1}^n \left( \sum_{i=1}^n \frac{g_{ij}^2}{\max_k g_{kj}^2} - 1 \right) \quad (1.43)$$

D'autre part, il est intéressant de souligner que, dans la partie 2.1.2 de l'état de l'art, les performances des algorithmes de l'ACI considérés seront jugées et comparées en fonction des deux indices  $E_1$  (principalement) et  $E_2$ . De plus, notre protocole se basera sur une version modifiée de cet indice  $E_1$ , comme nous le verrons au chapitre 4.

À présent, intéressons-nous à l'un des principaux sujets de l'ACI : son indépendance.

## 1.7 Indépendance

On note, en premier lieu, que toutes les informations importantes voire primordiales sont données dans cette partie. Cependant, tous les détails et explications, pertinents et intéressants, sont données en annexe 1. De nombreuses allusions y seront donc faites.

La notion d'indépendance statistique est abordée dans (Hyvärinen, 1999e; Stone, 2002). Il ne faut pas la confondre avec la notion moins restrictive de décorrélation. Nous avons vu précédemment que le letimotiv de l'ACI est l'indépendance des composantes indépendantes recherchées. Intéressons-nous alors à sa définition mathématique; la définition plus « littéraire » est disponible à l'annexe 1.

### 1.7.1 Définition mathématique

Soit  $y_1, y_2, \dots, y_n$  des variables aléatoires (centrées pour simplifier) de densité de probabilité conjointe  $p(y_1, \dots, y_n)$ . Les variables  $y_i$  sont *mutuellement statistiquement indépendantes* si leur densité de probabilité vérifie (Hyvärinen, 1999e) :

$$p(y_1, \dots, y_n) = p_1(y_1)p_2(y_2) \dots p_n(y_n) = \prod_{i=1}^n p_i(y_i) \quad (1.44)$$

où  $p_i(y_i)$  est la densité marginale de  $y_i$ .

L'indépendance ne doit pas être confondue avec la décorrélation. En effet (Hyvärinen et Oja, 2000), deux variables aléatoires  $y_i$  et  $y_j$  sont décorrélées (covariance nulle), si elles vérifient :

$$\forall i \neq j, E[y_i y_j] - E[y_i]E[y_j] = 0 \quad (1.45)$$

alors que si elles sont indépendantes, elles vérifient, avec  $g_1$  et  $g_2$  deux fonctions intégrables, la condition plus restrictive suivante :

$$\forall i \neq j, E[g_1(y_i)g_2(y_j)] - E[g_1(y_i)]E[g_2(y_j)] = 0 \quad (1.46)$$

Ainsi, indépendance et décorrélation sont des notions différentes. Plus précisément, la décorrélation est un cas particulier de l'indépendance (prendre  $g_1 = g_2 = id$ ). On en déduit la propriété suivante : *deux variables aléatoires indépendantes sont décorrélées; la réciproque est fausse*. Un exemple illustrant la différence entre décorrélation et indépendance est disponible en annexe 1, section 2.

En outre, on rappelle que, d'après (1.45), si  $Y$  est un vecteur de variables aléatoires  $y_i$  décorréliées, alors la matrice de covariance  $C_Y$  de  $Y$  est diagonale.

### 1.7.2 Mesure d'indépendance : l'information mutuelle

Revenons à la définition de l'ACI par Cardoso qui affirme que l'on recherche les composantes linéaires les plus *indépendantes* possibles. Cela sous-entend qu'il faut trouver une manière de juger, de mesurer cette indépendance; c'est-à-dire qu'il faut résoudre une expression du genre (Cardoso, 2002) :

$$\min_A I(A^{-1}X) \quad (1.47)$$

où, à partir de (1.19),  $Y = \tilde{S} \approx A^{-1}X$  et  $I(Y)$  est une mesure de dépendance. La problématique de l'ACI se résume donc à minimiser une fonction de dépendance statistique sur  $Y$ . La question est de savoir quelle fonction utiliser.

Le premier à avoir posé une telle mesure est Comon (Comon, 1994), qui a introduit l'*information mutuelle*  $I(\cdot)$ . Celle-ci est la *divergence de Kullback-Leibler* (DKL)  $\delta(\cdot|\cdot)$  entre la densité de probabilité  $P_Y$  de  $Y$  et du produit des densités marginales de chacune de ses composantes  $\prod_i P_{Y_i}$ , soit :

$$I(Y) = \delta(P_Y | \prod_i P_{Y_i}) \quad (1.48)$$

La DKL de deux distributions de probabilités  $f$  et  $g$  non nulles est définie par :

$$\delta(f|g) = \int f \log \left( \frac{f}{g} \right) \quad (1.49)$$

et assure l'équivalence suivante :

$$\delta(f|g) = 0 \Leftrightarrow f = g \quad (1.50)$$

Ainsi, en revenant à notre problème d'indépendance et d'après la définition (1.44), on a :

$$I(Y) = 0 \Leftrightarrow \text{les composantes de } Y \text{ sont indépendantes} \quad (1.51)$$

Ainsi, pour rendre les composantes de  $Y$  indépendantes, il faut minimiser  $I(Y)$ . Par conséquent, d'après (1.47), cette information mutuelle apparaît comme une mesure de dépendance  $I(Y)$  possible. Nous verrons dans la partie 1.8 qu'il en existe d'autres mais avant cela, nous devons introduire et présenter les liens existant entre dépendance, corrélation et « non-gaussianité », terme que nous définissons en 1.7.3.2 et qui est le second terme majeur de l'ACI, après l'indépendance.

### 1.7.3 Dépendance, corrélation et non-gaussianité

Toute la théorie de cette section est issue des articles (Cardoso, 2004a, 2002, 2004b). La dépendance a été introduite dans la partie précédente. Voyons les deux autres notions.

#### 1.7.3.1 Approximations

Pour ce faire, il faut au préalable définir les meilleures approximations de  $P_Y$  au sens de la DKL (minimum de la distance). Ainsi, on définit la meilleure approximation :

- a. du produit :  $P_Y^P = \prod_i P_{Y_i}$ ;
- b. gaussienne :  $P_Y^G = \mathcal{N}(E[Y], C_Y)$ ;
- c. gaussienne du produit :  $P_Y^{P \wedge G} = \mathcal{N}(E[Y], \text{diag}(C_Y))$ .

où  $\mathcal{N}(E[Y], C_Y)$  est la distribution gaussienne d'espérance  $E[Y]$  et de matrice de covariance  $C_Y$  de  $Y$ ;  $\text{diag}(C_Y)$  est la matrice diagonale de termes diagonaux, ceux de  $C_Y$ .

A priori les expressions de ces approximations sont logiques. En effet, pour le produit, cela semble évident et la *dépendance* (ou information mutuelle) est alors définie comme :

$$I(Y) = \delta(P_Y | P_Y^P) \tag{1.52}$$

Pour la gaussienne, il semble pertinent de penser que la meilleure approximation est la normale de même moyenne et matrice de covariance que l'approximé ! Quant à l'approximation gaussienne du produit, elle est l'« intersection » des deux.

### 1.7.3.2 Non-gaussianité

Sous ces conditions, la *non-gaussianité*  $G(Y)$  d'un vecteur aléatoire  $Y$  est alors définie comme la DKL de  $P_Y$  et de  $P_Y^G$ ; soit :

$$G(Y) = \delta(P_Y | P_Y^G) \quad (1.53)$$

En d'autres termes, cette non-gaussianité mesure l'écart entre une variable aléatoire et une variable gaussienne, de mêmes moments statistiques d'ordre 1 (moyenne) et 2 (variance). Plus cet écart est faible, plus la variable est gaussienne.

### 1.7.3.3 Corrélation

La *corrélacion*  $C(Y)$  d'un vecteur aléatoire  $Y$  peut être définie comme l'information mutuelle de l'approximation gaussienne  $Y^G$  de  $Y$ , soit la DKL de  $P_Y^G$  et de  $P_Y^{P \wedge G}$  :

$$C(Y) = I(Y^G) = \delta(P_Y^G | P_Y^{P \wedge G}) \quad (1.54)$$

### 1.7.3.4 Équation caractéristique

Il est alors possible d'exprimer le lien entre dépendance, corrélation et non-gaussianité par une égalité, appelée *équation caractéristique* :

$$I(Y) + \sum_i G(Y_i) = C(Y) + G(Y) \quad (1.55)$$

soit sous une autre forme :

$$I(Y) = C(Y) - \sum_i G(Y_i) + G(Y) \quad (1.56)$$

Les preuves algébrique et même géométrique sont données en annexe 1, sections 3 et 4.

En guise de premières remarques, on peut noter que cette équation (1.56) permet, tout d'abord, de quantifier l'écart existant entre indépendance et décorrélacion (entre dépendance et corrélation) : c'est la différence de gaussianité entre  $Y$  et ses composantes; et

d'autre part, de relever la propriété suivante : *dans le cas de variables aléatoires gaussiennes, l'indépendance est réduite à la décorrélation; c'est-à-dire que deux variables aléatoires gaussiennes décorrélées sont indépendantes.*

La preuve est simple : si les variables sont gaussiennes, alors  $P_Y = P_Y^G$  et  $G(Y) = G(Y_i) = 0$ ; d'où  $I(Y) = C(Y)$ . Ainsi, dépendance équivaut à corrélation et on obtient le théorème.

Intéressons nous à présent aux « affinités » existantes entre composantes indépendantes et composantes gaussiennes.

#### **1.7.4 Indépendance et gaussianité font mauvais ménage**

À partir de l'équation (1.56), nous pouvons énoncer deux idées majeures sur la nature des « affinités », plutôt frigides, devant exister entre indépendance et gaussianité pour résoudre un problème d'ACI :

- a. on recherche les sources les moins gaussiennes possibles;
- b. il faut qu'au plus une le soit (condition nécessaire, non suffisante).

##### **1.7.4.1 Sources les moins gaussiennes possibles**

Si l'on considère la propriété mathématique de la non-gaussianité, invariante par transformation linéaire (soit  $G(Y) = G(A^{-1}X) = G(X)$ ), on peut remarquer que, vu que  $X$  est constant,  $G(X)$  et donc  $G(Y)$  le sont également. On obtient ainsi la propriété suivante (Cardoso, 2002) : « *minimiser la dépendance  $I(Y)$  revient à minimiser la corrélation  $C(Y)$  tout en maximisant la non-gaussianité  $G(Y_i)$  de chacune des composantes  $Y_i$  de  $Y$  (non-gaussianités marginales)* ».

En d'autres termes (Cardoso, 2004a), « *les composantes linéaires indépendantes recherchées sont celles qui sont les plus décorrélées et les moins gaussiennes possibles !* ».

Ce résultat, qui représente l'idée essentielle de l'ACI, met en lumière la direction à suivre pour résoudre un problème d'ACI : maximiser la non-gaussianité. On en dira plus dans la partie 1.8 concernant les fonctions de contraste.

#### 1.7.4.2 Au plus une source gaussienne

On sait que dans le cas de variables gaussiennes, indépendance et décorrélation sont équivalentes. En découle alors une autre propriété importante de l'ACI : *il est impossible d'extraire les composantes indépendantes si plus d'une source est gaussienne.*

Deux critères permettent de démontrer ce théorème quand toutes les sources sont gaussiennes (la propriété *une au plus* vient de Linnik (Cardoso, 2004b)) : le théorème de la limite centrale et la faiblesse de la décorrélation face à l'indépendance. Les démonstrations sont disponibles en annexe 1, section 5.

En outre, pour appuyer, si besoin était, l'idée de l'antipathie existant entre gaussianité et ACI, les conditions d'existence et d'unicité de la décomposition ACI sont présentées en annexe 1, section 6. On y montre qu'il en existe une infinité quand  $X$  est gaussienne, alors qu'elle est unique quand  $X = AS$ , avec  $S$ , composée de sources indépendantes.

#### 1.7.5 Conclusion

En résumé de cette partie sur l'indépendance, l'idée primordiale à retenir est que les composantes indépendantes doivent être les moins gaussiennes possibles et qu'au plus une seule peut l'être. Cela signifie que le but sera de maximiser les non-gaussianités. Cependant, leur définition (1.53) faisant intervenir la DKL est difficilement exploitable car elle nécessite la connaissance des densités de probabilité, généralement inconnues. Une autre solution est l'utilisation de *fonctions de contraste* (terme issu du lexique du domaine des réseaux de neurones). Ce sont des fonctions à minimiser ou maximiser, discriminatoires envers un « critère » (ici la non-gaussianité), afin d'atteindre un but (ici l'indépendance).

Typiquement, l'information mutuelle en est un exemple; car la minimiser rend les composantes indépendantes. Nous présentons maintenant les autres fonctions de contraste les plus fréquemment utilisées.

Remarquons avant d'aborder cette nouvelle notion, qu'il peut être intéressant, par curiosité, de jeter un coup d'oeil à l'annexe 2, dans laquelle l'ACI est comparée à deux autres méthodes classiques d'extraction de données, que sont l'analyse en composantes principales (ACP), autre analyse en composantes (d'où un intérêt certain) et la projection-poursuite (PP).

## 1.8 Fonctions de contraste

Les fonctions de contraste (Hyvärinen et Oja, 2000; Hyvärinen, 1999e,a, 1997a) que nous allons étudier sont, en premier lieu et principalement, le kurtosis puis la négentropie. Figurent également, plus à titre d'information que d'utilisation dans ce mémoire, les méthodes basées sur la théorie de l'information, à savoir l'information mutuelle, le maximum de vraisemblance et l'infomax. Mais avant cela, il est nécessaire d'introduire la notion mathématique de *statistiques d'ordres plus élevés*.

### 1.8.1 Statistiques d'ordres plus élevés

D'un point de vue statistique, le fait que la décorrélation soit une condition plus faible que l'indépendance signifie que les moments d'ordre 1 (moyenne) et 2 (variance) sur lesquels « travaille » la décorrélation sont insuffisants pour résoudre l'ACI. Ainsi, il est besoin de recourir à des moments d'ordres plus élevés (Karhunen et al., 1997) pour avoir plus d'information statistique et pouvoir ainsi récupérer des composantes, indépendantes et pas seulement décorrélées. On parle de « statistiques d'ordres plus élevés » (en anglais « Higher-Order Statistics » (HOS)). La conséquence de ces statistiques est la prise de conscience que les moments sont parfois insuffisants à décrire les caractéristiques statistiques de variables aléatoires; c'est pourquoi les cumulants (Comon, 1994; Mendel,



1991; Hyvärinen, 1999e; Giannakopoulos, 1998; Cardoso et Comon, 1996; De Lathauwer et al., 2000) ont été introduits et notamment le kurtosis, qui est le cumulants d'ordre 4. La définition des cumulants, dont celle du kurtosis, est largement explicitée en annexe 3. Intéressons-nous plus précisément au kurtosis comme fonction de contraste.

### 1.8.2 Kurtosis

On définit le kurtosis (Hyvärinen et Oja, 2000; Hurri, 1997; Giannakopoulos, 1998) d'une variable aléatoire  $y$  centrée (si elle ne l'est pas, lui soustraire sa moyenne au préalable) comme le cumulants d'ordre 4,  $\kappa_4(y)$  :

$$\text{kurt}(y) = \kappa_4(y) = E[y^4] - 3 (E[y^2])^2 \quad (1.57)$$

Par définition, *le kurtosis d'une variable gaussienne est nul*. Ainsi, le kurtosis apparaît naturellement comme une fonction de contraste possible, car discriminatoire par rapport à la gaussianité des variables.

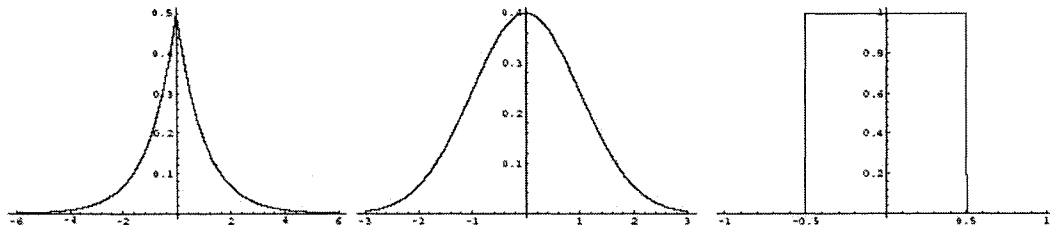


Figure 6 Supergaussienne, gaussienne et subgaussienne

On distingue deux types de variables : les *supergaussiennes*, de kurtosis positif et les *subgaussiennes*, de kurtosis négatif. Comme le montre la figure 6 (Hurri, 1997), les premières présentent une densité de probabilité, plutôt pointue (un exemple typique est la distribution laplacienne), alors que les dernières sont plus « allongées » (typiquement la distribution uniforme); au centre, la distribution gaussienne. On verra en 3.3.1.3 que leurs propriétés face à l'ACI sont différentes selon leurs natures.

Ainsi, le but est de maximiser un kurtosis positif ou de minimiser un kurtosis négatif. Pour ce faire, on recherche itérativement les directions  $b_i$ , telles que  $y_i = b_i X$ , qui maximisent ou minimisent (dépendamment du cas) le kurtosis de  $y_i$ , composante indépendante. Typiquement, des techniques du type « descente de gradient » sont utilisées.

De plus, le kurtosis présente l'avantage d'être simple, à la fois conceptuellement et pratiquement; car il est un opérateur semi-linéaire; à savoir, pour deux variables  $x$  et  $y$  et un scalaire quelconque  $\alpha$ , on a, par application de la définition :

$$\kappa_4(x + y) = \kappa_4(x) + \kappa_4(y) \quad (1.58)$$

$$\kappa_4(\alpha x) = \alpha^4 \kappa_4(x) \quad (1.59)$$

Par conséquent, dans le cas de l'ACI, le kurtosis d'une composante indépendante  $y_i$  s'exprime en fonction de ceux des mélanges, à partir de (1.14) :

$$\kappa_4(y_i) = \kappa_4 \left( \sum_{j=1}^m b_{ij} x_j \right) = \sum_{j=1}^m \kappa_4(b_{ij} x_j) = \sum_{j=1}^m b_{ij}^4 \kappa_4(x_j) \quad (1.60)$$

Cependant, il présente un énorme désavantage : celui d'être très sensible aux valeurs aberrantes (« outliers » en anglais). En effet, la puissance 4 dans sa définition ( $E[y^4]$ ) entraîne que si l'un des mélanges ou l'un des coefficients de la matrice de séparation extrait est erroné, alors la puissance 4 amplifie cet effet. Ainsi, le kurtosis, bien que (ou car) simple, apparaît insuffisamment efficace comme fonction de contraste; on lui préfère la « négentropie », anglicisme accepté du terme anglais « negentropy », qui se veut une généralisation du kurtosis.

### 1.8.3 Négentropie

Comme son nom l'indique, la négentropie ou « entropie normalisée négative » (Comon, 1994) découle de l'entropie dont nous rappelons à présent la définition.

### 1.8.3.1 Rappel sur l'entropie

L'entropie (Hyvärinen et Oja, 2000; Giannakopoulos, 1998; Stone et Porrill, 1998), une notion issue de la théorie de l'information et utilisée en théorie du codage, est une mesure du *désordre*. Typiquement, plus il y a du désordre, plus la valeur de l'entropie est élevée.

En effet, sa définition est, pour une variable aléatoire discrète  $y$  (typiquement  $y_i$  dans le modèle de l'ACI), de  $k$  valeurs possibles  $(y_l)_{l \in [1, k]}$  (typiquement  $y_{il}$ ) :

$$H(y) = - \sum_{l=1}^k p_l \log p_l \quad (1.61)$$

où  $p_l = P(y = y_l)$  est la probabilité que  $y$  soit égale à  $y_l$ .

Ainsi, dans le cas de la certitude (un  $p_l = 1$  et les autres à 0), on a  $H(y) = 0$  minimal; alors que dans le cas de l'incertitude (équiprobabilité  $p_l = 1/k$ ),  $H(y) = \log k$  est maximal.

Dans le cas d'un vecteur  $Y$  de variables aléatoires de densité de probabilité  $f(y)$ , l'entropie devient l'entropie différentielle, définie par :

$$H(Y) = - \int_Y f(y) \log f(y) dy = -E[\log f(y)] \quad (1.62)$$

Le résultat essentiel de l'entropie est le fait qu'à variance égale, de tous les types de distributions possibles, *c'est la distribution gaussienne qui présente l'entropie la plus élevée*. Cela signifie, d'une part, que la gaussienne est logiquement le symbole de l'aléatoire (difficulté à prédire la valeur de la variable) et, d'autre part, que l'entropie est une fonction de contraste pertinente (car discriminatoire vis-à-vis du caractère gaussien), à minimiser. Pour ce faire, on utilise une version légèrement modifiée de cette entropie : la *négentropie*.

### 1.8.3.2 Définition

En reprenant l'idée de la maximalité gaussienne de l'entropie, la négentropie est simplement définie comme la différence d'entropies entre une variable aléatoire gaussienne

$y_{\text{gauss}}$  et une variable quelconque  $y$  de même variance. Ainsi, la négentropie  $J(y)$  de cette variable  $y$  est :

$$J(y) = H(y_{\text{gauss}}) - H(y) \quad (1.63)$$

L'intérêt de cette définition est que  $J(y) = 0$  pour une gaussienne et est strictement positif dans tous les autres cas. Le but est donc de maximiser  $J(y)$  (maximiser l'écart, la non-gaussianité). La négentropie est la fonction de contraste utilisée dans l'algorithme FastICA, sur lequel s'est basé notre travail et qui sera présenté au chapitre 3.

Dans le cas d'un vecteur aléatoire  $Y$ , la négentropie s'exprime sous la forme adaptée de (1.63) :

$$J(Y) = H(Y_{\text{gauss}}) - H(Y) \quad (1.64)$$

où, en ce cas, ce ne sont plus les variances qui sont égales, mais les *matrices de covariance* de  $Y$  et  $Y_{\text{gauss}}$ , vecteur aléatoire gaussien.

Ainsi, le but est de maximiser la négentropie de  $Y$  ou de chaque  $y_i$  individuellement.

En outre, en développant l'expression de l'entropie dans l'équation ci-dessus, la négentropie apparaît alors comme l'expression de la divergence de Kullbak-Leibler (voir 1.7.2) entre  $Y$  et  $Y_{\text{gauss}}$ . Il y a donc bien la notion de mesure de distance à la gaussianité :

$$J(Y) = - \int_Y f(y) \log \frac{f(y)}{f(y_{\text{gauss}})} dy = \delta(Y|Y_{\text{gauss}}) \quad (1.65)$$

### 1.8.3.3 Approximations

L'expression de la négentropie (et de l'entropie aussi d'ailleurs) est très satisfaisante théoriquement parlant, par rapport aux statistiques. Cependant, elle est difficile à mettre en oeuvre pratiquement car nécessite la connaissance, peu aisée, des densités de probabilités; c'est pourquoi des approximations ont été utilisées pour parer ce problème.

Historiquement, la première utilisée (Jones et Sibson, 1987; Comon, 1994) est celle faisant intervenir les cumulants d'ordre 3 et 4 (voir l'annexe 3, section 2) :

$$J(y) \approx \frac{1}{12} \kappa_3^2 + \frac{1}{48} \kappa_4^2 \quad (1.66)$$

avec  $y$  variable centrée, de variance unitaire (analogue à notre problème, d'après (1.35)). On a bien  $J(y_{\text{gauss}}) = 0$  et  $J(y) > 0$  sinon. Cependant, cette approximation présente l'inconvénient du kurtosis : sa sensibilité aux valeurs aberrantes. Cette approximation n'est donc pas valide.

Hyvärinen (Hyvärinen, 1998b) a introduit une nouvelle forme :

$$J(y) \approx c [E\{G(y)\} - E\{G(\nu)\}]^2 \quad (1.67)$$

où :

- a.  $c$  est une constante;
- b.  $y$  et  $\nu$  sont deux variables respectivement quelconque et gaussienne, centrées et de variances unitaires; ce qui concorde avec la définition de l'entropie et nos hypothèses sur les composantes indépendantes;
- c.  $G$  est une fonction non quadratique (car sinon,  $J$  est nulle), robuste; c'est-à-dire insensible aux valeurs extrêmes.

En choisissant  $G(y) = y^4$ , on tombe sur le carré du kurtosis, défini en (1.57). Cette négentropie, telle qu'introduite, en est donc une généralisation; tout en étant meilleure. Nous en dirons plus au chapitre 3, où cette approximation qui sert de base à l'algorithme Fast-ICA, sur lequel nous nous sommes appuyés pour nos tests pratiques, sera plus amplement détaillée.

À noter, pour les curieux, que Girolami et Fyfe (Girolami et Fyfe, 1996) ont développé des algorithmes ACI basés sur le kurtosis et la négentropie pour solutionner des problèmes de projection-poursuite (voir annexe 2, section 2).

## 1.8.4 Autres

À présent, nous nous intéressons succinctement aux autres fonctions de contraste possibles (Hyvärinen et Oja, 2000; Hyvärinen, 1999e). Nous ne rentrerons pas dans les détails car ils ne seront pas utilisés ultérieurement.

### 1.8.4.1 Information mutuelle

La première de celle-ci est l'information mutuelle, que nous avons déjà présentée en 1.7.2, sous la forme d'une DKL, mais qui peut également s'exprimer sous une forme dérivée de l'entropie (Bell et Sejnowski, 1995a; Almeida, 2001). En effet, pour un vecteur  $Y$ , on a :

$$I(Y) = I(y_1, y_2, \dots, y_n) = \sum_{i=1}^n H(y_i) - H(Y) \quad (1.68)$$

En utilisant la définition (1.61) de l'entropie, on retombe sur l'expression (1.49) de la DKL; d'où la relation. Et comme nous l'avons dit, le but serait ici de minimiser cette information mutuelle.

Cette information mutuelle est également exprimable en fonction de la négentropie, avec les  $y_i$  de variances unitaires, sous la forme  $I(Y) = C - \sum_{i=1}^n J(y_i)$ , où  $C$  est une constante, dont l'expression exacte et les détails sont disponibles dans (Comon, 1994). Cette expression renforce l'équivalence entre l'information mutuelle (à minimiser) et la négentropie (à maximiser).

### 1.8.4.2 Maximum de vraisemblance

La deuxième présentée succinctement est le maximum de vraisemblance (ou « maximum of likelihood » (ML) en anglais), dont on calcule ici son « log de vraisemblance »  $L$  :

$$L = \sum_{l=1}^k \sum_{i=1}^n \log p_i(b_i x_l^c) + k \log |\det B| \quad (1.69)$$

où  $p_i$  est la densité de probabilité de  $s_i$ , supposée connue (!). Pour plus de détails et d'explications, se référer à (Pham et al., 1992).

### 1.8.4.3 Infomax

Enfin, la dernière fonction de contraste est le principe de l'infomax, qui découle de la théorie des réseaux de neurones. En effet, elle est basée sur la maximisation de l'entropie de sortie d'un réseau de neurones avec des sorties non-linéaires. Si l'on considère  $X$  comme l'entrée du réseau, les sorties sont alors de la forme  $g_i(b_i X)$ , où les  $g_i$  sont des fonctions scalaires non-linéaires (dérivées des  $G_i$ , succinctement présentées en 1.8.3.3) et les  $b_i$  sont les vecteurs de poids des neurones. Le but est alors de maximiser l'entropie des sorties  $H(g_1(b_1 X), \dots, g_n(b_n X))$ . Il a été montré (Cardoso, 1997; Pearlmutter et Parra, 1997) que cette fonction de contraste était équivalente au maximum de vraisemblance, équivalente à celles déjà présentées.

Cette revue de détails des fonctions de contraste conclut le premier chapitre sur la théorie de l'ACI. En résumé, il faut retenir principalement les hypothèses mathématiques vérifiées par  $X$ ,  $S$  et  $Y$  et le fait que c'est la négentropie approximée qui sera utilisée comme base de notre algorithme. Les raisons en sont notamment données dans l'état de l'art, que nous présentons dès maintenant.

## CHAPITRE 2

### ÉTAT DE L'ART

Après avoir présenté la théorie de l'analyse en composantes indépendantes, nécessaire à la bonne compréhension de la suite du mémoire, nous nous intéressons dans ce deuxième chapitre à l'état de l'art de l'ACI utilisée sur des signatures spectrales. Celui-ci se fera en deux temps : tout d'abord, nous présenterons et expliquerons le choix de l'algorithme FastICA comme algorithme de base de nos travaux; et dans un second temps, nous présenterons les travaux existant sur l'extraction des signatures spectrales par l'ACI.

#### 2.1 Choix de l'algorithme ACI

##### 2.1.1 Problématique

Comme l'affirme Cardoso (Cardoso, 2002) : « il n'y a pas d'algorithme ACI universel ». Il existe différents algorithmes pour différentes approches ! (voir à ce sujet, l'annexe 2, section 1.5, qui explique qu'il n'existe pas une seule méthode de l'ACI mais plusieurs, contrairement à l'ACP).

Plus précisément (Hyvärinen, 1999e), la *méthode ACI* choisie est l'association de deux éléments : la *fonction de contraste* et l'*algorithme d'optimisation*. Chacun de ces deux éléments a son importance. En effet, la fonction de contraste détermine les propriétés statistiques de la méthode ACI; alors que l'algorithme détermine ses propriétés matérielles (convergence, charge informatique, ressources mémoire), qui ne sont pas négligeables. Une bonne méthode ACI forme donc un tout.

Un choix de méthode optimale peut donc se résumer en deux questions :

- a. comment veut-on estimer les composantes indépendantes ?;
- b. quel type d'algorithme utiliser ?.



### 2.1.1.1 Méthode d'estimation

Pour la méthode d'estimation, il existe deux façons de calculer les composantes (voir 1.4.2.2) :

- a. *une par une* : c'est le mode déflation, soit extraire les  $b_i$  séparément;
- b. *toutes d'un coup* : c'est le mode symétrique, soit extraire  $B$  en entier.

Ces deux modes d'estimation nécessitent, en outre, une règle d'apprentissage récursive sur la valeur de  $B$  ou des  $b_i$  et un critère associé à minimiser ou maximiser (Giannakopoulos, 1998). Il existe plusieurs règles et plusieurs critères; d'où la seconde interrogation.

### 2.1.1.2 Types d'algorithmes

Trois types d'algorithmes existent (Hyvärinen et Oja, 1997; Hyvärinen, 1997a) :

- a. les algorithmes *par « lots »* (ou « batch-mode algorithm » en anglais);
- b. les algorithmes *adaptatifs*;
- c. les algorithmes du *point fixe* (ou « fixed-point algorithms » (FP) en anglais).

Les algorithmes par lots sont basés sur la minimisation ou maximisation de fonctions de contraste, ne faisant intervenir que les cumulants d'ordre plus élevés (Cardoso, 1992; Comon, 1994). Ceux-ci sont parmi les premiers à avoir été implantés historiquement, donc, bien que performants à l'époque, leurs capacités actuelles s'avèrent limitées. En effet, ils nécessitent des opérations matricielles difficiles (notions de tenseurs matriciels (Cardoso et Comon, 1996; De Lathauwer et al., 2000)), donc une charge informatique élevée et sont sensibles aux valeurs aberrantes; c'est pourquoi les algorithmes adaptatifs leur sont préférés.

Ces derniers sont basés sur des méthodes de gradient stochastiques (généralement, descente de gradient), dont les implémentations peuvent être assimilables à celles de réseaux de neurones; dans le sens où  $B$  ou  $b_i$  sont considérés comme les poids mis à jour du

réseau (Amari et al., 1996; Bell et Sejnowski, 1995b; Delfosse et Loubaton, 1995; Hyvärinen et Oja, 1996a; Jutten et Herault, 1991; Cichocki et al., 1996). Même si plus efficaces que les précédents, ces algorithmes présentent le défaut de converger lentement vers la solution et également d’être sensible aux valeurs aberrantes.

Pour y remédier, une troisième forme a été développée par Hyvärinen (Hyvärinen, 1999a; Hyvärinen et Oja, 1997; Hyvärinen, 1999b, 1997a; Hyvärinen et Oja, 1996a) : les algorithmes du point fixe, basés sur le kurtosis ou la négentropie. Son nom (point fixe) vient du fait que la règle récursive, comme nous le verrons en détail dans le chapitre 3, présente la forme caractéristique d’une équation du type point fixe (ou encore appelé « principe de contraction de Banach ») en  $B$ ; à savoir  $B = f(B)$ , avec  $f$  une fonction convergente.

Pour mener à bien nos expérimentations, nous avons décidé d’opter pour l’algorithme FastICA de Hyvärinen, qui est une forme de l’algorithme du point fixe. Nous en donnons les raisons dans la partie suivante, où les algorithmes principaux de l’ACI sont comparés avec l’algorithme du point-fixe.

## 2.1.2 Comparaison des différents algorithmes

Avant de commencer, notons que toutes les informations, qui vont être énoncées dans la suite, notamment la définition des algorithmes, leurs performances, sont expliquées plus en détail dans l’annexe 4 et dans (Giannakopoulos et al., 1999b; Giannakopoulos, 1998; Giannakopoulos et al., 1998, 1999a; Karhunen, 1996; Karhunen et al., 1997); c’est pourquoi seuls les grandes propriétés et résultats seront explicités dans cette partie.

### 2.1.2.1 Algorithmes considérés

Les algorithmes comparés dans ces articles sont :

- a. les algorithmes du *point fixe (FP)*, définis par une non-linéarité (dérivée de la fonction  $G$  (voir 1.8.3.3)) et une méthode. Ici, la non-linéarité généralement utilisée est

- la tangente hyperbolique ( $\tanh$ ) avec la méthode déflation (FPth) ou symétrique (FPsymth);
- b. l'algorithme *naturel du gradient ACY*, du nom de ses trois concepteurs Amari-Cichocki-Yang (Amari et al., 1996);
  - c. l'algorithme *étendu de Bell et Sejnowski, ExtBS* (« Extended Bell-Sejnowski » en anglais), algorithme de gradient basé sur le critère de l'infomax (Bell et Sejnowski, 1995b; Girolami et Fyfe, 1997);
  - d. l'algorithme *EASI* de Cardoso et Laheld (Cardoso et Laheld, 1996) pour « Equivariant Adaptive Separation via Independence » en anglais; soit, en français, algorithme qui réalise la séparation adaptative équivariante de sources indépendantes;
  - e. l'algorithme *NPCA-RLS* (Karhunen et Pajunen, 1997b; Pajunen et Karhunen, 1998; Karhunen et Pajunen, 1997a) pour « Nonlinear Principal Component Analysis by Recursive Least Squares » ; soit, en français, une ACP par une technique récursive des moindres carrés.

Nous rappelons que nous ne donnons ici pas plus de détails sur ces algorithmes (se référer à l'annexe 4 si besoin est). Après les avoir présentés, nous allons maintenant juger de leurs performances pures.

### 2.1.2.2 Performances des algorithmes

D'un point de vue protocolaire, les algorithmes sont comparés sur des données générées artificiellement, constituées de mélanges de signaux de base (sinusoïdes, dents de scie, ...). Les proportions des mélanges sont connues (en d'autres termes, la matrice de mélange  $A$  est connue) car le but est de comparer les algorithmes sur les propriétés de précision et de vitesse de calcul. Ainsi, en connaissant  $A$ , on connaît initialement  $S$  et il est alors facile de comparer les composantes indépendantes rendues par chaque algorithme avec les signaux de  $S$ . Le critère d'erreur utilisé est l'indice de performance  $E_1$ , tel que présenté en 1.6.4. Nous nous intéressons aux trois critères suivants : la précision, la charge informatique et

l'erreur en fonction du nombre de sources. Les figures sont issues de (Giannakopoulos et al., 1999b; Giannakopoulos, 1998).

### 2.1.2.2.1 Précision

Pour ce test de précision, les algorithmes testés sont ceux initialement présentés, auxquels ont été rajoutés l'algorithme BS (de Bell et Sejnowski) de base (dont ExtBS est une extension) et WACY (« White » ACY), qui est l'ACY mais dans le cas blanchi (différence peu importante à ce stade). De plus, pour 100 réalisations des données d'entrée, 10 sources supergaussiennes (kurtosis positif, voir 1.8.2) ont été considérées parce qu'il a été montré (Giannakopoulos, 1998) qu'avec ce type de sources, tous les algorithmes convergent constamment.

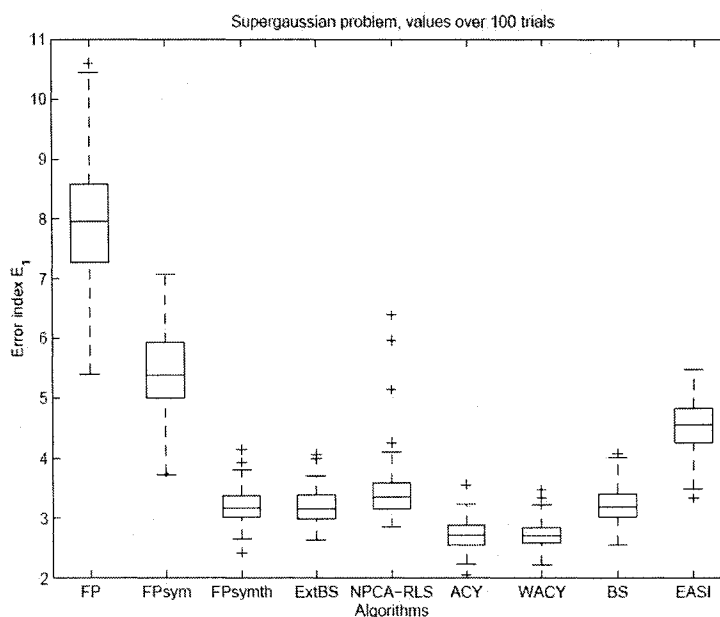


Figure 7 Erreur  $E_1$  en fonction de la précision

La figure 7 montre les résultats pour la précision : on y représente la valeur de  $E_1$  pour chaque algorithme. On rappelle, d'après 1.6.4, qu'une valeur nulle de  $E_1$  indique une extraction parfaite et que plus  $E_1$  est élevée, moins l'algorithme est performant; donc

moins, dans ce cas, il est précis. Les « boîtes » que l'on peut voir sont délimitées par le premier et troisième quartile des valeurs, avec la valeur moyenne au centre. L'intervalle en pointillés représente l'intervalle de confiance; c'est-à-dire l'intervalle des valeurs possibles et considérées comme correctes de  $E_1$  pour chaque algorithme. Cet intervalle est égal à 1,5 fois la taille de la « boîte » de chaque côté et centré en la moyenne. Ainsi, les croix représentent les valeurs aberrantes, car hors intervalle de confiance.

La conclusion que l'on peut tirer de cette figure est, tout d'abord, que l'algorithme naturel du gradient ACY (blanchi ou non) présente la meilleure précision. Globalement, on peut dire que les algorithmes FPsymth, ExtBS et NPCA-RLS sont dans le même ordre de performance que ceux du gradient, même si NPCA-RLS présente beaucoup plus de valeurs aberrantes. A l'extrême, l'algorithme à point fixe simple (FP) est le plus mauvais. Ceci s'explique par sa fonction non linéaire cubique, qui amplifie donc grossièrement les erreurs. Intéressons nous à présent à la charge informatique.

#### 2.1.2.2.2 Charge informatique

Pour mesurer celle-ci, on mesure l'erreur  $E_1$  et le nombre de calculs sur les flottants nécessaires à chaque algorithme pour converger vers les composantes indépendantes; le critère d'arrêt étant un seuil prédéfini, correspondant au taux d'erreur maximum acceptable. Les expérimentations sont similaires à celles de l'étude de la précision; à savoir que 10 sources ont été utilisées pour 100 réalisations. Les résultats sont présentés à la figure 8.

On remarque que les algorithmes à point fixe, quelque'ils soient, convergent avec le moins de ressources informatiques nécessaires; néanmoins, l'algorithme FP est loin d'être satisfaisant quant au nombre d'erreurs (large intervalle des valeurs de  $E_1$ ). Pour les autres, NPCA-RLS converge le plus vite. Les algorithmes du gradient (ACY, BS et ExtBS) offrent une erreur faible mais nécessitent 20 à 50 fois plus d'opérations sur les flottants que les FP pour converger. L'algorithme EASI n'est pas très bon dans ce domaine.

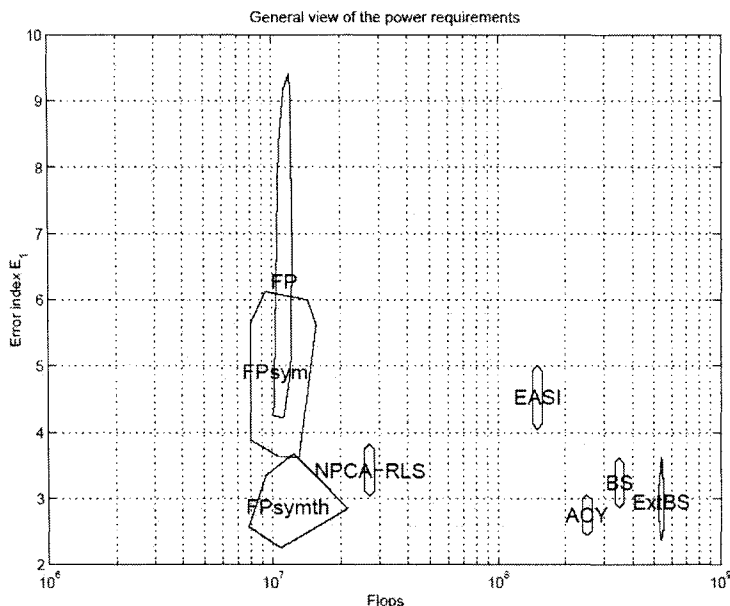


Figure 8 Erreur  $E_1$  en fonction de la charge informatique

Analysons le dernier critère de performance : le comportement de l'erreur quand le nombre de sources initiales est augmenté.

### 2.1.2.2.3 Erreur en fonction du nombre de sources

Dans cette partie, l'erreur est calculée comme la racine carrée de l'indice  $E_2$ , présentée en 1.6.4 et « cousin » quadratique de  $E_1$ . On la mesure pour différentes valeurs du nombre de sources. Les résultats obtenus, présentés à la figure 9, sont les valeurs médianes sur 50 réalisations des données d'entrée.

On constate globalement que les algorithmes ACY, ExtBS et FPSymth offrent le meilleur comportement et la meilleure précision. L'algorithme à point-fixe simple (FP) se comporte le plus mal du fait de sa non linéarité cubique, même si son erreur augmente peu à partir de sept sources. De plus, on constate un pic d'erreurs autour de 5-6 sources pour les algorithmes EASI et NPCA-RLS; la raison en est inconnue. En outre, pour un plus grand

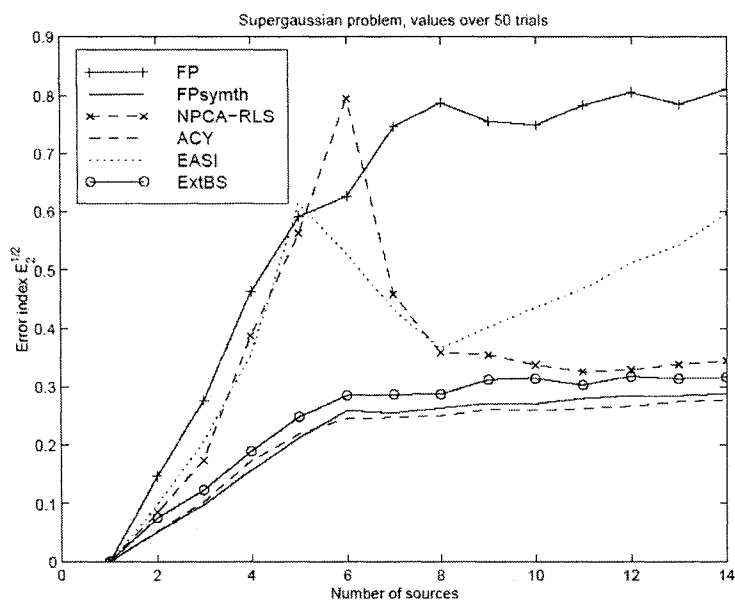


Figure 9 Erreur en fonction du nombre de sources

nombre de sources, la précision de NPCA-RLS est proche des meilleurs, alors que celle de EASI se dégrade plus ou moins linéairement avec le nombre de sources.

### 2.1.3 Conséquence : choix du FastICA

Si l'on récapitule toutes les propriétés déductibles des trois tests réalisés, on en arrive à la conclusion que, globalement, le meilleur algorithme est *l'algorithme à point fixe symétrique à non linéarité en tangente hyperbolique (FPsymth)*. Vu qu'il est le plus rapide des algorithmes considérés, son concepteur Hyvärinen lui a donné l'appellation « FastICA<sup>1</sup> ».

Ce FastICA apparaît donc finalement à la fois comme l'algorithme le plus rapide pour converger, nécessitant le moins d'opérations et relativement peu sensible à l'augmentation du nombre de sources. Ces trois raisons majeures justifient donc parfaitement le fait qu'il ait été choisi comme routine à notre programme Matlab sur lequel est basé notre projet de maîtrise. Nous en ferons une étude plus approfondie en 3.3.2.

<sup>1</sup>le code Matlab est disponible à l'adresse : <http://www.cis.hut.fi/projects/ica/fastica/>

De plus, anecdotiquement, nous avons choisi FastICA car, au préalable, probablement pour les raisons évoquées plus haut, Luis Da Costa, un étudiant en Ph.D. de l'ÉTS, qui travaillait sur ce projet avant moi, avait développé du code Matlab à partir de cette routine FastICA. Nous avons poursuivi et amélioré le travail de programmation, à partir de ses travaux et ceux d'Hyvärinen.

Maintenant que nous avons « réglé » la problématique matérielle de notre mémoire, nous pouvons nous intéresser à l'état de l'art des applications de l'ACI. On rappelle que, d'après la partie 1.3, il existe de nombreuses applications à l'ACI, principalement issues de la séparation aveugle de sources. Dans la suite, nous ne développons pas les applications déjà citées mais nous nous concentrons uniquement sur l'application concernant les signatures spectrales. Comme nous allons le constater, les papiers scientifiques ne sont pas légions sur ce sujet.

## 2.2 Travaux sur les signatures spectrales

Avant de vraiment nous intéresser à la revue de littérature de notre projet, il est important d'en présenter, correctement et précisément, la problématique qui l'entoure; à savoir la *ségrégation hyperspectrale* (ou « hyperspectral unmixing » en anglais). En effet, l'acquisition du lexique de ce domaine est essentielle à la bonne compréhension de notre projet.

### 2.2.1 Préambule : la ségrégation hyperspectrale

La ségrégation hyperspectrale consiste, à partir d'une image hyperspectrale, à être capable de déterminer pour chaque pixel de chacune des couches de l'image (définies individuellement par une longueur d'onde donnée) quels éléments (ou matériaux dans la littérature) le composent et en quelles proportions. Ainsi, si un pixel contient a priori trois éléments  $A$ ,  $B$  et  $C$ , en proportions respectives  $\alpha$ ,  $\beta$  et  $\gamma$ , la valeur de réflectance  $R$  du pixel  $(x, y)$  à la longueur d'onde  $\lambda$  est définie, en fonction des réflectances en  $\lambda$  des trois éléments, par :

$$R(x, y, \lambda) = \alpha R(A, \lambda) + \beta R(B, \lambda) + \gamma R(C, \lambda) \quad (2.1)$$



et le but est donc de connaître, à la fois,  $A$ ,  $B$  et  $C$  mais aussi les proportions  $\alpha$ ,  $\beta$  et  $\gamma$ , qui, par nature (somme de pourcentages), vérifient l'équation :

$$\alpha + \beta + \gamma = 1 \quad (2.2)$$

D'un point de vue protocole, les matériaux sont relativement connus car les expérimentateurs connaissent globalement la composition du terrain. Ainsi, la détermination des matériaux se fait généralement à partir d'une base de matériaux, dont le nombre est légèrement plus élevé que le nombre de matériaux présents dans le pixel. On parle alors de ségrégation hyperspectrale avec *connaissance a priori*. En ce cas, le problème principal consiste plutôt à déterminer les proportions.

En outre, il arrive que dans certains pixels, seul un matériau soit présent. Le pixel est alors appelé « *endmember* » (ou membre de fin) et le matériau est défini comme pur ou fondamental. En fait, une proportion supérieure à 80% permet d'affirmer qu'il est pur (Parra et al., 2000). Le pixel est appelé « *endmember* » car si l'on représente les pixels par des points dans un espace  $n$ -dimensionnel où  $n$  représente le nombre de matériaux et où chaque axe indique la proportion de chaque matériau pour le pixel (coordonnées de 0 à 1); alors l'« *endmember* » est représenté comme un point à la « fin » de l'axe de son matériau pur; d'où l'appellation « membre de fin ». Ces pixels « *endmembers* » sont recherchés en tant que références pour l'image hyperspectrale.

De plus, les matériaux présents dans les pixels sont appelés *abondances* (« *abundance* » en anglais) et leurs proportions relatives deviennent les *fractions d'abondance* (« *abundance fractions* » en anglais) à déterminer.

En reprenant l'équation (2.1), appliquée cette fois à toutes les longueurs d'onde (soit à toute l'image hyperspectrale) et à un nombre  $n$  de matériaux  $(M_i)_{i \in [1, n]}$  de proportions respectives  $\alpha_i$ , on obtient alors l'équation générale de la signature spectrale  $R(x, y)$  du

pixel de coordonnées  $(x, y)$  :

$$R(x, y) = \sum_{i=1}^n \alpha_i R(M_i) \quad (2.3)$$

où  $R(M_i)$  est la réflectance du matériau  $M_i$ , pour toutes les longueurs d'onde; soit finalement, sa signature spectrale et où les fractions d'abondance  $\alpha_i$  vérifient l'égalité :

$$\sum_{i=1}^n \alpha_i = 1 \quad (2.4)$$

Après avoir présenté la ségrégation hyperspectrale, il est maintenant intéressant d'en faire l'analogie avec l'ACI.

### 2.2.2 Analogie avec l'ACI

Dans la littérature, il existe, en fait, deux modèles basés sur l'ACI, qui permettent de représenter le phénomène de ségrégation hyperspectrale, telle que présentée. La différence provient de la façon dont on conçoit le problème; c'est-à-dire quel élément l'on considère comme composantes indépendantes.

En effet, les composantes indépendantes  $y_i$  à rechercher peuvent être considérées comme :

- a. soit les signatures spectrales  $R(M_i)$  des matériaux (Bayliss et al., 1997; Parra et al., 2000; Lennon et al., 2001);
- b. soit les fractions d'abondance  $\alpha_i$  (Chiang et al., 2000; Chang et al., 2002; Nascimento et Bioucas Dias, 2005; Kosaka et al., 2005; Kosaka et Kozugi, 2003; Shah et Varshney, 2004; Shah et al., 2002).

Par nature, ces deux approches s'avèrent conceptuellement radicalement différentes dans leur façon d'adapter l'ACI à la ségrégation hyperspectrale. En effet, elles sont différentes dans leur approche de l'hypothèse fondamentale d'indépendance des sources de l'ACI. Autant pour la première méthode, les réflectances sont supposables mutuellement indépendantes (encore qu'il faille négliger les interactions naturelles existant entre les matériaux; ce qui n'est pas si aisé; on y reviendra dans les recommandations) et le modèle de l'ACI

est applicable; autant pour la seconde, par le fait même que les fractions d'abondance sont liées entre elles, comme l'indique l'équation (2.4), l'ACI semble théoriquement non applicable.

