

2-1 Introduction :

La séparation aveugle de source consiste à estimer un jeu de p source inconnues à partir d'un jeu de m observation. ces observations sont des mélanges de ces sources et proviennent de capteurs (antennes, microphone, cameras par exemple). Le mélange entre ces sources , qui s'effectue pendant leur propagation jusqu'aux capteurs, est inconnu. La SAS est une disciplines qui permet de nombreuses applications [21] dans de nombreuses disciplines telles l'acoustique, les télécommunications, le génie biomédical [22], l'astrophysique [23].

Il existe dans la littérature plusieurs types de mélanges découpés en deux catégories : les mélanges linéaires et les mélanges non linéaires.

Il existe différentes méthodes permettant la séparation de sources dans le cadre des mélanges linéaires instantanés. Ces méthodes sont le plus souvent regroupées en trois catégories de méthodes. La première catégorie regroupe les méthodes fondées sur l'analyse en composantes indépendantes (ACI) (independent component analysis(ICA)). Dans la deuxième catégorie, on trouve les méthodes basées sur la factorisation en matrices non-négatives (FMN) (non-negative matrix factorisation (NMF)). Enfin, les méthodes fondées sur l'analyse en composantes dans la troisième et dernière catégorie.

Les contributions de cette thèse sont fondées sur la première catégorie qui est l'analyse en composantes indépendantes.

2-2 L'Analyse en composantes principales :

2-2-1 Introduction :

L'Analyse en composantes principales ("Principal Component Analysis"), appelée aussi «analyse géométrique des données » ou « analyse des corrélations » , en abrégé ACP est une méthode d'analyse des données la plus connue et la plus utilisée , qui fut introduite en 1901 par K.Pearson et développée par H. Hotelling en 1933. Nécessitant d'importants calculs numériques, L'ACP n'est devenue une technique opérationnelle qu'à partir des années 1960, avec le développement des moyens de calcul informatique [24].

L'Analyse en composantes principales permet d'effectuer une représentation simplifiée d'une série de variables inter corrélées. Cette technique utilise comme principe une transformation des variables initiales en de nouvelles variables non corrélées. Ces

nouvelles variables sont appelées les Composantes Principales. Chaque Composante principale est une combinaison linéaire des variables initiales [25].

2-2-2 Définition :

L'analyse en composantes principales (ACP) est une méthode descriptive qui permet d'effectuer une représentation simplifiée d'une série de variables inter corrélées. Cette technique utilise comme principe une transformation des variables initiales en de nouvelles variables non corrélées. Ces nouvelles variables sont appelées les Composantes Principales. Chaque composante principale est une combinaison linéaire des variables initiales. La mesure de la quantité d'informations représente sa variance. Les variances associées à chaque composante principale sont classées par ordre décroissant. La composante principale la plus informative est donc la première, et la moins informative la dernière [26]. La méthode est basée sur l'hypothèse qu'il existe de fortes corrélations entre les données étudiées. On passe d'un certain nombre de variables potentiellement corrélées à un plus petit nombre de variables non corrélées, les "Composantes Principales". La 1ère "Composante Principale" absorbe le plus de variance possible, la 2ème "Composante Principale" absorbe le plus de variance possible parmi la variance restante, etc. L'ACP Permet d'analyser des données multivariées et de les visualiser sous forme de nuages de points dans des espaces géométriques [27].

L'analyse en composantes principales est utilisée pour réduire la dimension (le nombre de variables) d'un problème. Cette diminution du nombre de variables doit s'effectuer en perdant un minimum d'informations. Le but de l'Analyse en Composantes Principales (ACP) est de condenser les données originelles en de nouveaux groupements, appelées nouvelles composantes [25,26].

2-2-2-1 L'ACP se divise en étapes :

- ✓ Normalisation des données pour être indépendants des unités des P paramètres.
- ✓ Calcul d'une matrice de similarité C (bien souvent la corrélation).
- ✓ Recherche des éléments propres de C, qui donnent les axes principaux.
- ✓ Représentation des individus dans le nouvel espace (en ne considérant que les valeurs propres) [27].

2-2-3 formulation mathématique d'une ACP :

Soit $\mathbf{x}$ un tableau à $\mathbf{n}$ lignes et $\mathbf{m}$ colonnes. La ligne $\mathbf{i}$ d'écrit la valeur prise par l'individu $\mathbf{i}$ par $\mathbf{m}$ variables quantitatives. Les données sont centrées et réduites, c'est-à-dire que chaque variable à une moyenne nulle et une variance égale à 1

$$\mathbf{X} = \begin{bmatrix} X_{11} & \dots & X_{1j} & \dots & X_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ X_{i1} & \dots & X_{ij} & \dots & X_{im} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & \dots & X_{nj} & \dots & X_{nm} \end{bmatrix} \dots \dots \dots (2.1)$$

On note $\mathbf{x}_j$ le vecteur – colonne constitué par les éléments de la colonne $\mathbf{j}$ de $\mathbf{x}$. $\mathbf{x}_{ij}$ désigne l'élément de $\mathbf{x}$ situé à l'intersection de la ligne $\mathbf{i}$ et de la colonne $\mathbf{j}$, c'est-à-dire la valeur de l'