## 2.2.3 Méthode des fractions d'abondance

### 2.2.3.1 Justification

Cependant, pour cette seconde méthode, deux éléments peuvent permettre de l'expliquer. D'une part, dans les conditions réelles expérimentales de prises de mesures, sous l'effet du bruit et de l'atmosphère, les fractions d'abondance des matériaux peuvent varier de pixel en pixel et peuvent donc être vues comme des phénomènes aléatoires; c'est pourquoi il peut apparaître plus réaliste de les considérer comme des variables aléatoires, plutôt que des constantes inconnues (Chiang et al., 2000). Si on les considère de plus indépendantes, elles sont assimilables aux signaux sources de l'ACI, que l'on applique alors sur les fractions d'abondance et non plus sur les matériaux. D'autre part, l'équation (2.4) exprime le fait que la famille  $(\alpha_1, \dots, \alpha_n)$  est liée, mais cela ne signifie pas mathématiquement que nécessairement la famille  $(\alpha_1, \dots, \alpha_p)$ , avec  $p < n$ , l'est également. Elle peut être indépendante. Ainsi, on se retrouve avec une « indépendance relative » et il apparaît alors qu'il est *peut-être* possible d'appliquer l'ACI (Nascimento et Bioucas Dias, 2005).

### 2.2.3.2 Modèle

Toutefois, une nouvelle forme de l'équation  $X = AS$ , appelé modèle LSMA (pour « Linear Spectral Mixture Analysis », soit l'analyse de mélanges linéaires de spectres), est même proposée dans (Chen et Zhang, 1999; Chang et al., 2002). En notant :

- a.  $L$  le nombre de bandes spectrales;
- b.  $r$ , de dimension  $[L, 1]$ , le vecteur colonne d'un pixel (noté abusivement  $r$ ) de l'image hyperspectrale, qui, autrement dit, représente la signature du pixel;

- c.  $M$ , de taille  $[L, p]$ , la matrice des signatures des  $p$  endmembers, telle que  $M = [m_1, m_2, \dots, m_p]$ , où chaque colonne  $m_i$ , de dimension  $[L, 1]$ , de  $M$  est la signature du  $i^{\text{ème}}$  matériau du pixel  $r$ ;
- d.  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_p]^T$  est le vecteur colonne, de taille  $[p, 1]$ , des fractions d'abondances du pixel  $r$ ; plus précisément,  $\alpha_i$  représente la proportion du  $i^{\text{ème}}$  matériau dans le pixel;

le modèle LSMA se traduit alors par la forme matricielle :

$$r = M\alpha + n \quad (2.5)$$

où  $n$  représente le bruit, assimilable aux erreurs de mesure et le but est d'estimer  $\alpha$ , en tant que composantes indépendantes. Cette équation est l'analogie de la classique  $x = As + n$ .

La LSMA nécessite ses hypothèses, s'adoptant à l'ACI :

- a. la connaissance a priori de  $M$  n'est pas nécessaire;
- b. les  $p$  fractions d'abondances,  $\alpha_1, \alpha_2, \dots, \alpha_p$ , sont considérées comme des signaux sources aléatoires inconnus (plutôt que comme des constantes inconnues) statistiquement mutuellement indépendants, dont au plus un peut être gaussien;
- c. la matrice  $M$  doit être de rang plein; ce qui signifie que les  $p$  signatures des matériaux  $m_1, m_2, \dots, m_p$  sont alors linéairement indépendantes.

Ces hypothèses sont analogues aux conditions théoriques de l'ACI, présentées en 1.5. Le but devient alors d'estimer  $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)^T$ , tel que  $\hat{\alpha} = W r$ , à associer à la forme  $y = \hat{s} = Bx$ , où  $\hat{\alpha}$  est le vecteur des composantes indépendantes.

### 2.2.3.3 Travaux

À titre d'information, nous citons quelques travaux utilisant cette méthode, avec leur fonction de contraste associée. Il y a notamment :

- a. Chang et Chiang (Chiang et al., 2000; Chang et al., 2002), utilisant les cumulants;

- b. Kosaka et al. (Kosaka et al., 2005; Kosaka et Kozugi, 2003), qui s'intéressent au phénomène physique du rayonnement pour justifier le modèle;
- c. Shah et al. (Shah et al., 2002; Shah et Varshney, 2004), qui utilisent les statistiques d'ordre plus élevés (voir 1.8.1), la théorie bayésienne et le maximum de vraisemblance.

Nous ne présentons pas les résultats de leurs recherches, disponibles dans les articles cités car Nascimento et Bioucas Dias (Nascimento et Bioucas Dias, 2005) se sont intéressés à la validité de cette méthode. Leur conclusion est qu'il y aura toujours des « endmembers » mal démélangés. C'est pour cette raison et le fait que nous avons travaillé depuis le début avec l'autre méthode, qui nous semble plus simple et surtout plus logique, que nous avons décidé de négliger cette méthode (et donc de ne pas rentrer dans les détails de la littérature), pour nous concentrer sur la seconde que nous présentons, plus exhaustivement, avec les écrits s'y rapportant, dès à présent.

## 2.2.4 Méthode des signatures spectrales

### 2.2.4.1 Modèle

À partir de (2.3), l'analogie avec l'ACI et notamment avec son équation symbolique (1.8), où :

$$x_j = \sum_{i=1}^n a_{ji} s_i \quad (2.6)$$

est évidente. En effet, on associe naturellement :

- a. la signature spectrale  $R(x, y)$  du pixel  $(x, y)$  avec le mélange  $x_j$  du modèle ACI;
- b. la fraction d'abondance  $\alpha_i$  avec le coefficient  $a_{ji}$  de la matrice de mélange  $A$ ;
- c. la signature spectrale  $R(M_i)$  du matériau  $M_i$  avec le signal source  $s_i$ .

## 2.2.4.2 Travaux

### 2.2.4.2.1 Bayliss et la cICA

La première à avoir directement travaillé sur les signatures spectrales est Jessica Bayliss (Bayliss et al., 1997). Une raison est qu'à cette époque, l'idée de considérer les fractions d'abondance comme composantes n'existait pas encore. Elle s'est basée sur une nouvelle méthode de l'ACI, à l'époque, appelée *ACI contextuelle* (ou « contextual ICA » en anglais), abrégée en « cICA » et issue des travaux de Pearlmutter et Parra (Pearlmutter et Parra, 1996, 1997).

En résumé, cette cICA part des règles d'apprentissage de l'algorithme BS (voir (4.13) et (4.14) de l'annexe 4, section 2) mais fait remarquer qu'il n'a aucune restriction sur la forme des densités de probabilités  $f_i(y_i(t))$  des composantes indépendantes  $y_i$  recherchées. Ainsi, partant de là, Bayliss pose l'hypothèse qu'en plus de  $B$ , celles-ci peuvent dépendre du contexte environnant (d'où l'appellation ACI contextuelle), typiquement les valeurs connues des composantes aux temps antérieurs ( $f_i(y_i(t-1)), f_i(y_i(t-2))$ ) et des valeurs de réflectance aux longueurs d'onde antérieures.

Bayliss utilise donc cette cICA car elle est spécialement adaptée aux mélanges linéaires et qu'elle considère le contexte extérieur (ici la théorie de l'imagerie hyperspectrale). Cependant, cette cICA présente l'inconvénient de ne pas être calculable analytiquement; ce qui augmente son temps de convergence, quand c'est le cas. Ceci est lié au fait que cet algorithme ne fait pas intervenir les statistiques d'ordre plus élevées (voir 1.8.1). Pour y remédier, Bayliss les fait intervenir en réalisant un prétraitement des données (les mélanges de signatures) : elle les blanchit (voir 3.1.2), c'est-à-dire qu'elle les décorrèle.

Son approche est la suivante : elle réalise le prétraitement précédent, applique la cICA sur les mélanges blanchis puis applique un post-traitement sur les signatures (composantes indépendantes) extraites, pour pouvoir les identifier. Pour ce faire, elle calcule la corréla-

tion (plus précisément, le coefficient de corrélation) existant entre ces signatures, une fois centrées et celles de la base de données de référence USGS; puis associe les couples de coefficients de corrélation maxima.

Cependant, les résultats expérimentaux, en premier lieu sur des données synthétiques, puis sur des données réelles, mettent plutôt en lumière l'imprécision et la prudence d'utilisation du critère qu'est la corrélation. En effet, il semblerait que, sur *un* exemple de données synthétiques (les signatures sont connues au départ, on les mélange et on observe ce que rend cICA), la corrélation fonctionne. Cependant, il n'y pas de résultat sur une grande base de données, seulement sur un exemple particulier ! Pour ce qui est des données réelles, issues d'une image AVIRIS (pour « Airborne Visible InfraRed Imaging Spectrometer ») du site géographique de Cuprite dans le Nevada, Bayliss explique elle-même que « bien que les signaux démélangés concordent avec les informations de la nature du terrain, les corrélations font naître des doutes sur la confiance des résultats ». Une raison invoquée pour expliquer ces problèmes est l'intervention du bruit. Mais de même, le test n'est fait que sur un seul groupe restreint de quatre, cinq signatures; ce qui n'est pas représentatif, statistiquement parlant. De plus, l'algorithme cICA, bien que semblant efficace à l'époque, présente l'inconvénient majeur de converger lentement, quand il converge ! À ce titre, l'algorithme FastICA représente donc, à présent, une alternative d'outil beaucoup plus efficace.

En résumé, nous pouvons affirmer qu'en 1997, Bayliss a été la première à utiliser une forme particulière de l'ACI (cICA), performante pour l'époque, et que l'idée d'utiliser la corrélation entre les composantes indépendantes et les signatures de la base de données semblait a priori très pertinente. Cependant, le protocole expérimental de l'époque nous semble ne pas permettre de conclure sur la validité de la méthode, basée sur la corrélation.

### 2.2.4.2.2 Parra et le MAP

Le deuxième (et second !) article pertinent est celui de l'équipe de Parra (Parra et al., 2000), qui utilise une approche probabiliste bayésienne non supervisée, basée sur le maximum (de probabilité) a posteriori (« MAP »). Avant de rentrer dans les détails du modèle, l'une des comparaisons qu'ils font eux-mêmes avec la méthode de Bayliss est que, comme nous le verrons par la suite, leur méthode tient plus compte des conditions expérimentales, à savoir les conditions sur  $A$ , du bruit ou de l'information a priori.

Le modèle MAP est le suivant. Tout d'abord, à partir de l'adaptation matricielle de l'équation typique (1.10) de l'ACI bruitée ( $X = AS + N$ , où  $N = (n_j)_{j \in [1,m]}$  représente la matrice de bruit, assimilé ici gaussien), Parra, en appliquant le théorème de Bayes, exprime la probabilité  $p(A, S|X)$  de connaître  $A$  et  $S$ , connaissant  $X$  a priori :

$$p(A, S|X) = \frac{p(X|A, S)p(A, S)}{p(X)} = \frac{p_n(N)p_a(A)p_s(S)}{p(X)} \quad (2.7)$$

où  $p_n(N)$ ,  $p_a(A)$  et  $p_s(S)$  sont les distributions respectives de  $N$ ,  $A$  et  $S$ . La méthode MAP consiste alors à chercher le couple  $(A, S)$  qui maximise  $p(A, S|X)$ .

Cette équation (2.7) met en lumière certaines hypothèses du modèle :

- a. la logique veut que :  $p(X|A, S) = p_n(N)$ ;
- b.  $p(A, S) = p_a(A)p_s(S)$  sous-entend l'indépendance entre  $A$  et  $S$ ; hypothèse somme toute logique physiquement si l'on considère l'image entière et non pixel par pixel;
- c. Tous les signaux de  $S$  et  $N$  sont considérés indépendants les uns des autres;
- d.  $A$  représente ici une matrice de mélange, représentative de la vérité du terrain; c'est-à-dire que comme physiquement, la somme des fractions d'abondance de chaque pixel vaut 1 (voir (2.4)), alors la somme des lignes de  $A$  vaut 1. Mathématiquement, cela se traduit par les égalités suivantes équivalentes :



$$\begin{cases} \sum_{i=1}^n a_{ji} = 1, a_{ji} \geq 0 \\ A\mathbf{1}_n = \mathbf{1}_m \end{cases} \quad (2.8)$$

où  $\mathbf{1}_n$  et  $\mathbf{1}_m$  sont des vecteurs colonnes de 1, de tailles respectives  $n$  et  $m$ , par analogie avec  $S$ ,  $X$  et  $AS = X$ .

Le point précédent montre que Parra prend en considération le contexte expérimental de l'analyse hyperspectrale pour déterminer les probabilités a priori. En effet, l'équation (2.8) entraîne la positivité et la normalisation de  $A$ , qui s'exprime mathématiquement par :

$$p_a(A) = \delta(A\mathbf{1}_n - \mathbf{1}_m)\Theta(A) \quad (2.9)$$

où  $\delta(\cdot)$  est la fonction de Kronecker (égale à 1 si son argument est nul; nulle sinon), qui symbolise la normalisation et  $\Theta(\cdot)$  est la fonction échelon (égale à 1 quand son argument est positif; nulle sinon), symbolisant, pour sa part, la positivité.

À partir de cette simplification de la densité de probabilité de  $A$ , Parra met en évidence le concept d'ACI *contrainte*, dans le sens où les solutions qui seront alors trouvées seront toujours contraintes aux conditions de positivité et de normalisation de  $A$ . Et c'est cette dernière idée de « contraintes » sur  $A$  qui nous semble la notion la plus intéressante ici. En effet, d'un point de vue protocole, à partir de l'outil FastICA que nous avons décidé de choisir, dans la première partie de cet état de l'art, il semblerait alors pertinent d'intégrer ces conditions dans la routine de FastICA.

Finalement, d'un point de vue expérience et résultats, selon le même procédé que Bayliss, Parra arrive, dans le cas non bruité, à retrouver les signatures de départ, mais de même, uniquement pour un seul exemple particulier. Il faudrait pouvoir voir les résultats sur un échantillon plus grand. En tout cas, il est intéressant de retenir le concept de contraintes sur  $A$  !

Vu que nous n'avons pas trouvé d'autres références intéressantes sur le sujet, notre état de l'art se termine donc ici. Ainsi, après avoir présenté l'état de l'art sur la question de l'ACI,

à la fois comme outil algorithmique dans une première partie et comme modèle pour la ségrégation hyperspectrale dans une seconde, il est temps de tirer les enseignements de celui-ci.

### 2.3 Conclusions

Les conclusions que l'on peut tirer de cet état de l'art sont de trois ordres :

- a. la partie sur l'analyse des algorithmes a prouvé que le meilleur outil pour l'ACI est le FastICA; c'est pourquoi nous l'utilisons dans notre protocole expérimental. Il est, à ce titre, le sujet du prochain chapitre;
- b. l'étude de la ségrégation hyperspectrale a permis, pour la méthode considérant les signatures spectrales comme composantes indépendantes (qui nous semble plus logique et plus proche de l'esprit de l'ACI que l'autre), de mettre en lumière l'idée intéressante pour la suite : les contraintes sur  $A$  (positivité et normalisation). Elles seront ainsi intégrées dans la routine du FastICA.
- c. l'idée de considérer les fractions d'abondances comme composantes indépendantes pourrait peut-être s'avérer une solution intéressante. En faire une étude approfondie serait une amélioration, une ouverture possible pour un travail futur (maîtrise ou doctorat) avec comme grand intérêt de pouvoir comparer les deux approches.

Nous nous intéressons à présent plus précisément à l'algorithme FastICA, par l'intermédiaire des algorithmes du point fixe, notion plus générale à laquelle il appartient.

## CHAPITRE 3

### ALGORITHMES DU POINT FIXE

Dans ce chapitre, nous allons étudier plus précisément les algorithmes du point fixe, en reprenant et développant, à partir des notations mathématiques du premier chapitre, les notions abordées en 1.8 sur les fonctions de contraste, que sont principalement le kurtosis et la négentropie. Nous allons montrer, tout d'abord, qu'il est pertinent de blanchir les mélanges, avant de leur appliquer ces fonctions; le but étant d'extraire des algorithmes de convergence vers les CIs. Nous étudierons, en premier lieu, l'algorithme du point fixe basé sur le kurtosis; puis, celui basé sur la négentropie, qui se veut une généralisation et une amélioration du précédent.

#### 3.1 Prétraitement

À des fins de simplifications et d'améliorations des algorithmes de l'ACI, il apparaît intéressant de réaliser un prétraitement des données (les mélanges  $x_j$  de la matrice  $X$ ) qui consiste à les centrer et les blanchir. En effet, ces deux opérations permettent de normaliser les moments d'ordre 1 (la moyenne) à 0 et d'ordre 2 (la matrice de covariance) à la matrice identité; afin de se préoccuper ensuite des moments d'ordres supérieurs pour atteindre le but d'indépendance des CIs (voir la partie 1.8.1 concernant les statistiques d'ordres plus élevés).

##### 3.1.1 Centrer $X$

La première opération de prétraitement est, si besoin est, de centrer les mélanges  $x_j$ , soit également  $X$ . Celle-ci est surtout d'ordre pratique : on centre les données brutes (dites du monde réel) que l'on veut séparer. On sait d'après (1.9) et (1.20) qu'en théorie, les mélanges sont déjà initialement centrés; c'est notamment le cas lors de l'utilisation de données générées artificiellement, dont on s'assure qu'elles respectent la théorie. En ce

cas, il n'y a donc aucun besoin de centrer (d'où l'utilisation de « si besoin est »). Si non, il suffit d'appliquer ces équations respectivement à  $x_j$  et  $X$  :

$$X \leftarrow X - E[X] \quad (3.1)$$

$$x_j \leftarrow x_j - E[x_j], \forall j \in [1, m] \quad (3.2)$$

Dans la suite,  $X$  sera considérée comme centrée !

### 3.1.2 Blanchiment

La seconde (et plus importante) opération est le *blanchiment* (« whitening » ou « sphering » en anglais), qui consiste à décorréliser les mélanges; c'est-à-dire à normaliser leur moment d'ordre 2, qu'est la variance. Ainsi, la variance de chaque mélange blanchi est égale à 1 et leur matrice de covariance devient donc l'identité. Typiquement, on transforme  $X$  en une matrice  $V$  dite *matrice des mélanges blanchis*, centrée et de matrice de covariance unitaire, telle que, selon (1.37) :

$$\begin{cases} E[V] = 0 \\ E[VV^T] = I \end{cases} \quad (3.3)$$

De plus, chaque ligne de  $V$  est un mélange blanchi  $v_i = (v_{il})_{(i,l) \in [1,n] \times [1,k]}$ , soit  $V = (v_1, v_2, \dots, v_n)^T$ , vérifiant selon le même principe que (1.5) et (1.12) :

$$\begin{cases} E[v_i] = 0 & , \forall i \in [1, n] \\ \text{cov}(v_i, v_j) = 0 & , \forall j \neq i, j \in [1, n] \\ \text{var}(v_i) = E[v_i^2] = 1 \end{cases} \quad (3.4)$$

On note, dès à présent, qu'avec le blanchiment, le nombre de mélanges est passé de  $m$  (nombre initial de mélanges) à  $n$ , nombre de composantes indépendantes. Ce fait est expliqué par la suite. On rappelle que  $m \geq n$ .

L'intérêt de ce blanchiment est, comme nous le verrons plus tard, d'améliorer la convergence des algorithmes, de réduire la dimensionnalité des données et le bruit (Hyvärinen et Oja, 1996a). Voyons comment cela se réalise mathématiquement.

Pratiquement, pour passer de  $X$  à  $V$ , on introduit la *matrice de blanchiment*  $U$ , de taille  $[n, m]$ , telle que :

$$V = UX \quad (3.5)$$

qui est obtenue par une ACP (voir annexe 2, section 1.2), dont les étapes successives sont :

- a. calcul de la matrice de covariance  $C$  de  $X$  :  $C = E[XX^T]$ ;
- b. calcul de la matrice des valeurs propres  $D_C$  et des vecteurs propres  $E$  de  $C$ , telles que  $CE = ED_C$ , avec :

$$\begin{cases} C = ED_C E^{-1} = ED_C E^T \\ D_C = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \\ E = (E_{\lambda_1}, E_{\lambda_2}, \dots, E_{\lambda_n}) \end{cases} \quad (3.6)$$

où :

- (a)  $\lambda_i$  est la  $i^{\text{ième}}$  plus élevée valeur propre de  $C$  et le  $i^{\text{ième}}$  élément diagonal de  $D_C$ , elle même diagonale, par construction;
  - (b)  $E_{\lambda_i}$  est le  $i^{\text{ième}}$  vecteur propre (associé à la valeur propre  $\lambda_i$ ) de  $C$  et la  $i^{\text{ième}}$  colonne de  $E$ , matrice orthogonale toujours par construction;
  - (c) conservation des  $n$  vecteurs propres correspondant aux  $n$  valeurs propres les plus élevées ou significatives. En d'autres termes, on rejette les  $m - n$  valeurs propres non significatives (typiquement, inférieures à un seuil prédéterminé). Généralement, elles correspondent à du bruit; d'où l'avantage évoqué plus haut de diminuer le bruit.
- c. calcul de  $U$  par l'opération matricielle :

$$U = D_C^{-1/2} E^T \quad (3.7)$$

d. calcul de  $V$  par  $V = UX$ .

La preuve de (3.3) devient alors évidente. En effet :

a.  $E[V] = UE[X] = 0;$

b.  $E[VV^T] = UE[XX^T]U^T = U(ED_C E^T)U^T = D_C^{-1/2}(E^T E)D_C(E^T E)D_C^{-1/2} = I$ , par orthogonalité de  $E$ .

L'intérêt de cette matrice des mélanges blanchis  $V$  est justement de passer d'un espace quelconque à un espace blanchi, défini principalement par une réduction de la dimensionnalité (de  $m$  à  $n$ ) et par une simplification des données (inhibition des moments). Ce nouvel espace permet donc une simplification globale de l'ACI. Plus particulièrement, on reprend les équations caractéristiques  $X = AS$  (1.19) et  $Y = BX$  (1.25), en les appliquant à l'espace blanchi. De belles surprises nous attendent; la principale étant que la matrice de séparation  $B$  va se simplifier.

En effet, en introduisant (ceci est purement théorique, vu que l'on ignore  $B$  que l'on recherche) une nouvelle matrice  $W^T$ , carrée d'ordre  $n$ , telle que :

$$B = W^T U \tag{3.8}$$

on obtient alors :

$$Y = BX = (W^T U)X = W^T(UX) = W^T V \tag{3.9}$$

Or, d'après (1.37),  $S$  et  $Y$  sont déjà blanchis. Leurs équivalents dans l'espace blanchi sont donc elles-mêmes. Ainsi, par analogie avec  $Y = BX$ , il apparaît alors que  $W^T$  est la matrice de séparation dans l'espace blanchi, soit la *matrice de séparation blanchie*; et présente la propriété intéressante d'être orthogonale.

En effet, d'après les équations (1.37) et (3.3), on a :

$$E[YY^T] = I = W^T E[VV^T]W = W^T W \tag{3.10}$$

Soit :

$$W^{-1} = W^T \quad (3.11)$$

Ainsi, grâce au blanchiment, le problème a été simplifié. En effet, au lieu de rechercher une matrice  $B$  quelconque à  $n^2$  inconnues (son nombre de termes), le « jeu » consiste maintenant à trouver une matrice  $W^T$  orthogonale, dont seuls  $\frac{n(n-1)}{2}$  éléments sont donc à déterminer (car orthogonale) ! On parle de degrés de liberté.

Cependant, la notation de la transposée dans  $W^T$  peut interpeller. En fait, cette notation découle directement des algorithmes expérimentaux déjà mis en place et que nous présenterons en 3.3. Ceux-ci extraient les composantes indépendantes de telle manière que chaque colonne  $w_i$  de  $W$  représente le vecteur des coefficients linéaires (ou poids) de chacun des mélanges blanchis à la composante indépendante  $y_i$ . La colonne est construite algorithmiquement de même. De ce fait, la transposée est la traduction mathématique du produit scalaire  $y_i = \langle w_i, V \rangle = w_i^T V$ .

Plus précisément, avec les équations, on a :

$$W = (w_1, w_2, \dots, w_n) \quad (3.12)$$

$$w_i = (w_{1i}, w_{2i}, \dots, w_{ni})^T \quad (3.13)$$

$$y_i = w_i^T V = \sum_{j=1}^n w_{ji} v_j, \forall (i, j) \in [1, n] \quad (3.14)$$

À noter que l'on retrouve (3.9) à partir de (3.14), comme expression matricielle de la forme vectorielle.

En outre, en multipliant chaque membre de cette équation (3.9) par  $W$ , inverse de  $W^T$  par orthogonalité, on obtient :

$$V = WY \quad (3.15)$$

et  $W$  apparaît alors, par comparaison avec (1.28), comme la *matrice de mélange blanchie*; c'est-à-dire l'estimée de  $A$ ,  $\tilde{A}$ , dans l'espace blanchi. Elles sont reliées par la relation :

$$\tilde{A} = U^{-1}W \quad (3.16)$$

obtenue selon  $V = UX = U\tilde{A}Y = (U\tilde{A}W^T) V$ .

Nous récapitulons toutes les analogies entre les variables et leurs équivalentes blanchies dans le tableau II.

Tableau II

Équivalence des variables dans les espaces normal et blanchi

Intitulé	Espace normal	Espace blanchi	Relation
Matrice des sources	$S$	$S$	$I$
Matrice des CIs	$Y$	$Y$	$I$
Matrice des mélanges	$X$	$V$	$V = UX$
Matrice de séparation	$B$	$W^T$	$B = W^T U$
Matrice de mélange	$\tilde{A}$	$W$	$\tilde{A} = U^{-1}W$

Nous allons voir dans la suite que les algorithmes de l'ACI présentés auront pour but de calculer cette matrice de séparation après blanchiment  $W^T$ , avec généralement  $V$  comme entrée.

### 3.2 Analogie avec les réseaux de neurones

Avant cela, nous pouvons faire une remarque sur le modèle de l'ACI. Une des branches référence en la matière vient de l'équipe de Äapo Hyvärinen, de l'Université de Technologie d'Helsinki (HUT) en Finlande, qui est spécialisée dans le domaine des réseaux de neurones. À ce titre, certains algorithmes de l'ACI (voir l'état de l'art, partie 2.1) sont des algorithmes neuronaux. D'ailleurs, les algorithmes du point fixe découlent directement de



ce côté « réseaux de neurones ». Nous pouvons ainsi brièvement présenter le modèle neuronal de l'ACI. Il est représenté par le réseau ci-dessous (Karhunen, 1996; Karhunen et al., 1997), en figure 10.

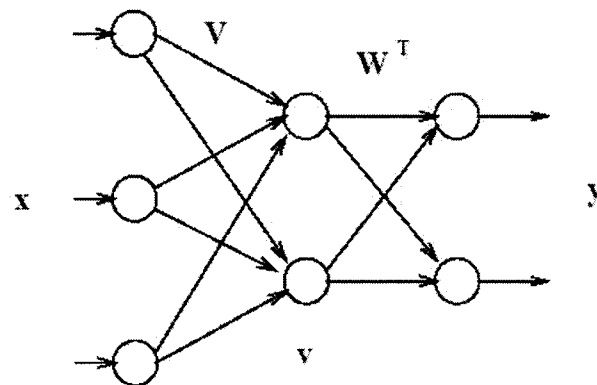


Figure 10 Réseau de neurones de l'ACI

On remarque que ce réseau possède trois couches de noeuds. La première est constituée des  $m$  noeuds de mélanges (matrice  $X$ ); la seconde des  $n$  ( $\leq m$ ) mélanges blanchis ( $V$ ); et enfin, la dernière des  $n$  composantes indépendantes ( $Y$ ). Concernant les connexions synaptiques, les poids synaptiques de la première à la deuxième couche sont les coefficients  $u_{ji}$  de la matrice de blanchiment  $U$ , tels que  $u_{ji}$  est le poids du noeud  $j$  de la première couche ( $x_j$ ) au noeud  $i$  de la deuxième ( $v_i$ ). De même, les poids entre les deux dernières couches sont les coefficients  $w_{il}$ , du noeud  $i$  de la deuxième au noeud  $l$  de la dernière ( $y_l$ ).

De plus, ce modèle met clairement en évidence les étapes de prétraitement accomplies : blanchiment des mélanges, avec réduction de la dimensionnalité (d'où réduction du nombre de noeuds entre les deux premières couches); et recherche de la matrice de séparation blanchie  $W^T$  pour obtenir les composantes indépendantes  $Y$ .

Pour plus de détails sur l'interprétation du modèle de l'ACI comme problème de type réseaux de neurones, il est possible de se référer à (Karhunen et al., 1997; Karhunen, 1996), dans lesquels les notions globales du modèle neuronal appliqué à l'ACI, avec notamment

certaines règles d'apprentissage, sont plus largement expliquées. En outre, sans rentrer dans les détails, quelques exemples de fonctions de contraste neuronales basées sur le kurtosis sont disponibles dans (Hyvärinen et Oja, 1996a,b) ou sur des règles d'apprentissage de type Hebbian/anti-Hebbian dans (Hyvärinen et Oja, 1998b), pour les connaisseurs. De plus, une comparaison entre les fonctions neuronales de l'ACI et de l'ACP est réalisée dans (Oja et al., 1995).

Intéressons-nous à présent aux algorithmes du point fixe, basées sur les fonctions de contraste de la partie 1.8.

### **3.3 Algorithmes du point fixe**

Dans cette partie, nous allons distinguer le cas de la recherche d'une composante, puis de toutes; en montrant que la seconde catégorie est une prolongation de la première.

#### **3.3.1 Pour une composante**

Il existe deux types de fonctions : celle basée sur le kurtosis et celle sur la négentropie. Dans les deux cas, on se place dans l'espace blanchi pour profiter des simplifications présentées en 3.1.2 et ce sont des fonctions du point fixe, fonctions de  $w$ , colonne de  $W$ , donc définies par une équation du type  $w = f(w)$ , qui sont utilisées. Avant de rentrer dans les détails, nous allons expliquer les conditions de convergence de ces fonctions.

##### **3.3.1.1 Convergence**

On rappelle que, localement, d'après 1.8.2, le kurtosis doit être maximisé ou minimisé et que, d'après 1.8.3, la négentropie doit être maximisée. Il a été montré dans (Hyvärinen et Oja, 1997; Hyvärinen, 1997a; Hyvärinen et Oja, 2000) que si c'est vrai localement, alors la composante indépendante  $y = w^T V$  trouvée, dans l'espace blanchi, converge au signe près vers une des sources  $s_i$  (rappel de l'équation (1.41) de la partie 1.6.3).

Ainsi, de part l'indétermination du signe, il s'avère qu'extraire  $w$  ou  $-w$  ne change rien. Ainsi, les algorithmes du point fixe sur  $w$  convergent quand, d'une itération ( $k$ ) à la suivante ( $k + 1$ ), les vecteurs  $w(k)$  et  $w(k + 1)$  pointent dans la même direction, peu importe le sens (signe). Or, on verra que, dans les algorithmes proposés, les vecteurs  $w$  sont normalisés à une norme unitaire, par respect de l'orthogonalité théorique de  $W$ .

Cela signifie donc deux conditions équivalentes de convergence :

- a.  $w(k)$  et  $w(k + 1)$  sont égaux au signe près :  $w(k + 1) = \pm w(k)$ ;
- b. leur produit scalaire est égal à  $\pm 1$  :  $\langle w(k + 1), w(k) \rangle = w^T(k + 1)w(k) = \pm 1$ .

qui s'expriment respectivement mathématiquement, avec  $\epsilon$  un seuil positif infiniment petit, par :

$$|w(k + 1) \pm w(k)| < \epsilon \quad (3.17)$$

$$1 - |w^T(k + 1)w(k)| < \epsilon \quad (3.18)$$

Après avoir vu les conditions de convergence, débutons l'étude des algorithmes à point fixe par celui basé sur le kurtosis.

### 3.3.1.2 Basé sur le kurtosis

Tout d'abord, rappelons que, d'après 1.8.2, le kurtosis est défini par :

$$\text{kurt}(y) = E[y^4] - 3(E[y^2])^2 \quad (3.19)$$

où  $y$  est une variable centrée; et présente la propriété d'être nul dans le cas d'une variable gaussienne. Le but est de maximiser ou minimiser ce kurtosis.

Dans le cas blanchi, le kurtosis d'une composante indépendante  $y = w^T V$  centrée devient :

$$\text{kurt}(w^T V) = E \{(w^T V)^4\} - 3 [E \{(w^T V)^2\}]^2 = E \{(w^T V)^4\} - 3 \|w\|^4 \quad (3.20)$$

car  $E \{(w^T V)^2\} = E \{w^T V V^T w\} = w^T E \{V V^T\} w = w^T w = \|w\|^2$ . Or, vu que  $W$  est orthogonale, ses colonnes le sont également deux à deux; ce qui signifie que  $w$  vérifie la contrainte :  $\|w\|^2 = w^T w = 1$ ; soit la norme unitaire :  $\|w\| = \sqrt{w^T w} = 1$ .

Dans la suite, pour les vecteurs, la norme est définie par  $\|w\| = \sqrt{w^T w}$  et normaliser consiste à faire l'opération  $w \leftarrow w / \|w\|$ .

La contrainte de normalité de  $w$  représente une condition de convergence de l'algorithme. Ainsi, pour quantifier cet écart à la convergence, Hyvärinen et Oja (Hyvärinen et Oja, 1996a, 1997) ont introduit un terme de pénalité  $F(\|w\|^2)$ , tel que la fonction de contraste à maximiser ou minimiser soit :

$$J(w) = E \{(w^T V)^4\} - 3 \|w\|^4 + F(\|w\|^2) \quad (3.21)$$

En ce cas, trouver les composantes indépendantes revient à annuler le gradient de  $J$  par rapport à  $w$ , soit :

$$\frac{\partial J}{\partial w} = 0 = E \{V(w^T V)^3\} - 3 \|w\|^2 w + f(\|w\|^2) w \quad (3.22)$$

Finalement, sous la contrainte  $\|w\| = 1$  (qui sera appliquée dans les algorithmes, comme mentionné dans la partie précédente),  $\|w\|^2 = 1$  et  $f(\|w\|^2) = f(1)$  est une constante, que l'on néglige (Hyvärinen et Oja, 1996a), car d'après 3.3.1.1, c'est la direction de  $w$  qui est véritablement pertinente.

On obtient ainsi *l'équation du point fixe du kurtosis pour une composante* :

$$w = E \{V(w^T V)^3\} - 3w \quad (3.23)$$

L'algorithme du point fixe de l'ACI, basé sur le kurtosis, est donc, avec  $k$  l'itération, l'algorithme 1 ci-dessous.

On remarque bien que cet algorithme respecte les deux caractéristiques importantes de  $w$  : il vérifie l'équation du point fixe du kurtosis ( $w(k) = E \{V(w^T(k-1)V)^3\} - 3w(k-1)$ ),

---

**Algorithme 1** : Algorithme du point fixe du kurtosis pour une composante
 

---

Choisir un seuil  $\epsilon$ ;

Construire aléatoirement un vecteur  $w(0)$  de norme 1;

$$w(1) = E \{V(w^T(0)V)^3\} - 3w(0);$$

$$w(1) \leftarrow w(1) / \|w(1)\|;$$

$$k = 1;$$

**tant que**  $1 - |w^T(k)w(k-1)| \geq \epsilon$  **faire**

$$\left[ \begin{array}{l} k \leftarrow k + 1; \\ w(k) = E \{V(w^T(k-1)V)^3\} - 3w(k-1); \\ w(k) \leftarrow w(k) / \|w(k)\|; \end{array} \right.$$


---

l'orthonormalité de  $w$ , en le normalisant ( $w(k)/\|w(k)\|$ ) et s'assure qu'à la convergence, la direction de  $w$  est constante ( $1 - |w^T(k)w(k-1)| \geq \epsilon$ , négation logique de la condition (3.18)). La preuve de la convergence de l'algorithme est disponible dans (Hyvärinen et Oja, 1997).

Cet algorithme permet de récupérer la première colonne  $w_1$  de  $W$ , donc la première composante indépendante  $y_1 = w_1^T V$ . Nous présenterons en 3.3.2 la façon d'obtenir les  $n - 1$  autres colonnes et composantes car elle est similaire à celle de l'algorithme de la négentropie.

### 3.3.1.3 Basé sur la négentropie

À partir des faiblesses relevées chez le kurtosis en 1.8.2, telles que, principalement, son manque de robustesse par rapport aux valeurs aberrantes, on rappelle que, pour les parler, la négentropie  $J(y)$ , présentée en 1.8.3, et son approximation (1.67) suivante ont été introduites :

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \quad (3.24)$$

où  $y$  et  $\nu$ , variable gaussienne, sont centrées et de variance unitaire. Cette approximation est basée sur l'utilisation d'une fonction de contraste  $G$ , qui permet de généraliser l'expression du kurtosis.

Concernant cette fonction  $G$ , il a été montré dans (Hyvärinen, 1999a, 1998b, 1997b) que, pour des raisons de robustesse, elle doit croître plus lentement (respectivement plus rapidement) qu'une fonction quadratique dans le cas de variables supergaussiennes (resp. subgaussiennes). De plus, elles doivent respecter la condition suivante pour garantir la convergence de l'algorithme présenté plus bas :

$$E\{y_i g(y_i) - g'(y_i)\} \neq 0 \quad (3.25)$$

où  $g = G'$  est la dérivée de  $G$ . Ainsi, trois formes préférentielles ont été retenues pour  $G$  (avec l'expression de leurs dérivées premières  $g$  et secondes  $g'$ ) :

$$\begin{aligned} G_1(y) &= \frac{1}{a_1} \log \cosh a_1 y & G_2(y) &= -\frac{1}{a_2} e^{-a_2 y^2/2} & G_3(y) &= \frac{1}{4} y^4 \\ g_1(y) &= \tanh(a_1 y) & g_2(y) &= y e^{-\frac{a_2 y^2}{2}} & g_3(y) &= y^3 \\ g'_1(y) &= a_1 (1 - \tanh^2(a_1 y)) & g'_2(y) &= (1 - a_2 y^2) e^{-\frac{a_2 y^2}{2}} & g'_3(y) &= 3y^2 \end{aligned} \quad (3.26)$$

avec  $1 \leq a_1 \leq 2$  et  $a_2 \approx 1$ , deux constantes, telles que :

- $G_1$  est efficace pour tout type de situation;
- $G_2$  est plus adaptée dans les cas de variables très fortement supergaussiennes (kurtosis positif élevé) et/ou quand la robustesse joue un rôle prépondérant;
- $G_3$  n'est à utiliser que dans le cas de subgaussiennes, sans valeurs aberrantes, car avec  $G_3$ , (3.24) devient exactement le carré du kurtosis.

Ainsi, c'est généralement  $G_1$  qui est utilisé ! Cette affirmation concorde avec la conclusion de la partie 2.1.3 de l'état de l'art qui stipulait que le meilleur algorithme était le point fixe à méthode symétrique (nous la traitons en 3.3.2.2) à non-linéarité en tangente hyperbolique ( $g_1$ ).

La problématique (Hyvärinen, 1999a,b) consiste alors à trouver les colonnes  $w$  maximisant  $J(y) = J(w^T V)$ , sous la contrainte de variance unitaire  $E\{y^2\} = E\{(w^T V)^2\} = 1$ . De par l'expression de  $J(y)$ , la maximiser revient à trouver les optimums de  $E\{G(w^T V)\}$ . D'après les conditions de Kuhn-Tucker, ceux-ci sont obtenus pour les solutions  $w$  telles que :

$$F(w) = E\{Vg(w^T V)\} - \beta w = 0 \quad (3.27)$$

avec  $\beta = E\{w^T Vg(w^T V)\}$ , constante, telle que  $w = w_0$ , la solution recherchée et  $g$  une (au choix) des trois fonctions  $(g_i)_{i \in [1,3]}$  de l'équation (3.26).

On résout alors (3.27) par la méthode de Newton, telle que, pour l'itération  $k$  :

$$w(k+1) = w(k) - JF^{-1}(w(k))F(w(k)) \quad (3.28)$$

où  $JF$  est la matrice jacobienne de  $F$ , telle que :

$$JF(w) = E\{VV^T g'(w^T V)\} - \beta I \approx (E\{g'(w^T V)\} - \beta) I \quad (3.29)$$

On obtient alors l'itération de Newton suivante :

$$w(k+1) = w(k) - \frac{E\{Vg(w^T(k)V)\} - \beta w(k)}{E\{g'(w^T(k)V)\} - \beta} \quad (3.30)$$

Finalement, en multipliant chaque membre de l'équation par  $E\{g'(w^T(k)V)\} - \beta$  et suite à quelques simplifications arithmétiques, on aboutit à l'équation du point fixe de la négentropie pour une composante :

$$w = E\{Vg(w^T V)\} - E\{g'(w^T V)\}w \quad (3.31)$$

Ainsi, selon la fonction  $g$  choisie, on obtient l'une des trois équations suivantes :

$$w_{g_1} = E\{V \tanh(a_1 w^T V)\} - E\{a_1 (1 - \tanh^2(a_1 w^T V))\}w \quad (3.32)$$

$$w_{g_2} = E\{V(w^T V)e^{-\frac{a_2(w^T V)^2}{2}}\} - E\{(1 - a_2(w^T V)^2)e^{-\frac{a_2(w^T V)^2}{2}}\}w \quad (3.33)$$

$$w_{g_3} = E\{V(w^T V)^3\} - 3E\{(w^T V)^2\}w = E\{V(w^T V)^3\} - 3w \quad (3.34)$$

À noter que la dernière équation avec  $g_3$  est celle de la kurtosis; ce qui confirme, conformément à 1.8.3.3, que l'algorithme du point fixe basé sur la négentropie est une généralisation améliorée (car plus robuste) de celui du kurtosis.

En conservant la notation  $g$  générale, l'algorithme du point fixe de la négentropie est représenté par l'algorithme 2 suivant, où, de même que pour le kurtosis, le vecteur  $w$  recherché est normalisé à chaque itération et la condition de convergence est bien (3.18).

---

**Algorithme 2** : Algorithme du point fixe de la négentropie pour une composante

---

Choisir un seuil  $\epsilon$ ;

Construire aléatoirement un vecteur  $w(0)$  de norme 1;

$$w(1) = E \{Vg(w^T(0)V)\} - E \{g'(w^T(0)V)\} w(0);$$

$$w(1) \leftarrow w(1) / \|w(1)\|;$$

$$k = 1;$$

**tant que**  $1 - |w^T(k)w(k-1)| \geq \epsilon$  **faire**

$$\left[ \begin{array}{l} k \leftarrow k + 1; \\ w(k) = E \{Vg(w^T(k-1)V)\} - E \{g'(w^T(k-1)V)\} w(k-1); \\ w(k) \leftarrow w(k) / \|w(k)\|; \end{array} \right.$$


---

À noter, en outre, que pour la négentropie, il existe une autre forme possible d'algorithme, appelée algorithme du point fixe stabilisé, qui fait intervenir le paramètre  $\mu$ , qui est le pas d'échantillonnage et qui permet de régler la convergence de l'équation du point fixe, qui présente alors une forme très semblable à (3.30) :

$$w(k+1) = w(k) - \mu \frac{E\{Vg(w^T(k)V)\} - \beta w(k)}{E\{g'(w^T(k)V)\} - \beta} \quad (3.35)$$

Pour plus de détails, voir (Hyvärinen, 1999a).

Nous venons de voir comment calculer les vecteurs  $w$  individuellement de deux méthodes; nous allons à présent voir comment toutes les calculer.



### 3.3.2 Pour toutes les composantes

Remarquons pour débiter que l'utilisation des deux algorithmes du point fixe précédents pour calculer toutes les composantes porte le nom général de méthode *FastICA*, qui est la méthode privilégiée à utiliser pour résoudre un problème d'ACI, d'après la première partie (2.1) de l'état de l'art. À ce titre, le code Matlab des algorithmes présentés ci-dessous est disponible sur le site du HUT<sup>1</sup> (voir partie 3.2).

Pour obtenir les  $n - 1$  autres colonnes et composantes restantes à partir de celle extraite par les algorithmes précédents, il existe deux méthodes différentes (Hyvärinen, 1999a, 1997a; Hyvärinen et Oja, 1997).

Pour la première méthode, la méthode *déflation*, il suffit d'appliquer l'algorithme du point fixe  $n - 1$  fois consécutivement, en s'assurant bien que les vecteurs  $w$  nouvellement obtenus soient bien différents les uns des autres. Pour ce faire, on les décorrèle les uns des autres en leur faisant vérifier leur propriété d'orthogonalité mutuelle; et ce, en réalisant une *orthogonalisation de Gram-Schmidt* entre chaque vecteur  $w$  extrait. La matrice  $W$  s'obtient alors par concaténations successives des  $w$ , comme cela est présenté en 3.3.2.1.

Pour la seconde méthode, appelée méthode *symétrique*, on travaille directement à partir de  $W$ , qui, à chaque itération, est orthogonalisée par Gram-Schmidt et actualisée. Les explications sont données en 3.3.2.2.

Dans la suite, bien que les deux algorithmes du point fixe soient valables, du fait que la négentropie soit une généralisation du kurtosis (voir l'équation (3.34)), nous allons présenter les deux méthodes en utilisant l'équation du point fixe de la négentropie. Il est tout à fait possible d'adapter les algorithmes présentés au kurtosis; il suffit pour ce faire d'utiliser  $g_3$  dans l'équation du point fixe.

---

<sup>1</sup><http://www.cis.hut.fi/projects/ica/fastica/>

### 3.3.2.1 Méthode déflation

La méthode déflation consiste à calculer les vecteurs colonnes  $w$  un par un. En d'autres termes, après avoir déjà extrait  $i$  vecteurs  $w_1, w_2, \dots, w_i$ , on applique l'équation du point fixe à  $w_{i+1}$ . On lui soustrait ensuite, à chaque itération, ses projections sur les  $i$  précédents vecteurs ( $w_{i+1}^T w_j = \langle w_{i+1}, w_j \rangle$ ), pour s'assurer de son orthogonalité avec eux, par une orthogonalisation de Gram-Schmidt, représentable par les deux formes équivalentes suivantes :

$$w_{i+1} \leftarrow w_{i+1} - \sum_{j=1}^i (w_{i+1}^T w_j) w_j \quad (3.36)$$

$$w_{i+1} \leftarrow w_{i+1} - WW^T w_{i+1} \quad (3.37)$$

où  $W = (w_1, w_2, \dots, w_i)$  se construit à chaque nouveau vecteur calculé (colonne supplémentaire). Enfin, on normalise le nouveau vecteur  $w_{i+1}$  ainsi obtenu. L'algorithme 3 reprend ce processus.

### 3.3.2.2 Méthode symétrique

Contrairement à la méthode déflation, la méthode symétrique (Hyvärinen, 1999a), quant à elle, ne privilégie aucun vecteur  $w$  et consiste à partir directement d'une matrice  $W$  quelconque, à l'orthogonaliser, selon Gram-Schmidt, par l'opération :

$$W \leftarrow (WW^T)^{-1/2} W \quad (3.38)$$

qui est l'équivalent matriciel de deux opérations vectorielles (3.36) et (3.37). Ensuite, la mise à jour se fait selon deux règles possibles :

- a. une simple trouvée computationnellement (Hyvärinen, 1999a) :

$$W \leftarrow \frac{3}{2} W - \frac{1}{2} (WW^T) W \quad (3.39)$$

---

**Algorithme 3** : Algorithme de la méthode déflation
 

---

 $W = ()$ ;

 Choisir un seuil  $\epsilon$ ;

**pour**  $i=1$  à  $n$  **faire**

 Construire aléatoirement  $w_i(0)$ , de norme 1;

$$w_i(1) = E \{ V g(w_i^T(0)V) \} - E \{ g'(w_i^T(0)V) \} w_i(0);$$

$$w_i(1) \leftarrow w_i(1) - WW^T w_i(1);$$

$$w_i(1) \leftarrow w_i(1) / \|w_i(1)\|;$$

 $k = 1$ ;

**tant que**  $1 - |w_i(k)^T w_i(k-1)| \geq \epsilon$  **faire**
 $k \leftarrow k + 1$ ;

$$w_i(k) = E \{ V g(w_i^T(k-1)V) \} - E \{ g'(w_i^T(k-1)V) \} w_i(k-1);$$

$$w_i(k) \leftarrow w_i(k) - WW^T w_i(k);$$

$$w_i(k) \leftarrow w_i(k) / \|w_i(k)\|;$$

 $W = (W, w_i)$ ;

---

b. l'équation du point fixe de la négentropie :

$$W \leftarrow E \{ V g(W^T V) \} - E \{ g'(W^T V) \} W \quad (3.40)$$

On préfère la seconde car elle fait référence aux algorithmes précédents et c'est la méthode utilisée dans le code Matlab de FastICA.

Le critère de convergence est l'application matricielle de (3.18); soit, avec  $k$  l'itération :

$$1 - \|W^T(k)W(k-1)\| < \epsilon \quad (3.41)$$

où, du fait que  $W$  est construite orthogonale unitaire,  $W^T(k)W(k-1)$  est une matrice diagonale de termes compris entre  $-1$  et  $1$ , car étant des valeurs de cosinus par produit scalaire. En ce sens, la norme utilisée ici  $\|\cdot\|$  est le minimum absolu de cette diagonale; soit :

$$\|W^T(k)W(k-1)\| = \min |\text{diag}(W^T(k)W(k-1))| \quad (3.42)$$

Finalement, la méthode symétrique est résumée par l'algorithme 4, qui récapitule toutes les notions abordées. La preuve de sa convergence est disponible dans (Hyvärinen, 1999a).

---

**Algorithme 4** : Algorithme de la méthode symétrique

---

Choisir un seuil  $\epsilon$ ;

Construire aléatoirement  $W(0)$ ;

Orthogonaliser  $W(0)$  suivant :  $W(0) \leftarrow (W(0)W^T(0))^{-1/2} W(0)$ ;

$W(1) = E \{Vg(W^T(0)V)\} - E \{g'(W^T(0)V)\} W(0)$ ;

$W(1) \leftarrow (W(1)W^T(1))^{-1/2} W(1)$ ;

$k = 1$ ;

**tant que**  $1 - \|W^T(k)W(k-1)\| \geq \epsilon$  **faire**

$k \leftarrow k + 1$ ;
$W(k) = E \{Vg(W^T(k-1)V)\} - E \{g'(W^T(k-1)V)\} W(k-1)$ ;
$W(k) \leftarrow (W(k)W^T(k))^{-1/2} W(k)$ ;

---

Selon la conclusion de la partie 2.1.3 de l'état de l'art, l'algorithme FastICA à utiliser est donc l'algorithme 4, avec la non-linéarité  $g_1 = \tanh$ , symbolisé par l'abréviation « FP-symth ». Nous présentons, de plus, quelques propriétés intéressantes générales de FastICA, pour corroborer cette conclusion.

### 3.3.2.3 Propriétés du FastICA

L'algorithme FastICA présente les propriétés suivantes (Hyvärinen, 1999a) :

- a. la convergence est cubique (ou du moins quadratique), contrairement aux descentes de gradient qui sont linéaires. La conséquence est une convergence très rapide (d'où l'appellation « FastICA »);
- b. il n'y pas de paramètre d'apprentissage à régler au préalable, contrairement aux algorithmes du gradient; ce qui le rend simple à utiliser;

- c. l'algorithme trouve directement les composantes indépendantes de n'importe quelle distribution non gaussienne, en utilisant n'importe quelle non-linéarité  $g$ . Cependant, la performance peut être améliorée en s'intéressant de plus près à  $g$ ;
- d. les composantes peuvent être estimées une à une (déflation) ou simultanément (symétrique).

### 3.4 Sortie du FastICA

Finalement, dans le cas pratique présenté succinctement en 1.4.2.3, une fois que l'on a extrait de façon équivalente, tous les  $w_i$  par la méthode déflation ou  $W$  par la méthode symétrique, on obtient, à partir du tableau I, en sortie de la méthode FastICA :

- a. les composantes indépendantes par  $Y = W^T V$  (3.9), que l'on peut alors comparer avec  $S$ , suivant  $Y = GS$ ;
- b. la matrice de séparation  $B$  par  $B = W^T U$  (3.8), à comparer avec l'inverse  $A^{-1}$  selon  $B = GA^{-1}$ ;
- c. l'estimée de la matrice de mélange  $\tilde{A}$  par  $\tilde{A} = U^{-1}W$  (3.16), à comparer avec la matrice de mélange réelle  $A$ , selon  $A = \tilde{A}G$ .

Il apparaît ainsi que dans les trois cas,  $G$  est à extraire analytiquement, afin de vérifier que celle-ci concorde bien avec sa forme théorique (1.38). C'est en ce sens (rappels de 1.6.4 et 2.1.2.2) que l'utilisation de l'indice  $E_1$  comme mesure de performance prend tout son sens.

Nous allons à présent nous intéresser au protocole expérimental de notre étude, qui reprend les algorithmes FastICA présentés précédemment.

## CHAPITRE 4

### PROTOCOLE EXPÉRIMENTAL

Dans ce chapitre, nous allons directement mettre en pratique les algorithmes du point fixe que nous venons de présenter en élaborant un protocole expérimental pour récupérer les composantes indépendantes à partir de signaux sources, qui seront ici des images, des signaux basiques de traitement de signal (tels dent de scie, sinusoïde, . . .) ou des signatures. Dans la suite, l'appellation « signal » désignera indifféremment ces trois catégories. Signalons dès à présent que le protocole puis les résultats présentés ont été obtenus à partir de la programmation de codes Matlab.

#### 4.1 Généralités

Nous débutons en rappelant selon 1.4.2.3 les différentes approches expérimentales possibles, avant d'énoncer leur problème inhérent majeur et la solution adoptée pour le contre-carrer.

##### 4.1.1 Approches

Il existe deux approches possibles : la théorique et la pratique. Dans la première,  $S$  et  $A$  sont inconnues et le but est de les connaître (et non estimer) à partir de  $X$  seule. Ainsi, la matrice  $Y$  obtenue, qui est simplement  $S$ , présente son ambiguïté de signe et les sources sont définies au signe près. Dans notre application des signatures spectrales, le but est de déterminer les éléments (donc les signatures) présents dans la parcelle, en utilisant l'ACI; la finalité étant de réussir cette action sans connaissance a priori sur les éléments potentiellement présents dans cette parcelle. Cependant, l'ACI n'en étant qu'à ses balbutiements par rapport à l'application hyperspectrale, elle est plutôt considérée ici comme un outil, dont on veut juger (car on l'ignore pour le moment) la fiabilité, l'efficacité pour résoudre

la problématique de l'imagerie hyperspectrale; c'est pourquoi la seconde approche, la pratique, sera privilégiée dans la suite et fin de ce mémoire.

Cette approche pratique consiste, typiquement, à partir d'une base de  $N$  signatures pures de la base USGS, à :

- a. en choisir  $n$  d'entre elles;
- b. former  $n$  mélanges à partir des  $n$  choisies (on prend le même nombre de mélanges que de sources, suivant la conclusion de 1.5.1);
- c. appliquer ces  $n$  mélanges en entrée de l'ACI;
- d. en sortie, associer chacune des  $n$  composantes indépendantes obtenues à la signature de la base initiale qui lui est la plus ressemblante, selon un certain critère.

Cette approche est appelée  *$n$  parmi  $N$* .

En choisissant  $N = n$ , une autre approche, intitulée à juste titre  *$n$  parmi  $n$* , est envisageable. Elle consiste à considérer les signatures choisies comme formant la base entière. Cette dernière, intrinséquement plus simple, consiste à s'intéresser exclusivement à la robustesse de l'outil ACI, en se demandant si elle est capable de retourner les signatures qu'on lui a présentées en entrée (sous forme de mélanges); alors que l'approche plus générale dépasse finalement le cadre de l'ACI en rajoutant la dimension « comparaison / association » des paires « signaux originaux / composantes indépendantes », tout en s'interrogeant sur la capacité de l'ACI à extraire les signatures et à les reconnaître, ultérieurement et à part, dans une base de signatures de cardinalité plus importante.

De la façon dont les deux approches  *$n$  parmi  $N$*  et  *$n$  parmi  $n$*  ont été formulées, elles apparaissent comme deux notions proches mais distinctes. Au début de notre raisonnement, ce fut le cas, mais ce n'est que relativement récemment que nous avons réalisé que, finalement, elles étaient complémentaires et indissociables. En effet, leurs natures ont évolué au fur et à mesure de la prise de conscience, de plus en plus prononcée, du problème de post-

traitement inhérent à l'ACI : son absence totale de moyens de vérification de la validité des résultats. Voyons en quoi consiste exactement ce problème.

#### 4.1.2 Validité des résultats

Initialement, nous utilisons l'approche  $n$  parmi  $n$  pour juger de la robustesse de l'ICA, en comparant physiquement les signatures initiales et celles rendues par l'ACI et en associant, par une mesure de distance classique, les couples les plus ressemblants. Contrairement à son concept extrêmement simple, la mise en application de cette approche a mis en lumière la difficulté rédhibitoire de l'ambiguïté d'ordre de l'ACI : elle ne permet pas un traitement automatique; dans le sens où nous ne disposons d'aucun critère permettant d'affirmer avec certitude que les couples trouvés sont bien les bons !

En effet, l'ordre et la forme d'une même composante indépendante peuvent tellement varier d'une application (test) de l'ACI à une autre que vouloir comparer signatures initiales et finales nécessiterait rigoureusement l'intervention d'un observateur externe qui permettrait de visu de valider les associations ou non; et encore là, par expérience, celles-ci demeurent fort subjectives d'un sujet à l'autre et d'un test à l'autre (c'est atrocement vrai dans le cas de signatures de formes initialement proches et/ou de proportions de mélanges déséquilibrées); ce qui n'est évidemment pas tolérable comme protocole et pour l'utilisateur en question.

À ce titre, examinons l'exemple visuel de la figure 11. La première ligne représente les trois sources initiales (dans ce cas, les images 3, 4 et 5 de la base de données d'images du protocole (présentée en 5.2.1, on y reviendra le moment venu) qui ont été centrées et décorréélées). Après formation des mélanges (ici, chaque image participe à 33 % à chacun des trois mélanges) et traitement par l'ACI, nous avons obtenu les trois composantes (centrées et décorréélées également) de la deuxième ligne et les paires ont été associées par ressemblance et disposées en colonnes. La notion de subjectivité dans le choix prend tout son sens avec cet exemple : il n'est pas dit que tous auraient formé ces paires-ci. Alors



que dire d'une machine qui ne possède pas les facultés de l'oeil humain ! L'exemple a été conçu de telle façon que l'ACI fonctionne piètrement; ce dysfonctionnement provenant de la nature même des images, semblables dans leurs formes.

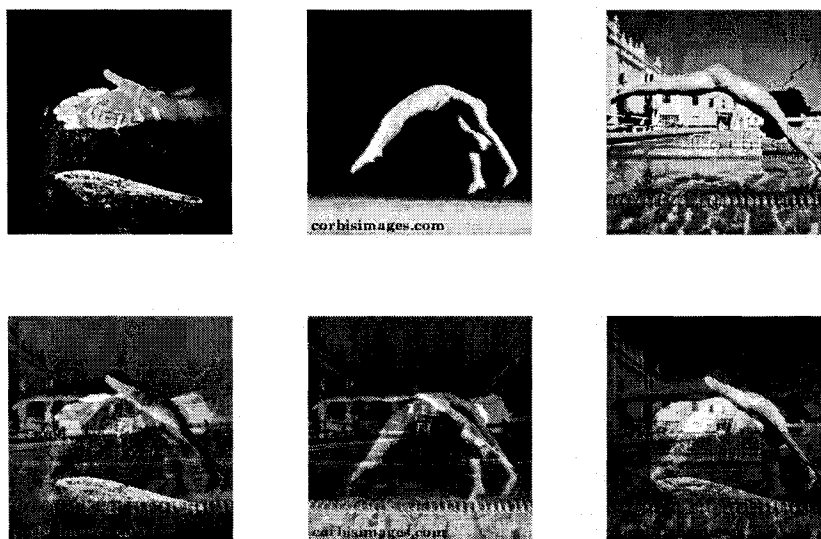


Figure 11 Exemple de subjectivité dans les associations signaux / composantes

Pour en revenir à la notion de certitude, nous nous « contentions » finalement d'associer les paires les plus ressemblantes selon un minimum de distance, sans pour autant être assuré que ces associations étaient correctes; une erreur de permutation des couples étant toujours possible, mais non décelable car il n'y a aucun critère de validation des résultats en mode automatique. Ce manque était évidemment encore plus manifeste dans le cas de l'approche  $n$  parmi  $N$ , où la seule chose que nous pouvions constater est l'absence ou la présence du signal de départ dans les associations : s'il était absent, cela représentait, à juste titre, une erreur mais s'il était présent, cela était comptabilisé comme une association correcte mais rien ne permettait d'en être persuadé; erreur de permutation possible oblige (encore et toujours).

Pour résoudre ce problème et tenter néanmoins d'automatiser un système qui, même totalement manuel, présentait des failles certaines, nous avons introduit la notion de *niveau de confiance* des résultats, pour juger de leur vraisemblance et ainsi passer en mode semi-automatique. Typiquement, le taux de bonnes associations est lié à un niveau de confiance : on traite les associations de façon automatique (non plus manuelle), tout en leur assignant un niveau de confiance; d'où un traitement semi-automatique et une démarche scientifique beaucoup plus rigoureuse du problème ! En d'autres termes, nous sommes (uniquement !) capables d'affirmer que pour tel ensemble de signatures, l'ACI et le nouveau système de post-traitement ont été capables de retrouver les signatures originales avec tel pourcentage d'erreur suivant une certaine confiance associée ! Les résultats auront donc la forme de taux d'erreurs, fonctions de niveaux de confiance.

#### 4.1.3 Mise au point lexicale

Ce qu'il est primordial de stipuler dès à présent est le fait que ce nous appelons (abusivement) *confiance* ici n'est en rien reliée à la notion d'*intervalle de confiance* en statistique. En fait, c'est en notre propre conscience que nous avons décidé d'utiliser ce terme car selon son expression et les explications données en 4.4.1.2, il nous semble logique, de par ce qu'elle représente, de parler de confiance.

Pour plus de détails, intéressons-nous à la façon dont, conceptuellement parlant, elle est calculée. En fait, nous allons succinctement expliquer qu'il existe deux types de confiances : la théorique et la pratique.

De même que ce qui vient d'être énoncé, l'emploi des expressions *théorique* et *pratique* est aussi un abus de langage dans le sens où, selon les approches décrites en 4.1.1, notre protocole se base uniquement sur une approche pratique, dans laquelle les signaux initiaux et les proportions de mélanges sont connus (contrairement à l'approche théorique) ! En ce sens, pour être rigoureux, il vaudrait mieux parler séparément d'approches :

- a. pratique théorique;

b. pratique pratique.

Cependant, afin d'éviter une lourdeur évidente et récurrente, nous postulons qu'à partir de maintenant les termes *approches théorique et pratique* représentent respectivement les approches pratique théorique et pratique pratique.

Maintenant que ces considérations de vocabulaire ont été clarifiées, précisons réellement ces notions de confiance théorique et pratique.

#### 4.1.4 Confiances

Dans la partie 4.1.1, nous avons décrit les deux approches  $n$  parmi  $n$  et  $n$  parmi  $N$ , en laissant sous-entendre finalement qu'elles étaient liées. En fait, comme nous allons le voir, chacune d'entre elles a un rôle spécifique : la première permet de calculer la confiance théorique tandis que la seconde calcule la confiance pratique et le taux d'erreur.

En effet, comme nous l'avons mentionné en 4.1.2, en mode automatique, nous n'avons pas ce critère de confiance. Pour l'obtenir, nous repensons le protocole de la façon suivante :

- a. l'approche  $n$  parmi  $n$  est redéfinie comme la phase *théorique* (de validation);
- b. l'approche  $n$  parmi  $N$  est redéfinie comme la phase *pratique* (de résultat).

Plus précisément, cela signifie que, dépendamment du cas, la nature de la matrice  $S$  des signaux de départ est différente. Dans le cas théorique, la matrice  $S$  considérée est celle des  $n$  signaux choisis parmi les  $N$  de la base totale. Dans le cas pratique, la matrice  $S$  est cette base totale que nous nommerons par la suite  $S^{\text{base}}$ . En effet, dans le cas réel, nous ne disposons que des  $n$  composantes et des  $N$  signaux de la base totale :  $S$  y est totalement inconnue, contrairement au cas théorique, par définition. De plus, vu que le nombre des composantes indépendantes rendues en sortie de l'ACI est égal à  $n$ , comparer  $S$  et  $Y$  présentera fort logiquement un taux d'erreur inférieur à une comparaison  $S^{\text{base}} / Y$ . Pour ces deux raisons, le premier type de comparaison sera qualifié de *théorique*, tandis que le suivant le sera de *pratique* : les paires formées à partir de  $S$  et  $Y$  seront considérées

comme les paires justes (théoriques), auxquelles nous comparerons les paires (pratiques) conçues à partir de  $S^{\text{base}}$  et  $Y$ .

Nous verrons en détail en 4.4.1.2 que les paires théoriques sont formées par ordre décroissant de confiances; elles-même calculées à partir de la matrice  $G$  (voir 1.6.3) reliant  $S$  et  $Y$ . Ces associations théoriques permettent ainsi de connaître les couples corrects, relativement à une confiance *théorique*. Pour le cas pratique, la même procédure est répétée en remplaçant  $S$  par  $S^{\text{base}}$ ; la nouvelle confiance associée sera elle qualifiée de *pratique*. Finalement, pour calculer le taux d'erreur, il suffit de comparer les couples pratiques aux couples théoriques. Ainsi, au lieu du seul critère du taux d'erreur utilisé initialement, nous disposons maintenant de deux critères supplémentaires (confiances théorique et pratique), permettant de pondérer les résultats en fonction de confiances.

Après avoir vu le protocole dans sa globalité avec l'expression des notions clés d'approche et de confiance, nous détaillons à présent les trois sections du protocole expérimental retenu : le prétraitement des données brutes, l'application de l'ACI et enfin, le post-traitement consistant en la formation des paires et aux calculs de confiances et du taux d'erreur associé à celles-ci. La structure du protocole expérimental est représentée à la figure 12.

## 4.2 Prétraitement

On adoptera dans la suite la notation suivante  $M_{\mu,\Sigma}$  désignant une matrice  $M$ , telle que :

- a. son vecteur moyenne est  $E[M_{\mu,\Sigma}] = \mu$ ;
- b. sa matrice de covariance est  $E[(M_{\mu,\Sigma} - \mu)(M_{\mu,\Sigma} - \mu)^T] = \Sigma$ .

Ouvrons une parenthèse pour signaler que cette notation est un abus dans le sens où deux matrices différentes, telles par exemple  $S_{\mu,\Sigma}$  et  $S_{\mu,\Sigma}^{\text{base}}$ , présentent les mêmes indices, alors que leurs moments peuvent être distincts. En fait, cette notation, comme nous le verrons ci-dessous, n'a d'intérêt que de distinguer les deux cas suivants :

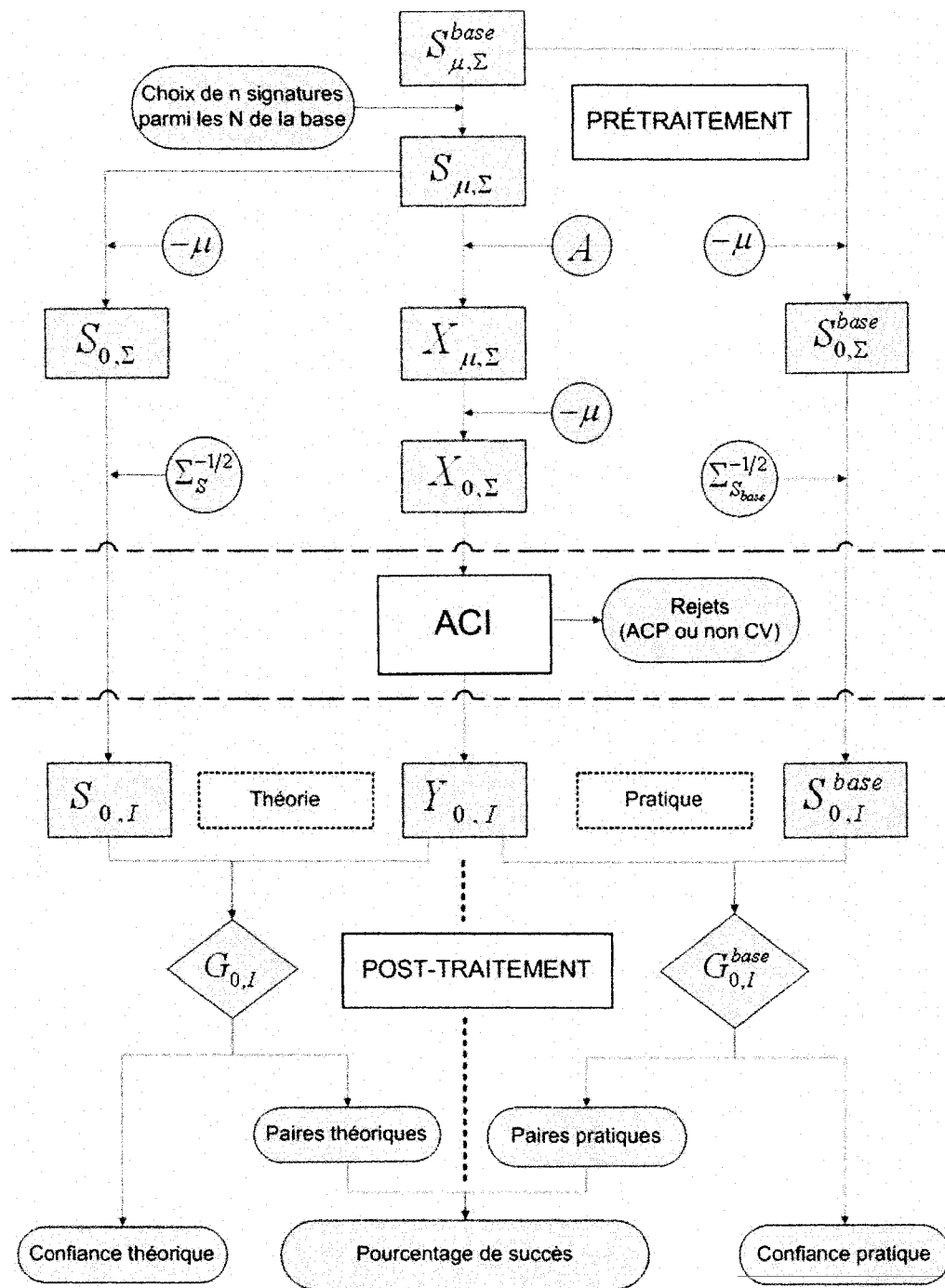


Figure 12 Protocole expérimental

- a. centrée ( $M_{0,\Sigma}$ ) ou non ( $M_{\mu,\Sigma}$ );
- b. décorrelée (matrice de covariance égale à la matrice identité ( $M_{\mu,I}$ )) ou non ( $M_{\mu,\Sigma}$ ).

Ainsi, dans la suite,  $M_{0,I}$  désignera une matrice  $M$  centrée et décorrelée. Refermons la parenthèse.

#### 4.2.1 Conformités statistiques avec le modèle

Le prétraitement, schématisé par la partie supérieure de la figure 12, englobe toutes les opérations statistiques permettant aux matrices  $S_{\mu,\Sigma}$  et  $X_{\mu,\Sigma}$  de satisfaire aux conditions théoriques, résumées en (1.20) et (1.37), imposées par l'algorithme FastICA; soit  $E[X] = 0$ ,  $E[S] = 0$  et  $E[SS^T] = I$ . Le but est donc de centrer les mélanges  $X$  et de décorreléer  $S$ .

En fait, le problème est plus subtil que cela. En effet, vu que dans la réalité, seule  $X$  (plus précisément  $X_{\mu,\Sigma}$ ) est connue, les seuls prétraitements permis expérimentalement sont ceux assurant uniquement que  $X$  est centrée (même si  $S$  est censée être décorrelée). Ainsi, vu que dans le protocole,  $X$  est formée à partir de  $S$ , par produit avec la matrice de mélange  $A$ , le prétraitement concernant  $X$  se présente sous la forme suivante :

$$X_{\mu,\Sigma} = A S_{\mu,\Sigma} \quad (4.1)$$

$$X_{0,\Sigma} \leftarrow X_{\mu,\Sigma} - \mu \quad (4.2)$$

où  $\mu = \mu_X = A \mu_S$ .

De fait,  $X$  sera alors de la forme correcte  $X_{0,\Sigma}$  centrée. Plus de détails sur la nature de  $A$  seront donnés dans la partie 4.2.2.

Le fait est que si nous avons formé  $X$ , tout d'abord, sous la forme  $X_{0,\Sigma} = A S_{0,\Sigma}$ , nous n'aurions pas suivi scrupuleusement la « vérité du terrain », où nous ne disposons réellement que des données brutes (non centrées); même si mathématiquement parlant, cela revient exactement au même puisque la moyenne de  $X$  est calculée à partir de celle

de  $S$ . En outre et indéniablement, il nous était impossible de former  $X$  sous la forme  $X_{0,I} = A S_{0,I}$  (par analogie avec  $Y_{0,I}$ ) car en ce cas, le protocole aurait été biaisé du fait qu'une partie de l'information théorique de  $S$  (principalement son moment d'ordre 2), censée être inconnue, aurait été déjà directement présente dans les mélanges; ce qui aurait faussé en bien les résultats. D'où une vigilance à avoir à l'esprit à ce niveau-là !

En ce sens, la forme  $X_{0,\Sigma}$  nous place en conformité avec l'entrée de l'ACI. Concernant la sortie, nous savons d'après (1.37) que les composantes  $Y$  sont de la forme  $Y_{0,I}$ . Ainsi, que ce soit dans le cas théorique avec  $S_{\mu,\Sigma}$  ou pratique avec  $S_{\mu,\Sigma}^{\text{base}}$ , on ne peut les comparer avec  $Y_{0,I}$  avec efficacité que si, justement, elles vérifient les mêmes propriétés statistiques pour les deux premiers moments; sinon, comme nous avons pu le constater empiriquement, des problèmes d'échelle liés aux variances des signaux  $S$  apparaissent et rendent impossible une comparaison pertinente.

Dans ce cas, il est toléré de modifier les moments de  $S$  car il s'agit de conditions *post-traitement*. Nous l'évoquerons plus en détail dans la partie 4.4, mais nous pouvons dès à présent noter que, partant de  $S_{\mu,\Sigma}$ , le post-traitement consistera, selon l'annexe 5, à réaliser les opérations suivantes (à adapter pour  $S_{\mu,\Sigma}^{\text{base}}$  avec notamment  $\Sigma_{S^{\text{base}}}$  au lieu de  $\Sigma_S$ ) :

$$S_{0,\Sigma} \leftarrow S_{\mu,\Sigma} - \mu \quad (4.3)$$

$$S_{0,I} \leftarrow \Sigma_S^{-1/2} S_{0,\Sigma} \quad (4.4)$$

Finalement, le post-traitement théorique (respectivement pratique) consistera à comparer  $S_{0,I}$  (resp.  $S_{0,I}^{\text{base}}$ ) avec  $Y_{0,I}$ , tel que représenté dans la partie inférieure de la figure 12. À ce moment, nous sommes bien en tout point en conformité avec les propriétés (1.20) et (1.37).

Il est important de souligner que, dans notre protocole,  $S_{\mu,\Sigma}$  est choisie, à partir de  $S_{\mu,\Sigma}^{\text{base}}$ , comme une des  $\binom{N}{n} = \frac{N!}{n!(N-n)!}$  combinaisons non ordonnées possibles de  $n$  éléments parmi  $N$ . On néglige l'ordre des signaux de départ, du fait de l'ambiguïté d'ordre en sortie

de l'ACI. Par exemple, la combinaison des signaux  $\{1, 2, 3\}$  désigne indifféremment les six ensembles ordonnés possibles suivants :  $\{1, 2, 3\}$ ,  $\{1, 3, 2\}$ ,  $\{2, 1, 3\}$ ,  $\{2, 3, 1\}$ ,  $\{3, 1, 2\}$  et  $\{3, 2, 1\}$ . Pour rendre valable notre étude, nous testons toutes les combinaisons non ordonnées possibles de  $n$  parmi  $N$ . Cela signifie algorithmiquement qu'une première boucle « for » sera utilisée pour déterminer la séquence  $S_{\mu,\Sigma}$  en question à partir de la base commune  $S_{\mu,\Sigma}^{\text{base}}$ , tel qu'indiqué à l'algorithme 5.

À ce stade, nous avons conceptuellement formé les mélanges, mais voyons à présent comment est précisément déterminée la matrice de mélange  $A$ .

#### 4.2.2 Formation de $A$

Concernant  $A$ , un de ses intérêts majeurs est de permettre de juger de la robustesse de l'ACI pour différentes matrices de mélange. En d'autres termes, le but est de déterminer l'effet des changements des proportions des signaux sources initiaux sur la capacité de l'ACI à obtenir les bonnes composantes indépendantes. Pour ce faire, pour un même ensemble de signaux sources (une combinaison particulière), nous faisons varier les différents pourcentages de  $A$ . En d'autres termes, cela signifie, de même que dans la partie précédente, la présence d'une boucle « for » dédiée aux différentes  $A$ . De par les conditions à remplir énoncées ci-bas et les deux boucles « for », nous avons décidé de restreindre notre étude à un nombre de signaux  $n = \{2, 3, 4\}$  au départ.

Les contraintes à respecter, conformément à ce qui a déjà été affirmé au cours de ce mémoire, sont les suivantes :

- a. un rang plein; c'est-à-dire un déterminant non nul;
- b. la somme des lignes égale à 1;
- c. avoir une matrice  $A$  dont les lignes sont « sensiblement » égales et non parfaitement égales (pour traduire l'analogie avec les constitutions proches mais différentes des parcelles de champ contiguës !), de manière à avoir des matrices quasi-uniformes et



surtout éviter que, pendant le blanchiment, l'ACP ne réduise cette matrice en une seule ligne et donc annule notre test;

- d. faire varier, pour chaque nouveau test, le pourcentage de la première source par rapport aux autres (que l'on considère égales), par un pas relativement faible.

Pour satisfaire toutes les conditions précédentes, nous introduisons le pas  $k$  ( $k \in [0, 1]$ ) du pourcentage  $p$  du premier signal et le pas correctif  $\epsilon$ , tels que :

- a.  $p$  varie de 0 à 1 par pas de  $k$ ; soit  $p = i \cdot k$ ,  $i \in [0, \frac{1}{k}]$ , où  $i$  est un entier naturel, ce qui signifie que  $k$  doit vérifier la condition  $1/k$  entier;
- b.  $\epsilon$  permet de rendre les lignes non égales pour ne pas être annulées par l'ACP mais suffisamment proches pour que le test soit valide.

Typiquement, pour  $n$ ,  $k$  et  $\epsilon$  donnés, la matrice  $A_{n,k,\epsilon}$  est un ensemble de matrices  $A_{n,p,\epsilon}$  telles que :

$$A_{n,k,\epsilon} = (A_{n,p,\epsilon})_{p=\{i \cdot k, i \in [0, \frac{1}{k}]\}} \quad (4.5)$$

avec, selon les trois valeurs de  $n$  :

$$A_{2,p,\epsilon} = \begin{pmatrix} p & 1-p \\ p+\epsilon & 1-p-\epsilon \end{pmatrix}$$

$$A_{3,p,\epsilon} = \begin{pmatrix} p-\epsilon & \frac{1-p}{2} + \epsilon & \frac{1-p}{2} \\ p & \frac{1-p}{2} & \frac{1-p}{2} \\ p+\epsilon & \frac{1-p}{2} & \frac{1-p}{2} - \epsilon \end{pmatrix} \quad (4.6)$$

$$A_{4,p,\epsilon} = \begin{pmatrix} p-\epsilon & \frac{1-p}{3} + \epsilon & \frac{1-p}{3} & \frac{1-p}{3} \\ p & \frac{1-p}{3} & \frac{1-p}{3} & \frac{1-p}{3} \\ p & \frac{1-p}{3} & \frac{1-p}{3} - \epsilon & \frac{1-p}{3} + \epsilon \\ p+\epsilon & \frac{1-p}{3} & \frac{1-p}{3} & \frac{1-p}{3} - \epsilon \end{pmatrix}$$

Du fait de la présence dans la première colonne de  $-\epsilon$  pour  $n = \{3, 4\}$  (respectivement  $+\epsilon$  pour  $n = \{2, 3, 4\}$ ), le cas  $i = 0$  (resp.  $i = \frac{1}{k}$ ) est impossible, car sinon, on se retrouverait avec des éléments négatifs (resp. supérieurs à 1); ce qui est impossible, par définition de la matrice de mélange. On préfère donc plutôt considérer  $i \in [1, \frac{1}{k} - 1]$  pour les trois valeurs de  $n$ ; soit  $\frac{1}{k} - 1$  matrices  $A_{n,p,\epsilon}$ .

Ainsi, par exemple, pour  $n = 3$ ,  $k = 10\%$  et  $\epsilon = 1\%$ , on obtient les premières matrices  $A_{3,p,0.01}$  suivantes :

$$A_{3,0.1,0.01} = \begin{pmatrix} 0.09 & 0.46 & 0.45 \\ 0.10 & 0.45 & 0.45 \\ 0.11 & 0.45 & 0.44 \end{pmatrix} \quad A_{3,0.2,0.01} = \begin{pmatrix} 0.19 & 0.41 & 0.40 \\ 0.20 & 0.40 & 0.40 \\ 0.21 & 0.40 & 0.39 \end{pmatrix}$$

$$A_{3,0.3,0.01} = \begin{pmatrix} 0.29 & 0.36 & 0.35 \\ 0.30 & 0.35 & 0.35 \\ 0.31 & 0.35 & 0.34 \end{pmatrix} \quad A_{3,0.4,0.01} = \begin{pmatrix} 0.39 & 0.31 & 0.30 \\ 0.40 & 0.30 & 0.30 \\ 0.41 & 0.30 & 0.29 \end{pmatrix}$$

et ainsi de suite, jusqu'à  $p = (\frac{1}{k} - 1)k = 1 - k = 0.9 = 90\%$ , pour un total de  $\frac{1}{k} - 1 = 9$  matrices  $A_{3,p,0.01}$ .

On remarque, d'après leurs formes, que ces matrices  $A_{n,p,\epsilon}$  ont été construites (par tâtonnements judicieux) de telle façon que ce soit la première source (première colonne) qui varie en pourcentage et que le reste des changements soient uniformément (à  $\epsilon$  près) répartis sur les  $n - 1$  restantes, tout en respectant les conditions énoncées plus haut.

En effet, en calculant les déterminants, on trouve que  $\det A_{n,p,\epsilon} = (-\epsilon)^n \neq 0$ ; d'où  $\text{rang}(A_{n,k,\epsilon}) = n$  est plein et la première condition est respectée dans les trois cas. De vue, la somme des lignes vaut bien 1. Concernant les valeurs de couples  $(k, \epsilon)$ , pour lesquelles l'ACP ne pose pas de problème de simplifications intempestives non désirées, vu

qu'elles varient en fonction de l'application considérée, nous ne les donnerons que dans le chapitre 5, dédié aux résultats.

Toutes les conditions sont ainsi bien respectées. On forme alors les mélanges (non centrés) par produit de  $A$  par  $S_{\mu,\Sigma}$ ; ce qui conclut l'étude du prétraitement. Passons à présent à celle de l'application de l'ACI.

### 4.3 Application de l'ACI

Dans cette partie, l'ACI rend, suivant la procédure détaillée au chapitre 3, les composantes indépendantes  $Y_{0,I}$ , la matrice de séparation  $B$  et l'estimée  $\tilde{A}$  de  $A$ , à partir des mélanges  $X_{0,\Sigma}$ . Nous avons opté, suivant les conclusions de la partie 2.1 de l'état de l'art, pour la méthode symétrique et la non-linéarité en tangente hyperbolique.

Il est à noter que, vu que nous n'exécutons qu'une seule fois l'algorithme ACI, il s'avère possible et judicieux de comparer les résultats d'association obtenus entre  $Y_{0,I}$  et  $S_{0,I}$  d'une part (théorie), et entre  $Y_{0,I}$  et  $S_{0,I}^{\text{base}}$  d'autre part (pratique). S'il avait fallu l'exécuter deux fois, de part son ambiguïté d'ordre, aucune comparaison n'eut été possible. Ce fait constitue un bon point positif de notre protocole.

En outre, un point important a dû être considéré à ce niveau : l'irrégularité de convergence d'une application à l'autre de l'algorithme ACI; on parlera de *test*. En effet, vu que le point de départ de l'algorithme 4 de la méthode symétrique est une matrice  $W$  aléatoire, il arrive qu'en certains cas isolés (signaux initiaux proches), l'algorithme ACI ne converge pas. Pour remédier à cette situation, nous réalisons plusieurs tests, de façon à moyenner cet effet. De fait, tel qu'indiqué à l'algorithme 5, une troisième boucle « for », dédiée aux différents tests, est utilisée. Sauf contre-indication, 50 tests sont réalisés par séquence (c'est-à-dire pour une matrice  $A$  et une combinaison  $S_{\mu,\Sigma}$  données). Nous avons fixé à 1 000, le nombre maximal d'itérations de l'algorithme FastICA permises pour converger.

Au-delà, le test est considéré comme non-convergeant et surtout *non valide* et ne subit donc pas le post-traitement ultérieur.

Ainsi, dans les résultats présentés dans le chapitre suivant, on distinguera les tests valides de ceux qui ne le sont pas, pour les deux raisons possibles : l'ACP et la non convergence. En effet, concernant l'ACP, l'expérience a montré qu'il est possible que, dépendamment de la combinaison  $S$  traitée, l'ACP réduise le nombre de signaux et rende donc invalide le test en question. Par conséquent, seuls les tests n'ayant subi ni la réduction de dimension par l'ACP, ni la non convergence par FastICA seront considérés comme valables et seront traités en post-traitement. On considérera donc dans notre analyse des résultats le *pourcentage de convergence* des tests de chaque matrice  $A$  de chaque combinaison. Intéressons nous à présent au post-traitement.

#### 4.4 Post-traitement

Tel que représenté à la figure 12, le post-traitement peut se décomposer en trois parties :

- a. la théorie comparant  $S_{0,I}$  et  $Y_{0,I}$  et fournissant confiance et paires théoriques;
- b. la pratique comparant  $S_{0,I}^{\text{base}}$  et  $Y_{0,I}$  et fournissant confiance et paires pratiques;
- c. le calcul du pourcentage de succès d'associations du test en comparant paires théoriques et pratiques.

Voyons en détail en quoi celles-ci consistent. À noter qu'un exemple complet de l'application du post-traitement du protocole est détaillé à la partie 4.7, dans laquelle toutes les notions qui vont être présentées maintenant sont concrètement explicitées.

##### 4.4.1 Partie théorique

Dans la partie théorique (ce sera la même chose adaptée pour la partie pratique), la procédure est la suivante : calcul de la matrice  $G_{0,I}$  puis calcul de la confiance théorique à partir de celle-ci.

#### 4.4.1.1 Matrice $G_{0,I}$

L'intérêt de cette matrice  $G_{0,I} = (g_{ij})_{(i,j) \in [1,n]^2}$  est qu'elle est le lien entre  $S_{0,I}$  et  $Y_{0,I}$  :  $g_{ij}$  est la contribution du signal  $S_j$  à la composante indépendante  $Y_i$ . D'après 1.6.3 et des simplifications matricielles, elle se calcule de la façon suivante :

$$G_{0,I} = Y_{0,I} S_{0,I}^{-1} \quad (4.7)$$

$$= (B X_{0,\Sigma}) \left( \Sigma_S^{-1/2} S_{0,\Sigma} \right)^{-1} = B (X_{0,\Sigma} S_{0,\Sigma}^{-1}) \Sigma_S^{1/2} = B A \Sigma_S^{1/2} \quad (4.8)$$

Bien que cela revienne algébriquement au même, pour éviter de calculer des pseudo-inverses ( $S_{0,I}^{-1}$ ), nous avons préféré, dans notre programme, calculer  $G_{0,I}$  comme produit de  $B$  par  $A$ , pondéré par  $\Sigma_S^{1/2}$ .

On rappelle, en outre, que, sous les conditions théoriques des deux premiers moments neutralisés (ce qui est explicitement le cas ici),  $G_{0,I}$  est interprétable comme le produit d'une matrice de permutation par une matrice diagonale de termes  $\pm 1$ . Or, vu que l'on veut associer à chaque composante le signal qui lui est le plus proche, la procédure logique consisterait à associer les paires par maximum absolu par ligne de  $G_{0,I}$ , qui est censé avoir une valeur proche de 1. Cependant, nous avons remarqué empiriquement que, dans certains cas litigieux (notamment plusieurs maxima dans la même colonne : plusieurs signaux assignés à la même composante), les associations ne satisfaisaient pas ce critère; c'est pourquoi nous avons introduit la notion de *confiance*, version adaptée de l'erreur  $E_1$ , présentée en 1.6.4.

On se rappellera par la suite la mise au point lexicale de la partie 4.1.3 concernant la nature propre de « notre » confiance, qui ne se veut en aucune façon représenter un intervalle de confiance statistique.

#### 4.4.1.2 Confiance théorique

Ce qui apparaît finalement comme le plus pertinent n'est pas tant le maximum que l'écart de ce dernier avec :

- a. la somme des contributions des autres signaux pour la composante  $Y_i$  (ligne  $i$ );
- b. la somme des contributions des autres composantes pour le signal  $S_j$  (colonne  $j$ ).

Ainsi, logiquement, plus l'écart entre ce maximum et ces deux sommes sera important, plus la probabilité de bonne association entre la composante et le signal lui correspondant sera élevée. La confiance représente justement cette probabilité et l'erreur  $E_1$ , telle que formulée en 1.6.4 selon (1.42)

$$E_1 = \sum_{i=1}^n \left( \sum_{j=1}^n \frac{|g_{ij}|}{\max_k |g_{ik}|} - 1 \right) + \sum_{j=1}^n \left( \sum_{i=1}^n \frac{|g_{ij}|}{\max_k |g_{kj}|} - 1 \right)$$

traduit mathématiquement cette notion d'écart, tout en demandant un travail d'adaptation à notre cas, présenté maintenant.

Pour ce faire, vu qu'ici, on ne s'intéresse qu'à une seule ligne et colonne par terme  $g_{ij}$ , au lieu des  $2n$  termes de  $E_1$ , on n'en calcule que deux, que nous nommons :

- a.  $E(S_j/Y_i)$ , l'erreur sur la ligne  $i$  du signal  $S_j$ , connaissant  $Y_i$ ;
- b.  $E(Y_i/S_j)$ , l'erreur sur la colonne  $j$  de la composante  $Y_i$ , connaissant  $S_j$ .

Ainsi,  $E(S_j/Y_i)$  représente le premier écart, tandis que  $E(Y_i/S_j)$  le second. Typiquement, si, par exemple,  $E(S_j/Y_i) = 0.1$ , cela signifie que, sur la ligne  $i$ , le reste (la somme absolue de toutes les valeurs, excepté  $g_{ij}$ ) représente 10 % de  $|g_{ij}|$  et on peut donc intuitivement que le couple  $(i, j)$  est a priori correct.

À partir de ces erreurs, on définit pour chaque couple  $(i, j)$  de composante/signal, les deux matrices d'erreurs suivantes  $E_{SY}$  et  $E_{YS}$ , de termes génériques :

$$E_{SY}(i, j) = E(S_j/Y_i) = \sum_{c=1}^n \frac{|g_{ic}|}{|g_{ij}|} - 1 = \sum_{c=1, c \neq j}^n \frac{|g_{ic}|}{|g_{ij}|} \quad (4.9)$$

$$E_{YS}(i, j) = E(Y_i/S_j) = \sum_{l=1}^n \frac{|g_{lj}|}{|g_{ij}|} - 1 = \sum_{l=1, l \neq i}^n \frac{|g_{lj}|}{|g_{ij}|} \quad (4.10)$$

Ainsi,  $E_{SY}(i, j)$  (respectivement  $E_{YS}(i, j)$ ) représente le pourcentage du reste par rapport à  $g_{ij}$  sur la ligne  $i$  (resp. la colonne  $j$ ). Si  $|g_{ij}|$  est effectivement le maximum absolu de la ligne  $i$  et colonne  $j$ , on retombe exactement sur les deux termes génériques de  $E_1$ ; ce qui montre bien que notre méthode est une version particularisée de  $E_1$  : *on considère l'erreur  $E_1$  dans un sens local, plutôt que dans son sens commun global!* Cette idée est également un point bénéfique de notre méthode.

L'erreur du couple  $(Y_i, S_j)$  est alors définie comme la moyenne de ces deux erreurs, soit :

$$E(Y_i, S_j) = \frac{E_{SY}(i, j) + E_{YS}(i, j)}{2} \quad (4.11)$$

On divise par 2 car, comme expliqué plus haut, seuls deux termes (une ligne et une colonne) sont considérés dans le calcul de  $E(Y_i/S_j)$  et  $E(S_j/Y_i)$ .

La *matrice de confiance* théorique (locale)  $C_L$ , de terme général  $(C_L)_{ij} = C_L(Y_i, S_j)$ , est alors définie en fonction de l'erreur moyenne selon :

$$(C_L)_{ij} = C_L(Y_i, S_j) = 1 - \frac{E(Y_i, S_j)}{n - 1} \quad (4.12)$$

Cette confiance  $C_L(Y_i, S_j)$  représente typiquement l'écart relatif du poids  $|g_{ij}|$  du couple  $(Y_i, S_j)$  par rapport aux sommes absolues des autres poids sur les ligne  $i$  et colonne  $j$ .

Vu que l'on souhaite logiquement une confiance comprise entre 0 et 1, avant de la calculer comme « réciproque » de l'erreur par rapport à 1, nous divisons celle-ci par  $n - 1$ , qui est le

maximum possible commun de  $E_{SY}(i, j)$  et  $E_{YS}(i, j)$  (la moyenne maintient cette valeur). Il est obtenu, selon 1.6.4, dans le cas (très) particulier où, par exemple pour  $E_{SY}$ , tous ses termes sont égaux (les maxima de chaque ligne sont les minima). En appliquant (4.9), on trouve bien  $n - 1$ . À noter que cette valeur maximale a été considérée à partir d'un  $|g_{ij}|$  au dénominateur de (4.9) maximum de sa ligne (ou colonne). S'il ne l'est pas, l'erreur  $E(Y_i, S_j)$  peut très bien s'avérer supérieure à  $n - 1$  (en fait, nous n'avons aucun contrôle en ce cas) et la confiance négative. Cela dit, outre la possible étrangeté logique d'une telle situation, elle apparaît finalement pertinente car s'il ne s'agit pas du maximum, il a peu de raisons d'être correct pour former le couple et se voit donc attribuer une confiance négative, signe de sa non-sélection.

La procédure est alors de calculer la confiance pour chaque terme de  $G_{0,r}$  et de former ensuite les couples par ordre décroissant des confiances théoriques : on obtient la première paire théorique, avec sa confiance associée; on retire alors de la matrice de confiance les ligne et colonne correspondant à cette paire et on forme la deuxième paire avec la confiance restante la plus élevée; et ainsi de suite jusqu'à traitement de toutes les paires. Il est important de noter que le signe de la composante de la paire  $(i, j)$  formée est le signe de  $g_{ij}$ , pour tenir compte de l'ambiguïté de signe des composantes.

Cette méthode permet ainsi de régler le problème des cas litigieux de plusieurs composantes pour le même signal : on choisit celle qui présente la plus grande confiance théorique; c'est-à-dire « celle en laquelle on a le plus confiance » (selon notre propre définition de la confiance, on le répète). Les restants auront une confiance négative (car non maximum); ce qui les élimine d'office. La sélection de paires en a donc été améliorée.

Ainsi, à la fin de la partie théorique du post-traitement, nous disposons des paires théoriques par ordre décroissant de confiances associées. Intéressons-nous à présent à la partie pratique du post-traitement.



#### 4.4.2 Partie pratique

Concernant la partie pratique, le concept est exactement le même. La seule et grande différence vient du fait que l'on considère  $S_{0,I}^{\text{base}}$ , plutôt que  $S_{0,I}$ ; soit  $G_{0,I}^{\text{base}}$ , plutôt que  $G_{0,I}$ . Là où  $G_{0,I}$  est une matrice carrée  $[n, n]$ ,  $G_{0,I}^{\text{base}}$  est de dimension  $[n, N]$  et il n'existe alors aucun moyen, tel qu'en (4.8), d'éviter le calcul d'une pseudo-inverse. Ainsi,  $G_{0,I}^{\text{base}}$  est définie par :

$$G_{0,I}^{\text{base}} = Y_{0,I} S_{0,I}^{\text{base}^{-1}} \quad (4.13)$$

ce qui pourrait éventuellement entraîner des soupçons quant à la capacité et la façon dont Matlab procède, spécialement avec des faibles valeurs; mais nous ne rentrons pas dans ces considérations purement logicielles.

En outre, il ne s'agit plus de l'association d'une des  $n$  composantes parmi  $n$  signaux possibles, mais parmi  $N$  ( $N \geq n$ ). De fait, le maximum possible pour la confiance pratique locale va différer. Plus précisément, selon le même principe de calcul que dans la partie précédente, le maximum possible de  $E_{SY}$  est  $N - 1$  (car  $G_{0,I}^{\text{base}}$  a  $N$  colonnes) et celui de  $E_{YS}$  demeure  $n - 1$  ( $n$  lignes). Ainsi, selon (4.11) qui reste valide, nous avons les inégalités suivantes, pour tout  $(i, j) \in [1, n] \times [1, N]$  :

$$0 \leq E_{SY}(i, j) \leq N - 1 \quad (4.14)$$

$$0 \leq E_{YS}(i, j) \leq n - 1 \quad (4.15)$$

$$0 \leq E(Y_i, S_j) \leq \frac{N - 1 + n - 1}{2} = \frac{N + n - 2}{2} \quad (4.16)$$

La *matrice de confiance* pratique (locale)  $C_L^{\text{base}}$ , de terme général  $(C_L^{\text{base}})_{ij} = C_L^{\text{base}}(Y_i, S_j)$ , est alors définie selon :

$$(C_L^{\text{base}})_{ij} = 1 - \frac{2}{N + n - 2} E(Y_i, S_j) \quad (4.17)$$

Cette définition de la confiance pratique locale appelle deux remarques. Tout d'abord, d'après l'équation (4.11), il aurait pu sembler préférable de normaliser chacune de deux

erreurs (suivant respectivement  $N - 1$  et  $n - 1$ ) avant de les moyenner; cependant, ce coefficient de normalisation  $\frac{N+n-2}{2}$  a l'avantage de « pondérer » différemment l'impact des deux erreurs. En effet, vu que  $N \geq n$ , il est plus pertinent d'accorder plus d'importance à l'erreur suivant les lignes ( $E_{SY}$ ) à  $N$  éléments que suivant les colonnes ( $E_{YS}$ ) à  $n$  éléments. Le fait de ne normaliser qu'au moment de définir la confiance permet de conserver judicieusement cette propriété !

En outre et comme conséquence de cette normalisation pondérée, il ne sera pas possible de comparer les confiances théoriques directement par rapport aux pratiques; dans le sens où, par exemple, une affirmation du genre « la pratique sera logiquement plus faible que la théorique, car répartie sur un plus grand nombre de signaux initiaux » n'est pas valable. Il y a du sens et de la logique à comparer théoriques entre elles et pratiques entre elles mais pas les unes par rapport aux autres.

Finalement, hormis ces considérations de cardinalité de signaux, la méthode d'associations des couples, basée sur la décroissance des confiances, reste exactement la même. Au final, on dispose de  $n$  paires formées, auxquelles sont associées leurs confiances pratiques.

Passons au calcul du taux d'erreur ou, de façon équivalente, du pourcentage de succès d'associations.

#### 4.4.3 Pourcentage de succès

Après avoir calculé paires et confiances, théoriques et pratiques, pour une combinaison  $S_{\mu,\Sigma}$  donnée, on forme sa matrice *resultat* (sans accent par conformité avec la notation Matlab) qui compile toutes ces informations et qui, au final, pour tous les tests et toutes les matrices  $A$ , est de dimension 3; la dernière étant dédiée à  $A$ . Concernant les deux premières dimensions, elle présente  $2n$  lignes et «  $2 \times$  nombre de tests » colonnes.

Plus précisément, pour chaque test, qui correspond à deux colonnes, on « écrit » à :

- a. la  $i^{\text{ème}}$  ligne ( $i \leq n$ ) de la *première* colonne, l'indice *signé* du signal  $(S_{j_t})_{j_t \in [1, n]}$  associé *théoriquement* à la composante  $Y_i$ . Par exemple, pour  $i = 2$ , si l'on détermine le couple  $(-Y_2/S_4)$ , on y inscrira la valeur  $-4$ ;
- b. la  $i^{\text{ème}}$  ligne de la *deuxième* colonne, la confiance théorique associée au couple théorique déterminé à la première colonne;
- c. la  $(n + i)^{\text{ème}}$  ligne de la *première* colonne, l'indice *signé* du signal  $(S_{j_p})_{j_p \in [1, N]}$  associé *pratiquement* à la composante  $Y_i$ ;
- d. la  $(n + i)^{\text{ème}}$  ligne de la *deuxième* colonne, la confiance pratique associée au couple pratique déterminé à la première colonne.

Vu que le nombre et la nature des composantes restent constants que l'approche soit théorique ou pratique (car nous n'avons réalisé qu'un seul et même test pour les deux), l'intérêt de cette méthode réside dans la simplicité de comparaison des paires théoriques et pratiques. En effet, de par la construction de la matrice *resultat*, on sait que les  $i^{\text{ème}}$  et la  $(n + i)^{\text{ème}}$  lignes désignent la même composante indépendante  $Y_i$ . Il y a donc erreur d'association du signal  $S_{j_t}$  pour ce test quand, à la première colonne, les indices signés de ces deux lignes ne coïncident pas; a contrario succès d'association quand ils coïncident. On traite de même tous les signaux  $(S_j)_{j \in [1, n]}$  pour chacun des tests.

On détermine ainsi, par matrice de mélange et par combinaison, le *pourcentage de succès* de chaque signal  $S_j$  comme la moyenne des succès d'association sur l'ensemble des tests; typiquement le rapport du nombre de bonnes associations sur le nombre total de tests.

#### 4.4.4 Mesures globales d'erreur

En plus du pourcentage de succès, nous avons calculé deux mesures d'erreur globales supplémentaires par test : l'erreur  $E_1$ , telle que rappelée plus tôt et l'erreur  $E_A$ , qui est la matrice de l'écart d'estimation entre  $A$  et  $\tilde{A}$ .

#### 4.4.4.1 Erreur $E_1$ et confiances globales

Conjointement aux erreurs  $E_1$  adaptées locales calculées pour chacun des termes  $g_{ij}$  pour obtenir les confiances locales, il est intéressant d'introduire l'erreur  $E_1$ , comme mesure globale d'erreur et pour déterminer, selon le même principe que 4.4.1.2, la confiance globale du test. Cette procédure va en fait se décomposer en deux parties :

- a. calcul de  $E_1$  à partir de  $G_{0,I}$ , qui donnera la confiance théorique globale du test;
- b. calcul de  $E_1^{\text{base}}$  à partir de  $G_{0,I}^{\text{base}}$ , qui donnera la confiance pratique globale du test.

Pour calculer ces deux erreurs, il suffit d'appliquer la définition (1.42) sur les deux matrices  $G$ . Les confiances théorique  $C_G$  et pratique  $C_G^{\text{base}}$  globales du test sont alors définies suivant l'équation suivante :

$$C_G = 1 - \frac{E_1}{2n(n-1)} \quad (4.18)$$

$$C_G^{\text{base}} = 1 - \frac{E_1^{\text{base}}}{n(N-1) + N(n-1)} \quad (4.19)$$

où les diviseurs  $2n(n-1)$  et  $n(N-1) + N(n-1)$  représentent, selon l'approche de 4.4.1.2, les maximums possibles de  $E_1$  et  $E_1^{\text{base}}$  car pour la première erreur, il y a  $2n$  calculs semblables à une ligne de maximum  $n-1$ ; alors que pour la seconde, il y a  $n$  lignes de maximum possible commun  $N-1$  et  $N$  colonnes de maximum  $n-1$ .

Il nous semble préférable d'utiliser la confiance  $C_G$  plutôt que  $E_1$  car celle-ci est beaucoup plus « parlante ». En effet, par exemple, pour  $n = 3$ , une confiance de 90 % est plus compréhensible que  $E_1 = 1.2$ , alors que ces deux nombres sont équivalents.

L'intérêt de ces confiances globales est qu'elles permettent de pondérer les confiances locales des paires déterminées. Plus précisément, la confiance locale est celle de la paire formée : elle indique uniquement la confiance que l'on peut avoir pour ce couple donné; alors que, dans le même temps, la confiance globale indique celle que l'on peut avoir en le test entier, en considérant  $G_{0,I}$  ou  $G_{0,I}^{\text{base}}$  dans son ensemble. Ainsi, les confiances globales

apparaissent comme une « pondération », une mise en relief complémentaire du résultat local : une paire locale possible possède une certaine confiance, *relativement* à la confiance globale du test !

Passons au second type d'erreur, l'erreur d'estimation de la matrice de mélange  $A$ .

#### 4.4.4.2 Erreur $E_A$

Cette erreur  $E_A$  est l'écart entre  $A$  et son estimée  $\tilde{A}$  déterminée par l'ACI. De par la nature de la matrice de mélange et selon (1.28), la  $i^{\text{ème}}$  colonne de  $A$  (respectivement  $\tilde{A}$ ) correspond aux proportions du  $i^{\text{ème}}$  signal (resp. composante) sur l'ensemble des mélanges  $X$ . Ainsi, du fait des permutations d'ordre et de signe inhérentes à l'ACI dont il faut tenir compte, comparer  $A$  et  $\tilde{A}$  ne revient pas simplement à les comparer terme à terme, mais colonnes à colonnes permutées et signées. Ainsi, les paires établies théoriquement donnent les associations et les colonnes à comparer. Par exemple, si nous sommes confrontés à la paire  $(-Y_2, S_1)$ , cela signifie que nous calculons l'erreur terme à terme entre la première colonne de  $A$  et « moins » la deuxième de  $\tilde{A}$ .

Matriciellement, ceci se traduit par l'intervention de la matrice « parfaite »  $G_P$  équivalente de  $G_{0,I}$ ; c'est-à-dire, selon  $G_P = PD$  (1.38), la matrice produit d'une matrice de permutation par une matrice diagonale (de termes  $\pm 1$ ). Typiquement, cette matrice  $G_P$  s'obtient directement à partir des paires théoriques  $(Y_i, S_j)$  (où  $j$  est l'indice de  $S$ , pas celui de  $S^{\text{base}}$  actualisé) trouvées selon :

$$\forall (i, k) \in [1, n]^2, G_P(i, k) = \text{signe}(Y_i) \delta_{kj} = \begin{cases} \text{signe}(Y_i) & , \text{ si } k = j \\ 0 & , \text{ sinon} \end{cases} \quad (4.20)$$

où  $\delta$  est le symbole de Kronecker.

L'erreur  $E_A$ , en tant qu'erreur relative, est alors définie par l'expression :

$$E_A = \left| A - \tilde{A}G_P \right| \div A \quad (4.21)$$

où  $\div$  désigne la division terme à terme. On calcule ici l'erreur relative; c'est-à-dire un pourcentage d'erreur par rapport aux valeurs de  $A$ . De fait,  $E_A$  est une matrice d'erreur.

Le défaut de cette définition est le problème d'échelle qu'elle engendre. En effet, selon notre protocole, la matrice de covariance de  $S$  vient jouer les trouble-fêtes car on a :

$$\tilde{A} Y_{0,I} = A S_{0,\Sigma} = A \Sigma_S^{1/2} S_{0,I} \quad (4.22)$$

Ainsi, on observe que  $\tilde{A}$  a « absorbé »  $\Sigma_S^{1/2}$ . D'où un facteur d'échelle perturbateur qui signifie que  $\tilde{A}$  n'aura probablement pas des pourcentages comme valeurs et que la somme des lignes ne fera surement pas 1, comme  $A$ . Pour résoudre ce problème et vu que nous n'avons en fait aucun contrôle sur  $\tilde{A}$ , au lieu de la comparer avec  $A$ , nous le faisons avec  $A_\Sigma = A \Sigma_S^{1/2}$  et l'expression de  $E_A$  actualisée est donc finalement :

$$E_A = \left| A_\Sigma - \tilde{A} G_P \right| \div A_\Sigma \quad (4.23)$$

Cette définition de  $E_A$  conclut la description du post-traitement.

## 4.5 Variables finales interprétables

Après avoir présenté l'ensemble de notre protocole expérimental, nous décrivons maintenant parmi toutes les variables précédemment citées celles que nous avons décidé de sélectionner en tant que données finales car elles synthétisent selon nous l'ensemble des notions abordées et permettent d'interpréter correctement et pertinemment les comportements local et global d'une combinaison donnée. Ce sont en d'autres termes les variables retenues pour l'analyse et la discussion des résultats dans le prochain chapitre.

### 4.5.1 Nature

Plus précisément, il s'agit par combinaison :

- a. du taux de convergence (mesure des effets négatifs de l'ACP et de la non-convergence de FastICA);

- b. des confiances globales théorique et pratique ( $C_G$  et  $C_G^{\text{base}}$ );
- c. des confiances locales théorique et pratique ( $C_L$  et  $C_L^{\text{base}}$ ) de chacun des  $n$  signaux;
- d. du pourcentage de succès (d'associations) de chaque signal, calculé à partir de la matrice *resultat*;
- e. de l'erreur des proportions de chaque signal dans la matrice de mélange  $A$  (matrice d'erreur  $E_A$ ).

Hormis évidemment le taux de convergence, il est important de noter que tous ces calculs sont effectués *uniquement* dans le cas de tests convergeants. Plus précisément, en dehors de la dernière variable, on transforme en premier lieu la matrice *resultat* en son équivalente « épurée » de ses doubles colonnes de zéros, symbolisant une non-convergence du (ou des) test, qui leur est relié. Ainsi, c'est à partir de cette matrice réduite que sont effectivement calculés les confiances et succès, tout en étant pondérés par le taux de convergence. Par exemple, un pourcentage de succès de 90 % perd toute sa force pour un taux de convergence de moins de 10 %.

Pour chacun de ces paramètres (hormis le taux de convergence), on calcule la moyenne  $\mu$  et l'erreur standard  $\sigma_n$  (voir définition ci-après en 4.5.2) sur tous les tests; c'est-à-dire que l'on se retrouve avec deux valeurs par paramètre, par matrice  $A$  et par combinaison  $S$ . À ce niveau, nous étudions si leur comportement est uniforme sur l'ensemble des matrices  $A$  (propriété de stabilité de l'ACI par rapport à la nature des mélanges !). Le cas échéant, on refait de même sur l'ensemble des matrices  $A$ , afin de récupérer deux valeurs (moyenne et erreur standard) par combinaison  $S$ .

#### 4.5.2 Erreur standard

Avant d'aller plus loin, définissons l'erreur standard. Pour des observations  $(x_i)_{i \in [1, n]}$ , où  $n$  est le nombre d'observations (typiquement, le nombre de tests quand on veut les valeurs pour une matrice  $A$  donnée), de moyenne  $\mu$ , l'erreur standard  $\sigma_n$  est définie à partir de

l'écart-type  $\sigma$  (non biaisé) selon :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (4.24)$$

$$\sigma_n = \frac{\sigma}{\sqrt{n}} \quad (4.25)$$

Dans ce genre d'exercice de considérations statistiques, on préfère généralement calculer l'erreur standard plutôt que l'écart-type car celle-ci représente les variations de la moyenne des données et non celles des données elles-mêmes symbolisées par l'écart-type. Vu que dans notre cas, nous nous intéressons aux valeurs pour l'ensemble des matrices  $A$  et non uniquement pour l'ensemble des tests (moyenne de moyennes), l'erreur standard a plus de sens que l'écart-type; c'est pourquoi dans la suite, on ne considèrera que ces deux paramètres statistiques : moyenne et erreur standard.

### 4.5.3 Explications de la sélection

Pour conclure cette partie, nous allons expliquer les raisons pour lesquelles les cinq groupes de variables susnommés ont été sélectionnés.

Tout d'abord, concernant le *taux de convergence*, la raison en est évidente : c'est le meilleur indicateur de stabilité de l'algorithme FastICA face à une combinaison donnée et, comme énoncé plus tôt, il permet de relativiser le pourcentage de succès.

Ensuite, pour l'*erreur d'estimation des proportions de mélange* de la matrice  $A$ , nous l'avons conservée uniquement à des fins de curiosité. En effet, nous montrerons dans le chapitre 5 quelques tracés de ces erreurs pour chaque signal de quelques combinaisons, afin de plutôt constater le comportement global de celles-ci en fonction du paramètre  $p$  (voir 4.2.2) que de faire une étude approfondie du phénomène. Intuitivement, nous nous attendons à une erreur décroissante pour le premier signal ( $p$  augmente de plus en plus) et inversement à une croissance pour les  $n - 1$  signaux restants. L'étude de ce paramètre n'ira pas plus loin; ce serait une recommandation possible de s'y intéresser plus en profondeur.



En outre, l'utilité du *pourcentage de succès* de chaque signal s'affirme d'elle-même, puisqu'il s'agit du paramètre auquel nous nous intéressons depuis le commencement. Il mesure la capacité de l'ACI à récupérer les signaux qu'on lui propose en entrée.

Cependant, ces succès sont à relativiser en fonction des confiances; ce qui constitue leur intérêt majeur. Sans trop anticiper sur les résultats où nous constaterons que les succès d'associations sont très bons dans la grande majorité des cas, il apparaît que les confiances peuvent être considérées comme des paramètres plus pertinents que les succès eux-mêmes. En effet, les *confiances globales* indiquent la confiance que l'on peut avoir en général, dans son ensemble, envers une combinaison donnée, à la fois en ne considérant que la base réduite (confiance théorique) ou la base entière (confiance pratique). Ce sont des indicateurs globaux, sachant que la théorique a l'avantage d'être indépendante de la base totale initiale donc en ce sens peut-être plus généralisable et interprétable que la pratique, qui a inversement néanmoins l'avantage de considérer la base entière. Ainsi, changer de base totale initiale modifie la confiance globale pratique, mais pas la théorique; d'où l'intérêt contradictoire mais réel (fonction du but et de l'utilisation recherchés) de chacune de ces deux confiances.

Quant aux *confiances locales*, elles indiquent à l'utilisateur les confiances propres à chacune des paires formées, à chacun des signaux de la combinaison considérés individuellement ! Elles permettent donc de juger de la capacité propre de l'ACI à correctement « récupérer » un signal particulier; ce sont donc d'excellents indicateurs locaux.

Plus précisément, les *théoriques* permettent de juger si les paires soi-disant correctes le sont effectivement. Pour ce faire, nous observons graphiquement si elles sont visuellement satisfaisantes et déterminons alors empiriquement un seuil de confiance théorique locale, commun à toutes les combinaisons, en-dessous duquel nous postulons ne pas pouvoir accepter cette paire et donc continuer le processus d'étude des associations : il y a *erreur*. Si la confiance théorique est bien supérieure à ce seuil, nous poursuivons en faisant de même

en considérant cette fois les confiances *pratiques* avec détermination toujours empirique du seuil (commun) de confiance pratique locale. Si la confiance lui est inférieure, il y a *rejet*; sinon, nous étudions le succès d'associations. Cette procédure est récapitulée à la figure 13.

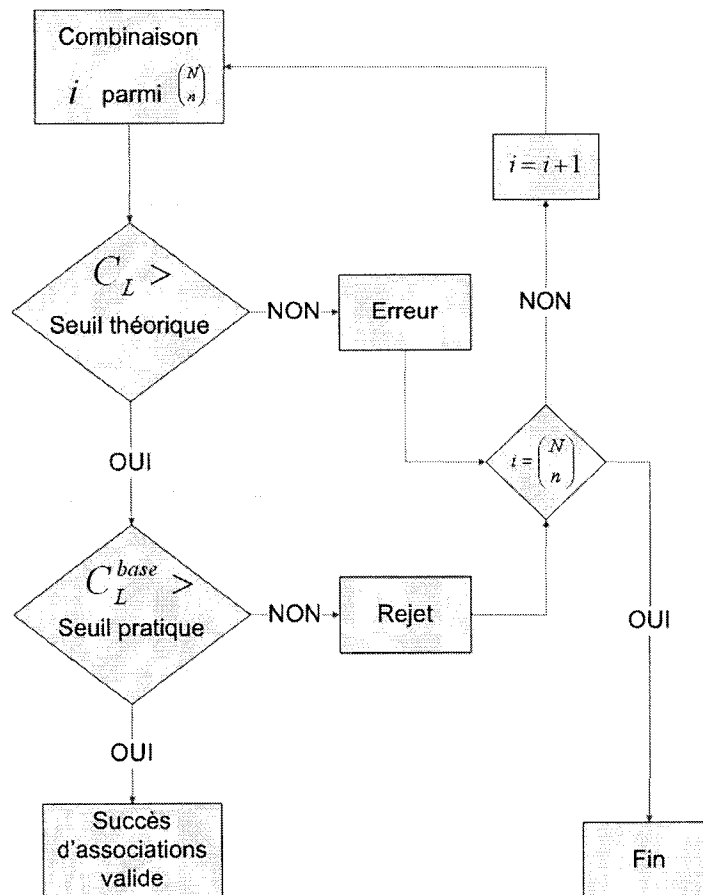


Figure 13 Processus de rejet des paires théoriques et pratiques en fonction des seuils de confiance locale

À noter que vu que les seuils sont déterminés de façon empirique, ceci signifie typiquement qu'ils vont varier d'un type de données à un autre; c'est ce que nous montrerons en 5.5.4; d'où une étude ciblée et au cas par cas en fonction de l'application considérée.

Avant d'étudier et analyser les résultats que toutes ces variables sont sujettes à apporter, nous terminons ce chapitre dédié au protocole en récapitulant, d'une part, notre algorithme; et, d'autre part, en l'appliquant sur un exemple particulier.

## 4.6 Algorithme

Notre méthode est résumée à l'algorithme 5 suivant, où la notation  $f(\cdot)$  désigne simplement l'expression « fonction de ».

## 4.7 Exemple d'application du protocole expérimental

### 4.7.1 Choix de $S_{\mu,\Sigma}$

On choisit les images  $\{1, 3, 4\}$  de la partie 5.2.1. Par conséquent,  $S_1 = S_1^{\text{base}}$ ,  $S_2 = S_3^{\text{base}}$  et  $S_3 = S_4^{\text{base}}$ .

### 4.7.2 Partie théorique

Sur un des 50 tests pour la matrice  $A$  avec  $p = 25\%$  (la première image représente 25% des mélanges), la matrice  $G_{0,I}$  avait la forme suivante :

$$G_{0,I} = \begin{pmatrix} -1.0000 & -0.0029 & 0.0103 \\ 0.0104 & -0.0188 & 0.9998 \\ -0.0027 & 0.9998 & 0.0189 \end{pmatrix}$$

ce qui signifie, d'une part, que ce test est bien convergent (il n'a subi ni la réduction par l'ACP, ni la non convergence par FastICA); et que d'autre part, par exemple pour la deuxième ligne :

$$Y_2 = 0.0104 S_1 - 0.0188 S_2 + 0.9998 S_3$$

soit intuitivement le couple associé  $(Y_2, S_3)$  et pour l'ensemble de  $G_{0,I}$ ,  $(Y_1, -S_1)$ ,  $(Y_2, S_3)$  et  $(Y_3, S_2)$ .

---

**Algorithme 5** : Algorithme de base du protocole expérimental
 

---

 Formation des  $N$  signaux de la base totale  $S_{\mu,\Sigma}^{\text{base}}$  et choix de  $n$  ( $n \leq N$ );

$$S_{0,\Sigma}^{\text{base}} \leftarrow S_{\mu,\Sigma}^{\text{base}} - \mu;$$

$$S_{0,I}^{\text{base}} \leftarrow \Sigma_{S^{\text{base}}}^{-1/2} S_{0,\Sigma}^{\text{base}};$$

**pour toutes les combinaisons de  $n$  parmi  $N$  faire**

 Formation de  $S_{\mu,\Sigma}$  pour la combinaison en question;

$$S_{0,\Sigma} \leftarrow S_{\mu,\Sigma} - \mu;$$

$$S_{0,I} \leftarrow \Sigma_S^{-1/2} S_{0,\Sigma};$$

**pour matriceA = 1 à nbTotalMatricesA faire**

$$A_\Sigma = A \Sigma_S^{1/2};$$

$$X_{\mu,\Sigma} = A S_{\mu,\Sigma};$$

$$X_{0,\Sigma} \leftarrow X_{\mu,\Sigma} - \mu;$$

**pour test = 1 à nbTotalTests faire**
**si réduction par ACP ou non convergence de FastICA alors**

⊥ return;

$$[Y_{0,I}, \tilde{A}, B] = \text{ACI}(X_{0,\Sigma});$$

$$G_{0,I} = B A_\Sigma;$$

$$G_{0,I}^{\text{base}} = Y_{0,I} (S_{0,I}^{\text{base}})^{-1};$$

% Confiances locales et globales, erreurs globales et sur A

$$[C_L, E_1] = f(G_{0,I}) \quad [C_L^{\text{base}}, E_1^{\text{base}}] = f(G_{0,I}^{\text{base}});$$

$$C_G = f(E_1) \quad C_G^{\text{base}} = f(E_1^{\text{base}});$$

$$E_A = f(A_\Sigma, \tilde{A}, G_P);$$

% Formation des paires théoriques et pratiques

$$\text{resultat}(1 : n, 2 \text{ test} - 1 : 2 \text{ test}, \text{matriceA}) = f(C_L, G_{0,I});$$

$$\text{resultat}(n+1 : 2n, 2 \text{ test} - 1 : 2 \text{ test}, \text{matriceA}) = f(C_L^{\text{base}}, G_{0,I}^{\text{base}});$$

% Pourcentage de succès d'associations

**pour signal = 1 à n faire**

⊥ succès(signal, matriceA) = f(resultat(:, :, matriceA));
 

---

S'en suivent alors les calculs d'erreurs  $E_{SY}(i, j)$  et  $E_{YS}(i, j)$ . Par exemple, pour  $g_{32} = 0.9998$ , on obtient :

$$E_{SY}(3, 2) = \sum_{c=1, c \neq 2}^3 \frac{|g_{3c}|}{|g_{32}|} = \frac{0.0027 + 0.0189}{0.9998} = 0.0216$$

$$E_{YS}(3, 2) = \sum_{l=1, l \neq 3}^3 \frac{|g_{l2}|}{|g_{32}|} = \frac{0.0029 + 0.0188}{0.9998} = 0.0217$$

Soit une erreur totale  $E(3, 2) = \frac{E_{SY}(3,2) + E_{YS}(3,2)}{2} = 0.02165$  et une confiance, pour  $n = 3$ ,  $C_L(3, 2) = 1 - \frac{E(3,2)}{2} = 0.9892 = 98.92\%$  envers le résultat.

On suit la même procédure pour toute la matrice  $G_{0,I}$  et on obtient la matrice de confiance théorique (locale)  $C_L$  associée suivante :

$$C_L = \begin{pmatrix} 0.9934 & -173.0823 & -47.9964 \\ -47.7401 & -25.7238 & 0.9854 \\ -185.5682 & 0.9892 & -25.6798 \end{pmatrix}$$

De par la forme proche du cas parfait de  $G_{0,I}$ , la matrice de confiance présente trois confiances très proches de 1 et toutes les autres « nettement » négatives, indiquant une confiance nulle, conséquence de leurs proportions négligeables dans la matrice globale. Conformément à 4.4.1.2, le coefficient de normalisation  $(n - 1)$  ayant été calculé dans le cas du maximum, il est normal de constater des confiances correctement comprises entre 0 et 1 pour les trois maximums trouvés et des valeurs négatives (non contrôlables) pour tous les autres éléments.

L'association des paires se fait alors récursivement :

- le maximum de  $C_L$  est  $C_L(1, 1) = 0.9934$  : on forme donc le couple  $(Y_1, S_1)$ . Mais vu que le signe de  $g_{11}$  (dans la matrice  $G_{0,I}$ ) est négatif, le véritable couple formé est  $(-Y_1, S_1)$ , auquel est associé une confiance théorique de 99.34%;
- on élimine les premières ligne et colonne;

- c. le nouveau maximum de  $C_L$  est alors  $C_L(3, 2) = 0.9892$  et comme  $g_{32}$  est positif, le deuxième couple théorique est donc  $(Y_3, S_2)$  avec une confiance de 98.92%;
- d. on élimine de même et reste alors le dernier couple  $(Y_2, S_3)$ , de confiance 98.54%.

On se retrouve finalement avec les couples théoriques suivants, en redonnant aux signaux  $S$  leurs indices initiaux de  $S^{\text{base}}$  :

- a.  $(-Y_1, S_1^{\text{base}})$ , de confiance 99.34%;
- b.  $(Y_3, S_3^{\text{base}})$ , de confiance 98.92%;
- c.  $(Y_2, S_4^{\text{base}})$ , de confiance 98.54%.

Les trois premières lignes (des deux colonnes relatives au test considéré) de la matrice *resultat* seront donc (par ordre croissant des indices de  $Y$ ) :

$$\begin{pmatrix} -1 & 0.9934 \\ 4 & 0.9854 \\ 3 & 0.9892 \end{pmatrix}$$

où le  $-1$  désigne le signe de  $Y_1$  transposée sur  $S_1$ , selon la notation  $(Y_1, -S_1)$  adoptée pour remplir *resultat*.

### 4.7.3 Partie pratique

On reprend le même principe dans la pratique, en considérant cependant  $S^{\text{base}}$  ( $N = 7$  images) dans sa totalité, plutôt que sa sélection  $S$  ( $n = 3$  images).

On obtient la matrice  $G_{0,I}^{\text{base}}$  suivante :

$$G_{0,I}^{\text{base}} = \begin{pmatrix} -0.9956 & -0.0207 & -0.0022 & 0.0066 & -0.0480 & 0.0780 & 0.0002 \\ 0.0135 & -0.0347 & -0.0194 & 0.9968 & -0.0231 & 0.0643 & 0.0014 \\ -0.0033 & -0.0504 & 0.9895 & 0.0174 & 0.1253 & 0.0480 & 0.0024 \end{pmatrix}$$

On remarque logiquement qu'en dehors des colonnes  $\{1, 3, 4\}$ , indices des images sélectionnées comme  $S$ , les valeurs sont faibles car les autres images n'ont pas participé à

l'ensemble théorique. Cependant, elles ne sont pas nulles; ce qui traduit tout de même un effet perturbateur logique.

Selon le même principe que précédemment, on obtient la matrice  $C_L^{\text{base}}$  de confiance pratique (locale) :

$$C_L^{\text{base}} = \begin{pmatrix} 0.9787 & -6.35 & -119.03 & -40.1 & -2.25 & -0.9 & -90.21 \\ -18.83 & -3.28 & -12.7 & 0.9774 & -6.05 & -1.38 & -105.08 \\ -82.78 & -2.08 & 0.9661 & -15.00 & -0.18 & -2.48 & -64.03 \end{pmatrix}$$

et les paires pratiques suivantes :

- a.  $(-Y_1, S_1^{\text{base}})$ , de confiance 97.87%;
- b.  $(Y_2, S_4^{\text{base}})$ , de confiance 97.74%;
- c.  $(Y_3, S_3^{\text{base}})$ , de confiance 96.61%.

On complète alors les trois dernières lignes (des deux colonnes relatives au test considéré) de la matrice *resultat* et on obtient pour ce test unique la forme suivante :

$$\begin{pmatrix} -1 & 0.9868 \\ 4 & 0.9708 \\ 3 & 0.9783 \\ -1 & 0.9787 \\ 4 & 0.9774 \\ 3 & 0.9661 \end{pmatrix}$$

où les lignes (1, 4), (2, 5) et (3, 6) sont à comparer.

#### 4.7.4 Succès d'associations

Selon 4.4.3, nous pouvons affirmer que le pourcentage de succès est ici de 100% pour chacune des trois images car les indices signés se correspondent bien pour chaque couple de lignes. Nous constaterons dans le chapitre 5 que c'est très souvent le cas; ce qui renforce

la pertinence de notre niveau de confiance associé. De plus, selon 4.5.3, il n'y a ni erreur ni rejet vu les valeurs très élevées de confiances.

#### 4.7.5 Erreurs $E_1$ et $E_A$

Suivant les principes de 4.4.4, pour cet exemple, ses erreurs globales sont respectivement  $E_1 = 0.1280$  et  $E_1^{\text{base}} = 4.3714$ ; ce qui donne des confiances associées théorique et pratique globales de :

$$C_G = 1 - \frac{E_1}{2n(n-1)} = 1 - \frac{0.1280}{12} = 0.9893 = 98.93\%$$

$$C_G^{\text{base}} = 1 - \frac{E_1^{\text{base}}}{n(N-1) + N(n-1)} = 1 - \frac{4.3714}{32} = 0.8634 = 86.34\%$$

Tel que mentionnée en 4.4.4.1, la confiance globale peut être vue comme une « pondération » du résultat local. Typiquement, en considérant  $G_{0,I}^{\text{base}}$  dans son ensemble, on peut affirmer que les résultats pratiques globaux auront une confiance de 86.34 %; alors que, par exemple, la paire  $(Y_2, S_4^{\text{base}})$  a une confiance de 97.74 % : la confiance locale est à relativiser par rapport à la confiance globale du test considéré.

Quant à l'erreur  $E_A$ , pour les matrices  $A$ ,  $\Sigma_S^{1/2}$ , donc  $A_\Sigma$  et  $\tilde{A}$  suivantes :

$$A = \begin{pmatrix} 0.249 & 0.376 & 0.375 \\ 0.25 & 0.375 & 0.375 \\ 0.251 & 0.375 & 0.374 \end{pmatrix} \quad \Sigma_S^{1/2} = \begin{pmatrix} 73.24 & 2.16 & -2.66 \\ 2.16 & 67.56 & -3.98 \\ -2.66 & -3.98 & 67.22 \end{pmatrix}$$

$$A_\Sigma = \begin{pmatrix} 18.0504 & 24.4431 & 23.0493 \\ 18.1214 & 24.3777 & 23.0507 \\ 18.1973 & 24.3838 & 22.9808 \end{pmatrix} \quad \tilde{A} = \begin{pmatrix} 17.7327 & 23.9151 & 22.8004 \\ 17.8043 & 23.8511 & 22.8017 \\ 17.8807 & 23.8596 & 22.7326 \end{pmatrix}$$

et la matrice  $G_P$ , déduite des paires théoriques, suivante :



$$G_P = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

on obtient la matrice d'erreur  $E_A$  :

$$E_A = \begin{pmatrix} 0.0176 & 0.0216 & 0.0108 \\ 0.0175 & 0.0216 & 0.0108 \\ 0.0174 & 0.0215 & 0.0108 \end{pmatrix}$$

Si l'on compare  $A$  et  $\tilde{A}$ , on remarque bien l'effet d'échelle provoqué par  $\Sigma_S$ . De plus,  $E_A$  indique des erreurs relatives moyennes d'estimation respectives de 1.75 %, 2.16 % et 1.08 % par colonne (c'est-à-dire par signal  $S_j$ ); ce qui est en soi un excellent résultat.

Ce calcul d'erreur conclut cet exemple et, dans le même temps, notre protocole expérimental. Passons à présent au dernier chapitre de ce mémoire consacré aux résultats.

## CHAPITRE 5

### DISCUSSION ET INTERPRÉTATION DES RÉSULTATS

Nous présentons dans ce chapitre les résultats obtenus à partir du protocole expérimental du chapitre précédent. Nous débutons en rappelant les valeurs des paramètres de notre méthode et les différents types de données sur lesquels nous avons testé notre protocole.

#### 5.1 Valeurs des paramètres

Suite aux conclusions des chapitres précédents, nous avons décidé de fixer les paramètres suivants :

- a. le nombre de signaux choisis variera entre 2 et 4, soit  $n = \{2, 3, 4\}$ ;
- b.  $S$  représente toutes les combinaisons non ordonnées de  $S^{\text{base}}$ , qui symbolise les données détaillées dans la partie suivante; leur nombre est  $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ ;
- c. le nombre de mélanges  $m$  est égal à  $n$ , soit  $m = n$ ;
- d. l'algorithme du point fixe FastICA choisi est celui basé sur la méthode symétrique, à non-linéarité en tangente hyperbolique;
- e. le nombre de tests par combinaison  $S$  et par matrice  $A$  est de 50;
- f. le nombre maximal d'itérations de FastICA pour converger est fixée à 1 000;
- g. les valeurs  $k$  et  $\epsilon$ , variant en fonction de  $n$  et des données, sont détaillées en 5.2.4.

#### 5.2 Données

Nous appliquons notre protocole sur trois types de données :

- a. 7 images de niveaux de gris;
- b. 5 signaux (unidimensionnels) temporels classiques;
- c. 7 signatures spectrales.

Nous les détaillons l'un après l'autre.

### 5.2.1 Images en niveaux de gris

La base de données des images en niveaux de gris utilisée comme  $S^{\text{base}}$  sont les 7 images ordonnées représentées à la figure 14 et décrivant dans l'ordre (entre parenthèses, leur indice respectif dans  $S^{\text{base}}$ ) un boxeur (1), une coureuse (2), une main pleine de craie (3), une gymnaste (4), un nageur en train de plonger (5), un alpiniste (6) et du bruit gaussien (7). Ce sont toutes des images de  $256 \times 256 = 65\,536$  pixels pour 256 niveaux de gris, dont les six premières sont issues du site Web « [www.corbisimages.com](http://www.corbisimages.com) ».



Figure 14 Base de données des sept images en niveaux de gris

Dans le protocole, les images sont considérées comme des signaux unidimensionnels. Pour ce faire, on concatène leurs vecteurs lignes, de façon à ce que chaque image apparaisse en fait comme un profil de 65 536 valeurs de niveaux de gris.

En outre, l'intérêt d'utiliser des images et de débiter avec celles-ci est le fait que, contrairement à l'imagerie hyperspectrale, beaucoup de travaux ont été réalisés sur l'application de l'ACI sur des images. Et la conclusion est toujours la même : l'outil ACI est particulièrement efficace et robuste (notamment par rapport à différentes matrices  $A$ ) dans cette situation; c'est ce que nous voulons entre autres démontrer dans ce chapitre.

De plus, on a voulu s'intéresser, selon 1.7.4.2, à la propriété indiquant qu'au plus une source gaussienne doit être présente pour la validité de l'ACI; c'est pourquoi nous avons introduit, parmi notre panel de sept images, l'image 7 de bruit gaussien et l'image 4 non gaussienne mais proche, comme l'indique le calcul ci-dessous des kurtosis (voir 1.8.2) de toutes les images. En fait, selon l'annexe 3, section 3, nous avons décidé de plutôt considérer le kurtosis normalisé car il présente l'intérêt d'être minoré par  $-2$  et d'être donc en ce sens « plus parlant » que le kurtosis classique, car accentuant (grâce à la division par le carré du moment d'ordre 2) les faibles écarts de valeurs possiblement initialement présents.

On rappelle que le kurtosis normalisé est défini pour une variable centrée  $s$  et un kurtosis  $\kappa_4(s)$  par  $\gamma_4(s) = \frac{\kappa_4(s)}{E^2\{s^2\}}$ . En appliquant cette formule pour chacune des sept images, nous obtenons les valeurs suivantes de kurtosis normalisés :

$$-1.2846 \quad -0.3857 \quad 3.0455 \quad -0.0929 \quad -0.2931 \quad -0.9625 \quad 0.0175 \quad (5.1)$$

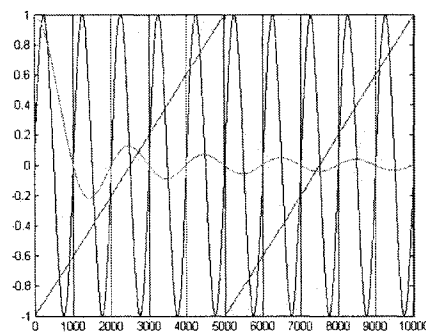
On remarque bien la valeur très proche de zéro pour l'image gaussienne (ce n'est pas exactement 0, car nous ne disposons que de 65 536 valeurs et non d'une infinité !) et la valeur proche pour la quatrième. Il sera donc intéressant de constater les différences de résultats, si elles existent, quand aucune, une ou ces deux images sont présentes dans  $S$ .

Pour les cinq images restantes, quatre sont subgaussiennes (kurtosis négatif) et une, la troisième, est supergaussienne (kurtosis positif).

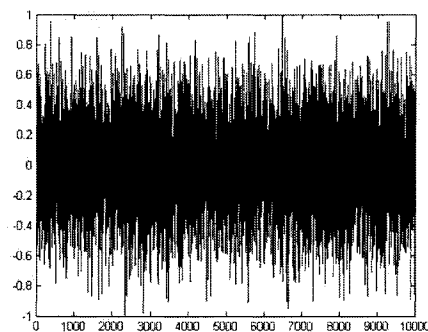
### 5.2.2 Signaux de base

Le deuxième ensemble de données utilisé est un ensemble de 5 signaux de base de 10 000 points chacun représentés à la figure 15, à savoir :

- $S_1(t) = \sin \frac{2\pi t}{1000}$ , une sinusoïde de période 1 000;
- $S_2(t) = \begin{cases} 1, & t \in [2000k, 1000 \cdot (2k + 1)], k \in [[0, 4]] \\ -1, & \text{sinon} \end{cases}$  ; une fonction créneau unitaire de période 2 000;
- $S_3(t) = \begin{cases} \frac{2t}{5000} - 1, & t \in [0, 5000] \\ S_3(t - 5000), & t \in ]5000, 10000] \end{cases}$  ; une fonction rampe de période 5 000;
- $S_4(t) = \text{sinc} \frac{2\pi t}{2000}$ ; une fonction « sinus cardinal » de pseudo-période 2 000;
- $S_5(t)$  un signal normal (gaussien) de 10 000 échantillons.



(a) Signaux hors gaussien



(b) Signal gaussien

Figure 15 Base de données des cinq signaux temporels classiques

De même que précédemment, nous distinguerons les cas gaussien / non gaussien en présentant deux séries de résultats : une avec ce signal normal  $S_5$  dans  $S^{\text{base}}$  ( $N = 5$ ), l'autre sans ( $N = 4$ ).

En outre, leurs kurtosis normalisés sont respectivement :

$$-1.5 \quad -2 \quad -1.2 \quad 9.0374 \quad 4.10^{-7} \quad (5.2)$$

où l'on peut constater que trois signaux sont (fortement) subgaussiens, contre un supergaussien et un (effectivement très !) gaussien.

### 5.2.3 Signatures spectrales

Le dernier ensemble considéré et évidemment le plus intéressant par rapport à notre champ d'étude est l'ensemble suivant de 7 signatures de la base USGS (entre parenthèses leur appellation officielle), représentées à la figure 16 :

- a.  $S_1$  la mauvaise herbe (« cheatgrass ANP92-11A »);
- b.  $S_2$  l'herbe sèche (« dry long grass AV87-2 (Brown) »);
- c.  $S_3$  la feuille d'érable (« maple leaves DW92-1 »);
- d.  $S_4$  l'amphibole (« amphibole NMNH78662 »);
- e.  $S_5$  l'hectorite (« hectorite SHCa-1 »);
- f.  $S_6$  l'halloysite (« halloysite NMNH106236 »);
- g.  $S_7$  la kaolinite (« kaolinite CM9 »).

En premier lieu, leurs kurtosis normalisés valent :

$$-1.3801 \quad -0.2933 \quad -1.2093 \quad -0.0233 \quad 8.1272 \quad 0.9515 \quad 2.0413 \quad (5.3)$$

soit trois supergaussiennes et quatre subgaussiennes, dont la quatrième ayant « malencontreusement » un kurtosis proche de zéro, pouvant sous-entendre qu'elle est gaussienne. Nous verrons plus tard en quoi cette propriété pourrait jouer un rôle dans les résultats.

En outre, en se basant sur l'annexe 6, qui décrit précisément l'arborescence des signatures spectrales dans la nomenclature USGS (classée par ordre croissant de précision en catégorie, type, groupe et famille), le tableau III donne l'arborescence des sept signatures que

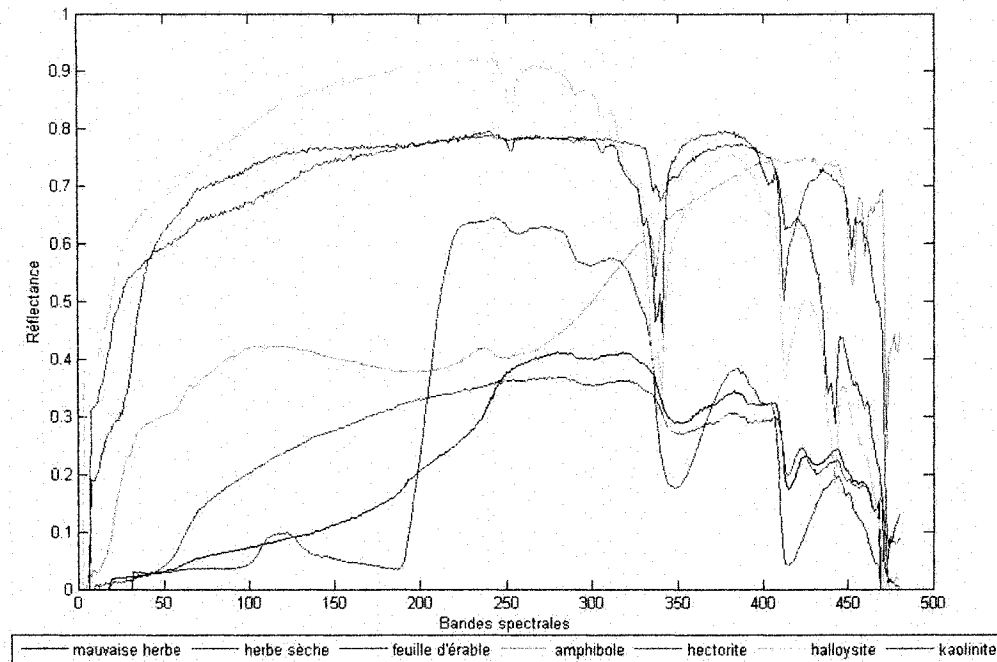


Figure 16 Base de données des sept signatures USGS choisies (réflectance de 0.2 à 3  $\mu\text{m}$ )

nous avons choisies (les tirets indiquent soit l'absence de caractéristiques plus lointaines dans l'arborescence ou seulement un inintérêt pour notre étude future).

Il met en lumière le fait qu'il existe des affinités entre celles-ci et que simultanément, certaines sont au contraire très disparates; ce qui se traduit graphiquement à la figure 16 par des signatures distinctes et d'autres proches. Il apparaît ainsi que l'ensemble choisi est suffisamment hétéroclite avec des catégories différentes (herbe versus minéral) et en même temps suffisamment homogène (même type voire même groupe), pour pouvoir fournir des résultats intéressants en plusieurs points sur un cas possiblement réel; d'où une certaine pertinence de notre base de choix.

Tableau III

Arborescence de la base de données des sept signatures choisies

Numéro	Signature	Catégorie	Type	Groupe	Famille
1	mauvaise herbe	plante	herbe	–	–
2	herbe sèche	plante	herbe	–	–
3	feuille d'érable	plante	arbre	–	–
4	amphibole	minéral	inosilicate	–	–
5	hectorite	minéral	phyllosilicate	montmorillonite	–
6	halloysite	minéral	phyllosilicate	kaolinite	halloysite
7	kaolinite	minéral	phyllosilicate	kaolinite	kaolinite

De plus, ce tableau III montre de façon claire, de par sa structure, les différentes comparaisons intéressantes que nous serons amenés à réaliser dans la partie 5.6.3; à savoir les comparaisons entre des signatures :

- a. de catégories différentes (plante / minéral);
- b. de même catégorie mais de types différents (herbe / arbre (plante); inosilicate / phyllosilicate (minéraux));
- c. de mêmes catégorie et type mais de groupes différents (les deux herbes (plantes); montmorillonite / kaolinite (phyllosilicate));
- d. de mêmes catégorie, type et groupe, mais de familles différentes (halloysite / kaolinite (kaolinite)).

#### 5.2.4 Valeurs de $(k, \epsilon)$ en fonction de la nature des données

Telles que précisées au dernier point de 5.1, les valeurs de couples  $(k, \epsilon)$ , déterminant la forme et le nombre de  $A$  (voir 4.2.2), pour lesquelles l'ACP ne pose pas de problème de simplifications intempestives non désirées, dépendent des données. De façon empirique,



en lançant les tests sous Matlab, nous avons fixé les valeurs suivantes résumées dans le tableau IV variant selon  $n$  et la nature des données.

Tableau IV

Valeurs de  $k$ ,  $\epsilon$  et du nombre de matrices  $A$  en fonction de  $n$  et des données

$n$	Images			Signaux						Signatures		
	$k\%$	$\epsilon\%$	$A$	$N=5$			$N=4$			$k\%$	$\epsilon\%$	$A$
				$k\%$	$\epsilon\%$	$A$	$k\%$	$\epsilon\%$	$A$			
<b>2</b>	1	0.1	99	1	0.2	99	1	0.5	99	2	1	49
<b>3</b>	1	0.1	99	1	0.5	99	1	0.5	99	2.5	1	39
<b>4</b>	1	0.1	99	2.5	1	39	2	0.5	49	5	1.5	19

Les commentaires que l'on peut faire à partir de ce tableau sont, tout d'abord, que les valeurs de  $k$  et  $\epsilon$  sont généralement faibles, d'où un nombre assez élevé de matrices  $A$ ; ce qui permet de faire une étude réaliste, complète et pertinente des effets de  $A$  sur l'ACI. C'est particulièrement vrai dans le cas des images, où, quelque soit le nombre  $n$  considéré, les paramètres restent constants (étude au pourcent près); ce qui est une preuve supplémentaire de la robustesse de l'ACI dans le cas préférentiel des images. En revanche, on remarque que les paramètres croissent pour les signaux et encore plus pour les signatures en fonction de  $n$  (étude de 2 % à 5 % seulement); signe au contraire d'une certaine instabilité vis-à-vis de l'ACP et probablement générale.

### 5.3 Commentaires de forme sur les tableaux de résultats

À présent que les trois bases de données sont connues et que tous les paramètres sont correctement fixés, avant de discuter concrètement des résultats obtenus, présentons la forme générale des tableaux sous laquelle les valeurs pertinentes décrites en 4.5 ont été compilées. Un exemple est illustré par le tableau V, qui indique les résultats (en pourcentages) des combinaisons d'images 1-2-5 et 1-2-6 pour  $n = 3$  images.

Tableau V

Exemple de tableau de résultats pour des combinaisons d'images

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-5	100	$83.10 \pm 0.02$	80.48	$89.25 \pm 0.04$	$92.14 \pm 0.02$	100
				$82.14 \pm 0.02$	$88.45 \pm 0.01$	100
				$77.90 \pm 0.01$	86.86	100
1-2-6	100	91.20	81.81	$97.83 \pm 0.01$	97.10	100
				$86.99 \pm 0.01$	91.67	100
				88.79	93.26	100

Concernant les colonnes « Convergence », «  $C_G$  » et «  $C_G^{\text{base}}$  », leurs natures sont claires et correspondent à la combinaison entière. Pour les confiances locales théoriques  $C_L$  et pratiques  $C_L^{\text{base}}$ , leur ordre (leur ligne) correspond à l'ordre des numéros des signaux utilisés (indifféremment images, signaux de base et signatures) dans la combinaison en question. Par exemple, pour la combinaison 1-2-5, les confiances locales du signal 1 sont inscrites à la première ligne et valent  $89.25 \pm 0.04\%$  en théorie et  $92.14 \pm 0.02\%$  en pratique (qui est supérieure mais incomparable du fait de la normalisation, tel qu'expliqué en 4.4.2). Le même principe s'applique au taux de bonnes associations. Plus généralement, chaque valeur est accompagnée de son erreur standard. Quand cette dernière n'est pas indiquée, comme pour 1-2-6, cela signifie qu'elle est au maximum de l'ordre de  $10^{-4}$ , donc négligeable et donc négligée par commodité. Enfin, les combinaisons pour lesquelles il y a des « problèmes » (tels non convergence ou associations différentes de 100 %) sont indiquées en gras (ce n'est pas le cas ici).

Soulignons en outre que vu le nombre de combinaisons assez conséquent dans certains cas (par exemple, pour  $n = 3$  et  $N = 7$ , il y en a  $\binom{7}{3} = \frac{7!}{3!4!} = 35$ ), les tableaux de résultats sont souvent denses et tiennent sur plusieurs pages; c'est pourquoi nous avons préféré tous les disposer en annexe 7. Ainsi, pour appuyer notre analyse, nous ferons référence à chaque fois à la partie pertinente en annexe.

## 5.4 Remarques générales

Intéressons-nous à présent aux résultats proprement dits en commençant par des constatations générales qui se sont vérifiées de façon commune (certes à un degré moindre pour les signatures, comme nous le montrerons notamment en 5.6.3) pour nos trois types de données. Elles concernent l'étude :

- a. du cas particulier  $n = 2$ ;
- b. de la convergence;
- c. de la matrice de mélange avec, d'une part, l'étude de l'uniformité en  $A$  et, d'autre part, à titre de curiosité, l'étude, à partir de  $E_A$ , des erreurs d'estimations des proportions de mélange par le tracé et l'analyse succincte de leurs comportements;
- d. de la dépendance à la nature de la combinaison considérée.

### 5.4.1 Cas particulier de $n = 2$

La première remarque générale que nous pouvons faire est que, pour  $n = 2$ , quelle que soit l'application, les confiances théoriques locales sont égales entre elles et à la confiance théorique globale. Ceci s'explique mathématiquement par les natures particulières ici de  $Y$  et  $S$  et les définitions de ces confiances et des erreurs qui permettent de les calculer. En reprenant ces définitions, on se rend compte que, pour  $n = 2$ , c'est toujours le cas, quelle que soit l'application. On le constate aisément aux tableaux XVI, XIX, XXII et XXV de l'annexe 7.

### 5.4.2 Convergence

En outre, concernant la convergence de l'ACI, on remarque que dans la grande majorité, quel que soit le cas, elle est de 100 %; ce qui prouve bien l'efficacité de l'algorithme FastICA choisi (méthode symétrique à non-linéarité en tanh). Les cas particuliers de

non convergence totale seront abordés séparément dans chacune des parties 5.6.1 à 5.6.3 concernées.

### 5.4.3 Résultats sur $A$

Les résultats sur  $A$  consistent en deux points : l'étude de l'uniformité du comportement des confiances en fonction des proportions des mélanges et l'erreur d'estimation (récapitulée par  $E_A$ ) de celles-ci.

#### 5.4.3.1 Uniformité en $A$

Selon le but énoncé en 4.2.2, nous jugeons ici du comportement de l'ACI en fonction des proportions de mélange, en se demandant principalement si elle est indépendante de celles-ci. Pour ce faire, on considère les variations des confiances globales en fonction du paramètre  $p$ . En se basant sur les faibles valeurs générales d'erreurs standards, nous pouvons présumer que l'ACI est uniforme, indépendante envers les mélanges.

Pour le prouver, nous présentons à la figure 17 les meilleur (colonne de gauche), moyen (centre) et pire (droite) comportements des confiances en fonction de  $p$  pour nos trois types de données : les images (première ligne), les signaux de base (milieu) et les signatures (dernière). Comme titres des sous-figures, sont inscrites les erreurs standards globales théorique puis pratique. La confiance théorique (respectivement pratique) correspond au trait le plus (resp. moins) gras. En abscisses, est indiqué le numéro de la matrice  $A$  dont le nombre est, on le rappelle, indiqué au tableau IV car dépendant des valeurs des couples  $(k, \epsilon)$ . Quant aux ordonnées, peu importe les valeurs des confiances associées, c'est véritablement la forme des courbes qui nous préoccupe ici.

Globalement (hormis le pire comportement des signatures), on remarque bien que les deux types de confiances sont quasiment constants en fonction de la proportion  $p$  du premier signal dans les mélanges; ce qui montre de façon indéniable à quel point l'ACI est stable

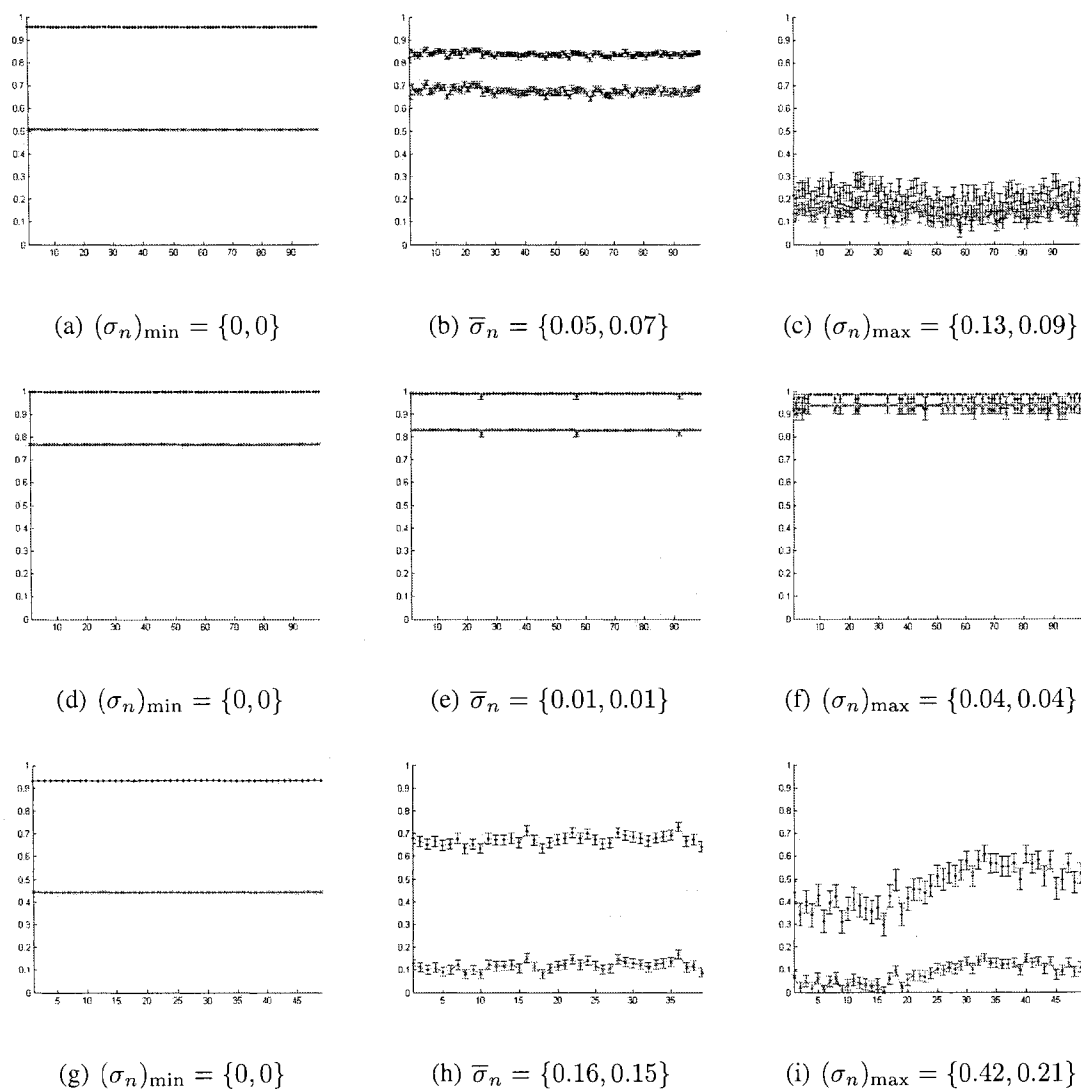


Figure 17 Meilleur, moyen et pire comportements en  $A$  pour les images, les signaux et les signatures (confiance théorique en gras, pratique en normal)

et robuste vis-à-vis de  $A$ . De fait, nous pouvons affirmer que *l'ACI est indépendante de la nature des mélanges*; ce qui constitue en soi un résultat essentiel de notre étude.

En outre, en comparant les lignes, nous constatons que ce sont les signaux de base qui présentent le comportement le plus uniforme, puis les images dont les résultats sont fort satisfaisants et, enfin, les signatures dont les variations sont plus chaotiques mais tout

de même acceptables, excepté évidemment celles du pire cas. Cet ordre dans les performances se retrouvera quand nous aborderons les résultats particuliers en 5.6. À noter que les meilleurs (que l'on peut qualifier de parfaits ici) et pires comportements n'ont pour seul but que de montrer les cas extrêmes; la deuxième colonne de la figure 17 est à ce titre plus révélatrice.

#### 5.4.3.2 Erreur $E_A$

Tel qu'indiqué en 4.5.3, nous terminons l'étude sur  $A$  en nous intéressant succinctement aux erreurs d'estimation des proportions, symbolisées par la matrice  $E_A$ . On rappelle avant de débiter que, d'après 4.4.4.2, le problème d'échelle, inhérent à la division terme à terme, entraîne ici l'absence d'interprétation des valeurs en ordonnées et que seul le comportement global de l'erreur nous préoccupe.

De par une visualisation empirique des courbes obtenues, nous avons observé :

- a. que le comportement moyen était « sensiblement » le même pour les trois types de données; c'est pourquoi nous ne différencions les résultats trouvés qu'en fonction de la valeur de  $n$  et non de la nature des données. Rigoureusement, de même que précédemment, le comportement des signatures est plus chaotique (erreurs standards plus élevées et contrairement au point suivant, certaines des courbes sont plus linéaires qu'exponentielles), mais peu importe ici;
- b. deux comportements généraux :
  - (a) un sans discontinuité marqué par des variations exponentielles d'erreurs : négative pour le premier signal (paramètre  $p$  variable (voir 4.2.2)) et positive(s) pour les  $n - 1$  restants;
  - (b) un avec discontinuités car pour certaines proportions, la moyenne et l'erreur standard de l'erreur divergent vertigineusement telle une discontinuité.

À la figure 18, sont représentées, en fonction de  $p$ , les erreurs d'estimation de proportions de chacun des signaux constitutifs de la base partielle. Nous avons distingué en colonnes

les variations de  $n$  (2 pour la colonne de gauche, 3 pour celle du milieu et 4 pour la dernière). Sur la première ligne, sont illustrés les exemples sans discontinuité, contrairement à la seconde ligne.

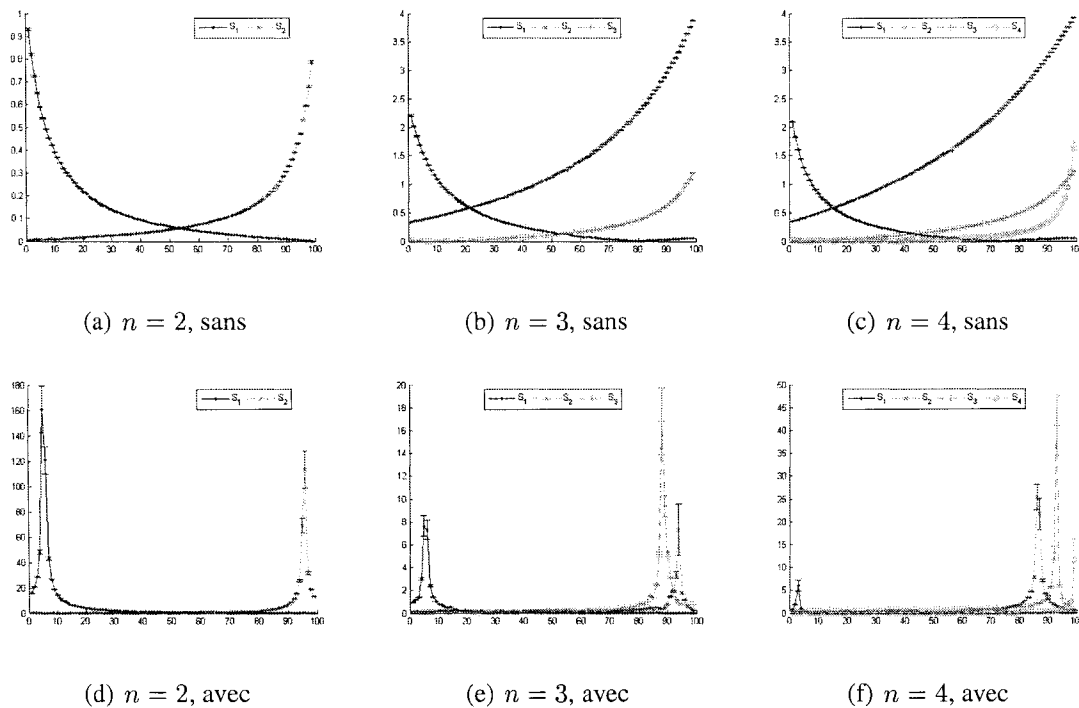


Figure 18 Erreurs d'estimation de la matrice de mélange pour  $n = \{2, 3, 4\}$ , sans et avec discontinuités, en fonction de la proportion  $p$  du premier des  $n - 1$  signaux

Nous remarquons, concernant la première ligne, un comportement somme toute assez logique et intuitif :

- pour le premier signal, l'erreur diminue de façon exponentielle au fur et à mesure que son poids dans le mélange augmente;
- pour les autres, c'est la réciproque. Le point intéressant, laissé pour les recommandations, serait de s'intéresser de plus près aux variations des  $n - 1$  autres signaux entre eux.

L'explication du premier point semble effectivement simple dans ce cas là. En effet, pour de faibles valeurs de  $p$  (faibles proportions), l'erreur du premier signal est plus élevée car dûe à sa faible connaissance et, a contrario, l'erreur est moindre pour les autres du fait d'une connaissance quasi-totale. Plus la proportion augmente, plus l'équilibre se rétablit pour le premier signal et donc plus son erreur diminue. À partir d'une certaine valeur, il se passe le phénomène inverse de celui décrit; à savoir que  $p$  devient trop grand donc les autres signaux ne sont pas assez représentatifs et leur erreur augmente.

Un résultat intéressant est de constater que, contrairement aux confiances indépendantes en  $A$ , l'erreur d'estimation des proportions varie logiquement en fonction de la représentativité présente de chacun des signaux. La recherche de l'explication serait une recommandation possible.

Pour les discontinuités, le principal intérêt, après la découverte de leur existence (a priori pas forcément soupçonnable), est d'en trouver la raison. Il semble empiriquement que les discontinuités apparaissent sensiblement aux mêmes proportions. Si ces discontinuités apparaissaient pour  $p \approx 0$  et/ou  $p \approx 100\%$ , l'explication évoquée plus haut aurait été fort valable, mais ce n'est pas le cas. En effet, comme nous pouvons le constater à la seconde ligne, celles-ci apparaissent plutôt pour des intervalles de proportions de l'ordre de 5 à 15 % et de 80 à 90 %. Au moins, une certaine symétrie peut être notée; ce qui semble logique. Sans trop s'avancer, une raison possible viendrait peut-être de l'équation du point fixe utilisée (voir 3.3.1.3). En effet, la présence dans sa définition de  $\tanh$  pourrait poser des problèmes de continuité car sa réciproque  $\tan$ , qui n'est pas définie en  $\frac{\pi}{2} [\pi]$ , est utilisée dans l'algorithme 4 de la méthode symétrique lors de l'orthogonalisation de Gram-Schmidt (calcul de la transposée de  $W$  et inversion du produit  $WW^T$ ). Mais ce n'est qu'une supputation; il faudrait s'y atteler de plus près : une nouvelle recommandation.



#### 5.4.4 Dépendance à la nature de la combinaison

La remarque suivante va consister à montrer que, dépendamment de la base initiale à laquelle appartient une signature particulière originale (donc non décorrélée), sa forme une fois décorrélée ainsi que de celle de la composante indépendante qui lui est accouplée peuvent fortement varier. À noter que le même phénomène se produit également pour les deux autres types de données mais pour l'illustrer, nous ne considérons que les signatures.

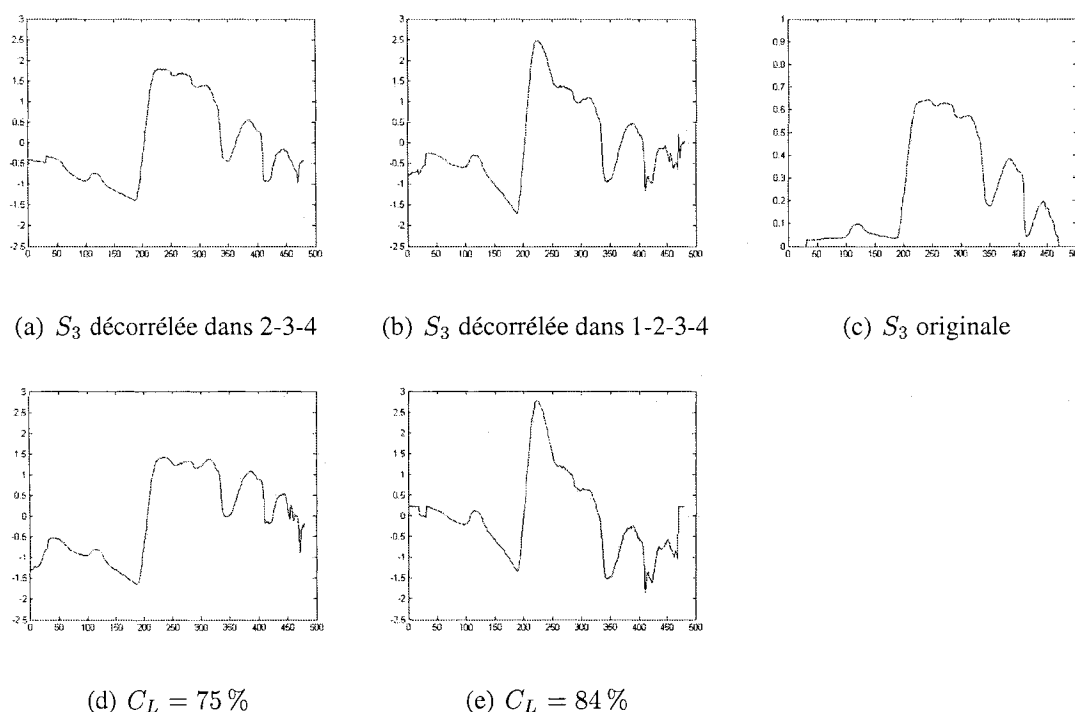


Figure 19 Effet de la nature de la base (décorrélacion)

À ce titre, est représentée, aux deux premières positions de la première ligne de la figure 19, la signature  $S_3$  respectivement décorrélée à partir des bases 2-3-4 et 1-2-3-4. En effet, selon l'équation (4.4) du prétraitement de notre protocole expérimental en 4.2, la décorrélation est réalisée sur l'ensemble des signatures formant la base partielle  $S$  (et non  $S^{\text{base}}$ ). Ceci signifie typiquement qu'une même signature n'aura jamais son équivalent décorrélé rigoureusement identique d'une combinaison à une autre dont elle fait partie; ce

qui explique les différences obtenues entre les figures 19(a) et 19(b). De fait, la seconde ligne montre les composantes indépendantes rendues dans les deux cas (correspondance par colonne) et a pour intérêt de montrer que même si, en comparant avec la forme de la signature originale  $S_3$  représentée en 19(c) (valeurs comprises entre 0 et 1), la composante 19(d) semble beaucoup plus acceptable que sa voisine de droite, il n'en est rien car la comparaison se fait uniquement par rapport aux signatures décorréelées. Ainsi, dans ce cas, en comparant les deux premières lignes, il apparaît que les deux estimées sont relativement correctes avec un avantage certain pour celle de droite; comme leurs valeurs de confiances respectives tendent à le souligner.

Cet exemple permet donc de se rendre compte de la dépendance des résultats en fonction de la nature de la combinaison initialement choisie. Constatant que la source du « problème » était la façon de décorréler, nous avons essayé de l'appliquer en premier lieu sur la base entière  $S^{\text{base}}$  et sélectionner ensuite chacune des signatures particulières d'une combinaison donnée à partir de cette base totale décorrélée. Outre le fait que cette méthode présente le désavantage certain de considérer l'ensemble des signatures et donc d'être moins théorique que prévu (si l'on souhaite par exemple estimer la capacité de démêlage de l'ACI pour la combinaison 1-2, il semble naturel de ne considérer que les signatures 1 et 2 dans la décorrélation initiale), après l'avoir testée empiriquement, nous n'avons noté aucune amélioration notable dans les résultats. Nous avons donc décidé de nous borner à notre méthode initiale.

Avant d'aborder le compte-rendu de chaque type de données séparément, nous détaillons et expliquons la difficulté générale empêchant une interprétation parfaitement correcte des résultats; à savoir l'interprétation des valeurs de confiances.

### **5.5 Interprétation des valeurs de confiances**

Celle-ci va consister à soulever et expliquer les deux points pertinents et fondamentaux suivants concernant les confiances :

- a. leur nécessité pour distinguer les (nombreux) cas d'associations parfaites (100 %);
- b. leur non suffisance par la présentation de quelques cas ambigus, qui participeront à l'élaboration d'une stratégie, d'une philosophie de « soumission » aux valeurs intrinsèques des confiances, pondérées par un intervalle de confiance.

### 5.5.1 Nécessité pour la distinction des cas d'associations parfaites

Pour aborder le premier point, la remarque importante à formuler est le fait qu'en se basant sur les tableaux de résultats de l'annexe 7, le taux d'associations est très souvent de 100 % (hormis pour  $n = \{3, 4\}$  chez les signatures; on y reviendra); ce qui est la preuve de l'efficacité de l'ACI et/ou de la pertinence de l'introduction par notre protocole de la notion de confiances. En effet, en se basant uniquement sur les associations, il apparaît impossible d'affirmer avec une totale certitude si les composantes indépendantes rendues en sortie de l'ACI correspondent effectivement bien aux signaux d'entrée. Pour s'en persuader (ou au contraire s'en dissuader), il faut analyser les valeurs de confiances. Pour ce faire, nous le montrons sur un exemple basé sur des images. Ce serait de même pour les deux autres applications, mais les images présentent l'avantage indéniable d'être visuelles.

De fait, comparons à la figure 20 les résultats obtenus pour l'image  $S_2$ , quand elle fait partie d'une part de la combinaison 2-4 (figure 20(a)) et d'autre part de 2-5 (figure 20(b)). Malgré le fait que celles-ci présentent toutes deux un pourcentage de succès d'associations parfait de 100 %, on note la nette différence de capacité à juger si la paire formée est effectivement la bonne entre les deux combinaisons : pour 2-5, il n'y a pas de problème (colonne du milieu); en revanche, pour 2-4 (première colonne), c'est beaucoup moins évident. Cette impression est validée numériquement par les valeurs de confiances (82 % pour la paire 2-5 et 25 % pour 2-4); ce qui exprime leur pertinence.

Plus précisément, pour 2-4, suivant les constatations amères que nous faisons en 4.1.2, on ne peut franchement pas conclure sur la véracité de l'association. L'utilisateur finalement est laissé seul juge de la situation (à lui de décider si, à 25 %, cette paire le satisfait). Dans

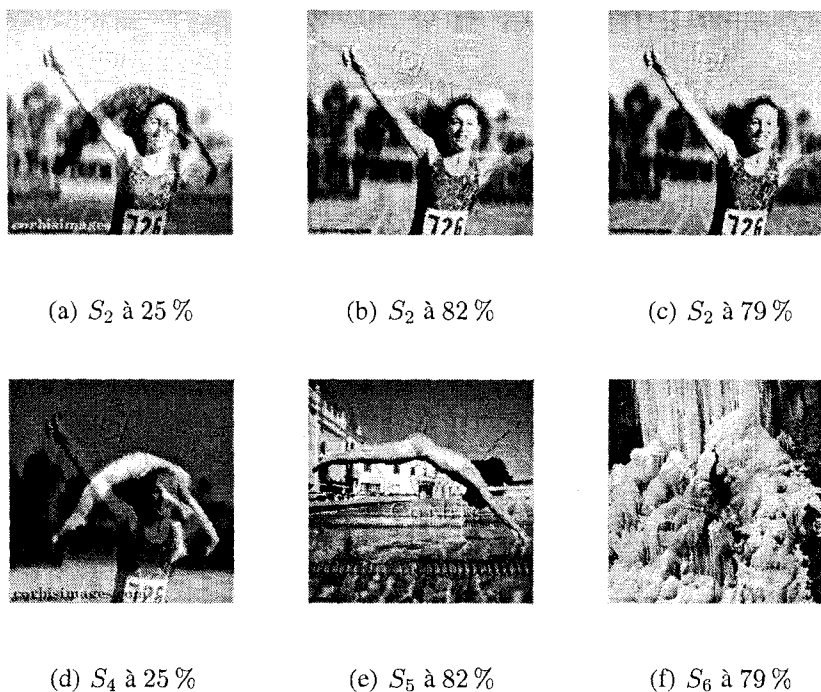


Figure 20 Illustration de la pertinence des confiances pour une même image présente dans trois combinaisons différentes ayant toutes un taux d'associations de 100 % (ici  $S_2$  dans 2-4, 2-5 et 2-6, par colonne de gauche à droite)

ce cas, on ne peut pas apporter une réponse formelle, mais au moins la confiance participe à ce questionnement et à la prise de décision finale, en donnant des valeurs indiquant la réalité de la situation.

Cet exemple particulier met donc indéniablement en évidence l'intérêt et la pertinence des confiances. Le taux d'associations seul n'a finalement aucun intérêt, si évidemment il n'est pas initialement mauvais; auquel cas il se suffit malheureusement à lui même. La confiance apparaît donc comme une *condition nécessaire* à la future prise de décision.

La question est alors de savoir si elle est également une condition suffisante. En d'autres termes, la confiance constitue-t-elle finalement une mesure linéaire croissante irréfutable de bon jugement (auquel cas elle est suffisante) ou présente-t-elle tout de même des failles ?

### 5.5.2 Condition non suffisante

Pour répondre à cette question, intéressons-nous aux images 20(b) et 20(c). Bien que celle de droite présente une confiance théorique locale (légèrement) inférieure à celle de gauche, il semble, si l'on se fie à notre jugement visuel (naturellement et malheureusement subjectif !), qu'elle « ressemble *légèrement* plus » à l'originale (voir la figure 14(b)) que son homologue de gauche. Cette impression vient de la nature même des deux images (5 et 6) qui complètent la base dans ces deux cas. En effet, la glace de l'alpiniste (6) étant plus claire que le corps du plongeur (5), ce dernier ressort plus au deuxième plan, contrairement à cette glace qui elle se fond mieux.

Cet exemple souligne ouvertement les différences naturelles existant entre l'homme et la machine. L'homme, de par son acuité, différencie mieux et dans ce cas-ci autrement que la machine, qui elle traite numériquement les données. Les valeurs de confiances discordantes (le terme étant peut-être excessif) traduisent exactement ce phénomène.

En tout cas, on peut tirer deux conclusions importantes de cette constatation; à savoir que :

- a. un même signal (indépendamment image, signal ou signature), dépendamment de ceux (ou celui) qui forment la base conjointement avec lui, n'obtiendra pas forcément la même confiance théorique locale (ce qui est tout à fait logique et justifie l'appellation « locale »);
- b. il est possible que, malgré une confiance (légèrement) supérieure dans une combinaison donnée, son association puisse sembler (c'est encore là évidemment subjectif !) moins correcte ou, du moins, moins convaincante, que celle formée dans une autre combinaison présentant une confiance inférieure.

En d'autres termes, cela signifie que la validité des résultats dépend certes des confiances mais pas seulement. Elle dépend également de la base considérée; c'est pourquoi les confiances sont une condition nécessaire mais malheureusement *non suffisante*, en cas de doute (confiances pas assez élevées, le « pas assez » dépendant de la base et de l'ap-

plication désirée; on y reviendra en 5.5.4). En effet, *seulement* en ce cas, les confiances apparaissent finalement comme des indicateurs, certes utiles, mais à vérifier à chaque cas. Cependant, relativisons en rappelant qu'en absence de doute (que ce soit une bonne ou une mauvaise association), d'après la partie précédente, leur nécessité n'est absolument pas à remettre en cause car elles sont alors évidemment indispensables pour distinguer des taux d'associations parfaits, qui seuls sont complètement inutiles.

### 5.5.3 Philosophie de la prise de décision

Pour appuyer cette idée récurrente de subjectivité en certaines occasions litigieuses, considérons un exemple basé sur des signatures spectrales avec les deux évolutions parallèles (de  $S_3$  dans différentes combinaisons) de la figure 21 où, pour la première ligne (respectivement la seconde), sont illustrés des exemples de confiances locales croissantes de 56 (resp. 40) à 100 %, dont la figure 21(d) (resp. 21(h)) est typiquement 19(a) (resp. 19(b)).

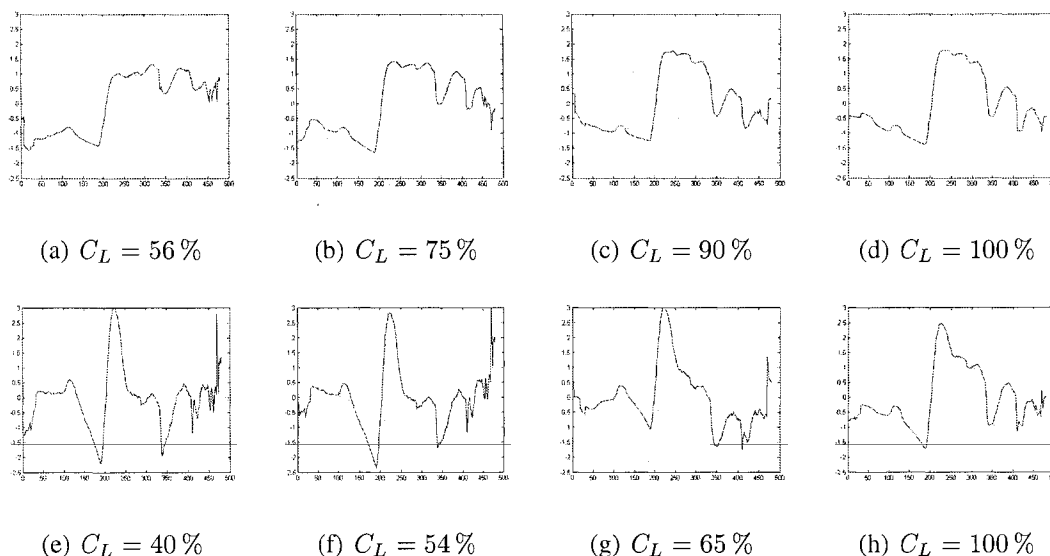


Figure 21 Exemple de subjectivité pour les signatures

Tout d'abord, on peut clairement affirmer que, par ligne, plus la confiance augmente, plus la forme de la signature tend vers sa forme finale et correcte : il y a donc bien une amélio-

ration notoire, qui apporte donc une preuve supplémentaire de la nécessité et l'efficacité de notre confiance.

En outre, en comparant les deux lignes l'une par rapport à l'autre et de façon plus précise, la figure 21(a) aux 21(f) et 21(g), il *semblerait* subjectivement que 21(a) est « meilleure » (une notion bien relative !) que 21(f) (c'est-à-dire qu'elle ressemble plus à 21(d) que 21(f) à 21(h)); ce que confirment les valeurs de confiances. Cependant, rien ne permet de l'affirmer avec certitude car contrairement au cas des images où notre cerveau est capable et habitué à discerner les différences visuelles (bien que souvent subjectives !), nous ne possédons ici aucun critère permettant de mesurer l'importance des différences entre les signatures originales et leurs estimées, issues de combinaisons différentes (conséquence de 5.4.4).

Par conséquent, devant ce petit aveu de faiblesse et pour y remédier, deux hypothèses ont été posées :

- a. les confiances sont jugées correspondre à un critère efficace permettant de mesurer ces différences et auquel on ne peut que « se soumettre » et se baser;
- b. tout en étant « polies » par une solution palliative consistant en l'introduction d'un écart-type autour des seuils locaux de confiance théorique et pratique (introduits en 4.5.3 et illustrés à la figure 13) dans le but d'englober les quelques ambiguïtés de subjectivité inhérentes à cette problématique d'associations de signaux. Il est donc davantage question d'un *intervalle de confiance* qu'un seuil de confiance lui-même.

Ainsi, appliquées à cet exemple de la figure 21, nos hypothèses nous permettent d'affirmer, d'une part, selon le premier point, que nous ne pouvons que considérer que 21(a) est plus ressemblant que 21(f), mais sans signifier pour autant que 21(f) est faux; car, en confirmation du second point, la présence d'un écart-type pour les seuils permettrait justement d'encadrer cette ambiguïté de distinction.

En y regardant de plus près, ces hypothèses s'appliquent parfaitement au cas litigieux des images 20(b) et 20(c), où rigoureusement c'est en la première que nous avons le plus confiance mais les deux appartiendraient à un intervalle de confiance centré en 80 % et d'écart-type  $\pm 2$  %.

#### 5.5.4 Valeurs des seuils empiriques locaux

En complément et terminaison au développement des deux parties précédentes, nous donnons dans le tableau VI les valeurs des seuils locaux théorique  $S_L$  et pratique  $S_L^{\text{base}}$  choisies pour chacun des trois types de données. Il est important de bien insister sur le fait que celles-ci ont été déterminées *empiriquement*; c'est-à-dire de visu, selon notre propre perception. Devant une telle subjectivité, il apparaît tout à fait prévisible voire normal qu'une autre personne puisse préférer des seuils différents : c'est effectivement à l'appréciation de chacun. Ceci étant dit, toute l'analyse qui s'en suit est fondée sur ces valeurs !

Tableau VI

Seuils empiriques de confiances locales

$n$	Images		Signaux		Signatures	
	$S_L$	$S_L^{\text{base}}$	$S_L$	$S_L^{\text{base}}$	$S_L$	$S_L^{\text{base}}$
2	$40.00 \pm 1.00$	$79.00 \pm 1.00$	$< 72.70$	$< 77.17$	$13.99 < . < 33.24$	$62.00 \pm 2.00$
3	$68.50 \pm 2.00$	$82.50 \pm 1.50$	$< 75.77$	$< 78.10$	$57.00 \pm 3.00$	$68.50 \pm 3.00$
4	$75.50 \pm 3.00$	$83.50 \pm 2.00$	$< 79.09$	$< 80.88$	$65.00 \pm 4.50$	$71.50 \pm 4.50$

Les signaux dont les valeurs de confiances locales sont comprises à l'intérieur de l'écart-type sont considérés comme litigieux. Plus précisément, si la confiance est :

- inférieure à la borne inférieure de l'intervalle, il y a erreur ou rejet, le cas échéant;
- dans l'intervalle, c'est à l'appréciation de l'utilisateur;
- supérieure à la borne supérieure de l'intervalle, il y a succès.



Concernant les remarques de forme, nous pouvons affirmer, d'une part, que, dans le cas des signaux de base, nous utilisons des signes « < » car tous les cas fonctionnaient parfaitement; ce qui signifie que le seuil est inférieur aux minimums constatés pour chacun des  $n$ . D'autre part, pour les signatures et  $n = 2$ , seule la combinaison 4-6 semble n'avoir pas fonctionné avec une confiance locale théorique de 13.99 %, contrairement à toutes les autres, dont le minimum est 33.24 %; d'où l'intervalle.

Sur le fond, nous pouvons tout d'abord constater que les seuils de confiances augmentent en fonction de  $n$ ; ce qui est logique : plus il y a de signaux et plus il devient difficile de correctement les séparer et discerner. De même, suivant le même raisonnement, les écart-types des seuils ont aussi tendance à augmenter car il faut avouer qu'il devient vraiment de moins en moins évident (voire impossible) de réussir à vraiment séparer les cas de confiances proches quand le nombre de signaux considérés augmente; c'est pourquoi des écarts relativement élevés ont été soumis; et encore, honnêtement, rien ne nous garantit que tous les cas soient correctement catégorisés; la correctitude étant de plus ici sujette à forte subjectivité !

À titre d'illustration, les seuils  $S_L$  de signatures de 57 % pour  $n = 3$  et de 65 % pour  $n = 4$  traduisent numériquement les conclusions visuelles énoncées en 5.5.3 concernant les figures 21(a) et 21(g) : elles sont toutes deux dans la zone d'incertitude et représentent un exemple de signal litigieux.

En outre, les confiances des signatures sont globalement plus basses que celles des images. Cela vient probablement du fait, comme nous l'affirmions dans la partie précédente, que nous sommes naturellement plus aptes à distinguer des images que des signaux et en ce sens, nous avons peut-être été inconsciemment plus tolérants avec les signatures, du fait de l'absence de critère naturel physiologique de comparaison. C'est une nouvelle fois la preuve malheureuse de la trop grande importance de la subjectivité à ce niveau.

Le but est alors de comptabiliser, pour chacun des signaux de chacune des trois applications, ses pourcentages moyens d'erreur, de rejet et de taux d'associations valides (c'est-à-dire sans erreur ni rejet). Pour ce faire, il suffit de comparer les valeurs des confiances locales récapitulées dans tous les tableaux de l'annexe 7 avec les seuils proposés ci-haut. Nous présentons dans la suite les résultats obtenus pour chaque signal séparément en distinguant deux approches :

- a. le « bénéfice du doute », où les cas litigieux sont considérés comme corrects; cela signifie que nous prenons la borne inférieure des seuils comme référence;
- b. l'absence de doute, où, au contraire, ils sont traités comme faux et c'est la borne supérieure qui fait office de juge.

Plus précisément, le pourcentage d'erreur du signal  $i$  est calculé comme le rapport du nombre de fois où  $C_L(i)$  est inférieure à  $S_L$  sur le nombre total de fois où il intervient; soit,  $\binom{N-1}{n-1} = \frac{(N-1)!}{(N-n)!(n-1)!}$  fois (c'est-à-dire pour  $n = \{2, 3, 4\}$ ,  $\{3, 3, 1\}$ ,  $\{4, 6, 4\}$  et  $\{6, 15, 20\}$  fois pour respectivement  $N = \{3, 4, 7\}$ ). À partir des combinaisons valides restantes (celles sans erreur) auxquelles  $i$  appartient, on applique la même méthode en comparant cette fois  $C_L^{\text{base}}(i)$  à  $S_L^{\text{base}}$ . Finalement, sur les combinaisons finales correctes (supérieures au seuil local pratique), on calcule la moyenne du taux d'associations de ce signal  $i$ .

En outre, comme nous l'avons montré en 5.4.4, il apparaît qu'en fonction de la nature de la base à laquelle appartient un signal, son comportement de succès peut changer. En effet, il est tout à fait possible que, pour un signal donné, faire partie d'une certaine base rende plus difficile son extraction correcte en tant que composante indépendante (et fasse donc diminuer son succès) que dans une autre où cela pourrait se faire sans aucune difficulté. En ayant ce point important à l'esprit, il devient alors plus pertinent de regarder plus précisément les tableaux de l'annexe 7, qui ont l'avantage de distinguer toutes les combinaisons possibles. En effet, ces tableaux de succès (en fonction du seuil) ont le défaut de leur avantage : ils compilent toutes les données propres à un seul signal (d'où leur

aspect simple et agréable); c'est pourquoi ils ont cependant tendance à moyenniser les effets et donc à effacer de possibles flagrantes disparités en fonction de la base. Ainsi, ce type de tableau peut entraîner de fausses interprétations : il vaut donc mieux le considérer comme un appui supplémentaire à l'analyse, complémentaire des tableaux de résultat brut de l'annexe 7. En fait, il permet une première analyse du comportement global d'un signal donné, à pondérer avec ses comportements locaux propres à chacune des combinaisons dont il fait partie.

Ce point conclut les remarques générales. Nous abordons à présent l'analyse des résultats particuliers à chaque valeur de  $n$  pour chaque type de données séparément. Typiquement, outre les commentaires particuliers issus du tableau de l'annexe 7 qui lui est relié, dans chaque sous-cas, nous donnerons son tableau récapitulatif des seuils, tel que nous venons de le présenter. À ce titre, il est important de noter que seules les images et les signatures sont concernées car, pour les signaux de base, tout est parfait. Cette remarque montre, comme intuité en 5.4.3.1 par l'étude de l'uniformité des confiances en  $A$ , que l'ACI est la plus efficace avec les signaux de base : nouvelle illustration de la pertinence de l'utilisation de l'ACI en traitement de signal.

## **5.6 Résultats particuliers**

Nous débutons par les images en niveaux de gris, dont les résultats sont récapitulés aux tableaux XVI à XVIII inclus (annexe 7).

### **5.6.1 Images en niveaux de gris**

#### **5.6.1.1 Cas $n = 2$**

Pour  $n = 2$ , selon le tableau XVI de l'annexe 7, une première constatation que nous pouvons faire est le problème de la non convergence. Dans ce cas, seules trois combinaisons (4-6, 4-7 et 5-7) n'ont pas convergé à 100 % et principalement 4-6, avec seulement 7.92 %

de tests convergents. La raison est à première vue peu évidente; peut-être une opposition forte du nombre élevé de pixels noirs pour la quatrième image et blancs pour la sixième. Globalement, ces trois cas de non convergence font intervenir séparément ou ensemble les deux images gaussiennes 4 et 7; ce qui est intéressant et montre bien l'effet perturbateur d'une nature gaussienne des signaux.

Tableau VII

Erreur et rejet pour  $n = 2$  images

	Erreur		Rejet		Succès		Associations	
	inf	sup	inf	sup	inf	sup	inf	sup
<b>1</b>	0	0	0	0	100	100	99.99	99.99
<b>2</b>	16.67	16.67	0	0	83.33	83.33	99.99	99.99
<b>3</b>	0	0	0	0	100	100	99.99	99.99
<b>4</b>	16.67	16.67	0	0	83.33	83.33	100	100
<b>5</b>	16.67	16.67	0	0	100	100	100	100
<b>6</b>	0	0	0	0	83.33	83.33	99.99	99.99
<b>7</b>	16.67	16.67	0	0	83.33	83.33	99.98	99.98

En outre, concernant le comportement des confiances envers les seuils, hormis pour les combinaisons 2-4 et 5-7, on remarque, d'une part, que les confiances théoriques sont très élevées et suffisent à se persuader de la correctitude des associations par l'ACI; ce qui est clairement démontré dans le tableau VII par un succès parfait, un rejet et une erreur communs nuls, hormis pour les quatre signaux précédemment cités présentant une erreur non nulle de  $1/6$  car sur les six combinaisons dont ils font chacun partie, une ne fonctionne pas; d'où cette erreur de 16.67%. En d'autres termes, pour  $n = 2$ , seules les paires 2-4 et 5-7, plutôt gaussiennes, ne sont pas correctes; pour les autres, c'est parfait !

### 5.6.1.2 Cas $n = 3$

Pour  $n = 3$ , les résultats, disponibles au tableau XVII, sont logiquement moins bons que pour  $n = 2$ , en ce qui concerne notamment les confiances (donc le succès), les erreurs standards, mais pas les associations toutes à (quasiment) 100 %.

Un premier point intéressant est le fait que les problèmes de convergence apparaissent uniquement pour des combinaisons faisant intervenir les images 4, 6 et 7 (à savoir 1-4-6/7 (raccourci de 1-4-6 et 1-4-7), 3-4-6/7 et surtout 4-6-7 avec seulement 5 % de convergence); ce qui confirme la conclusion de la partie précédente sur l'effet des images gaussiennes.

Tableau VIII

Erreur et rejet pour  $n = 3$  images

	Erreur		Rejet		Succès		Associations	
	inf	sup	inf	sup	inf	sup	inf	sup
<b>1</b>	0	0	0	0	100	100	99.99	99.99
<b>2</b>	33.33	33.33	0	0	66.67	66.67	99.99	99.99
<b>3</b>	13.33	26.67	0	18.18	86.67	60	99.99	99.99
<b>4</b>	46.67	53.33	0	28.57	53.33	33.33	99.99	99.99
<b>5</b>	06.67	53.33	14.25	28.57	80	33.33	100	100
<b>6</b>	13.33	20	0	0	86.67	80	99.99	99.99
<b>7</b>	0	6.67	0	0	100	93.33	99.99	99.99

Concernant cette quatrième image, elle participe à toutes les combinaisons que l'on peut assurément considérer comme fausses; à savoir 1-2-4, 2-3-4, 2-4-5/6/7, 3-4-5 et 4-6-7. La conséquence immédiate est donc, à partir du tableau VIII, la dégradation normale des résultats avec principalement l'impact négatif de la quatrième image, qui présente le succès le plus faible. Cette constatation renforce une nouvelle fois l'effet de cette image pseudo-gaussienne sur les autres.

À noter au contraire que la septième image seule (la véritablement gaussienne) ne procure pas un tel effet. Au contraire, elle affiche des résultats quasi-parfaits. Deux explications sont possibles dans ce cas :

- a. ou bien l'image 7 est justement « trop » parfaitement gaussienne que l'ACI la différencie facilement des autres;
- b. ou bien tel que l'indique le théorème de Linnik énoncé en 1.7.4.2, au maximum une source gaussienne est acceptable et les mauvais résultats de la quatrième image rencontrés ici en seraient une excellente et concrète mise en relief.

En outre, pour accentuer la thèse de la nécessité de nos confiances énoncée en 5.5.1, la colonne des « Associations » se suffit à elle-même.

Enfin, il est intéressant de souligner l'importance de l'intervalle de confiance pour les seuils. Typiquement, si l'on considère la cinquième image, on observe que son erreur passe de 7 à 53 % selon la borne, pour un succès final oscillant entre 1/3 et 80 %; ce qui illustre la subjectivité et la délicatesse à faire preuve dans la détermination des bornes.

### 5.6.1.3 Cas $n = 4$

Pour  $n = 4$ , les résultats, disponibles au tableau XVIII, sont principalement le reflet en pire des résultats précédents :

- a. les non-convergences des combinaisons présentant les mêmes paires d'images défavorables que les deux parties précédentes; à savoir 4-6 et 4-7 pour 1-4-6-7 et 3-4-6-7;
- b. des erreurs standards plus élevées;
- c. des confiances généralement plus basses, une nouvelle fois marquées par l'effet néfaste de la quatrième image, illustré au tableau IX.

En effet, on remarque, à partir de celui-ci, les mêmes points que dans la partie précédente mais accentués; à savoir un succès parfait pour la première et dernière (gaussienne) images mais faible pour la quatrième de l'ordre de 15 à 30 %; et surtout une assez grande variation

Tableau IX

Erreur et rejet pour  $n = 4$  images

	Erreur		Rejet		Succès		Associations	
	inf	sup	inf	sup	inf	sup	inf	sup
1	0	0	0	0	100	100	99.99	99.99
2	50	55	0	0	50	45	99.99	99.99
3	35	40	0	41.67	65	35	100	100
4	70	80	0	25.00	30	15	100	100
5	25	70	0	33.33	75	20	100	100
6	0	10	0	5.56	100	85	100	100
7	0	0	0	0	100	100	100	100

générale en fonction de la nature des bornes. Cela ne signifie pas nécessairement des mauvais résultats, mais peut-être plutôt une sélection trop critique dans les seuils.

En outre, comme expliqué en 5.5.4, la nature de la base à laquelle appartient une image peut faire varier son comportement de succès et c'est donc pour cela qu'il faut être vigilant avec l'interprétation des tableaux de succès. Typiquement, en ce cas, le fait pour une image (en dehors de 1 et 7) d'être en compagnie de la quatrième image a tendance à diminuer son succès. Pour s'en convaincre, il suffit, par exemple, de regarder au tableau XVIII que le comportement de la deuxième image n'est mauvais qu'en présence de la quatrième (ou plus généralement les combinaisons 1-2-3-4, 1-2-4-5/6, 1-3-4-5, 2-3-4-5/6/7, 2-4-5-6/7, 2-4-6-7 et 3-4-5-6/7 pour s'en persuader). Ceci montre qu'elle est véritablement l'élément perturbateur car deuxième image gaussienne.

En résumé, les résultats sont logiquement nettement moins bons que les cas précédents mais, fait intéressant, suivent la même logique (« emplacement » des mauvais résultats) que  $n = 2$ . De plus, dans l'ensemble, pour tout  $n$ , les résultats sont satisfaisants, hormis pour certaines combinaisons faisant intervenir la quatrième image : la dépendance au choix

de la base et le fait, selon 1.7.4.2, qu'au plus un signal gaussien doit être présent pour rendre valide l'ACI sont une nouvelle fois mis en lumière ici.

## 5.6.2 Signaux de base

Intéressons-nous à présent aux résultats concernant les signaux temporels de base. Nous débutons avec la base de 5 signaux, comprenant le signal gaussien.

### 5.6.2.1 5 signaux

Contrairement au cas précédent des images, les résultats sont si uniformes, en fonction de  $A$  (erreurs standards faibles) et de  $n$ , que nous avons décidé de ne pas distinguer les cas mais faire en contrepartie des remarques d'ordre général pour tout  $n$ . Les résultats sont compilés dans les tableaux XIX à XXI (annexe 7).

Tout d'abord, on peut affirmer que globalement, à la fois de visu des paires et de par les confiances élevées et nettement suffisantes, les résultats sont excellents; c'est pourquoi comme nous l'avons énoncé précédemment en 5.5.4, il n'y a pas de tableau de seuils pour les signaux car tout est parfait ! Ceci permet d'aisément conclure que l'ACI est efficace dans le cas des signaux de base temporels. Ce fait s'explique notamment par la nature même des signaux, très caractéristique (dents de scie, sinusoïdes, créneaux...) et qui constitue donc à ce titre un cadre théorique idéal de test de l'ACI.

Le seul problème à relever est celui de la présence de « paliers de convergence » uniquement pour les combinaisons faisant intervenir le quatrième signal (le sinus cardinal), pour  $n = 2$ ; ce qui pourrait signifier un manque de stabilité de la part de l'ACI, mais vu les pourcentages en question (de l'ordre de 90 %), ce serait un peu excessif et naïf de penser cela. Peut-être est-ce plutôt dû à un problème de discontinuité, auquel nous nous intéressons dès à présent.



En effet, un autre point intéressant à soulever est, de par 5.4.4, le problème de discontinuités apparaissant du fait de la décorrélation entre ces signaux temporels aux natures très caractéristiques et discontinues pour les fonctions dents de scie et créneaux. La figure 22 illustre nettement ce point, où est successivement représenté le sinus cardinal (signal 4) sous sa forme d'origine, décorrélé avec la fonction créneaux ( $S_{0,I}$ ), au sein de la base entière ( $S_{0,I}^{\text{base}}$ ) et la composante rendue ( $Y_{0,I}$ ). On y observe l'effet discontinu dont la fonction créneaux est responsable.

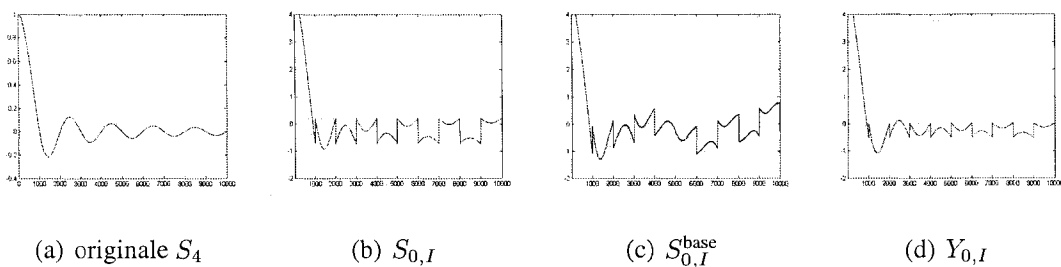


Figure 22 Effet de discontinuités à cause de la décorrélation pour les signaux

Cet exemple illustre clairement les variations dans la nature des signaux qu'entraîne la décorrélation. Dans ce cas, malgré le problème récurrent usuel d'échelles différentes (entre  $Y_{0,I}$  et  $S_{\mu,\Sigma}$  original), il est évident que, de par leurs formes très caractéristiques et différentes, les résultats auraient été meilleurs sous leurs formes premières (non décorrélés). Cependant, vu la perfection des résultats dans le cas décorrélé, on peut aisément s'en satisfaire.

Finalement, ces excellents résultats viennent donc corroborer ce que nous expliquions en 1.3 où les différentes applications efficaces de l'ACI étaient explicitées. Parmi elles, celles du domaine du traitement de signal (notamment détection d'artefacts dans les électroencéphalogrammes, étude des séries économiques temporelles, ...) tenaient une place de choix car l'ACI y était particulièrement efficace. Ces tests avec des signaux classiques de traitement du signal confirment cette affirmation.

Intéressons-nous à présent à l'effet du signal gaussien sur la base des signaux, en considérant celle de cardinalité  $N = 4$ .

### 5.6.2.2 4 signaux

En fait, concernant ce signal gaussien, on peut affirmer qu'il n'a aucun effet. Il suffit pour s'en convaincre d'observer les résultats, récapitulés aux tableaux XXII à XXIV, où aucune modification franche ni aucun effet notable ne sont à signaler.

En fait, de par la nature des tableaux de résultats, les seules modifications qui seraient possiblement à noter, si tel était le cas, concerneraient les confiances pratiques et le taux d'associations car pour les trois variables restantes (convergence et confiances théoriques globale et locale), elles ne dépendent que de la base réduite des signaux de la combinaison : il est donc tout à fait normal de retrouver les mêmes résultats (aux variations des erreurs standards près) pour ces variables; le contraire eût été surprenant !

Ce résultat de non-effet confirme, d'une part, la conclusion de la partie 5.6.1.2 où nous constatons pas de perturbation à la seule présence de l'image 7 « parfaitement » gaussienne, comme le signal en question ici; et atteste, une nouvelle fois, au risque de se répéter, qu'au plus une source gaussienne est acceptable pour s'assurer de la validité d'action de l'ACI. L'introduction de ce signal gaussien dans notre protocole pour ces données avait justement pour but d'illustrer ce phénomène; ce qui est chose bien faite !

Passons à présent à la dernière partie de cet ultime chapitre en nous intéressant à la capacité de l'ACI à correctement séparer les signatures spectrales.

### 5.6.3 Signatures spectrales

Nous nous intéressons aux résultats particuliers pour chacune des trois valeurs de  $n$  en nous focalisant principalement sur les comparaisons hiérarchiques décrites en 5.2.3.

### 5.6.3.1 Cas $n = 2$

Pour  $n = 2$  (tableau XXV), on constate, tout d'abord, qu'à l'exception de 2-6, il y a convergence dans tous les cas. En outre, ici (et pour la première fois), certains des taux d'associations sont loin d'être parfaits et présentent des valeurs plus conformes à la réalité. Cependant, après avoir retiré les combinaisons rejetées théoriquement puis pratiquement selon les seuils du tableau VI, on constate, selon le tableau X, que ces taux de bonnes associations ont augmenté; ce qui semble logique : le risque d'erreur de l'association est inversement proportionnel à sa confiance. Mais cela se produit au détriment du nombre de combinaisons valides après seuillage, qui diminue.

Par exemple, on vérifie que des problèmes d'associations, qui apparaissaient pour la quatrième signature gaussienne (voir 5.2.3) dans les couples 2-4 et 4-6 (8.90 % pour 2 dans 2-4 et 0 pour 6 dans 4-6), ont disparu du fait de l'élimination de ces deux couples lors du seuillage (4-6 en théorie et 2-4 en pratique). Leur succès a toutefois diminué (de 100 à respectivement 66.67 % et 83.33 %).

En d'autres termes (et on le vérifie pour tout  $n$ ), le fait d'avoir un seuil plus élevé permet d'éliminer certaines combinaisons où le taux d'associations est mauvais. Cela permet globalement d'augmenter la moyenne du taux de bonnes associations; en contrepartie, il fait diminuer le pourcentage de combinaisons valides. Il y a donc un compromis à faire ou un optimum à trouver.

Pour ce faire, il serait intéressant et à propos d'utiliser la technique des *courbes ROC* (pour « Receiver Operating Characteristic » en anglais), qui permettent justement de trouver ce point optimum. Nous ne l'évoquons que brièvement ici; nous la laissons comme future recommandation. Sans rentrer dans les détails disponibles dans (Fawcett, 2003), le principe est le suivant : une association peut être formée (acceptée) ou non (refusée), en fonction du seuil de confiance (indifféremment ici seuil théorique ou pratique). Si elle est acceptée (confiance supérieure au seuil) et effectivement bonne, on parle d'un *vrai positif*; si non

(acceptée mais mauvaise), on parle de *fausse alarme* ! L'idée est de tracer les variations des vrais positifs en fonction des fausses alarmes, en faisant varier la troisième variable qu'est le seuil de confiance. Cette courbe est par convention comprise dans un carré de coordonnées extrêmes  $([0, 0], [1, 1])$  et le but est de se rapprocher au maximum de  $[0, 1]$  qui traduit un maximum de vrais positifs pour un minimum de fausses alarmes : un maximum de succès pour un minimum d'incertitude !

Tableau X

Erreur et rejet pour  $n = 2$  signatures

	Erreur		Rejet		Succès		Associations		Confusions	
	inf	sup	inf	sup	inf	sup	inf	sup	inf	sup
<b>1</b>	0	0	16.67	33.33	83.33	66.67	80.19 ± 0.51	83.98 ± 0.28	3 (1-2)	3 (1-2)
<b>2</b>	0	0	33.33	33.33	66.67	66.67	93.86 ± 0.24	93.86 ± 0.24	3 (2-3)	3 (2-3)
<b>3</b>	0	0	0	16.67	100	83.33	95.91 ± 0.16	95.09 ± 0.19	2 (2-3)	2 (2-3)
<b>4</b>	16.67	16.67	20	20	66.67	66.67	100	100		
<b>5</b>	0	0	0	50	100	50	100	100		
<b>6</b>	16.67	16.67	0	0	83.33	83.33	93.00 ± 0.28	93.00 ± 0.28	3 (1-6)	3 (1-6)
<b>7</b>	0	0	16.67	16.67	83.33	83.33	100	100		

De retour au tableau X, on peut poursuivre en constatant tout d'abord que nous avons rajouté une colonne en fin de tableau, intitulée « Confusions », qui indique, uniquement pour les combinaisons valides (confiances supérieures aux deux seuils), la nature de la signature et de la combinaison pour laquelle, s'il y a lieu, une mauvaise association a été repérée. Par exemple, pour la sixième, la confusion « 3 (1-6) » signifie que pour la combinaison 1-6, dans 7 % des cas, 3 a été malencontreusement choisie à la place de 6. Notons que nous n'avons pas ajouté cette colonne supplémentaire pour les images car vu leurs taux d'associations de l'ordre de 99.99 %, ceci était sans grand intérêt.

On remarque de plus que les erreurs d'associations produites entre les trois premières signatures les font justement et seulement intervenir (3(1-2) pour 1, 3(2-3) pour 2 et 2(2-

3) pour 3)); ce qui signifie une (légère, de l'ordre de 8 %) difficulté ici à les séparer en certaines occasions. Ceci peut s'expliquer par le fait que ces trois signatures appartiennent à la catégorie des plantes; ce qui pourrait être une mise en évidence des problèmes liés à une hiérarchie trop proche.

Finalement, en regardant chaque combinaison séparément dans le tableau XXV, celles dont on est persuadés, selon nos critères de seuillage, que les signatures qui les constituent ont été correctement (au moins 95 %) associées, sont, pour les bornes supérieures, 1-3/5/7, 2-5/7, 3-5/6/7 et 6-7, auxquelles on ajoute, pour les bornes inférieures, 3/5-4 et 5-6/7; soit 9 et 13 combinaisons sur 21 : 42.86 % et 61.90 %, pourcentages bien moyens. À noter que toutes celles-ci présentent la caractéristique intéressante d'avoir une confiance globale théorique bien (de l'ordre de 10 %) supérieure à leur confiance globale pratique; ce qui est tout à fait logique, vu, d'après 4.5.3, les définitions et les rôles que nous voulions faire jouer à ces variables.

Si on analyse ces « bonnes » combinaisons en fonction de la hiérarchie énoncée au tableau III de 5.2.3, ceci signifie concrètement que :

- a. les signatures de plantes (1 à 3) sont correctement dissociables de celles des minéraux (1/2/3-5/7), hormis la sixième;
- b. la troisième (plante de type arbre) est la plus « conciliante » (ce que confirme le tableau X) : séparabilité parfaite avec les minéraux;
- c. avec la borne inférieure, on observe également d'excellents résultats pour les trois signatures de phyllosilicate (5 à 7);
- d. la quatrième a le comportement le plus mauvais : aucune combinaison pleinement correcte avec la borne supérieure.

Concernant ce dernier point, en retirant les six combinaisons comprenant cette quatrième signature, le nombre de correctes passe à 9 (borne supérieure) et 11 (inférieure) sur 15 possibles, pour des pourcentages de succès atteignant 60 % et 73.33 %.

### 5.6.3.2 Cas $n = 3$

Pour  $n = 3$ , dont le tableau XXVI compile tous les résultats, selon le même schéma d'analyse que la partie précédente, nous pouvons nous intéresser en premier lieu à la convergence. Fait intéressant ici : on observe des combinaisons qui n'ont jamais convergé à cause de l'ACP qui a annulé tous ces tests. Ceci signifie typiquement que les combinaisons en question, à savoir 1-2-5/6/7 et 4-6-5/7, sont plus « sensibles » que les autres, signe évident d'un comportement mauvais. Il s'agit, de même qu'en 5.6.3.1, des deux signatures d'herbe ensemble (1 et 2) face aux minéraux et de l'inosilicate (4) face aux trois minéraux restants. En outre, sans être nulle, on note des problèmes de convergence pour 1-2-3 (plantes entre elles), 1-3-6 et 2-5-6; signe possible d'une certaine opposition entre les plantes et le phyllosilicate.

Ensuite, en observant les taux d'associations, on remarque que, dans de nombreux cas, ils se détériorent sérieusement; ce qui est logique, vu qu'une signature a été ajoutée. Cependant, tel que mentionné dans la partie précédente, l'effet du seuillage annule la plupart de ces associations faibles.

Tableau XI

Erreur et rejet pour  $n = 3$  signatures

	Erreur		Rejet		Succès		Associations		Confusions	
	inf	sup	inf	sup	inf	sup	inf	sup	inf	sup
<b>1</b>	40	53.33	22.22	42.86	46.67	26.67	99.97 ± 0.01	100	3 (1-2-4)	
<b>2</b>	40	46.67	55.56	62.50	26.67	20	99.95 ± 0.03	99.93 ± 0.03	7 (2-3-7)	7 (2-3-7)
<b>3</b>	26.67	40	0	0	73.33	60	100	100		
<b>4</b>	80	93.33	0	0	20	6.67	100	100		
<b>5</b>	13.33	13.33	7.69	30.77	80	60	100	100		
<b>6</b>	66.67	80	40	100	20	0	74.46 ± 0.34	<b>X</b>	3 (1-6-7)	<b>X</b>
<b>7</b>	13.33	20	15.38	50	73.33	40	82.55 ± 0.38	87.18 ± 0.15	4 (2-4-7)	4 (2-4-7)
									4 (3-4-7)	
									4 (5-6-7)	4 (5-6-7)

Le tableau XI en apporte la preuve évidente : le taux minimal qui y est relevé est 74.46 % pour la sixième signature, alors que dans le même temps, 3-4-6 présente des taux de l'ordre de 24 %, mais éliminés par une confiance locale théorique insuffisante. En fait, il s'avère que les quinze combinaisons auxquelles 6 appartient ont été considérées comme non valides en pratique; d'où un succès pratique nul et une croix d'invalidité pour le taux d'associations de cette signature. Un autre exemple significatif est le taux de correctes associations de 99.93 % pour 2, pondéré par un maigre succès de 20 %.

En outre, en observant ce tableau XI plus attentivement, on note, tout d'abord, de gros écarts de résultats selon la borne comme par exemple l'erreur de 3 ou le rejet de 6. Ceci est une conséquence directe de la subjectivité longuement évoquée en 5.5. En effet, plus la zone d'incertitude est large, plus l'écart-type est important et donc plus la différence de résultats entre les bornes inférieure et supérieure est grande !

Concernant les confusions, on se rend compte, d'une part, que celles de 1 et 2 ne sont pas significatives, du fait de leurs hauts taux d'associations de l'ordre de 99.95 %; et, d'autre part, que la septième signature a été assimilée à la quatrième pour les combinaisons 2/3-4-7 et 5-6-7; ce qui indique un litige entre ces deux minéraux.

Enfin, en détaillant chaque combinaison, on relève que celles qui sont totalement correctes (aucune erreur ni rejet pour chacune des trois signatures la constituant) sont, pour la borne supérieure, 1/2-3-5 et 1/2-5-7, auxquelles on ajoute pour la borne inférieure 1-2-3, 1-3-6/7, 3/4-5-7 et 3-6-7; soit 4 et 10 combinaisons sur 35 : 11.43 % et 28.57 %, pourcentages faibles. Globalement, ceci confirme les conclusions de  $n = 2$ , avec comme différence d'ajouter la sixième signature à la quatrième comme plus mauvais éléments. Typiquement, en retirant ces deux signatures, il reste 10 combinaisons dont respectivement 4 et 7 « victorieuses », soit des pourcentages de 40 % et 70 %; ce qui constitue une amélioration notoire. On note de plus le même écart de 10 %, caractéristique du succès entre confiances globales théorique et pratique.

### 5.6.3.3 Cas $n = 4$

Toujours suivant la même stratégie, à partir du tableau XXVII, débutons l'analyse du cas  $n = 4$  par l'étude de la non-convergence. De même que précédemment, deux combinaisons, 1-2-5-6/7, présentent une convergence nulle, effet de l'ACP. Ce résultat confirme et est la conséquence de la même propriété relevée chez 1-2-5/6/7 pour  $n = 3$ . En outre, comme en 5.6.2.1, on note la présence de paliers de convergence qui, contrairement à ceux des signaux de base qui étaient de 99 %, sont ici de l'ordre de 48 % et concernent 1-2-3-5, 1-2-4-6 et 4-5-6-7, signe de même des difficultés dues à la proximité dans la hiérarchie. Finalement, nous relevons des taux de convergence non parfaits pour les combinaisons 1-2-3-6/7 et 2-3-6-5/7, pour la même conclusion.

De même, on remarque de façon encore plus évidente la dégradation des taux d'associations (de moyenne  $86.63 \pm 7.51$  %), cependant, clairement annulée par seuillage, tel que l'indique le tableau XII, où le taux minimal relevé est 91.38 % pour la septième signature. À ce titre, les confusions n'ont donc qu'un rôle mineur. On peut néanmoins noter que persistent toujours celles entre, d'une part, les premières signatures de plante, avec notamment la troisième qui offre le plus large panel de diversité; et, d'autre part, entre les quatre minérales avec surtout les erreurs entre 4 et 7. Ces constatations confirment celles des deux sections précédentes.

Concernant les taux d'erreur, de rejet et de succès, on constate que :

- a. les résultats sont mauvais : seule la cinquième signature offre des valeurs plus ou moins acceptables;
- b. trois des sept signatures ont un rejet total et donc un succès nul dans le cas de la borne supérieure; signe de leur nature litigieuse;
- c. une nouvelle fois (mais de façon encore plus convaincante), il existe un écart réel entre bornes inférieure et supérieure pour ce qui est du rejet; ce qui traduit vrai-



Tableau XII

Erreur et rejet pour  $n = 4$  signatures

	Erreur		Rejet		Succès		Associations		Confusions	
	inf	sup	inf	sup	inf	sup	inf	sup	inf	sup
<b>1</b>	30	55	35.71	100	45	<b>0</b>	$97.58 \pm 0.44$	<b>X</b>	4 (1-2-3-6) 3 (1-2-4-5/7) 3 (1-2-6-7)	<b>X</b>
<b>2</b>	40	65	25	42.86	45	20	$98.90 \pm 0.20$	$97.71 \pm 0.37$	3 (2-3-5-6/7) 3 (2-3-6-7)	3 (2-3-5-6/7)
<b>3</b>	35	45	7.69	27.27	60	40	$93.00 \pm 0.54$	$94.48 \pm 0.51$	1 (2-3-5-6) 1 (3-4-6-7) 2 (2-3-5-7) 4 (2-3-4-5) 6 (2-3-6-7) 1/4 (3-5-6-7)	1 (2-3-5-6) 2 (2-3-5-7) 6 (2-3-6-7) 1/4 (3-5-6-7)
<b>4</b>	75	95	40	100	15	<b>0</b>	$99.97 \pm 0.04$	<b>X</b>	1 (1-4-5-6)	<b>X</b>
<b>5</b>	10	15	11.11	29.41	80	60	$99.84 \pm 0.04$	$99.80 \pm 0.05$	6 (1/3-5-6-7)	6 (1/3-5-6-7)
<b>6</b>	60	70	25	100	30	<b>0</b>	$99.83 \pm 0.11$	<b>X</b>	7 (1/2/3-4-5-6) 2 (2-3-6-7)	<b>X</b>
<b>7</b>	40	55	25	77.78	45	10	$91.38 \pm 0.17$	$99.95 \pm 0.06$	4 (1-4-6-7) 4 (2-3-6-7)	4 (1-4-6-7)

ment la difficulté de détermination des seuils (notre seuil est peut-être trop critique (subjectif)) et la faible marge de manoeuvre dont nous disposons.

Enfin, si l'on comptabilise à partir du tableau XXVII les combinaisons parfaitement correctes (aucune erreur, ni rejet pour aucune des quatre signatures), le calcul est vite fait. En effet, on en note une seule pour la borne supérieure 2-3-5-7 (regroupant logiquement les quatre signatures présentant un succès supérieur non nul), accompagnée pour la borne inférieure par six autres, 1-2-3-6/7, 1-3-5/6-7 et 2/3-4-5-6; soit sur les 35 combinaisons totales, des pourcentages de 2.86 % et 20 %. De même que pour  $n = 2$ , si nous éliminons la quatrième signature, une nouvelle fois la plus mauvaise, nous considérons 15 combinaisons pour des pourcentages améliorés respectifs de 6.67 % et 46.67 %, signe évident de l'effet néfaste de cette signature. Si l'on retire de plus la sixième, on passe à 20 et

60 %; ce qui vient appuyer la remarque analogue pour  $n = 3$ . Globalement, ce n'est pas extraordinaire.

Avant de conclure hâtivement sur ces mauvais résultats, il est important de bien reconsidérer les trois points suivants permettant de relativement les adoucir :

- a. le critère très strict qu'est celui de considérer uniquement les combinaisons dont toutes les signatures sont parfaitement correctes;
- b. la subjectivité dans la détermination des seuils;
- c. la dépendance de la base, avec la présence ou non de signatures perturbatrices.

Ceux-ci expliquent en partie ces mauvais résultats; mais pas seulement, car on ne note pas de si mauvais résultats pour les images ou les signaux de base, alors que nous avons considéré ces mêmes trois points précédents, même si possiblement la dépendance est plus marquée ici.

#### **5.6.4 Bilan**

Ainsi, pour conclure partiellement sur les résultats particuliers obtenus, nous pouvons affirmer qu'ils sont excellents pour les signaux de base, bons pour les images et (même) meilleurs si l'on ne considère qu'une seule image gaussienne, mais mauvais pour les signatures.

Les deux principales raisons que nous voyons pour expliquer ce dernier point sont :

- a. l'effet négatif de la décorrélation de la base, qui crée, d'une part, une trop grande dépendance de la signature à la combinaison à laquelle elle appartient; et, d'autre part, inévitablement des déformations de formes entre la base partielle et totale; ce qui fausse les résultats. Cependant, cette décorrélation est rendue nécessaire comme solution au problème inhérent de l'ACI du changement d'échelle (matrice de covariance égale à la matrice identité pour les composantes indépendantes rendues);

- b. la subjectivité présente pour les trois types de données mais véritablement moins mesurable et analysable chez les signatures.

À noter que ces deux effets, décorrélation et subjectivité, sont présents pour chacun des trois types de données. Cependant, de par leur forme caractéristique, les signaux de base ne sont pas affectés; alors que pour les images, notre oeil est (relativement) habitué à noter des différences et donc à correctement distinguer les bonnes des mauvaises associations, pour une détermination plus précise et adéquate des seuils de confiance. Malheureusement, ces deux constatations ne sont plus valables dans le cas de signaux si particuliers que sont les signatures spectrales. Il s'agit donc aussi d'un problème de nature des données; c'est exactement ce qui nous intéressait dans ce travail.

### **5.7 Note positive finale**

Pour tout de même terminer ce travail sur une note positive, il est intéressant de rappeler les points suivants, issus de l'analyse, sous un autre angle, des résultats précédents. En effet, selon la mise en garde de 5.5.4 qui stipulait de bien jauger l'importance et la pertinence toutes *relatives* des tableaux de seuils face aux tableaux complets finaux, du fait de l'effet néfaste de certaines signatures perturbatrices, si l'on décide de considérer uniquement les tableaux de l'annexe 7, alors ce sont les conclusions beaucoup plus optimistes suivantes, mais tout aussi vraies, qui en découlent.

Tout d'abord, le démêlage de signatures est un problème beaucoup plus complexe (car beaucoup moins bien maîtrisé) que pour les images ou les signaux de base.

En outre, l'ACI a montré d'excellents résultats pour différencier, d'une part, les herbes fraîches (signature 1) des arbres (3) ou des minéraux (4 à 7); et, d'autre part, les arbres des minéraux. Elle s'avère donc pertinente pour distinguer les cultures, les arbres et les minéraux; ce qui apparaît très intéressant pour des applications telles que l'agriculture, la foresterie et l'exploration minière.

Il est néanmoins vrai que plusieurs indices, tels que l'indice de végétation par différence normalisée (NDVI), permettent déjà de faire de telles distinctions. Cependant, contrairement à ces indices, l'ACI ne nécessite aucun prétraitement de sélection des bandes pertinentes car elle les considère simplement toutes directement; ce qui est un avantage indéniable. Il serait donc intéressant de comparer ACI et NDVI; idée soumise en recommandations.

Enfin, il est important de souligner, en conclusion des trois points précédents, qu'il n'existe pas véritablement de méthode parfaite pour différencier les trois éléments distincts que sont les cultures, les arbres et les minéraux. Il est fort possible qu'en ayant choisi (très !) judicieusement (que des signatures dont on sait qu'elles sont séparables, ou même uniquement certaines bandes, plutôt que toutes) notre base de signatures de tests, nos résultats eussent été excellents... mais biaisés !

## CONCLUSION

Pour conclure ce mémoire, rappelons les différentes étapes de ce long processus. Le but de ce travail était de vérifier s'il était réalisable de démêler des signatures spectrales dans le but de connaître la constitution d'un couvert végétal; c'est pourquoi nous avons décidé d'utiliser la méthode statistique dite de l'Analyse en Composantes Indépendantes (ACI) dont la finalité est justement d'extraire de mélanges linéaires les signaux indépendants susceptibles de les constituer.

Ainsi, dans le premier chapitre, nous avons exhaustivement détaillé la théorie de l'ACI avec notamment toutes les notations vectorielles et matricielles de ce modèle, ses conditions d'application et ambiguïtés, les notions d'indépendance et de fonctions de contraste basées sur la répulsion naturelle envers les signaux gaussiens que présente l'ACI.

Ensuite, nous avons abordé l'état de l'art divisé en deux parties. Dans la première, nous avons étudié les différentes méthodes et algorithmes existants pour synthétiser cette notion globale d'ACI en concluant que le meilleur était celle d'Hyvärinen, FastICA, du fait de sa vitesse de convergence et son efficacité. Dans un second temps, nous avons présenté les travaux antérieurs (très légers en nombre) sur l'application de l'ACI à la ségrégation hyperspectrale. À notre grand désarroi, peu d'articles ont été trouvés; ce qui constituait d'ailleurs une des curiosités de ce travail devant le peu d'intérêt qu'il avait précédemment suscité.

Nous avons ensuite présenté dans le détail les algorithmes du point fixe, sur lesquels s'appuient FastICA, dont principalement celui à méthode symétrique à non-linéarité tangente hyperbolique que nous utilisons comme routine.

Dans l'avant-dernier chapitre, notre protocole expérimental a été longuement exposé avec la notion nouvelle mais adéquate de confiance, propre à notre système, consistant à un double niveau de formation des paires : un dit théorique à partir de la base partielle et un

autre dit pratique pour la base totale. En comparant ces deux types de paires, nous obtenons le succès d'associations, pondéré par le taux de convergence et ses caractéristiques de confiance : locale et globale, théorique et pratique.

Finalement, dans le dernier chapitre, nous avons présenté nos résultats obtenus à partir des trois types de données sur lesquels nous avons testé notre méthode : sept images en niveaux de gris, puis cinq signaux temporels classiques et enfin sept signatures présentant la caractéristique intéressante et pertinente d'appartenir à divers catégories (plante contre minéral) et types (herbe contre arbre et inosilicate contre phyllosilicate). Nous avons tout d'abord énoncé les résultats globaux, communs à ces trois types, à savoir l'uniformité de comportement de l'ACI envers les proportions de mélange et le problème inhérent à la décorrélation des bases totale et partielle. Quant aux résultats particuliers, nous avons montré que l'ACI démélange parfaitement les signaux de base, du fait de leurs formes caractéristiques et a donc mis en lumière le pourquoi de l'utilisation soutenue de l'ACI dans le domaine du traitement de signal. De même, avec les images, nous avons montré, qu'en excluant la quatrième image qui était la deuxième gaussienne dans l'échantillon (qui entraînait de mauvais résultats; ce qui confirmait la théorie que nous avons clairement démontrée), les résultats étaient bons; illustration de l'application de l'ACI aux images.

Enfin, pour les signatures, pour être honnête, les résultats sont (très) moyens car du fait de la subjectivité inhérente à la détermination de la séparabilité des signaux, nous avons été contraints de nous fier uniquement à nos valeurs de confiances, pondérées par un seuil de tolérance, qui s'avère être purement subjectif. Celui-ci a probablement joué un rôle négatif, mais toujours est-il que nous avons dans ce cas aussi relevé la présence de signatures « perturbatrices » qui dégradèrent fortement la capacité de démélange des autres signatures l'accompagnant. Sans celles-ci, les résultats étaient meilleurs mais non encore satisfaisants. Finalement, nous pouvons affirmer que, globalement, de par la forte subjectivité et le problème notoire d'échelle, tenté d'être réduit par la décorrélation, l'application de l'ACI s'avère trop imprécise dans le cas justement trop précis des signatures (importance

capitale des emplacements des pics et creux, . . . ). Cependant, pour certaines combinaisons particulières, l'ACI a merveilleusement bien fonctionné; ce qui signifierait typiquement un traitement cas par cas pour les signatures, du fait de la dépendance à la base. Plus précisément, elle a permis de correctement distinguer cultures, arbres et minéraux, les uns envers les autres; ce qui est un point fort, intéressant pour les applications agricole, forestière et d'exploration minière. Elle apporte ainsi une solution de rechange (ou du moins complémentaire; une comparaison y répondrait) aux indices classiques, tels NDVI, qui présentent le désavantage, par rapport à l'ACI, de préalablement devoir sélectionner les bandes « pertinentes ».

En d'autres termes, par rapport à notre but initial de juger de la capacité et pertinence de l'utilisation de l'ACI dans le démêlage des signatures, nous pouvons conclure en affirmant qu'elle dépend vraiment trop, d'une part, de la nature de la base; et, d'autre part, de la subjectivité humaine, pour être véritablement un critère satisfaisant dans le cas général. Elle peut cependant parfaitement l'être en particulier, en considérant des (groupes de) signatures pertinentes. Finalement, dépendamment de la nature de la base, nos résultats auraient pu être excellents mais évidemment biaisés; c'est pourquoi nous considérons ce travail comme un premier pas; surtout que nous soumettons dans la suite bon nombre de recommandations, qui, à notre humble avis, vont permettre d'améliorer et préciser les résultats.

## RECOMMANDATIONS

Il reste encore beaucoup d'améliorations et de travail à faire sur ce sujet d'études. De nombreux champs d'investigation en perspective. Les principales recommandations que nous voyons et proposons sont les suivantes, classées par ordre décroissant de priorité, d'intérêt et de rapidité de mise en exécution :

- a. utiliser la *méthode ROC* pour l'analyse plus approfondie des taux d'erreur et de rejet en fonction des seuils de confiances locales (voir 5.6.3.1). Celle-ci devrait permettre de déterminer les seuils optimaux et réduire les intervalles de confiances, afin de diminuer possiblement la subjectivité liée à toutes les observations;
- b. revoir la façon de décorrélérer les bases partielle  $S_{\mu,\Sigma}$  et totale  $S_{\mu,\Sigma}^{\text{base}}$  car c'est un point critique de l'étude (voir 5.4.4). Il serait peut-être pertinent d'appliquer une *ACP* (sur  $S_{\mu,\Sigma}$  et/ou  $S_{\mu,\Sigma}^{\text{base}}$ ) telle que dans le blanchiment des mélanges;
- c. dans le même ordre d'idées : dans le protocole, au lieu de considérer  $S_{0,I}^{\text{base}}$  dans son ensemble pour la comparaison pratique avec  $Y_{0,I}$ , comparer ce dernier à *chaque test* (d'où le défaut d'accroître considérablement le temps de calcul) avec l'ensemble de toutes ses ( $S_{0,I}^{\text{base}}$ ) combinaisons décorrélées possibles de  $n$  parmi  $N$  et adopter une stratégie de *vote*, afin de déterminer au final les couples corrects. Cette méthode devrait permettre d'améliorer les résultats ou au moins d'en savoir plus sur les erreurs;
- d. augmenter le nombre d'échantillons (observations) de chaque signal par, par exemple, une simple régression linéaire sur les échantillons déjà existants, pour constater ou non s'il pourrait s'agir d'une possibilité simple et rapide d'améliorer les résultats;
- e. refaire les tests des images avec seulement six images : éliminer la quatrième, la pseudo-gaussienne. Les résultats devraient être quasi-parfaits;
- f. réaliser une étude de cas d'une culture (base) avec et/ou sans mauvaise herbe pour juger de l'efficacité de l'ACI dans un cas particulier de signatures pertinentes;
- g. comparer l'ACI à l'indice NDVI et autres équivalents;



- h. réaliser une étude plus exhaustive qu'en 5.4.3.2 des comportements d'erreurs de  $E_A$  avec principalement l'origine des discontinuités et les variations entre les  $n - 1$  signaux restants;
- i. étudier le modèle bruité de l'ACI, énoncé en 1.4.1.2, et notamment, l'effet d'un pourcentage variable de bruit de différents types (gaussien, « sel et poivre », uniforme, ...);
- j. étudier plus en profondeur les autres algorithmes de l'ACI, présentés à l'annexe 4, tel notamment ACY; car ses propriétés sont tout de même proches de celles de Fast-ICA (voir 2.1.2.2);
- k. se consacrer à l'aspect statistique de l'ACI avec notamment l'étude des algorithmes dédiées à la minimisation de l'information mutuelle (liens avec la BSS) avec des auteurs tels que Comon, Cardoso, ...;
- l. considérer les fractions d'abondances comme composantes indépendantes tel que présenté succinctement en 2.2.3;
- m. intégrer les hypothèses de :
  - (a) dépendance à la base des signatures car issues du sol;
  - (b) pertes dues aux mesures en champs : distance avec l'avion;
  - (c) possible non perfection des signatures que l'on tente d'extraire, contrairement à celles obtenues en laboratoire. Typiquement, deux signatures proches auront deux signatures différentes en laboratoire (quasi-similaires mais différentes), alors qu'en champ, il y a de fortes chances qu'elles soient égales.

mais ces concepts sont loin d'être la priorité car non applicables dans un avenir proche.

## **ANNEXE 1**

### **Indépendance**

Cette annexe a pour but principal de démontrer les résultats ou théorèmes de la partie 1.7 du mémoire, tout en apportant certaines précisions pertinentes, si le besoin est.

## 1 Définition littéraire

Deux variables entre elles peuvent être de trois natures, concernant leur indépendance : corrélées, décorrélées ou indépendantes.

Deux variables sont dites *corrélées* si la connaissance d'une variable entraîne la connaissance *totale* de l'autre variable (typiquement toute fonction bijective de type  $y = f(x)$  : connaître  $x$  détermine totalement et exclusivement  $y$ ).

Deux variables sont dites *décorrélées* si la connaissance d'une variable entraîne la connaissance *partielle* de l'autre variable.

Enfin, deux variables sont dites *indépendantes* si la connaissance d'une variable n'entraîne *aucune* connaissance de l'autre. Par exemple (Stone, 2002), si deux personnes parlent en même temps, l'amplitude de la voix de l'une ne renseigne en aucune façon sur celle de l'autre : les amplitudes sont indépendantes. La représentation classique est un nuage de points très peu dense; typiquement, une distribution uniforme (voir figure 23(c)).

## 2 Exemple

On illustre la différence entre « corrélées », « décorrélées » et « indépendantes » avec l'exemple suivant (Stone et Porrill, 1998; Stone, 2002) : on considère deux variables aléatoires continues définies par :  $x = \cos z$  et  $y = \sin z$ , avec  $z \in [0, 2\pi]$ .

On sait que  $x^2 + y^2 = 1$ , soit  $y = \pm\sqrt{1 - x^2}$ . Ainsi,  $x$  et  $y$  sont décorrélées car à partir de  $x$ , seule une information partielle de  $y$  est potentiellement disponible. De plus, il est évident que  $E[x] = E[y] = 0$  et  $E[xy] = E\left[\frac{1}{2}\sin(2z)\right] = 0$ ; soit bien  $E[xy] - E[x]E[y] = 0$  et cela confirme le résultat. En revanche, si l'on considère les variables  $X$  et  $Y$  telles que

$X = x^2$  et  $Y = y^2$ ; alors,  $X + Y = 1$  et  $Y = 1 - X$  est une fonction bijective, qui définit de façon unique  $Y$  à partir de  $X$  :  $X$  et  $Y$  sont corrélées. Finalement,  $x^2$  et  $y^2$  sont corrélées, alors que  $x$  et  $y$  sont décorréliées; comme le montrent les figures 23(a) et 23(b). On rappelle que la figure 23(c) montre un ensemble de variables indépendantes, typiquement un nuage de points.

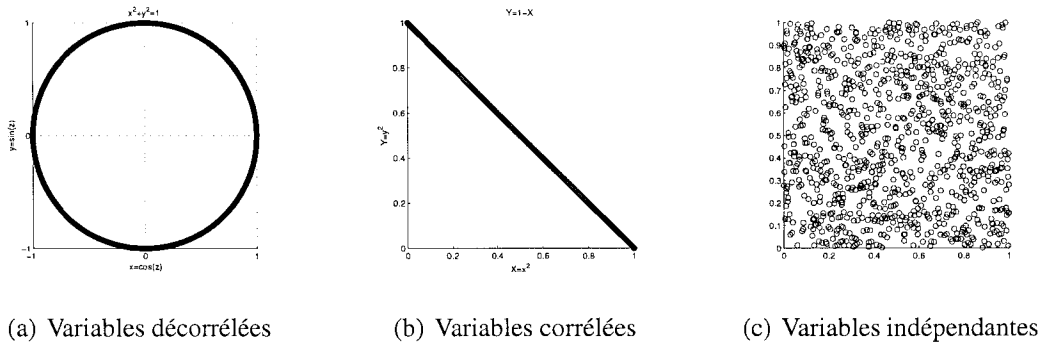


Figure 23 Variables décorréliées, corrélées et indépendantes

En outre,  $x$  et  $y$  ne sont donc pas indépendantes car pour une variable  $x$  donnée, on obtient de l'information partielle pour  $y$  (deux possibilités de valeurs) et mathématiquement, on peut notamment vérifier pour  $g_1(x) = g_2(x) = x^2$  que  $E[x^2 y^2] \neq E[x^2]E[y^2]$ . En effet, par les règles classiques de trigonométrie,  $E[x^2] = E[\cos^2 z] = E\left[\frac{1+\cos 2z}{2}\right] = \frac{1}{2}$ . De même,  $E[y^2] = \frac{1}{2}$  et  $E[x^2 y^2] = \frac{1}{8}$ ; d'où le résultat.

### 3 Équation caractéristique

De par sa nature mathématique, la divergence de Kullback-Leibler (DKL) vérifie la propriété de transitivité. Soit  $X$ ,  $Y$  et  $Z$  trois variables aléatoires, on a :

$$\delta(X|Z) = \delta(X|Y) + \delta(Y|Z) \quad (1.1)$$

Ainsi, de même, on a :

$$\delta(P_Y|P_Y^{P \wedge G}) = \delta(P_Y|P_Y^G) + \delta(P_Y^G|P_Y^{P \wedge G}) = G(Y) + C(Y) \quad (1.2)$$

$$\delta(P_Y|P_Y^{P \wedge G}) = \delta(P_Y|P_Y^P) + \delta(P_Y^P|P_Y^{P \wedge G}) = I(Y) + \sum_i G(Y_i) \quad (1.3)$$

On a le résultat  $\delta(P_Y^P|P_Y^{P \wedge G}) = \sum_i G(Y_i)$  car on fait l'analogie de  $\delta(P_Y|P_Y^G) = G(Y)$  pour chaque  $Y_i$  et car le log du produit des  $P_{Y_i}$  dans la DKL se transforme en somme.

En égalant les deux équations précédentes, on obtient bien (1.55) :

$$I(Y) + \sum_i G(Y_i) = C(Y) + G(Y) \quad (1.4)$$

#### 4 Géométrie de la dépendance

À partir de la théorie de l'information, on peut créer un espace de représentation des distributions où le théorème de Pythagore est applicable. Sans rentrer dans les détails, consultables en (Cardoso, 2004a), il est possible de démontrer que la propriété de transitivité de la DKL est l'application du théorème de Pythagore dans ce nouvel espace. Ainsi, chaque membre de l'équation caractéristique est interprétable comme l'égalité de Pythagore et il y a égalité car les deux ont une hypothénuse commune. La figure 24 (Cardoso, 2004a) illustre ce propos. On parle de « géométrie de la dépendance » (Cardoso, 2004a,b).

#### 5 Au plus une source gaussienne

Il est impossible d'extraire les composantes indépendantes si plus d'une source est gaussienne.

Ce théorème est démontrable, quand toutes les sources sont gaussiennes, grâce au théorème de la limite centrale et la faiblesse de la décorrélation face à l'indépendance.

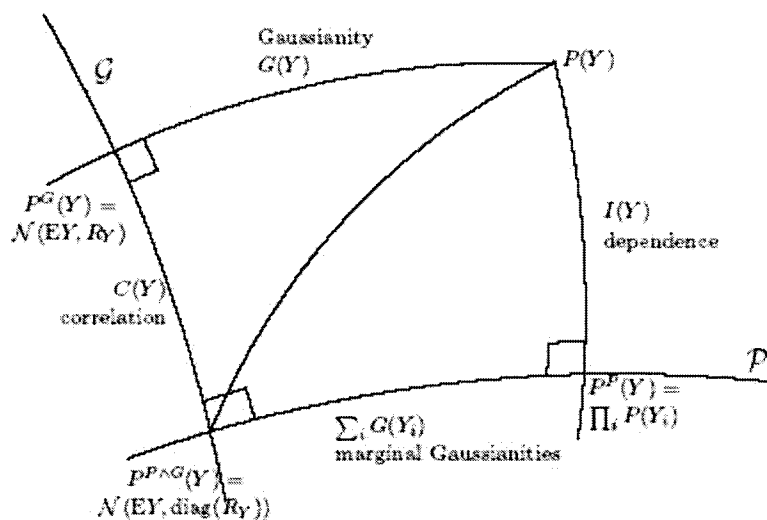


Figure 24 Géométrie de la dépendance

### 5.1 Théorème de la limite centrale

La première façon est de considérer le théorème de la limite centrale (Hurri, 1997). En effet, ce théorème de statistique s'exprime de la façon suivante :

Soit une suite  $(S_n)_{n \in \mathbb{N}}$  de variables aléatoires indépendantes de même moyenne  $\mu$  et de variance  $\sigma^2$ . Si l'on considère la variable aléatoire  $X_n = \sum_{i=1}^n S_i$ , alors :

$$\lim_{n \rightarrow \infty} \frac{X_n - E[X_n]}{\sigma_{X_n}} = \mathcal{N}(0, 1) \quad (1.5)$$

avec  $E[X_n] = \mu$  et  $\sigma_{X_n} = n\sigma$ .

En d'autres termes, la distribution d'une somme de  $n$  variables aléatoires indépendantes converge vers une distribution gaussienne, quand  $n$  tend vers l'infini; comme le montre la figure 25 (Hurri, 1997), où sont représentées la densité de probabilité (ou distribution) d'une variable aléatoire uniforme (à gauche) et les distributions de la somme de deux (au milieu) et de trois (à droite) telles variables uniformes. On observe bien la tendance vers

une distribution gaussienne. Cela deviendrait encore plus flagrant au fur et à mesure que le nombre de variables sommées augmenterait.

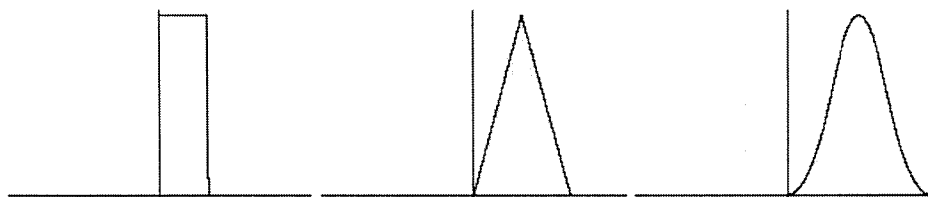


Figure 25 Illustration du théorème de la limite centrale

Dans le cas de l'ACI, on sait, par hypothèse, que les sources  $s_i$  sont de moyennes nulles, de variances unitaires et de matrice de covariance unitaire. Ainsi, d'après ce théorème, chaque mélange  $x_j$  présente approximativement une distribution gaussienne normalisée. « Approximativement » car premièrement, nous ne sommes pas en l'infini; et deuxièmement, car les mélanges sont en réalité des combinaisons linéaires de sources indépendantes; de ce fait, des termes scalaires sont présents mais ne modifient pas fondamentalement la nature gaussienne des mélanges.

On raisonne par l'absurde en cherchant une situation où il serait impossible de trouver les signaux sources. On pose l'hypothèse (Hyvärinen et Oja, 2000) que la matrice de proportion des mélanges  $A$  est orthogonale (soit  $AA^T = I$ ). Dans ce cas, la matrice de covariance  $C_X$  de  $X$  devient, après calcul,  $C_X = AA^T = I$ , l'identité. Par conséquent, chaque mélange  $x_j$  est décorrélé et car gaussien, est aussi indépendant. Finalement, on présente à l'ACI des signaux (les mélanges  $x_j$ ) qui sont indépendants et gaussiens, exactement comme les signaux sources recherchés. L'ACI renverra donc les mélanges comme sources et le problème n'est pas résolu.

## 5.2 Faiblesse de la corrélation

Il existe une autre démonstration du théorème « au plus une source gaussienne ». En effet, considérons toujours  $n$  variables aléatoires gaussiennes indépendantes  $(Y_i)_{i \in [1, n]}$ . Elles sont donc décorrélées et leur matrice de covariance est diagonale. De par la symétrie de la matrice de covariance, elles vérifient les  $n(n - 1)/2$  conditions :  $\text{cov}(Y_i, Y_j) (= \text{cov}(Y_j, Y_i)) = 0$ , pour  $i \neq j$ . Or, pour déterminer la transformation linéaire de  $X$  à  $Y$  grâce à  $B$ ,  $n^2$  conditions sont nécessaires (nombre d'éléments de  $B$ ). On se retrouve ainsi avec plus d'inconnues que d'équations; d'où une infinité de solutions possibles. Par conséquent, dans le cas de variables gaussiennes, la décorrélation est une condition trop faible pour résoudre le problème de l'ACI; d'où le théorème.

## 6 Existence et Unicité

Nous nous intéressons ici aux conditions d'existence et d'unicité de la décomposition ACI : celle d'un vecteur aléatoire représenté comme une combinaison linéaire des sources indépendantes ( $X = AS$ ). En général, une telle décomposition n'existe pas. Il existe cependant deux exceptions (Cardoso, 2002) : quand le vecteur est gaussien et dans le cas trivial où la décomposition est d'emblée vraie. Explications.

### 6.1 Gaussianité et infinité

Pour un vecteur gaussien, toute transformation qui décorrèle  $X$  rend ses composantes ( $Y$ ) indépendantes. Or, il existe une *infinité* de telles transformations. Elles sont de la forme  $B = \Lambda UC_X^{-1/2}$ , avec  $\Lambda$  diagonale,  $U$  orthogonale et  $C_X$  la matrice de covariance de  $X$ .

La preuve tient au fait qu'en calculant l'expression de  $C_Y$ , matrice de covariance de  $Y = BX$ , on trouve, après simplification de produits matriciels, que  $C_Y = \Lambda^2$ , une matrice diagonale. Donc les composantes de  $Y$  sont décorrélées et, car gaussiennes, indépendantes !



Ce résultat met, une fois de plus, en évidence l'insuffisance de la décorrélation, à trouver des solutions à l'ACI.

## 6.2 Trivialité et unicité

La seconde exception est celle où originellement  $X = AS$ , avec  $S$  un vecteur de composantes indépendantes avec au plus une gaussienne (le cas parfait !). En ce cas, si  $Y = BX$  est également un vecteur de composantes indépendantes, alors :

- a. la décomposition est *unique*;
- b.  $Y$  et  $S$  sont *essentiellement égaux* (voir 1.6.3).

De cette seconde exception résulte *le théorème d'unicité de l'ACI*, énoncé par Cardoso (Cardoso, 2002) :

« Si des données résultent effectivement d'un mélange linéaire de composantes non gaussiennes indépendantes, il est alors possible d'extraire ces composantes de façon essentiellement unique en ne s'appuyant que sur un critère d'indépendance mutuelle. »

Le terme « essentiellement unique » fait allusion à « essentiellement égaux ». Quant au « critère d'indépendance mutuelle », il désigne l'équation (1.56), où le but est de maximiser les non-gaussianités, qui permet alors d'extraire les composantes indépendantes. Pour ce faire, on introduit des fonctions de contraste, notion abordée en 1.8.

## **ANNEXE 2**

### **ACI versus ACP et PP**

Dans cette annexe, l'ACI est successivement comparée à l'analyse en composantes principales (ACP), puis à la projection poursuite (PP).

## 1 ACI versus ACP

Comme leurs noms l'indiquent, ACI et ACP (ou PCA en anglais pour « Principal Component Analysis ») sont deux méthodes d'analyse en composantes. À ce titre, elles sont à la fois similaires et différentes et méritent donc une comparaison.

### 1.1 Indépendance versus décorrélation

Là où l'ACI recherche des composantes indépendantes, l'ACP recherche des composantes *décorrélées*. Ainsi, au même titre que l'ACI est *la* décomposition d'un mélange linéaire en composantes indépendantes, l'ACP est *une* décomposition d'un mélange linéaire en composantes décorréliées (De Lathauwer et al., 2000).

On sait, d'après l'annexe 1, section 6.2, qu'il existe une unique décomposition de l'ACI (quand elle existe). En revanche, d'après la section 6.1, il existe une infinité de décompositions décorréliées (d'où « une » décomposition et non « la »). Par souci d'unicité, l'ACP recherche, en plus, des composantes orthogonales (Cardoso, 2004b). Ces composantes décorréliées et orthogonales sont appelées les composantes *principales*.

Les notations du modèle mathématique de l'ACP sont les mêmes que celles de l'ACI :  $X = AS$  (1.19) et  $Y = BX$  (1.25). Les seules différences (Cardoso, 2002) sont que, dans le cadre de l'ACP, la matrice de proportion des mélanges  $A$  est orthogonale, soit  $A^T A = I$  (car on recherche des composantes orthogonales) et que la corrélation de  $Y$ , estimée de  $S$  est nulle, soit  $C(A^{-1}X) = 0$  (fonction de contraste possible).

## 1.2 Décomposition en valeurs singulières

Pour obtenir ces directions orthogonales, l'ACP est en fait une *décomposition en valeurs singulières* (DVS) de la matrice de covariance  $C_X$  de  $X$ . Grossièrement, cela signifie que l'on décompose  $C_X$  sous la forme  $C_X = USV^T$ , avec  $U$  et  $V$  deux matrices orthogonales et  $S$  une matrice diagonale. De par la symétrie de  $C_X$ , cette équation se réduit alors à  $C_X = EDE^T$ , qui est la décomposition en valeurs/vecteurs propres (DVP) de  $C_X$ , avec :

- a.  $D$  la matrice diagonale des valeurs propres de  $C_X$ ;
- b.  $E$  la matrice orthogonale des vecteurs propres (associés aux valeurs propres) de  $C_X$ ; chaque colonne est un vecteur propre.

L'ACP consiste finalement en une DVP dans laquelle les directions orthogonales sont définies par les vecteurs propres. Ceux-ci sont deux à deux orthogonaux. Les composantes principales sont obtenues par ordre décroissant de variances, qui sont les valeurs des valeurs propres. En d'autres termes, la première composante principale extraite sera déterminée par le vecteur propre de  $C_X$ , présentant la valeur propre la plus élevée; et ainsi de suite, par décroissance, pour les autres. La figure 26 illustre la première composante principale, qui maximise la variance; la ligne indique la direction de la première composante principale, qui optimise, au sens des moindres carrés, le passage de dimension 2 à 1. La deuxième composante serait orthogonale à la première (Hyvärinen, 1999e).

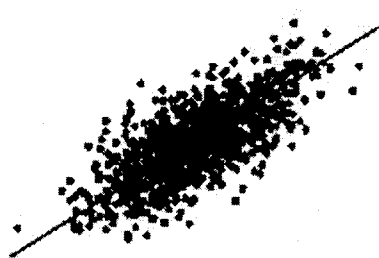


Figure 26 ACP d'un nuage de point

Plus de précisions sur l'ACP sont disponibles dans le tutoriel (Smith, 2002).

### 1.3 L'ACI comme extension de l'ACP

En outre, d'après l'annexe 1, section 6.1, on sait que la décorrélation est insuffisante à résoudre l'ACI car cette condition est moins forte que celle de l'indépendance. L'ACI apparaît donc comme une méthode plus perfectionnée que l'ACP. La figure 27 (Karhunen et al., 1997) illustre ce point : on considère un nuage de données uniformément concentrées et distribuées dans un parallélogramme; on applique l'ACI et l'ACP pour déterminer la meilleure base (les meilleures directions : les vecteurs  $b_i$ ), représentative des données. On remarque que les directions de l'ACP (trait pointillé) sont bien orthogonales mais représentent une plus mauvaise base que celles de l'ACI (trait plein), qui permettent d'exprimer parfaitement les données car suivent les médianes des côtés du quadrilatère.

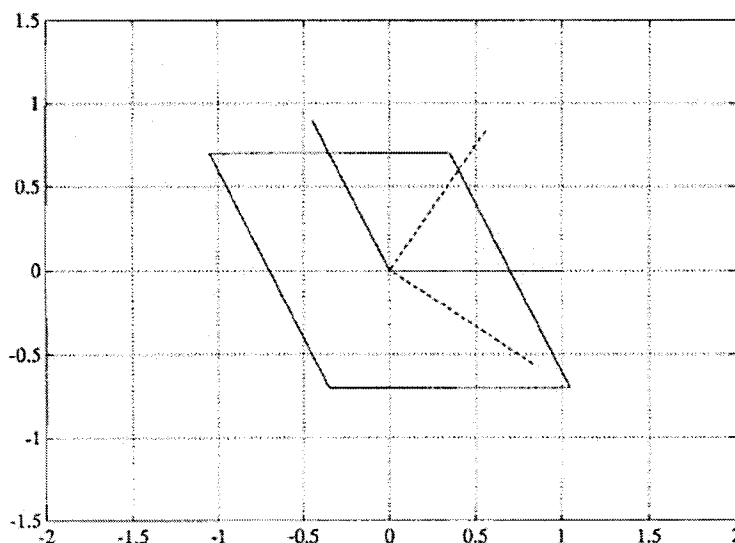


Figure 27 ACI versus ACP

Malgré ses défauts, l'ACP est tout de même utile à l'ACI. En effet, la première opération de l'ACI (après le centrage) est le *blanchiment* (voir 3.1.2), qui consiste à « blanchir » (décorrélérer) les données; ce qui est le rôle d'une ACP. Par conséquent, l'ACI peut être

considérée comme une extension (Karhunen et al., 1997; Hurri et al., 1996), un affinement (De Lathauwer et al., 2000) de l'ACP.

#### 1.4 Réduction de la dimensionnalité

De plus, outre le blanchiment, l'autre but de l'ACP (avec la décorrélation des mélanges  $x_j$ ) est de réduire la dimensionnalité des données. En effet, on a vu que les composantes principales sont déterminées par décroissance de la valeur des valeurs propres. Celles-ci, toutes réunies, représentent le contenu initial de l'information. Si l'on ignore certaines valeurs propres, on perd de l'information. Toutefois, si les valeurs des ignorées sont faibles ou négligeables (par rapport aux autres), alors on ne perd que peu d'information et elles peuvent souvent correspondre à du bruit (Hyvärinen et Oja, 1996a). On réduit ainsi la dimension des données sans perte majeure d'information : la dimensionnalité passe du nombre de valeurs propres initiales au nombre de conservées; c'est pourquoi cette technique d'ACP est notamment utilisée dans la compression de données. Concernant notre problème d'ACI, on voit bien en 3.1.2 que l'on passe de  $m$  mélanges à  $n$  mélanges blanchis (où initialement  $n$  est le nombre de sources et  $m \geq n$ ).

#### 1.5 Propriétés diverses

Pour terminer la comparaison entre ACI et ACP, nous pouvons résumer toutes les propriétés qui les lient ou au contraire les distinguent (Cardoso, 2002) :

- a. l'ACI, en tant qu'extension de l'ACP, ne peut rien lui apporter car celle-ci est trop limitée par rapport à l'ACI;
- b. la méthode d'application de l'ACP est unique et simple de par ses propriétés d'orthogonalité et de décorrélation (il s'agit toujours d'une DVS); contrairement à l'ACI où il existe de nombreuses méthodes, nombreux algorithmes pour converger à la solution (voir 1.8 et la partie 2.1 de l'état de l'art);

- c. l'ACP extrait les composantes principales par ordre décroissant de variances; d'où une hiérarchisation, contrairement à l'ACI, invariante par changement d'échelle;
- d. une ACP est toujours possible et réussie alors qu'une bonne ACI n'est possible que si l'hypothèse d'indépendance physique des données est valide. La solution de l'ACP existe toujours mais peut n'avoir aucune signification physique ! Elle est en fait purement mathématique ! On le repète, c'est une « simple » DVS.

À noter que dans (Oja et al., 1995), ACI et ACP sont comparées d'un point de vue charge informatique, à l'aide de réseaux de neurones, sur leur faculté à extraire leurs composantes respectives.

## 2 ACI versus projection poursuite

Une autre méthode intéressante à étudier et à lier avec l'ACI est la *projection poursuite* (PP) (ou « Projection pursuit » en anglais). La projection poursuite (Hyvärinen et Oja, 2000; Cardoso, 2002, 2004b; Stone et Porrill, 1998) consiste à rechercher des directions « intéressantes » dans un nuage de points multidimensionnel. Or, il est dit (Hyvärinen et Oja, 2000; Cardoso, 2002) que les directions intéressantes sont les moins gaussiennes possibles. C'est exactement, selon 1.7.4.1, le principe de l'ACI : le lien est évident.

Le but de la PP est donc de trouver une direction  $u$  telle qu'elle maximise la non-gaussianité  $G$  des nouvelles composantes (projetées sur ces directions privilégiées), soit :

$$\max_u G(u^T X) \quad (2.1)$$

Les composantes suivantes sont calculées de la même façon tout en s'assurant qu'elles soient orthogonales les unes avec les autres (Cardoso, 2002).

Finalement, la PP apparaît comme une variante (Hyvärinen et Oja, 2000) de l'ACI, dans sa recherche de non-gaussianité. Cependant, la PP ne postule aucune condition d'indépendance entre ses composantes. Mais si on la suppose initialement, alors, à partir des composantes extraites par la PP, on obtient les composantes indépendantes.

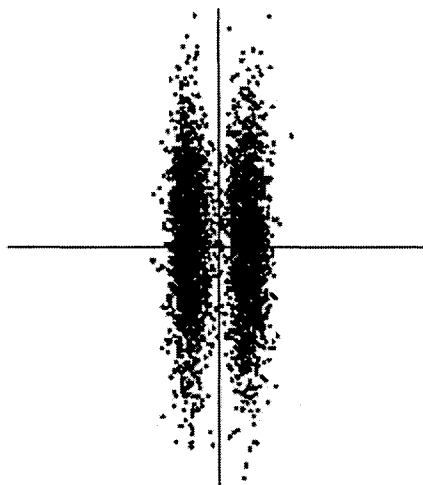


Figure 28 PP versus ACP

Ainsi, au même titre que l'ACI, la PP apparaît comme une méthode « supérieure » à l'ACP. Cette affirmation est illustrée par la figure 28, qui montre deux directions : celle verticale correspond à la première composante principale (maximisation de la variance) et l'autre, horizontale, à la direction de la PP. Il est clair que dans un but de séparation de classes, projeter sur la composante de la PP est pertinent et optimal; contrairement à la direction de l'ACP, qui annule la séparabilité initiale des deux groupes.



## **ANNEXE 3**

### **Moments et cumulants**

On rappelle qu'en abordant d'un point de vue statistique l'idée selon laquelle la décorrélation est une condition plus faible que l'indépendance, on peut introduire la notion de statistiques d'ordres plus élevés et plus précisément de moments d'ordres plus élevés.

## 1 Moments

On définit le *moment d'ordre  $r$*   $\mu'_r$  (respectivement *moment centré d'ordre  $r$*   $\mu_r$ ) d'une variable aléatoire  $x$  quelconque par :

$$\mu'_r(x) = \mu'_r = E[x^r] \quad (3.1)$$

$$\mu_r(x) = \mu_r = E[(x - E[x])^r] \quad (3.2)$$

alors les premiers moments sont connus et caractéristiques des statistiques :

$$\begin{aligned} \mu'_0 &= E[x^0] = 1 & \mu_0 &= 1 \\ \mu'_1 &= E[x] & \mu_1 &= 0 \\ \mu'_2 &= E[x^2] & \mu_2 &= E[(x - E[x])^2] = \sigma_x^2 = \mu'_2 - \mu_1'^2 \end{aligned} \quad (3.3)$$

Dans la suite, pour alléger les expressions, les moments et cumulants seront départis de leur « fonction de  $x$  ». Ainsi,  $\mu'_r$  désigne  $\mu'_r(x)$  et de même, comme nous le verrons, on remplace  $\kappa_r(x)$  par  $\kappa_r$ .

Il existe une autre façon de définir les moments, grâce à la *fonction génératrice des moments*  $M_x(t)$ , définie par :

$$M_x(t) = E[e^{tx}] \quad (3.4)$$

Plus précisément, en se rappelant du développement en série de Taylor de l'exponentielle au voisinage de 0 ( $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ ), on a, dans le cas de  $M_x(t)$ , au voisinage de 0 :

$$M_x(t) = E[e^{tx}] = \sum_{r=0}^{\infty} E[x^r] \frac{t^r}{r!} = \sum_{r=0}^{\infty} \mu'_r \frac{t^r}{r!} \quad (3.5)$$

Ainsi,  $\mu'_r$  s'obtient directement à partir de  $M_x(t)$  sous la forme :

$$\mu'_r = M_x^{(r)}(0) \quad (3.6)$$

et apparaît donc comme la valeur en 0 de la dérivée  $r^{\text{ième}}$  de cette fonction génératrice des moments  $M_x(t)$ ; soit également le  $r^{\text{ième}}$  coefficient du développement en série de MacLaurin (série de Taylor en 0) de  $M_x(t)$ .

## 2 Cumulants

Les statistiques d'ordres plus élevés mettent en lumière le fait que les moments sont parfois insuffisants à décrire les caractéristiques statistiques de variables aléatoires; c'est pourquoi les cumulants ont été introduits; et, ce, à partir de la *fonction génératrice des cumulants*  $K_x(t)$ , définie à partir de  $M_x(t)$ , telle que :

$$K_x(t) = \ln M_x(t) \quad (3.7)$$

Suivant le même principe que le moment d'ordre  $r$ , on appelle  $\kappa_r$  le *cumulant d'ordre  $r$* , défini à partir de l'équation précédente comme le  $r^{\text{ième}}$  coefficient du développement en série de MacLaurin de la fonction génératrice des cumulants  $K_x(t)$ , soit :

$$K_x(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} \quad (3.8)$$

$$\kappa_r = K_x^{(r)}(0) \quad (3.9)$$

À partir du développement en série de MacLaurin de  $\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n$ , et, à partir de (3.5), en développant le logarithme népérien de  $K_x(t)$  en 0, on obtient :

$$K_x(t) = \ln M_x(t) = \ln \left( 1 + \sum_{n=1}^{\infty} \mu'_n \frac{t^n}{n!} \right) \quad (3.10)$$

$$= \sum_{r=1}^{\infty} \left[ \frac{(-1)^{r+1}}{r} \left( \sum_{n=1}^{\infty} \mu'_n \frac{t^n}{n!} \right)^r \right] \quad (3.11)$$

Ainsi, en comparant (3.11) à (3.8), on peut en déduire, au moins pour les premières valeurs de  $r$ , l'expression des  $\kappa_r$ , en fonction des  $\mu'_r$ . Pour  $k$  de 1 à 4, on trouve successivement en développant :

$$\kappa_1 = \mu'_1 = E[x] \quad (3.12)$$

$$\kappa_2 = \mu'_2 - \mu_1'^2 = \mu_2 = E[(x - E[x])^2] \quad (3.13)$$

$$\kappa_3 = 2\mu_1'^3 - 3\mu_1'\mu_2' + \mu_3' = \mu_3 \quad (3.14)$$

$$\kappa_4 = -6\mu_1'^4 + 12\mu_1'^2\mu_2' - 3\mu_2'^2 - 4\mu_1'\mu_3' + \mu_4' = \mu_4 - 3\mu_2^2 \quad (3.15)$$

### 3 Kurtosis et kurtosis normalisés

Soit, plus précisément pour  $k = \{3, 4\}$  et en posant  $y = x - E[x]$ , variable centrée, on obtient :

$$\kappa_3 = \mu_3 = E[y^3] \quad (3.16)$$

$$\kappa_4 = \mu_4 - 3\mu_2^2 = E[y^4] - 3(E[y^2])^2 \quad (3.17)$$

Les cumulants d'ordre 3 et 4, soit  $\kappa_3$  et  $\kappa_4$ , sont appelés respectivement « skewness » et « kurtosis ».

Plus précisément, concernant le kurtosis, il présente la propriété intéressante d'être nul dans le cas d'une variable gaussienne. Cependant, pour savoir si une variable est effectivement gaussienne ou non, on préfère calculer le « kurtosis normalisé »  $\gamma_4$ , défini selon :

$$\gamma_4 = \frac{\kappa_4}{\mu_2^2} = \frac{\mu_4}{\mu_2^2} - 3 \quad (3.18)$$

Son avantage est qu'il est toujours minoré par  $-2$  (valeurs comprises dans  $[-2, +\infty[$ ). Il offre également une meilleure plage de comparaison car la division permet de « séparer » des variables présentant des valeurs de  $\mu_4$  initialement proches. Cette remarque est particulièrement vraie dans notre cas où, au chapitre 5, nous utilisons des données (signatures)

de natures proches statistiquement et de valeurs comprises entre 0 et 1; d'où un faible écart de moments d'ordre 4 et des kurtosis non normalisés semblables et proches de zéro; ce qui fausse donc notre perception de leur nature gaussienne (ou pas). Mais la division par leurs moments d'ordre 2 respectifs, faibles et proches de zéro, augmente au contraire leurs différences et permet ainsi un meilleur jugement de leur possible caractère gaussien.

## **ANNEXE 4**

### **Algorithmes de l'ACI**

Dans cette annexe, nous donnons plus de précisions sur les algorithmes suivants de l'ACI, présentés en 2.1.2.1, excepté le point-fixe (FP) de Hyvärinen (Hyvärinen, 1999a), qui est le sujet du chapitre 3 et qui n'est donc pas expliqué ici :

- a. Amari-Cichocki-Yang (ACY) (Amari et al., 1996);
- b. Bell-Sejnowski (BS) (Bell et Sejnowski, 1995b);
- c. « Equivariant adaptive source separation » (EASI) de Cardoso et Laheld (Cardoso et Laheld, 1996);
- d. « Nonlinear Principal Component Analysis by Recursive Least Squares » (NPCA-RLS) de Karhunen et Pajunen (Karhunen et Pajunen, 1997b; Pajunen et Karhunen, 1998; Karhunen et Pajunen, 1997a).

Nous allons nous intéresser à chacun d'entre eux, les uns à la suite des autres. Tous les détails, tels que les preuves mathématiques, sont disponibles dans (Giannakopoulos et al., 1999b) et dans leurs références propres, énoncées dans la liste ci-dessus.

À noter, en guise de préambule, que, dépendamment des algorithmes présentés, certains travaillent sur  $B$ , alors que d'autres travaillent sur  $W^T$ , version blanchie de  $B$  (voir le chapitre 3.1.2, concernant le blanchiment). En tout cas, tous utilisent une règle d'apprentissage récursive sur ce paramètre ( $B$  ou  $W$ ), du genre :

$$B(t+1) = B(t) + \Delta B \quad \text{ou} \quad W(t+1) = W(t) + \Delta W \quad (4.1)$$

avec  $t$  l'itération et  $\Delta B$  (ou  $\Delta W$ ), le paramètre de mise à jour, qui pourra avoir des formes différentes selon les algorithmes. Cette expression peut se résumer sous la forme plus simple :

$$B = B + \Delta B \quad \text{ou} \quad W = W + \Delta W \quad (4.2)$$

## 1 ACY

À partir de la notation temporelle (accréditée « traitement de signal ») du problème de l'ACI (voir partie 1.2), l'algorithme ACY (Amari et al., 1996) cherche à minimiser l'in-

formation mutuelle définie en (1.68) ( $I(Y) = \sum_{i=1}^n H(y_i) - H(Y)$ ), grâce à une descente de gradient sur  $B$ , du genre  $B_{n+1} = B_n + \Delta B = B_n + \frac{\partial B}{\partial t}$ , où, vu que  $I(Y) = I(B)$  car  $X$  est constant,  $\Delta B$  est de la forme :

$$\frac{\partial B}{\partial t} = -\eta(t) \frac{\partial I}{\partial B} B^T B \quad (4.3)$$

avec  $\eta(t)$  la fonction d'apprentissage, définie, par exemple, par  $\eta(t) = 250 \exp(-5t)$ , dans la simulation de (Amari et al., 1996).

Après développements mathématiques, il s'avère que :

$$\frac{\partial B}{\partial t} = \eta(t) [I - f(Y)Y^T] B \quad \text{où} \quad f(Y) = [f(y_1), f(y_2), \dots, f(y_n)]^T$$

avec  $f(y)$  la fonction définie par :

$$f(y) = \frac{3}{4} y^{11} + \frac{25}{4} y^9 - \frac{14}{3} y^7 - \frac{47}{4} y^5 + \frac{29}{4} y^3 \quad (4.4)$$

En faisant intervenir  $X$  dans l'équation précédente, la règle devient alors :

$$\frac{\partial B}{\partial t} = \eta(t) [I - f(BX)(BX)^T] B \quad (4.5)$$

et, après simplification pratique, on retrouve la forme de (Giannakopoulos et al., 1999b), avec  $\lambda$ , paramètre d'apprentissage constant :

$$\Delta B = \lambda [I - f(BX)(BX)^T] B \quad (4.6)$$

## 2 BS

De même que l'ACY, à partir de (1.68), l'algorithme BS (Bell et Sejnowski, 1995b) cherche à maximiser, par une croissance de gradient, l'entropie jointe  $H(Y)$ , définie selon (1.62) par :

$$H(Y) = -E[\log f(y)] = - \int_Y f(y) \log f(y) dy \quad (4.7)$$



Ici, la relation liant  $Y$  à  $X$  est de la forme :

$$Y = g(U) = g(BX + B_0) \quad (4.8)$$

où, dans le lexique des réseaux de neurones,  $B$  est assimilable à la matrice des poids et  $B_0$  une matrice de biais;  $g$  est typiquement une sigmoïde, définie par  $g(u) = \frac{1}{1+e^{-u}}$ .

D'après (4.8), on en déduit que :

$$f(y) = \frac{f(x)}{|J|} \quad (4.9)$$

où  $f(x)$ ,  $f(y)$  sont les densités de probabilités respectives de  $x$ ,  $y$  et  $|J|$  est la valeur absolue du jacobien  $J$  de  $Y = g(U)$ , défini par :

$$J = \det \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{pmatrix} \quad (4.10)$$

En injectant (4.9) dans (4.7), on obtient :

$$H(Y) = -E[\log f(y)] = E[\log |J|] - E[\log f(x)] = E[\log |J|] + H(X) \quad (4.11)$$

Ainsi, la règle d'apprentissage devient, à  $X$  constant :  $\Delta B = \frac{\partial H}{\partial B} = \frac{\partial}{\partial B} \log |J|$ . Or, on montre que  $J = (\det B) \prod_{i=1}^n y'_i$ , où  $y'_i = \frac{\partial y_i}{\partial u_i} = y_i(1 - y_i)$ , avec  $y_i = g(u_i)$ . Ainsi, la règle d'apprentissage devient :

$$\Delta B = \frac{\partial}{\partial B} \log |\det B| + \frac{\partial}{\partial B} \log \prod_i |y'_i| \quad (4.12)$$

Après développements mathématiques des deux termes du membre de droite, on obtient les règles recherchées, avec  $\lambda$  le taux d'apprentissage, définies par :

$$\Delta B = \lambda \left[ (B^T)^{-1} + (\mathbf{1} - 2Y)X^T \right] \quad (4.13)$$

$$\Delta B_0 = \lambda [\mathbf{1} - 2Y] \quad (4.14)$$

où  $\mathbf{1}$  est une matrice de 1, de même taille que  $Y$ .  $(\mathbf{1} - 2Y)X^T$  est appelé le *terme anti-Hebbien*, alors que  $(B^T)^{-1}$  est le *terme anti-redondance*.

Finalement, en introduisant  $X$ , la règle d'apprentissage se réécrit pratiquement :

$$\Delta B = \lambda \left[ (B^T)^{-1} + (\mathbf{1} - 2g(BX + B_0)) X^T \right] \quad (4.15)$$

$$\Delta B_0 = \lambda [\mathbf{1} - 2g(BX + B_0)] \quad (4.16)$$

### 3 EASI

L'algorithme EASI (Cardoso et Laheld, 1996), pour « Equivariant Adaptive Separation via Independence algorithm », réalise la séparation adaptative équivariante de sources indépendantes. Basiquement, il y a situation d'*équivariance* quand une transformation sur les données est équivalente à une transformation d'un paramètre de ces données.

À partir de cette idée de départ, l'algorithme EASI est basé sur l'idée de la *mise à jour en série*, définie par une règle d'apprentissage de type :

$$\Delta B = -\lambda H(Y(t))B(t) \quad (4.17)$$

où  $H(\cdot)$  est la fonction de mise à jour, différente de l'entropie, à déterminer. Ainsi, la transformation des données  $X$ , par  $Y = BX$ , devient équivalente à la transformation de  $B$  seule.

En se basant sur le produit scalaire euclidien des matrices, tel que  $\langle M|N \rangle = \text{Trace}[M^T N]$  et  $\langle M|M \rangle = \|M\|^2$ , et à partir du blanchiment de l'ACI (voir 3.1.2), il apparaît que  $H$  est la somme de deux contributions :

- a. la première,  $H_1$ , provient de la matrice de blanchiment  $U$  et est définie par :

$$H_1(Y) = YY^T - I;$$

- b. la deuxième,  $H_2$ , provient de la matrice orthogonale  $W^T$ , vérifiant  $Y = W^T V$ , telle que, après développements mathématiques, avec  $f(\cdot)$  une fonction impaire :
- $$H_2(Y) = f(Y)Y^T - Yf(Y^T).$$

Finalement, on obtient la règle d'apprentissage (de paramètre  $\lambda$ ) de l'algorithme EASI :

$$\Delta B = -\lambda [YY^T - I + f(Y)Y^T - Yf(Y^T)] B \quad (4.18)$$

Il existe également une forme normalisée de la règle précédente, telle que :

$$\Delta B = -\lambda \left[ \frac{YY^T - I}{1 + \lambda Y^T Y} + \frac{f(Y)Y^T - Yf(Y^T)}{1 + \lambda |Y^T f(Y)|} \right] B \quad (4.19)$$

#### 4 NPCA-RLS

L'algorithme NPCA-RLS (Karhunen et Pajunen, 1997a) réalise une analyse en composantes principales en utilisant une approche récursive non linéaire, basée sur la minimisation de l'erreur quadratique moyenne (ou des moindres carrés). Ici, la règle d'apprentissage concerne non plus  $B$ , mais  $W^T$ .

En effet, d'après 3.1.2, on sait que  $Y = W^T V$  et  $V = WY$  (car  $W$  est orthogonale ( $W^{-1} = W^T$ )), la fonction de coût à minimiser devient donc dans le cas simple du NPCA :

$$J_1(W) = E [\|V - Wg(Y)\|^2] = E [\|V - Wg(W^T V)\|^2] \quad (4.20)$$

et dans le cas amélioré du NPCA-RLS :

$$J_2(W(t)) = \sum_{i=1}^t \beta^{t-i} \|V(i) - W(i)g(W^T(i-1)V(i))\|^2 \quad (4.21)$$

où  $g$  est une fonction non-linéaire impaire (typiquement  $\tanh$ ) et  $\beta$  est appelé le « paramètre d'oubli » et représente le taux de diminution de l'influence des anciennes itérations de  $W$  sur les nouvelles.

La mise en pratique de cet algorithme s'avère quelque peu complexe aux premiers abords. En effet, bien que le taux d'apprentissage s'entraîne automatiquement et que le seul paramètre à considérer soit  $\beta$ , sa règle d'apprentissage nécessite l'intervention de 5 règles

supplémentaires, en plus de (4.2), qui sont :

$$\left\{ \begin{array}{l} Z(t) = g(W^T(t-1)V(t)) \\ H(t) = P(t-1)Z(t) \\ M(t) = H(t) / (\beta + Z^T(t)H(t)) \\ P(t) = \frac{1}{\beta} U [P(t-1) - M(t)H^T(t)] \\ E(t) = V(t) - W(t-1)Z(t) \\ W(t) = W(t-1) + E(t)M^T(t) \end{array} \right. \quad (4.22)$$

avec :

- a.  $Z(t) = g(Y(t))$ ;
- b.  $0 \leq \beta < 1$ , doit être proche de 0;
- c.  $U$  est telle que sa partie triangulaire supérieure est celle de l'argument calculée et que sa partie inférieure est la transposée de la précédente. Ainsi,  $U$  est définie symétrique;
- d.  $E(t)$  est la fonction d'erreur;
- e.  $\Delta W = E(t)M^T(t)$ ;
- f. à l'initialisation,  $P(0) = W(0) = I$ .

Finalement, le critère  $J_2$  à minimiser peut se réécrire sous la forme :

$$J_2(W(t)) = \sum_{i=1}^t \beta^{t-i} \|V(i) - W(i)Z(i)\|^2 = \sum_{i=1}^t \beta^{t-i} \|E(i)\|^2 \quad (4.23)$$

avec une erreur  $E(t)$  à minimiser; d'où l'appellation RLS, pour application récursive des moindres carrés.

## ANNEXE 5

### Neutralisation des deux premiers moments d'une matrice

Soit  $A$  une matrice rectangulaire de dimension  $[m, n]$ , de vecteur moyenne  $E[A]$  et de matrice de covariance  $E[(A - E[A])(A - E[A])^T]$ . Le but est de transformer cette matrice  $A$  en sa matrice équivalente décorrélée (moyenne nulle et de matrice de covariance unitaire (neutralisation des deux premiers moments)); soit :

- a.  $E[A] = 0$ ;
- b.  $E[(A - E[A])(A - E[A])^T] = E[AA^T] = I$ .

Pour ce faire, deux opérations sont nécessaires :

- a. centrer  $A$  :  $\bar{A} \leftarrow A - E[A]$ ;
- b. « diviser »  $\bar{A}$  par la « racine carrée » de sa matrice de covariance :

$$A_+ \leftarrow E [\bar{A}\bar{A}^T]^{-1/2} \bar{A}$$

Finalement, on obtient :  $E[A_+] = 0$  et  $E[A_+A_+^T] = I$ .

En résumé, à partir d'une matrice rectangulaire  $A$ , on obtient sa matrice équivalente de moyenne nulle et de matrice de covariance unitaire par la transformation :

$$A \leftarrow E [AA^T]^{-1/2} (A - E[A]) \quad (5.1)$$

*Preuve :*

Centrer  $A$  entraîne que  $E[\bar{A}] = 0$  et la deuxième opération entraîne que :

$$\begin{aligned} E [(A_+ - E[A_+])(A_+ - E[A_+])^T] &= E[A_+A_+^T] \\ &= E \left[ E [\bar{A}\bar{A}^T]^{-1/2} \bar{A}\bar{A}^T E [\bar{A}\bar{A}^T]^{-1/2} \right] \\ &= E [\bar{A}\bar{A}^T]^{-1/2} E [\bar{A}\bar{A}^T] E [\bar{A}\bar{A}^T]^{-1/2} \\ &= I \end{aligned}$$

## **ANNEXE 6**

### **Signatures spectrales USGS**

## 1 Source

Toutes les signatures utilisées sont issues de la base de données « splib04a » du laboratoire de spectroscopie de USGS (United States Geological Survey)<sup>1</sup>. Elles ont été obtenues à partir d'un spectromètre modifié de Beckman 5270, pour une bande de 0.2 à 3.0  $\mu\text{m}$  et corrigées en réflectance absolue (normalisation en faisant le rapport de ce qui est reçu par ce qui est envoyé). De plus, elles ont été mesurées pour un rapport signal à bruit d'au moins 500 et un niveau de réflectance de référence à 1.0.

## 2 Classement de la bibliothèque

La bibliothèque contient des signatures réparties en catégories et sous-catégories. On peut y distinguer la catégorie, le type, le groupe et la famille, que nous présentons maintenant.

### 2.1 Catégorie et types

Les spectres ou signatures de la base USGS sont au nombre de 444 répartis en deux catégories : les minéraux (423) et les « autres » (21). La catégorie des minéraux compte 18 types et celle des autres 3. Le tableau XIII montre la répartition des signatures de la base en fonction de leur catégorie et de leur type.

La figure 29 montre, quant à elle, un exemple graphique des écarts entre, d'une part, deux signatures de catégories différentes (minéraux versus autres, figure 29(a)); et, d'autre part, de même catégorie mais de types différents (chloride versus nesosilicate, figure 29(b)).

### 2.2 Groupes

Quand deux signatures font partie de la même catégorie et du même type, la subdivision suivante possible est celle du groupe.

---

<sup>1</sup>[http://speclab.cr.usgs.gov/spectral.lib04/clark1993/spectral\\_lib.html](http://speclab.cr.usgs.gov/spectral.lib04/clark1993/spectral_lib.html)



Tableau XIII

Répartition des signatures de la base USGS par catégorie et type

Minéraux	Nombre	Minéraux	Nombre	Minéraux	Nombre
Borate	8	Hydroxyde	13	Phyllosilicate	108
Carbonate	16	Inosilicate	45	Sorosilicate	5
Chloride	1	Nesosilicate	63	Sulfate	33
Cyclosilicate	7	Nitrate	1	Sulfide	23
Élément	2	Oxyde	24	Sulfosate	2
Halide	4	Phosphate	4	Tectosilicate	64
<b>Total : 18</b>					<b>423</b>
<b>Autres</b>					<b>Nombre</b>
Vernis du désert (Desert Varnish)					3
Kérogène (organique)					1
Plantes :					
Herbe (Grass)					3
Arbuste (Shrub)					6
Arbre (Tree)					8
<b>Total : 3</b>					<b>21</b>
<b>TOTAL</b>					<b>444</b>

La partie inférieure du tableau XIII indique que les plantes présentent trois groupes : ceux de l'herbe, de l'arbuste et de l'arbre. Quant aux minéraux, sur les 18 types, il en existe 7 sans groupe et 11 avec groupes. Le tableau XIV recense les groupes des 11 types avec leur nombre de signatures. Le rapport sous le nom de chaque type indique le nombre de signatures appartenant aux groupes de ce type par rapport au nombre de signatures total de ce type.

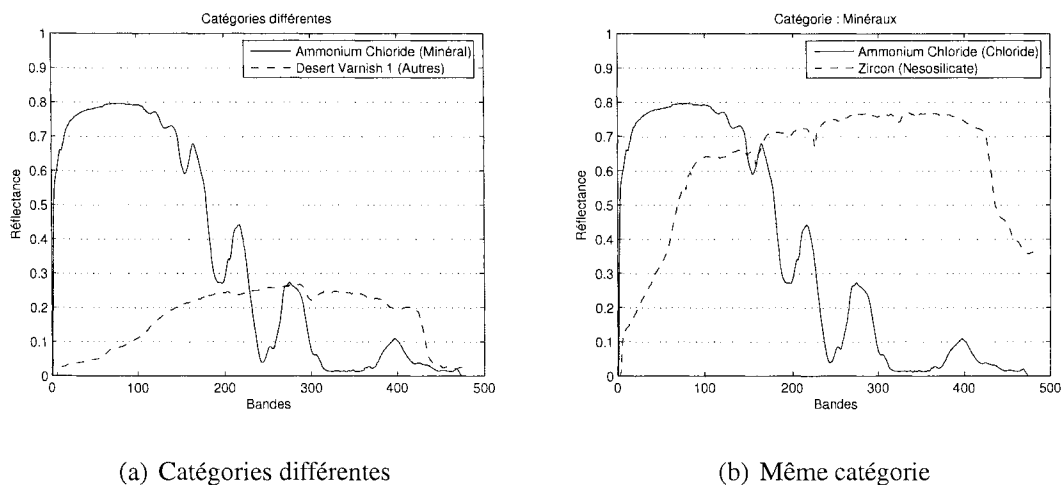


Figure 29 Signatures de même catégorie ou non

Tableau XIV

Répartition des signatures de minéraux par groupes

Type	Groupes		Type	Groupes	
Carbonate 9 / 16	Aragonite	Calcite	Phyllosilicate 83 / 108	Chlorite	Kaolinite
	Dolomite			Mica	Montmorillonite
Cyclosilicate 3 / 7	Axinite	Tourmaline	Sorosilicate 5 / 5	Epidote	
Inosilicate 39 / 45	Amphibole	Pyroxene	Sulfate 19 / 33	Alunite	Barite
				Copiapite	
Nesosilicate 36 / 63	Garnet	Olivine	Sulfide 9 / 23	Arsenopyrite	Chalcopyrite
				Cobaltite	Pyrite
Oxyde 13 / 24	Hematite	Rutile	Tectosilicate	Feldspar	Nepheline
	Spinel	56 / 64		Scapolite	Sodalite
Phosphate 3 / 4	Apatite	Olivine		Zeolite	
			<b>TOTAL :</b>	<b>275 / 391</b>	

Les figures 30(b) et 30(a) illustrent respectivement deux signatures de même catégorie, de même type et de même groupe ou non. Il apparaît clair visuellement que plus l'arborescence des signatures est semblable, plus les profils des signatures se ressemblent.

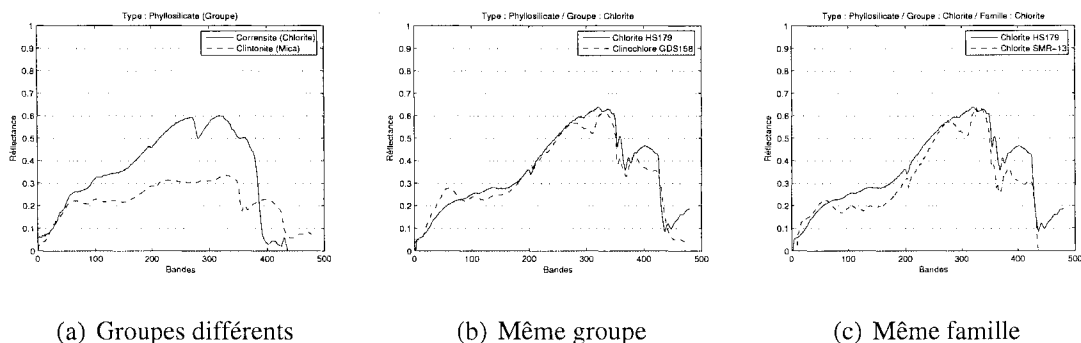


Figure 30 Signatures de même groupe / famille ou non

### 2.3 Familles

La complexité va d'ailleurs encore une fois augmenter car il existe une autre subdivision pour les signatures appartenant au même groupe : la famille.

De même que pour les groupes, certaines signatures appartiennent ou non à des familles. Le tableau XV montre la répartition des familles à l'intérieur des groupes de minéraux. Les rapports dans la colonne « Total » indiquent le nombre de signatures des familles du groupe considéré par rapport au nombre de signatures total de ce groupe. Dans la catégorie « autres », seul le groupe « Arbre » du type « Plantes » présente une famille « Aspen Leaf ».

Selon les figures 30(c) et 30(b) montrant des signatures de même famille ou non, il apparaît difficile de parvenir à véritablement les distinguer.

**Tableau XV**  
**Répartition des signatures de minéraux par familles**

Type	Groupes	Total	Familles	
Carbonate	Calcite	3 / 5	Calcite	
	Dolomite	2 / 2	Dolomite	
Inosilicate	Amphibole	15 / 21	Actinolite	Hornblende
			Richterite	Riebeckite
Tremolite				
	Pyroxène	11 / 18	Augite	Diopside
			Hedenbergite	Hypersthène
Nesosilicate	Garnet	20 / 23	Almandine	Andradite
			Grossular	Spessartine
	Olivine	11 / 13	Olivine	
Oxyde	Hematite	6 / 8	Hematite	
	Spinel	2 / 3	Magnetite	
Phyllosilicate	Chlorite	7 / 11	Chlorite	Clinocllore
	Kaolinite	20 / 25	Antigorite	Dickite
			Halloysite	Kaolinite
			Kaolin-Smectite	
Mica	23 / 30	Annite	Lepidolite	
		Muscovite	Phlogopite	
Montmorillonite	14 / 17	Kaolin-Smectite	Montmorillonite	
		Montmorillonite + Illite	Montronite	
Sorosilicate	Epidote	2 / 5	Epidote	
Sulfate	Alunite	14 / 16	Alunite	Jarosite
Sulfide	Chalcopyrite	2 / 2	Chalcopyrite	
	Pyrite	5 / 5	Pyrite	
Tectosilicate	Feldspar	23 / 28	Albite	Anorthite
			Buddingtonite	Labradorite
			Microcline	Oligoclase
			Orthoclase	Sanidine
Scapolite	6 / 8	Meionite	Mizzonite	
Zeolite	9 / 18	Clinoptilolite	Heulandite	
		Natrolite	Stilbite	

## ANNEXE 7

### Tableaux de résultats

Dans cette annexe, nous présentons, sous forme de tableaux, tous les résultats complets et exhaustifs de notre étude, concernant respectivement les images en niveaux de gris, les signaux de base et les signatures, pour  $n = \{2, 3, 4\}$ . Aucune analyse n'est réalisée dans cette annexe. Pour ce faire, il est conseillé de se reporter au chapitre 5. Nous débutons par les résultats pour les images en niveaux de gris.

## 1 Images en niveaux de gris

Tableau XVI

Moyennes et erreurs standard pour  $n = 2$  images

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2	100	$96.39 \pm 0.13$	$79.55 \pm 0.01$	$96.39 \pm 0.13$	$95.51 \pm 0.03$	<b><math>99.98 \pm 0.02</math></b>
				$96.39 \pm 0.13$	$95.98 \pm 0.04$	<b><math>99.98 \pm 0.02</math></b>
1-3	100	99.61	84.06	99.61	96.57	100
				99.61	95.32	100
1-4	100	$95.25 \pm 0.07$	$84.79 \pm 0.02$	$95.25 \pm 0.07$	$96.04 \pm 0.02$	100
				$95.25 \pm 0.07$	$96.06 \pm 0.02$	100
1-5	100	$88.03 \pm 0.01$	$85.82 \pm 0.06$	$88.03 \pm 0.01$	$93.32 \pm 0.01$	100
				$88.03 \pm 0.01$	93.61	100
1-6	100	$98.79 \pm 0.11$	$85.58 \pm 0.02$	$98.79 \pm 0.11$	$97.28 \pm 0.03$	<b><math>99.98 \pm 0.02</math></b>
				$98.79 \pm 0.11$	$96.83 \pm 0.03$	<b><math>99.98 \pm 0.02</math></b>
1-7	100	$98.68 \pm 0.16$	$96.23 \pm 0.07$	$98.68 \pm 0.16$	$95.80 \pm 0.04$	<b><math>99.96 \pm 0.03</math></b>
				$98.68 \pm 0.16$	$99.40 \pm 0.05$	<b><math>99.96 \pm 0.03</math></b>
2-3	100	96.39	83.82	96.39	96.34	100
				96.39	94.80	100
2-4	100	$24.54 \pm 0.10$	$71.58 \pm 0.02$	$24.54 \pm 0.10$	$73.79 \pm 0.03$	100
				$24.54 \pm 0.10$	$76.71 \pm 0.03$	100
2-5	100	81.75	79.39	81.75	91.56	100
				81.75	91.30	100
2-6	100	$78.83 \pm 0.08$	$80.35 \pm 0.01$	$78.83 \pm 0.08$	$91.34 \pm 0.02$	100
				$78.83 \pm 0.08$	$91.34 \pm 0.02$	100
2-7	100	$98.57 \pm 0.12$	$96.50 \pm 0.05$	$98.58 \pm 0.12$	$96.11 \pm 0.04$	100
				$98.58 \pm 0.12$	$99.38 \pm 0.04$	100
3-4	100	$90.87 \pm 0.24$	$80.85 \pm 0.07$	$90.87 \pm 0.24$	$93.49 \pm 0.07$	100
				$90.87 \pm 0.24$	$94.58 \pm 0.07$	100
3-5	100	$46.25 \pm 0.05$	76.64	$46.25 \pm 0.05$	$81.68 \pm 0.01$	100

Tableau XVI (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				46.25 ± 0.05	81.68 ± 0.02	100
3-6	100	94.19	83.56	94.19	93.99	100
				94.19	95.14	100
3-7	100	98.93 ± 0.13	97.08 ± 0.06	98.93 ± 0.13	94.56 ± 0.04	99.96 ± 0.03
				98.93 ± 0.13	99.56 ± 0.04	99.96 ± 0.03
4-5	100	48.89 ± 0.01	76.14	48.89 ± 0.01	80.82	100
				48.89 ± 0.01	81.95	100
4-6	07.92 ± 0.47	97.09 ± 0.15	85.55 ± 0.05	97.09 ± 0.15	88.70 ± 0.15	100
				97.09 ± 0.15	87.96 ± 0.16	100
4-7	84.28 ± 0.52	96.90	96.04	96.90	95.31	100
				96.90	98.89	100
5-6	100	95.61 ± 0.13	86.21 ± 0.03	95.61 ± 0.14	95.08 ± 0.04	100
				95.61 ± 0.14	94.68 ± 0.04	100
5-7	65.15 ± 0.69	38.05 ± 0.12	69.59 ± 0.06	38.05 ± 0.12	78.56 ± 0.02	100
				38.05 ± 0.12	80.16 ± 0.02	100
6-7	100	99.44	92.77	99.44	95.80	100
				99.44	99.62	100

Tableau XVII

Moyennes et erreurs standard pour  $n = 3$  images

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-3	100	97.85 ± 0.01	86.64	98.73 ± 0.01	96.76	100
				96.93 ± 0.01	96.54	100
				97.87	95.77	100
1-2-4	100	71.29	78.23	96.01	96.27	100
				60.22 ± 0.01	77.06	100
				57.65 ± 0.01	77.83	100
1-2-5	100	83.10 ± 0.02	80.48	89.25 ± 0.04	92.14 ± 0.02	100
				82.14 ± 0.02	88.45 ± 0.01	100
				77.90 ± 0.01	86.86	100
1-2-6	100	91.20	81.81	97.83 ± 0.01	97.10	100
				86.99 ± 0.01	91.67	100
				88.79	93.26	100
1-2-7	100	98.69 ± 0.02	88.43	98.35 ± 0.02	96.14 ± 0.01	100
				98.22 ± 0.02	96.51 ± 0.01	100

Tableau XVII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				$99.50 \pm 0.02$	$99.59 \pm 0.01$	100
1-3-4	100	$92.54 \pm 0.11$	$83.92 \pm 0.105$	$96.25 \pm 0.06$ $91.76 \pm 0.13$ $89.60 \pm 0.16$	$96.21 \pm 0.203$ $93.00 \pm 0.07$ $93.27 \pm 0.08$	100 <b>99.98</b> $\pm 0.02$ <b>99.98</b> $\pm 0.02$
1-3-5	100	77.34	80.39	93.63 71.45 66.96	94.55 83.39 82.33	100 100 100
1-3-6	100	97.68	86.76	99.51 96.91 96.61	98.05 95.11 96.56	100 100 100
1-3-7	100	$99.69 \pm 0.02$	$90.50 \pm 0.01$	$99.66 \pm 0.02$ $99.98 \pm 0.01$ $99.74 \pm 0.03$	$96.93 \pm 0.01$ 95.87 $99.77 \pm 0.02$	100 100 100
1-4-5	100	$77.16 \pm 0.01$	78.28	89.85 $71.30 \pm 0.02$ $70.34 \pm 0.01$	92.79 $82.52 \pm 0.01$ 83.26	100 100 100
<b>1-4-6</b>	<b><math>37.66 \pm 0.67</math></b>	$94.50 \pm 0.04$	$88.57 \pm 0.02$	$88.73 \pm 0.09$ $83.99 \pm 0.13$ $94.20 \pm 0.08$	$95.97 \pm 0.01$ $94.77 \pm 0.02$ $97.07 \pm 0.02$	<b>99.96</b> $\pm 0.04$ 100 <b>99.96</b> $\pm 0.04$
<b>1-4-7</b>	<b><math>17.37 \pm 0.59</math></b>	$97.17 \pm 0.02$	$89.97 \pm 0.01$	$98.26 \pm 0.08$ $92.04 \pm 0.06$ $92.69 \pm 0.08$	97.37 95.70 97.94	100 100 100
1-5-6	100	$93.37 \pm 0.03$	$86.91 \pm 0.01$	$93.12 \pm 0.02$ $90.40 \pm 0.04$ $96.59 \pm 0.04$	95.05 $92.75 \pm 0.02$ $95.78 \pm 0.02$	100 100 100
1-5-7	100	$91.97 \pm 0.06$	$90.00 \pm 0.05$	$93.59 \pm 0.02$ $88.40 \pm 0.08$ $93.92 \pm 0.08$	$93.89 \pm 0.01$ $91.61 \pm 0.04$ $96.68 \pm 0.04$	100 100 100
1-6-7	100	$99.49 \pm 0.02$	91.58	$99.47 \pm 0.03$ 99.44 $99.55 \pm 0.03$	$97.67 \pm 0.01$ 97.24 $99.60 \pm 0.01$	<b>99.98</b> $\pm 0.02$ 100 <b>99.98</b> $\pm 0.02$
2-3-4	100	$36.52 \pm 0.09$	$63.94 \pm 0.04$	$47.23 \pm 0.04$ $51.80 \pm 0.12$ <b>-1.45</b> $\pm 0.21$	$70.45 \pm 0.02$ $73.57 \pm 0.06$ $47.03 \pm 0.11$	100 100 100
2-3-5	100	$77.69 \pm 0.05$	$79.81 \pm 0.02$	$94.65 \pm 0.09$ $69.17 \pm 0.02$ $69.23 \pm 0.07$	$95.22 \pm 0.04$ 83.11 $82.03 \pm 0.04$	100 100 100
2-3-6	100	$90.10 \pm 0.01$	84.53	$88.17 \pm 0.02$	92.63	<b>99.98</b> $\pm 0.02$



Tableau XVII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				95.30 ± 0.01 86.82	94.62 91.57	99.98 ± 0.02 100
2-3-7	100	98.48	90.97	97.84 98.03 99.48 ± 0.01	96.63 95.39 99.66	100 100 100
2-4-5	100	48.78 ± 0.11	67.61 ± 0.03	51.47 ± 0.07 25.46 ± 0.18 68.10 ± 0.09	71.64 ± 0.38 61.00 ± 0.09 81.42 ± 0.05	100 100 100
2-4-6	100	55.08 ± 0.19	70.90 ± 0.07	49.48 ± 0.06 45.87 ± 0.19 69.89 ± 0.32	72.23 ± 0.03 71.85 ± 0.09 83.46 ± 0.16	100 100 100
2-4-7	100	74.42 ± 0.01	84.15	61.74 ± 0.02 61.69 ± 0.02 99.83 ± 0.02	76.80 79.37 99.81	100 100 100
2-5-6	100	83.75 ± 0.02	78.06 ± 0.03	77.98 ± 0.02 87.63 ± 0.02 85.65 ± 0.02	86.87 ± 0.01 90.01 ± 0.01 90.70 ± 0.01	100 100 100
2-5-7	100	93.32	90.00	90.80 90.08 99.09	92.59 92.00 99.42	100 100 100
2-6-7	100	92.47	88.35	89.19 ± 0.01 89.13 ± 0.01 99.09	92.31 92.31 99.36	100 100 100
3-4-5	100	41.46 ± 0.07	68.89 ± 0.02	39.30 ± 0.14 50.96 ± 0.11 34.13 ± 0.23	68.37 ± 0.09 73.05 ± 0.06 64.89 ± 0.12	100 100 100
3-4-6	96.20 ± 0.29	88.29 ± 0.13	82.72 ± 0.03	93.05 ± 0.07 85.16 ± 0.19 86.65 ± 0.21	93.83 ± 0.04 91.04 ± 0.09 91.51 ± 0.11	100 100 100
3-4-7	92.12 ± 0.36	94.35 ± 0.08	88.45 ± 0.05	94.71 ± 0.09 91.68 ± 0.12 96.67 ± 0.10	93.95 ± 0.04 93.39 ± 0.06 98.17 ± 0.05	100 100 100
3-5-6	100	77.33	77.47	69.26 ± 0.01 69.60 ± 0.01 93.13 ± 0.01	82.62 82.56 94.05	100 100 100
3-5-7	100	81.45 ± 0.01	86.27	72.53 ± 0.01 72.49 99.32 ± 0.01	83.73 ± 0.01 83.63 99.60	100 100 100

Tableau XVII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
3-6-7	100	97.81	90.28	96.98 $\pm$ 0.01	94.71	100
				96.85	95.63	100
				99.61 $\pm$ 0.01	99.71	100
4-5-6	100	77.74 $\pm$ 0.01	79.61	67.50 $\pm$ 0.02	81.15	100
				71.62	83.22	100
				94.10 $\pm$ 0.02	94.53	100
4-5-7	100	77.10	83.28	72.35	82.18	100
				66.75	80.33	100
				92.19 $\pm$ 0.01	95.83	100
4-6-7	04.98 $\pm$ 0.53	58.88 $\pm$ 1.43	73.89 $\pm$ 0.60	42.23 $\pm$ 0.53	63.04 $\pm$ 0.42	100
				56.51 $\pm$ 0.32	75.71 $\pm$ 0.44	100
				73.00 $\pm$ 0.46	86.76 $\pm$ 0.36	100
5-6-7	100	77.09 $\pm$ 0.03	79.36 $\pm$ 0.02	66.75 $\pm$ 0.05	80.13 $\pm$ 0.02	100
				97.80 $\pm$ 0.02	95.36 $\pm$ 0.01	100
				66.72 $\pm$ 0.05	81.42 $\pm$ 0.02	100

Tableau XVIII

Moyennes et erreurs standard pour  $n = 4$  images

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-3-4	100	66.68 $\pm$ 0.04	72.69 $\pm$ 0.03	97.70	97.15	100
				64.88 $\pm$ 0.02	74.72 $\pm$ 0.01	100
				67.14 $\pm$ 0.07	76.08 $\pm$ 0.05	100
				32.31 $\pm$ 0.10	53.80 $\pm$ 0.08	100
1-2-3-5	100	84.57 $\pm$ 0.06	81.11 $\pm$ 0.02	94.21 $\pm$ 0.02	94.34 $\pm$ 0.01	100
				92.36 $\pm$ 0.14	93.62 $\pm$ 0.09	100
				78.50 $\pm$ 0.02	84.46 $\pm$ 0.01	100
				73.22 $\pm$ 0.10	81.01 $\pm$ 0.07	100
1-2-3-6	100	94.21	87.92	98.65	97.76	100
				90.61	92.82	100
				96.76	95.54	100
				90.83	93.11	100
1-2-3-7	100	98.73	89.74	99.10 $\pm$ 0.01	97.09	99.98 $\pm$ 0.02
				97.75	96.80	99.98 $\pm$ 0.02
				98.49	96.20	100
				99.61	99.68	100

Tableau XVIII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-4-5	100	71.16 ± 0.05	76.28 ± 0.01	95.03	95.60	100
				64.82 ± 0.05	73.92 ± 0.03	100
				47.47 ± 0.10	64.25 ± 0.06	100
				76.46 ± 0.05	82.49 ± 0.03	100
1-2-4-6	100	73.81 ± 0.14	78.91 ± 0.07	96.52 ± 0.01	96.93	100
				65.19 ± 0.04	75.42 ± 0.03	100
				58.62 ± 0.18	71.78 ± 0.12	100
				74.91 ± 0.33	82.80 ± 0.22	100
1-2-4-7	100	85.51	85.51	97.24	96.62	100
				73.40	79.58	100
				71.64	80.23	100
				99.78	99.76	100
1-2-5-6	100	85.88	81.49	92.46 ± 0.02	93.91 ± 0.01	100
				77.87 ± 0.01	83.53	100
				83.36	87.06	100
				89.82	92.33	100
1-2-5-7	100	91.35	88.37	92.86	93.02	100
				88.08	89.74	100
				84.92	88.09	100
				99.56	99.61	100
1-2-6-7	100	95.22	88.43	98.44	97.34	100
				91.00	92.39	100
				92.18	93.78	100
				99.26	99.36	100
1-3-4-5	100	69.41 ± 0.08	76.83 ± 0.04	93.84 ± 0.04	94.64 ± 0.03	100
				52.54 ± 0.24	66.80 ± 0.17	<b>99.80 ± 0.07</b>
				69.38 ± 0.13	78.94 ± 0.08	100
				61.82 ± 0.15	73.87 ± 0.11	<b>99.80 ± 0.07</b>
1-3-4-6	100	85.59 ± 0.07	84.24 ± 0.04	94.21 ± 0.03	95.34 ± 0.17	100
				90.99 ± 0.10	91.60 ± 0.07	100
				74.68 ± 0.13	82.18 ± 0.09	100
				82.46 ± 0.20	87.56 ± 0.13	100
1-3-4-7	100	94.11 ± 0.05	88.78 ± 0.21	96.16 ± 0.03	95.69 ± 0.24	100
				92.03 ± 0.08	92.16 ± 0.05	100
				89.12 ± 0.08	91.38 ± 0.05	100
				99.12 ± 0.01	99.37	100
1-3-5-6	100	85.70	81.89	95.15	95.78	100
				78.63	84.24	100

Tableau XVIII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				74.74	82.29	100
				94.29 ± 0.01	94.79	100
1-3-5-7	100	88.45	86.41	95.71	95.13	100
				80.72 ± 0.01	85.10	100
				77.77	84.16	100
				99.60	99.67	100
1-3-6-7	100	98.69	90.91	99.60	98.22	100
				87.88	95.64	100
				97.59	96.85	100
				99.71	99.73	100
1-4-5-6	100	83.82 ± 0.06	83.37 ± 0.02	94.28 ± 0.03	95.41 ± 0.02	100
				72.08 ± 0.12	80.04 ± 0.07	100
				75.83 ± 0.07	81.90 ± 0.06	100
				93.11 ± 0.07	94.14 ± 0.05	100
1-4-5-7	100	85.49	83.63	92.92	93.39	100
				79.69	83.68	100
				75.18	81.73	100
				94.19	95.99	100
<b>1-4-6-7</b>	<b>40.12 ± 0.75</b>	95.50 ± 0.11	91.32 ± 0.06	96.10	96.31	100
				91.78 ± 0.20	93.37 ± 0.14	100
				95.62 ± 0.17	95.80 ± 0.11	100
				98.50	98.87 ± 0.05	100
1-5-6-7	100	94.99 ± 0.01	90.20 ± 0.02	95.24	95.49	100
				90.70 ± 0.02	91.63 ± 0.02	100
				97.34 ± 0.02	95.97 ± 0.01	100
				96.69 ± 0.05	97.59 ± 0.04	100
<b>2-3-4-5</b>	100	50.29 ± 0.04	65.79 ± 0.02	61.71 ± 0.02	72.03 ± 0.02	100
				47.57 ± 0.11	64.70 ± 0.07	<b>99.98 ± 0.02</b>
				<b>-0.42 ± 0.10</b>	32.40 ± 0.06	<b>99.98 ± 0.02</b>
				64.57 ± 0.07	75.24 ± 0.05	100
2-3-4-6	100	64.25 ± 0.04	70.79 ± 0.04	59.61 ± 0.03	71.90 ± 0.02	100
				71.88 ± 0.19	79.44 ± 0.13	100
				33.97 ± 0.17	54.54 ± 0.12	100
				85.90 ± 0.19	89.66 ± 0.13	100
2-3-4-7	100	67.96 ± 0.04	77.72 ± 0.02	64.69 ± 0.02	73.67 ± 0.01	100
				67.47 ± 0.07	76.26 ± 0.05	100
				31.68 ± 0.13	52.48 ± 0.09	100
				99.77	99.82	100

Tableau XVIII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
2-3-5-6	100	$83.40 \pm 0.03$	$79.96 \pm 0.01$	$89.16 \pm 0.07$	$91.71 \pm 0.05$	100
				$76.25 \pm 0.02$	$83.28 \pm 0.01$	100
				$77.67 \pm 0.08$	$83.24 \pm 0.05$	100
				$90.50 \pm 0.02$	$92.31 \pm 0.01$	100
2-3-5-7	100	88.77	86.64	$96.90 \pm 0.02$	$96.05 \pm 0.021$	100
				79.09	84.78	100
				$79.71 \pm 0.02$	$84.20 \pm 0.01$	100
				99.37	99.53	100
2-3-6-7	100	94.67	89.72	91.79	93.24	100
				96.79	95.18	100
				90.87	92.29	100
				99.24	99.41	100
2-4-5-6	100	$64.19 \pm 0.10$	$69.65 \pm 0.04$	$59.80 \pm 0.04$	$71.08 \pm 0.03$	100
				$41.69 \pm 0.12$	$59.98 \pm 0.08$	100
				$76.35 \pm 0.04$	$81.98 \pm 0.03$	100
				$78.56 \pm 0.26$	$84.56 \pm 0.18$	100
2-4-5-7	100	$72.72 \pm 0.05$	$78.29 \pm 0.02$	$66.98 \pm 0.05$	$74.34 \pm 0.03$	100
				$48.50 \pm 0.12$	$64.13 \pm 0.08$	100
				$76.33 \pm 0.05$	$81.88 \pm 0.03$	100
				$98.09 \pm 0.01$	98.67	100
2-4-6-7	100	$75.65 \pm 0.13$	$79.63 \pm 0.08$	$65.52 \pm 0.06$	$74.79 \pm 0.04$	100
				$61.75 \pm 0.17$	$73.64 \pm 0.11$	100
				$75.99 \pm 0.29$	$82.61 \pm 0.20$	100
				$99.32 \pm 0.01$	99.44	100
2-5-6-7	100	91.21	86.49	$85.12 \pm 0.01$	$88.21 \pm 0.01$	100
				91.02	91.53	100
				$90.18 \pm 0.01$	91.57	100
				$98.53 \pm 0.02$	$98.87 \pm 0.01$	100
3-4-5-6	100	$68.21 \pm 0.04$	$74.93 \pm 0.02$	$57.93 \pm 0.09$	$71.07 \pm 0.06$	100
				$65.27 \pm 0.07$	$75.69 \pm 0.05$	100
				$56.01 \pm 0.15$	$69.59 \pm 0.10$	100
				$93.63 \pm 0.10$	$94.28 \pm 0.07$	100
3-4-5-7	100	$69.96 \pm 0.03$	$78.55 \pm 0.01$	$58.86 \pm 0.09$	$71.37 \pm 0.07$	100
				$66.48 \pm 0.07$	$75.55 \pm 0.06$	100
				$56.18 \pm 0.15$	$68.90 \pm 0.10$	100
				98.34	98.85	100
<b>3-4-6-7</b>	<b><math>60.40 \pm 0.70</math></b>	$92.45 \pm 0.08$	$88.49 \pm 0.03$	$94.22 \pm 0.05$	$93.75 \pm 0.04$	100
				$87.03 \pm 0.16$	$90.10 \pm 0.10$	100

Tableau XVIII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				$89.81 \pm 0.19$	$91.48 \pm 0.13$	100
				$98.73 \pm 0.02$	$99.09 \pm 0.01$	100
3-5-6-7	100	88.33	84.38	$79.26 \pm 0.01$	84.50	100
				$79.42 \pm 0.01$	84.29	100
				95.26	94.61	100
				$99.37 \pm 0.02$	$99.52 \pm 0.01$	100
4-5-6-7	100	85.71	83.62	$76.86 \pm 0.02$	$82.27 \pm 0.01$	100
				75.78	81.53	100
				$95.73 \pm 0.02$	$94.92 \pm 0.01$	100
				94.46	96.10	100

## 2 Signaux de base

### 2.1 $N = 5$

Tableau XIX

Moyennes et erreurs standard pour  $n = 2$  avec  $N = 5$  signaux

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2	100	$99.82 \pm 0.06$	$92.82 \pm 0.02$	$99.82 \pm 0.06$	$98.02 \pm 0.02$	100
				$99.82 \pm 0.06$	$98.04 \pm 0.03$	100
1-3	100	$91.16 \pm 0.08$	$89.35 \pm 0.03$	$91.16 \pm 0.08$	$95.93 \pm 0.03$	100
				$91.16 \pm 0.08$	$94.43 \pm 0.03$	100
1-4	100 si $p \leq 0.97$ , 0 sinon	87.48	86.85	87.48	$93.57 \pm 0.12$	100
				87.48	$79.39 \pm 0.22$	100
1-5	100	99.22	92.68	99.22	97.80	100
				99.22	99.54	100
2-3	100	$99.50 \pm 0.10$	$89.16 \pm 0.04$	$99.50 \pm 0.10$	$97.87 \pm 0.04$	<b><math>99.88 \pm 0.05</math></b>
				$99.50 \pm 0.10$	$96.00 \pm 0.04$	<b><math>99.88 \pm 0.05</math></b>
2-4	100 si $p \leq 0.83$ , 0 sinon	$83.09 \pm 0.01$	84.93	$83.09 \pm 0.01$	$90.74 \pm 0.02$	100
				$83.09 \pm 0.01$	$84.76 \pm 0.03$	100
2-5	100	99.97	87.53	99.97	98.13	100
				99.97	99.83	100
3-4	100 si $p \leq 0.93$ , 0 sinon	$72.70 \pm 0.01$	74.89	$72.70 \pm 0.01$	$84.85 \pm 0.03$	100
				$72.70 \pm 0.01$	$77.17 \pm 0.01$	100

Tableau XIX (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
3-5	100	$97.57 \pm 0.13$	$96.04 \pm 0.07$	99.22	97.80	100
				99.22	99.54	100
4-5	0 pour $p \leq 0.02$ , 100 sinon	98.98	92.92	98.98	$82.53 \pm 0.02$	100
				98.98	$99.48 \pm 0.01$	100

Tableau XX

Moyennes et erreurs standard pour  $n = 3$  avec  $N = 5$  signaux

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-3	100	97.17	89.22	95.75	96.71	100
				<b>100</b>	98.34	100
				95.75	95.42	100
1-2-4	100	89.20	89.25	92.25	94.85	100
				90.83	92.75	100
				84.53	84.47	100
1-2-5	100	99.72	97.21	99.61	98.15	100
				99.97	98.42	100
				99.59	99.62	100
1-3-4	100	83.60	84.05	89.20	91.87	100
				$81.67 \pm 0.01$	85.54	100
				79.92	78.10	100
1-3-5	100	$96.41 \pm 0.01$	94.44	95.33	96.45	100
				$95.04 \pm 0.02$	$95.02 \pm 0.01$	100
				$98.85 \pm 0.02$	$99.13 \pm 0.01$	100
1-4-5	100	95.23	92.20	93.39	94.45	100
				93.19	82.58	100
				99.10	99.30	100
2-3-4	100	81.91	82.87	88.68	92.06	100
				81.28	86.60	100
				75.77	82.26	100
2-3-5	100	$99.49 \pm 0.02$	$93.73 \pm 0.01$	$99.97 \pm 0.01$	98.34	100
				$99.25 \pm 0.02$	$96.31 \pm 0.02$	100
				$99.24 \pm 0.02$	$99.37 \pm 0.02$	100
2-4-5	100	93.99	91.95	91.49	92.77	100
				91.06	87.36	100

Tableau XX (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				99.43	99.60	100
3-4-5	100	$89.32 \pm 0.02$	$84.94 \pm 0.01$	$84.75 \pm 0.03$	$86.51 \pm 0.02$	100
				85.53	80.69	100
				$97.68 \pm 0.03$	$98.29 \pm 0.02$	100

Tableau XXI

Moyennes et erreurs standard pour  $n = 4$  avec  $N = 5$  signaux

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-3-4	100	87.09	87.98	92.10	93.22	100
				92.21	93.30	100
				84.96	87.08	100
				79.09	81.97	100
1-2-3-5	100	98.19	93.72	96.88	96.95	100
				99.98	98.56	100
				$96.69 \pm 0.02$	$95.69 \pm 0.01$	100
				$99.22 \pm 0.02$	$99.26 \pm 0.02$	100
1-2-4-5	100	94.25	93.22	94.53	95.33	100
				93.84	93.77	100
				89.33	86.49	100
				99.29	99.37	100
1-3-4-5	100	91.00	88.90	92.61	92.88	100
				86.88	86.86	100
				86.10	80.88	100
				98.42	98.53	100
2-3-4-5	100	90.26	88.91	92.41	93.18	100
				86.64	87.78	100
				83.35	84.47	100
				98.65	98.82	100



2.2  $N = 4$ 

Tableau XXII

Moyennes et erreurs standard pour  $n = 2$  avec  $N = 4$  signaux

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2	100	$99.70 \pm 0.08$	$94.87 \pm 0.04$	$99.70 \pm 0.08$	$97.47 \pm 0.04$	100
				$99.70 \pm 0.08$	$97.49 \pm 0.04$	100
1-3	100	$91.10 \pm 0.09$	$90.27 \pm 0.05$	$91.10 \pm 0.09$	$94.89 \pm 0.05$	100
				$91.10 \pm 0.09$	$93.04 \pm 0.05$	100
1-4	100 si $p \leq 0.97$ , 0 sinon	87.48	83.84	87.48	$91.43 \pm 0.16$	100
				87.48	$74.33 \pm 0.22$	100
2-3	100	$99.60 \pm 0.09$	$87.61 \pm 0.04$	$99.60 \pm 0.09$	$97.34 \pm 0.05$	$99.98 \pm 0.02$
				$99.60 \pm 0.09$	$95.02 \pm 0.05$	$99.98 \pm 0.02$
2-4	100 si $p \leq 0.83$ , 0 sinon	$83.10 \pm 0.01$	82.86	$83.10 \pm 0.01$	$88.80 \pm 0.12$	100
				$83.10 \pm 0.01$	$81.09 \pm 0.15$	100
3-4	100 si $p \leq 0.93$ , 0 sinon	$72.70 \pm 0.01$	69.35	$72.70 \pm 0.01$	$81.53 \pm 1.12$	100
				$72.70 \pm 0.01$	$72.42 \pm 0.52$	100

Tableau XXIII

Moyennes et erreurs standard pour  $n = 3$  avec  $N = 4$  signaux

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-3	100	$97.15 \pm 0.01$	$89.48 \pm 0.01$	$95.74 \pm 0.02$	$96.05 \pm 0.01$	100
				$99.97 \pm 0.03$	$97.98 \pm 0.03$	100
				$95.74 \pm 0.01$	$94.52 \pm 0.01$	100
1-2-4	100	89.20	88.18	92.25	93.83	100
				90.82	91.33	100
				84.53	81.51	100
1-3-4	100	83.62	81.10	89.21	90.26	100
				81.70	82.70	100
				79.94	73.87	100
2-3-4	100	81.93	80.85	88.69	90.51	100
				81.29	83.96	100
				75.79	78.87	100

Tableau XXIV

Moyennes et erreurs standard pour  $n = 4$  avec  $N = 4$  signaux

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2-3-4	100	87.09	87.09	92.10	92.10	100
				92.22	92.22	100
				84.95	84.95	100
				79.09	79.09	100

### 3 Signatures spectrales

Tableau XXV

Moyennes et erreurs standard pour  $n = 2$  signatures

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
1-2	100	65.16 ± 0.59	62.29 ± 0.22	65.16 ± 0.59	75.48	<b>35.92 ± 1.13</b>
				65.16 ± 0.59	64.48 ± 0.15	100
1-3	100	61.90 ± 0.35	68.37 ± 0.12	61.90 ± 0.35	64.37	100
				61.90 ± 0.35	79.73 ± 0.08	100
1-4	100	42.04	63.17	42.04	53.28	100
				42.04	69.80	100
1-5	100	97.77 ± 0.01	70.22	97.77 ± 0.01	72.62	100
				97.76 ± 0.01	73.57	100
1-6	100	47.65 ± 1.00	56.73 ± 0.23	47.65 ± 1.00	61.40	<b>65.02 ± 1.42</b>
				61.40	66.31	<b>65.02 ± 1.42</b>
1-7	100	86.21 ± 0.01	65.15	86.21 ± 0.01	65.33	100
				86.21 ± 0.02	69.87	100
2-3	100	70.32 ± 0.66	67.21 ± 0.24	70.32 ± 0.66	65.72	<b>75.43 ± 0.95</b>
				70.32 ± 0.66	84.52	<b>75.43 ± 0.95</b>
2-4	100	33.24 ± 0.45	52.21 ± 0.09	33.24 ± 0.45	54.90 ± 1.63	<b>08.90 ± 0.70</b>
				33.24 ± 0.45	57.52 ± 1.72	100
2-5	100	93.39	73.64	93.39	72.91	100
				93.39	77.87	100
2-6	<b>99.80 ± 0.09</b>	57.25 ± 0.22	52.35 ± 1.01	52.35 ± 1.01	47.72 ± 0.39	100
				52.35 ± 1.01	65.19 ± 0.32	100
2-7	100	83.72	69.59	83.72	69.45	100

Tableau XXV (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				83.72	72.14	100
3-4	100	42.09 ± 0.04	71.16	42.09 ± 0.04	62.98 ± 0.02	100
				42.09 ± 0.04	74.05	100
3-5	100	98.61 ± 0.02	77.61	98.61 ± 0.02	88.67	100
				98.61 ± 0.02	70.85	100
3-6	100	76.11 ± 0.07	71.97 ± 0.03	76.11 ± 0.07	81.36 ± 0.02	100
				76.11 ± 0.07	72.96 ± 0.03	100
3-7	100	71.45 ± 0.04	74.04 ± 0.02	71.45 ± 0.04	84.51 ± 0.01	100
				71.45 ± 0.04	64.58 ± 0.02	100
4-5	100	74.90	68.38	74.90	81.83	100
				74.90	62.09	100
<b>4-6</b>	100	13.99	57.75	13.99	71.14	100
				13.99	<b>0</b>	<b>0</b>
4-7	100	53.97	64.30	53.97	73.74	100
				53.97	58.29	100
5-6	100	56.84	62.68	56.84	60.28	100
				56.84	69.59	100
5-7	100	50.98 ± 0.05	66.49	50.98 ± 0.05	61.50 ± 0.04	100
				50.98 ± 0.05	72.13	100
6-7	100	76.91 ± 0.28	66.59 ± 0.05	76.91 ± 0.28	78.33 ± 0.04	100
				76.91 ± 0.28	67.70 ± 0.13	100

Tableau XXVI

Moyennes et erreurs standard pour  $n = 3$  signatures

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
<b>1-2-3</b>	<b>62.51 ± 0.98</b>	72.47 ± 0.27	67.66 ± 0.12	70.58 ± 0.35	73.94 ± 0.16	100
				80.06 ± 0.33	68.09 ± 0.18	100
				66.78 ± 0.40	72.58 ± 0.26	100
<b>1-2-4</b>	100	46.77 ± 0.03	57.82 ± 0.06	66.92 ± 0.19	67.96 ± 0.12	<b>99.80 ± 0.10</b>
				49.49 ± 0.25	55.15 ± 0.06	<b>98.46 ± 0.28</b>
				20.68 ± 0.06	47.90	<b>18.36 ± 0.83</b>
<b>1-2-5</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>1-2-6</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>1-2-7</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
1-3-4	100	41.95 ± 0.02	67.92	29.05 ± 0.04	50.17 ± 0.02	100

Tableau XXVI (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				40.16 ± 0.02	69.57	100
				56.64 ± 0.04	69.76 ± 0.02	100
1-3-5	100	84.27 ± 0.20	72.78 ± 0.12	78.71 ± 0.33	75.85 ± 0.11	100
				80.91 ± 0.27	82.66 ± 0.23	100
				93.19 ± 0.02	75.30 ± 0.03	100
<b>1-3-6</b>	<b>55.08 ± 0.92</b>	76.60 ± 0.04	72.36 ± 0.01	73.61 ± 0.05	67.97 ± 0.02	100
				83.43 ± 0.05	86.26 ± 0.02	100
				72.77 ± 0.01	66.49	100
1-3-7	100	80.32	72.49	89.91 ± 0.02	75.01	100
				74.73 ± 0.01	84.34 ± 0.01	100
				76.33	66.71	100
<b>1-4-5</b>	100	55.03 ± 0.09	57.15 ± 0.05	45.91 ± 0.14	53.41 ± 0.07	<b>99.28 ± 0.17</b>
				36.79 ± 0.14	48.01 ± <b>5.38</b>	<b>02.62 ± 0.38</b>
				82.20	51.48 ± <b>5.57</b>	100
<b>1-4-6</b>	100	38.62 ± 0.14	50.60 ± 0.03	49.39 ± 0.19	59.78 ± 0.04	<b>27.08 ± 1.21</b>
				07.86 ± 0.79	57.53 ± 0.03	<b>27.08 ± 1.21</b>
				40.58 ± 0.26	42.15 ± 0.01	<b>26.97 ± 1.20</b>
<b>1-4-7</b>	100	46.29 ± 0.06	53.49 ± 0.01	54.91 ± 0.19	60.21 ± 0.02	<b>37.80 ± 1.15</b>
				20.95 ± 0.12	41.97 ± 0.03	<b>62.26 ± 1.15</b>
				60.51 ± 0.34	68.28	<b>62.26 ± 1.15</b>
<b>1-5-6</b>	100	72.67 ± 0.08	61.98 ± 0.04	79.77 ± 0.21	68.52	<b>93.44 ± 0.63</b>
				77.32 ± 0.09	75.90 ± 0.04	100
				60.91 ± 0.13	57.57 ± 0.03	100
1-5-7	100	84.60 ± 0.01	72.23	97.67 ± 0.02	75.49 ± 0.01	100
				77.49 ± 0.01	77.40	100
				78.63 ± 0.01	77.12 ± 0.01	100
<b>1-6-7</b>	100	66.68 ± 0.10	56.07 ± 0.09	59.79 ± 0.22	51.85 ± 0.16	<b>99.64 ± 0.15</b>
				56.90 ± 0.19	69.17 ± 0.10	<b>23.39 ± 1.03</b>
				83.36 ± 0.10	74.84 ± 0.06	100
2-3-4	100	53.38 ± 0.28	62.58 ± 0.19	66.52 ± 0.06	60.03 ± 0.05	100
				57.11 ± 0.41	67.33 ± 0.27	100
				28.58 ± 0.48	54.89 ± 0.25	100
2-3-5	100	91.34 ± 0.02	78.85	91.20 ± 0.04	82.98 ± 0.01	100
				89.47	88.92	100
				93.36 ± 0.04	79.46 ± 0.04	100
<b>2-3-6</b>	100	38.73 ± 0.14	60.19 ± 0.02	16.54 ± 0.30	39.62 ± 0.14	<b>98.92 ± 0.25</b>
				51.43 ± 0.07	61.30 ± 0.05	<b>92.92 ± 0.25</b>
				46.32 ± 0.09	54.03 ± 0.12	<b>46.72 ± 1.13</b>

Tableau XXVI (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
2-3-7	100	$74.11 \pm 0.38$	$72.01 \pm 0.16$	$80.17 \pm 0.43$	$77.99 \pm 0.26$	<b>99.80</b> $\pm 0.10$
				$70.47 \pm 0.32$	$77.54 \pm 0.20$	100
				$69.66 \pm 0.63$	$65.46 \pm 0.34$	<b>99.80</b> $\pm 0.10$
2-4-5	100	$69.60 \pm 0.19$	$65.90 \pm 0.04$	$64.80 \pm 0.27$	$60.39 \pm 0.10$	100
				$58.68 \pm 0.35$	$65.74 \pm 0.17$	100
				$85.32 \pm 0.09$	$79.99 \pm 0.04$	100
<b>2-4-6</b>	100	$46.89 \pm 0.01$	$57.08 \pm 0.07$	$59.75 \pm 0.23$	$59.89 \pm 0.13$	100
				$36.74 \pm 0.35$	$53.93 \pm 0.11$	100
				$44.17 \pm 0.16$	$55.72 \pm 0.02$	<b>38.62</b> $\pm 1.09$
<b>2-4-7</b>	100	$53.52 \pm 0.29$	$56.48 \pm 0.10$	$64.48 \pm 0.16$	$59.77 \pm 0.10$	100
				$22.06 \pm 0.69$	54.46	<b>55.23</b> $\pm 1.07$
				$56.45 \pm 0.44$	72.42	<b>55.23</b> $\pm 1.07$
<b>2-5-6</b>	<b>47.80</b> $\pm 1.06$	$64.94 \pm 0.35$	$62.87 \pm 0.21$	$64.88 \pm 0.61$	$60.16 \pm 0.35$	100
				$79.17 \pm 0.13$	$79.44 \pm 0.06$	100
				$50.77 \pm 0.56$	$54.06 \pm 0.35$	100
2-5-7	100	81.33	76.74	93.74	81.00	100
				72.81	77.90	100
				77.44	76.62	100
<b>2-6-7</b>	100	$48.08 \pm 0.17$	$53.22 \pm 0.08$	$38.22 \pm 0.25$	$50.43 \pm 0.05$	<b>88.56</b> $\pm 0.81$
				$30.03 \pm 0.33$	$52.24 \pm 0.64$	<b>88.56</b> $\pm 0.81$
				$71.41 \pm 0.19$	$65.36 \pm 0.14$	100
3-4-5	100	$65.57 \pm 0.07$	$68.15 \pm 0.02$	$54.94 \pm 0.12$	$65.79 \pm 0.07$	100
				$52.06 \pm 0.10$	$68.20 \pm 0.04$	100
				89.70	75.04	100
<b>3-4-6</b>	100	$40.79 \pm 0.07$	$59.48 \pm 0.03$	$46.23 \pm 0.19$	$68.40 \pm 0.26$	<b>24.62</b> $\pm 0.82$
				$28.61 \pm 0.26$	$52.18 \pm 0.22$	<b>24.62</b> $\pm 0.82$
				$46.57 \pm 0.33$	$50.33 \pm 0.01$	<b>22.36</b> $\pm 0.79$
<b>3-4-7</b>	100	$49.62 \pm 0.15$	$61.83 \pm 0.28$	$49.13 \pm 0.20$	$64.15 \pm 0.14$	<b>95.59</b> $\pm 0.47$
				$33.83 \pm 0.30$	$63.02 \pm 0.04$	<b>63.03</b> $\pm 1.03$
				$64.37 \pm 0.36$	$69.18 \pm 0.03$	<b>67.44</b> $\pm 1.07$
3-5-6	100	68.66	66.41	75.64	74.61	100
				74.27	69.54	100
				56.06	62.30	100
3-5-7	100	74.62	72.80	$85.35 \pm 0.01$	84.33	100
				72.96	71.25	100
				$65.56 \pm 0.01$	69.63	100
3-6-7	100	$72.97 \pm 0.03$	$66.55 \pm 0.02$	$69.77 \pm 0.09$	$73.36 \pm 0.10$	100
				$64.62 \pm 0.06$	$67.98 \pm 0.07$	100

Tableau XXVI (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				$84.53 \pm 0.12$	$72.93 \pm 0.10$	100
<b>4-5-6</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
4-5-7	100	$70.07 \pm 0.14$	$70.03 \pm 0.06$	$75.20 \pm 0.25$	$80.68 \pm 0.07$	100
				$75.24 \pm 0.10$	$72.71 \pm 0.06$	100
				$59.76 \pm 0.27$	$67.43 \pm 0.24$	100
<b>4-6-7</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>5-6-7</b>	100	$53.03 \pm 0.15$	$65.86 \pm 0.07$	$60.82 \pm 0.05$	$70.49 \pm 0.10$	100
				$38.00 \pm 0.46$	$58.76 \pm 0.16$	<b>99.74</b> $\pm 0.11$
				$60.15 \pm 0.10$	$78.36 \pm 0.03$	<b>23.13</b> $\pm 0.88$

Tableau XXVII

Moyennes et erreurs standard pour  $n = 4$  signatures

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
<b>1-2-3-4</b>	100	$49.56 \pm 0.08$	$54.23 \pm 0.06$	$58.86 \pm 0.11$	$64.69 \pm 0.04$	100
				$50.93 \pm 0.40$	$58.84 \pm 0.12$	<b>55.16</b> $\pm 1.40$
				$51.07 \pm 0.32$	$65.99 \pm 0.23$	100
				$16.24 \pm 1.03$	$37.90 \pm 0.65$	100
<b>1-2-3-5</b>	<b>0 si <math>p \leq 0.47</math></b> 100 sinon	$87.55 \pm 0.03$	$84.43 \pm 0.04$	$84.96 \pm 0.07$	$40.66 \pm 9.09$	100
				$76.59 \pm 0.03$	$43.06 \pm 9.63$	100
				$89.72 \pm 0.08$	$46.42 \pm 10.38$	100
				91.08	$42.24 \pm 9.45$	100
<b>1-2-3-6</b>	<b>58.53</b> $\pm 1.38$	$69.89 \pm 0.09$	$70.50 \pm 0.09$	$61.67 \pm 0.40$	$68.13 \pm 0.21$	<b>99.81</b> $\pm 0.19$
				$75.61 \pm 0.65$	$78.80 \pm 0.19$	100
				$65.19 \pm 0.18$	$70.98 \pm 0.15$	100
				$77.03 \pm 0.17$	$70.67 \pm 0.16$	100
<b>1-2-3-7</b>	<b>77.90</b> $\pm 1.24$	$79.48 \pm 0.24$	$74.09 \pm 0.13$	$75.51 \pm 0.09$	$71.40 \pm 0.05$	100
				$80.95 \pm 0.36$	$81.60 \pm 0.28$	100
				$86.12 \pm 0.35$	$88.05 \pm 0.26$	100
				$73.73 \pm 0.58$	$67.70 \pm 0.39$	100
1-2-4-5	100	$68.38 \pm 0.07$	$71.47 \pm 0.28$	$65.51 \pm 0.08$	$67.46 \pm 0.19$	<b>99.58</b> $\pm 0.19$
				$68.64 \pm 0.05$	$75.21 \pm 0.15$	100
				$41.79 \pm 0.20$	$46.00 \pm 0.12$	100
				$89.41 \pm 0.09$	$81.02 \pm 0.05$	100
<b>1-2-4-6</b>	<b>0 si <math>p \leq 0.48</math></b> 100 sinon	$54.13 \pm 0.09$	$55.32 \pm 0.28$	$65.56 \pm 0.30$	$35.95 \pm 8.04$	<b>71.80</b> $\pm 2.03$
				$59.69 \pm 0.10$	$35.42 \pm 7.92$	<b>81.20</b> $\pm 2.15$

Tableau XXVII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				$32.37 \pm 0.48$ $42.54 \pm 0.79$	$27.67 \pm 6.19$ $24.15 \pm 7.53$	<b>58.60</b> $\pm 2.46$ <b>01.40</b> $\pm 0.31$
<b>1-2-4-7</b>	100	$54.58 \pm 0.02$	$59.64 \pm 0.08$	$70.14 \pm 0.46$ $53.40 \pm 0.13$ $04.81 \pm 0.49$ $56.71 \pm 0.39$	$70.50 \pm 0.04$ $56.67 \pm 0.27$ $08.11 \pm 4.41$ $09.57 \pm 5.21$	<b>79.16</b> $\pm 1.44$ <b>42.32</b> $\pm 1.66$ <b>00.32</b> $\pm 0.17$ <b>99.79</b> $\pm 0.14$
<b>1-2-5-6</b>	0	0	0	0	0	0
<b>1-2-5-7</b>	0	0	0	0	0	0
<b>1-2-6-7</b>	100	$62.75 \pm 0.07$	$61.86 \pm 0.06$	$84.07 \pm 0.48$ $33.32 \pm 0.14$ $53.88 \pm 0.04$ $72.57 \pm 0.17$	$71.77 \pm 0.39$ $60.98 \pm 0.16$ $49.97 \pm 0.25$ $66.29 \pm 0.16$	<b>99.68</b> $\pm 0.17$ <b>15.79</b> $\pm 1.23$ <b>15.79</b> $\pm 1.23$ 100
<b>1-3-4-5</b>	100	$57.13 \pm 0.08$	$60.49 \pm 0.07$	$22.66 \pm 0.44$ $57.69 \pm 0.19$ $46.03 \pm 0.15$ $83.32 \pm 0.10$	$46.68 \pm 0.15$ $71.71 \pm 0.09$ $56.01 \pm 0.15$ $74.89 \pm 0.07$	<b>67.90</b> $\pm 1.58$ 100 <b>67.90</b> $\pm 1.58$ 100
<b>1-3-4-6</b>	100	$49.49 \pm 0.19$	$56.71 \pm 0.11$	$37.51 \pm 0.49$ $56.17 \pm 0.44$ $34.12 \pm 0.120$ $53.11 \pm 0.13$	$55.90 \pm 0.34$ $72.37 \pm 0.19$ $49.95 \pm 0.57$ $53.64 \pm 0.12$	<b>54.63</b> $\pm 1.94$ <b>78.42</b> $\pm 1.52$ <b>61.37</b> $\pm 2.09$ <b>51.90</b> $\pm 2.09$
<b>1-3-4-7</b>	100	$51.56 \pm 0.05$	$57.17 \pm 0.09$	$46.77 \pm 0.59$ $61.19 \pm 0.41$ $25.86 \pm 0.95$ $57.64 \pm 0.42$	$56.76 \pm 0.47$ $74.92 \pm 0.17$ $49.03 \pm 0.60$ $66.16 \pm 0.11$	<b>80.53</b> $\pm 1.37$ <b>80.53</b> $\pm 1.37$ <b>46.42</b> $\pm 2.07$ <b>36.42</b> $\pm 1.82$
1-3-5-6	100	$75.47 \pm 0.20$	$68.30 \pm 0.15$	$71.52 \pm 0.43$ $80.21 \pm 0.29$ $79.95 \pm 0.02$ $70.21 \pm 0.07$	$70.87 \pm 0.22$ $82.96 \pm 0.26$ $74.21 \pm 0.04$ $64.81 \pm 0.06$	100 100 100 100
1-3-5-7	100	$81.62 \pm 0.08$	$73.00 \pm 0.06$	$89.90 \pm 0.26$ $78.86 \pm 0.12$ $83.41 \pm 0.04$ $74.30 \pm 0.12$	$79.43 \pm 0.11$ $80.43 \pm 0.15$ $78.51 \pm 0.04$ $70.09 \pm 0.12$	100 100 100 100
1-3-6-7	100	$75.99 \pm 0.08$	$70.66 \pm 0.02$	$76.27 \pm 0.25$ $78.79 \pm 0.18$ $68.21 \pm 0.12$ $80.68 \pm 0.13$	$72.01 \pm 0.13$ $81.83 \pm 0.11$ $71.22 \pm 0.07$ $74.08 \pm 0.05$	100 100 100 100
1-4-5-6	100	$71.69 \pm 0.17$	$65.58 \pm 0.15$	$73.82 \pm 0.32$ $66.49 \pm 0.32$	$65.09 \pm 0.24$ $68.51 \pm 0.24$	<b>99.79</b> $\pm 0.14$ <b>99.90</b> $\pm 0.11$

Tableau XXVII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
				72.81 ± 0.16	72.44 ± 0.14	100
				73.62 ± 0.10	70.94 ± 0.05	<b>99.90 ± 0.11</b>
1-4-5-7	100	64.15 ± 0.19	60.98 ± 0.14	68.64 ± 0.33	60.67 ± 0.22	100
				44.93 ± 0.51	56.41 ± 0.40	100
				77.31 ± 0.11	76.97 ± 0.07	100
				64.33 ± 0.16	57.01 ± 0.16	100
<b>1-4-6-7</b>	100	58.35 ± 0.06	61.65 ± 0.03	67.77 ± 0.27	62.87 ± 0.03	<b>52.53 ± 1.60</b>
				34.54 ± 0.37	53.64 ± 0.37	<b>48.53 ± 1.29</b>
				39.06 ± 0.62	59.15 ± 0.06	<b>60.42 ± 1.76</b>
				76.94 ± 0.26	82.38 ± 0.18	<b>99.90 ± 0.11</b>
<b>1-5-6-7</b>	100	65.96 ± 0.20	63.07 ± 0.21	82.41 ± 0.28	67.31 ± 0.24	100
				74.86 ± 0.17	77.72 ± 0.08	<b>99.26 ± 0.27</b>
				47.57 ± 0.22	57.02	<b>54.00 ± 1.49</b>
				58.71 ± 0.36	65.56 ± 0.11	<b>97.90 ± 0.39</b>
<b>2-3-4-5</b>	100	71.29 ± 0.20	70.84 ± 0.14	75.06 ± 0.21	74.36 ± 0.21	100
				67.00 ± 0.38	77.86 ± 0.03	<b>66.95 ± 1.52</b>
				44.77 ± 0.54	62.15 ± 0.02	<b>66.95 ± 1.52</b>
				90.48 ± 0.03	82.15	100
<b>2-3-4-6</b>	100	54.45 ± 0.18	61.68 ± 0.12	67.01 ± 0.34	70.61 ± 0.23	100
				52.82 ± 0.20	65.49 ± 0.09	<b>39.79 ± 1.38</b>
				<b>-0.02 ± 0.91</b>	59.08 ± 0.01	<b>39.79 ± 1.38</b>
				63.27 ± 0.08	<b>0</b>	<b>0</b>
2-3-4-7	100	55.31 ± 0.03	61.15 ± 0.05	58.89 ± 0.11	66.11 ± 0.03	100
				61.24 ± 0.47	67.59 ± 0.44	<b>97.26 ± 0.58</b>
				15.33 ± 0.53	31.20 ± 0.51	<b>97.26 ± 0.58</b>
				70.07 ± 0.36	69.22 ± 0.24	100
<b>2-3-5-6</b>	<b>58.74 ± 1.73</b>	69.73 ± 0.26	74.57 ± 0.55	80.08 ± 1.07	77.63 ± 0.63	<b>97.03 ± 0.83</b>
				72.67 ± 0.51	76.85	<b>73.60 ± 2.34</b>
				83.11 ± 0.06	79.78 ± 0.06	100
				46.56 ± 1.99	65.36	<b>73.60 ± 2.34</b>
2-3-5-7	100	82.47 ± 0.15	76.79 ± 0.11	89.28 ± 0.36	88.82	<b>93.79 ± 0.65</b>
				84.95 ± 0.16	86.05	<b>93.79 ± 0.65</b>
				82.15	80.08	100
				73.32 ± 0.12	70.39 ± 0.12	100
<b>2-3-6-7</b>	<b>99.68 ± 0.17</b>	75.20 ± 0.07	70.42 ± 0.04	70.08 ± 0.12	73.56 ± 0.06	<b>99.26 ± 0.35</b>
				76.38 ± 0.12	77.25 ± 0.08	<b>99.26 ± 0.35</b>
				72.18 ± 0.09	69.83 ± 0.05	<b>99.37 ± 0.27</b>
				82.18 ± 0.07	73.72 ± 0.04	<b>99.37 ± 0.27</b>



Tableau XXVII (suite)

S	Convergence	$C_G$	$C_G^{\text{base}}$	$C_L$	$C_L^{\text{base}}$	Associations
2-4-5-6	100	$69.88 \pm 0.03$	$68.56 \pm 0.07$	$65.87 \pm 0.14$	$63.42 \pm 0.09$	100
				$66.42 \pm 0.13$	$66.20 \pm 0.01$	100
				$75.91 \pm 0.14$	$79.90 \pm 0.11$	100
				$71.34 \pm 0.07$	$70.20 \pm 0.10$	<b>99.79</b> $\pm 0.14$
2-4-5-7	100	$67.46 \pm 0.12$	$64.37 \pm 0.06$	$68.86 \pm 0.01$	$59.68 \pm 0.02$	100
				$56.69 \pm 0.34$	$60.47 \pm 0.16$	100
				$76.16 \pm 0.03$	$79.11 \pm 0.02$	100
				$68.13 \pm 0.14$	$63.43 \pm 0.20$	100
2-4-6-7	100	$52.23 \pm 0.23$	$59.46 \pm 0.15$	$61.18 \pm 0.19$	$56.73 \pm 0.12$	100
				$49.28 \pm 0.06$	$55.96 \pm 0.06$	100
				<b>-58.84</b> $\pm 2.30$	$60.63 \pm 0.08$	<b>22.84</b> $\pm 1.59$
				$60.21 \pm 0.40$	$69.39 \pm 0.31$	100
2-5-6-7	100	$56.84 \pm 0.06$	$61.21 \pm 0.03$	$62.53 \pm 0.03$	$61.40 \pm 0.03$	100
				$76.22 \pm 0.06$	$80.37 \pm 0.03$	100
				$27.43 \pm 0.20$	$42.17 \pm 0.06$	100
				$61.13 \pm 0.09$	$57.57 \pm 0.14$	100
3-4-5-6	100	$70.97 \pm 0.21$	$68.35 \pm 0.11$	$71.63 \pm 0.39$	$68.98 \pm 0.36$	100
				$64.94 \pm 0.44$	$70.35 \pm 0.26$	100
				$76.05 \pm 0.07$	$74.19 \pm 0.08$	100
				$71.22 \pm 0.09$	$72.57 \pm 0.03$	<b>99.90</b> $\pm 0.11$
3-4-5-7	100	$66.69 \pm 0.17$	$66.35 \pm 0.11$	$59.77 \pm 0.57$	$67.34 \pm 0.34$	100
				$63.71 \pm 0.25$	$63.94 \pm 0.36$	100
				$81.74 \pm 0.02$	$78.76 \pm 0.03$	100
				$61.37 \pm 0.16$	$62.82 \pm 0.05$	<b>64.42</b> $\pm 1.44$
3-4-6-7	100	$60.04 \pm 0.16$	$66.79 \pm 0.11$	$71.53 \pm 0.19$	$71.55 \pm 0.13$	<b>93.26</b> $\pm 0.87$
				$31.57 \pm 0.42$	$52.59 \pm 0.26$	<b>93.26</b> $\pm 0.87$
				$59.49 \pm 0.14$	$61.08 \pm 0.13$	100
				$74.38 \pm 0.34$	$80.06 \pm 0.25$	100
3-5-6-7	100	$64.74 \pm 0.16$	$67.08 \pm 0.11$	$79.72 \pm 0.32$	$78.66 \pm 0.30$	<b>89.16</b> $\pm 0.72$
				$71.16 \pm 0.23$	$77.27 \pm 0.06$	<b>98.21</b> $\pm 0.37$
				$42.49 \pm 0.17$	$55.80 \pm 0.04$	<b>70.42</b> $\pm 1.46$
				$63.94 \pm 0.34$	$70.35 \pm 0.46$	<b>23.16</b> $\pm 1.18$
4-5-6-7	100 si $p \leq 0.31$ 0 sinon	$57.88 \pm 0.06$	$65.76 \pm 0.04$	$76.34 \pm 0.09$	$24.87 \pm 8.63$	100
				$63.23 \pm 0.29$	$21.72 \pm 7.54$	<b>96.33</b> $\pm 0.80$
				$23.48 \pm 0.20$	$15.04 \pm 5.22$	<b>93.00</b> $\pm 1.53$
				$59.89 \pm 0.11$	$22.32 \pm 7.75$	<b>93.00</b> $\pm 1.53$

## BIBLIOGRAPHIE

- Almeida, L. (2001). ICA of linear and nonlinear mixtures based on mutual information. *Proceedings 2001 International Joint Conference on Neural Networks*, Washington, D.C.
- Amari, S., Cichocki, A., et Yang, H. (1996). A new learning algorithm for blind signal separation. *Proceedings of the 1995 Conference on Advances in Neural Information Processing Systems 8 (NIPS'8) : Natural and Synthetic, 27-30 Nov. 1995*, pages 757–763, Denver, Colorado. MIT Press.
- Back, J. et Weigend, A. (1997). A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, 8(4).
- Bayliss, J., Gualteri, J., et Crompton, R. (1997). Analyzing hyperspectral data with independent component analysis. Selander, J. M., editor, *Proceedings of the SPIE AIPR Workshop*, volume 9, P.O. Box 10, Bellingham WA 98227-0010. SPIE.
- Bell, A. et Sejnowski, T. (1995a). Blind separation and blind deconvolution : an information-theoretic approach. *Proceedings of the 1995 20th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 3415–3418, Detroit, Michigan.
- Bell, A. et Sejnowski, T. (1995b). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6) :1129–1159.
- Bell, A. et Sejnowski, T. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*, 37 :3327–3338.
- Cardoso, J. (1992). Iterative techniques for blind source separation using only fourth-order cumulants. *Proceedings EUSIPCO*, pages 739–742.
- Cardoso, J. et Comon, P. (1996). Independent component analysis, a survey of some algebraic methods. *1996 IEEE International Symposium on Circuits and Systems (ISCAS'96), 12-15 May 1996*, volume 2, pages 93–96, Atlanta, Georgia.
- Cardoso, J. et Laheld, B. (1996). Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12) :3017–3030.
- Cardoso, J.-F. (1997). Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4(4) :112–114.
- Cardoso, J.-F. (2002). Analyse en composantes indépendantes. *Proceedings of XXXIV Journées de Statistique (JSBL'2002)*, Bruxelles, Belgique.

- Cardoso, J.-F. (2004a). Dependence, correlation and gaussianity in independent component analysis. *Journal of Machine Learning Research*, 4(7-8) :1177–1203.
- Cardoso, J.-F. (2004b). La cocktail party et autres problèmes compliqués. *Bulletin de liaison de la Société de Mathématiques Appliquées et Industrielles*.
- Chang, C., Chiang, S., Smith, J., et Ginsberg, I. (2002). Linear spectral random mixture analysis for hyperspectral imagery. *Proceedings IEEE Transactions on Geoscience and Remote Sensing*, 40(2) :375–392.
- Chen, C. et Zhang, X. (1999). Independent component analysis for remote sensing study. *EOS/SPIE Symposium on Remote Sensing, Conference and Image and Signal Processing for Remote Sensing V*, volume 3871, pages 150–158, Florence, Italy.
- Chiang, S., Chang, C., et Ginsberg, I. (2000). Unsupervised hyperspectral image analysis using new independent component analysis. *Proceedings IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 7, Hawaii.
- Cichocki, A., Amari, S., Siwek, K., et Tanaka, T. (2003). ICALAB toolboxes. <http://www.bsp.brain.riken.jp/icalab>.
- Cichocki, A., Amari, S., et Thawonmas, R. (1996). Blind signal extraction using self-adaptive non-linear hebbian learning rule. *Proceedings NOLTA'96*, pages 377–380.
- Comon, P. (1994). Independent component analysis, a new concept ? *Signal Processing*, 36(3) :287–314.
- De Lathauwer, L., De Moor, B., et Vandewalle, J. (2000). An introduction to independent component analysis. *Journal of Chemometrics*, 14 :123–149.
- Delfosse, N. et Loubaton, P. (1995). Adaptive blind separation of independent sources : a deflation approach. *Signal Processing*, 45 :59–83.
- Fawcett, T. (2003). Roc graphs : Notes and practical considerations for data mining researchers. Technical report, HP Labs Tech Report HPL-2003-4.
- Field, D. (1994). What is the goal of sensory coding ? *Neural Computation*, 6 :559–601.
- Giannakopoulos, X. (1998). Comparison of adaptive independent component analysis algorithms. Master's thesis, École Polytechnique Fédérale de Lausanne (EPFL), Helsinki University of Technology, Laboratory of Computer and Information Science, Finland.

Giannakopoulos, X., Karhunen, J., et Oja, E. (1998). An experimental comparison of neural ICA algorithms. *Proceedings International Conference on Artificial Neural Networks*, Skövde, Sweden.

Giannakopoulos, X., Karhunen, J., et Oja, E. (1999a). Comparison of neural ICA algorithms using real-world data. *International Joint Conference on Neural Networks (IJCNN'99)*, volume 2, pages 888–893, Washington, DC, USA. Institute of Electrical and Electronics Engineers Inc., USA.

Giannakopoulos, X., Karhunen, J., et Oja, E. (1999b). An experimental comparison of neural algorithms for independent component analysis and blind separation. *International Journal of Neural Systems*, 9(2) :99–114.

Girolami, M. et Fyfe, C. (1996). Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms. *Advances in Neural Information Processing Systems Workshop 9*.

Girolami, M. et Fyfe, C. (1997). Generalized independent component analysis through unsupervised learning with emergent bussgang properties. *Proceedings 1997 IEEE International Conference on Neural Networks (ICNN'97)*, pages 2147–2152, Houston, Texas.

Hurri, J. (1997). Independent component analysis of image data. Master's thesis, Helsinki University of Technology, Department of Computer Science and Engineering.

Hurri, J., Hyvärinen, A., Karhunen, J., et Oja, E. (1996). Image feature extraction using independent component analysis. *Proceedings of IEEE 1996 Nordic Conference on Signal Processing (NORSIG'96)*.

Hyvärinen, A. (1997a). A family of fixed-point algorithms for independent component analysis. *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)*, volume 5, pages 3917–3920, Munich, Germany.

Hyvärinen, A. (1997b). One-unit contrast functions for independent component analysis : a statistical analysis. *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*, pages 388–397.

Hyvärinen, A. (1998a). Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22 :49–67.

Hyvärinen, A. (1998b). New approximations of differential entropy for independent component analysis and projection pursuit. *Advances in Neural Information Processing Systems 10*, pages 273–279. MIT Press.

- Hyvärinen, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, 10(3) :626–634.
- Hyvärinen, A. (1999b). The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Processing Letters*, 10 :1–5.
- Hyvärinen, A. (1999c). Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters*, 6(6) :145–147.
- Hyvärinen, A. (1999d). Sparse code shrinkage : Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11(7) :1739–1768.
- Hyvärinen, A. (1999e). Survey on independent component analysis. *Neural Computing Surveys*, 2.
- Hyvärinen, A., Hoyer, P., et Hurri, J. (2003). Extensions of ICA as models of natural images and visual processing. *Proceedings of the International Symposium on Independent Component Analysis and Blind Source Separation*, pages 963–974, Nara, Japan. IEEE Signal Processing Society Press.
- Hyvärinen, A., Karhunen, J., et Oja, E. (2001). *Independent Component Analysis*, chapter 1, pages 1–12. New York : John Wiley & Sons.
- Hyvärinen, A. et Oja, E. (1996a). One-unit learning rules for independent component analysis. *Advances in Neural Information Processing Systems 9 (NIPS'96)*, pages 480–486, Denver, Colorado. MIT Press.
- Hyvärinen, A. et Oja, E. (1996b). Simple neuron models for independent component analysis. *International Journal of Neural Systems*, 7(6) :671–687.
- Hyvärinen, A. et Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7) :1483–1492.
- Hyvärinen, A. et Oja, E. (1998a). Image feature extraction by sparse coding and independent component analysis. *Proceedings International Conference on Pattern Recognition (ICPR'98)*, pages 1268–1273, Brisbane, Australia.
- Hyvärinen, A. et Oja, E. (1998b). Independent component analysis by general non-linear hebbian-like learning rules. *Signal Processing*, 64(3) :301–313.
- Hyvärinen, A. et Oja, E. (2000). Independent component analysis : algorithms and applications. *Neural Networks*, 13(4-5) :411–430.

- Jones, M. et Sibson, R. (1987). What is projection pursuit? *Journal of the Royal Statistical Society, Series A*, 150(1) :1–37.
- Jutten, C. et Herault, J. (1991). Blind separation of sources, part I. an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1) :1–10.
- Karhunen, J. (1996). Neural approaches to independent component analysis and source separation. *Proceedings of the 4th European Symposium on Artificial Neural Networks (ESANN'96)*, pages 249–266, Bruges, Belgium.
- Karhunen, J., Oja, E., Wang, L., Vigario, R., et Joutsensalo, J. (1997). A class of neural networks for independent component analysis. *IEEE Transactions on Neural Networks*, 8(3) :486–504.
- Karhunen, J. et Pajunen, P. (1997a). Blind source separation and tracking using nonlinear PCA criterion : A least-squares approach. *Proceedings 1997 IEEE International Conference on Neural Networks (ICNN'97)*, pages 2147–2152, Houston, Texas.
- Karhunen, J. et Pajunen, P. (1997b). Blind source separation using least-squares type adaptive algorithms. *Proceedings 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, pages 3048–3051, Munich, Germany.
- Kiviluto, K. et Oja, E. (1998). Independent component analysis for parallel financial time series. *Proceedings ICONIP'98*, volume 2, pages 895–898, Tokyo, Japan.
- Kosaka, N. et Kozugi, Y. (2003). ICA aided linear spectral mixture analysis of agricultural remote sensing images. *Proceedings ICA*, pages 221–226, Nara, Japan.
- Kosaka, N., Uto, K., et Kosugi, Y. (2005). ICA-aided mixed-pixel analysis of hyper-spectral data in agricultural land. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5.
- Lennon, M., Mercier, G., Mouchot, M., et Hubert-Moy, L. (2001). Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. *Proceedings 2001 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'01)*, volume 6, pages 2896–2898.
- Makeig, S., Bell, A., Jung, T., et Sejnowski, T. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems 8*, pages 145–151. MIT Press.
- Mendel, J. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory : theoretical results and some applications. *Proceedings of the IEEE*, 79(3) :278–305.

- Nascimento, J. et Bioucas Dias, J. (2005). Does independent component analysis play a role in unmixing hyperspectral data? *Proceedings IEEE Transactions on Geoscience and Remote Sensing*, 43(1) :175–187.
- Oja, E., Karhunen, J., Wang, L., et Vigario, R. (1995). Principal and independent components in neural networks-recent developments. *Proceedings 7th Italian Workshop on Neural Networks (WIRN-95)*, pages 20–26, Vietri sul Mare, Italy.
- Pajunen, P. et Karhunen, J. (1998). Least-squares methods for blind source separation based on nonlinear PCA. *International Journal of Neural Systems*, 8 :601–612.
- Parra, L., Spence, C., Sajda, P., Ziehe, A., et Müller, K. (2000). Unmixing hyperspectral data. *Advances in Neural Information Processing (Proc. NIPS'99)*, volume 12, pages 942–948. MIT Press.
- Pearlmutter, B. et Parra, L. (1996). A context-sensitive generalization of ICA. *Proceedings International Conference on Neural Information Processing (ICONIP'96)*, pages 151–156, Hong Kong.
- Pearlmutter, B. et Parra, N. (1997). Maximum likelihood blind source separation : A context-sensitive generalization of ICA. *Advances in Neural Information Processing Systems*, volume 9, pages 613–619.
- Pham, D., Garrat, P., et Jutten, C. (1992). Separation of a mixture of independent sources through a maximum likelihood approach. *Proceedings EUSIPCO*, pages 771–774.
- Ristaniemi, T. et Joutsensalo, J. (1999). On the performance of blind source separation in CDMA downlink. *Proceedings International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 437–441, Aussois, France.
- Shah, C., Arora, M., Robila, S., et Varshney, P. (2002). ICA mixture model based unsupervised classification of hyperspectral imagery. *IEEE Proceedings of 31st International Workshop on Applied Imagery and Pattern Recognition*, pages 29–35.
- Shah, C. et Varshney, P. (2004). A higher order statistical approach to spectral unmixing of remote sensing imagery. *Proceedings 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'04)*, volume 2, pages 1065–1068.
- Smith, L. (2002). A tutorial on principal component analysis. Technical report, University of Otago, New Zealand.
- Stone, J. (2002). Independent component analysis : An introduction. *Trends in Cognitive Sciences*, 6(2) :59–64.

Stone, J. et Porrill, J. (1998). Independent component analysis and projection pursuit : A tutorial introduction.

Vigario, R. (1997). Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph. clin. Neurophysiol.*, 103(3) :395–404.

Vigario, R., Jousmäki, V., Hämäläinen, M., Hari, R., et Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. *Advances in Neural Information Processing 10 (Proc. NIPS'97)*, pages 229–235, Cambridge, MA. MIT Press.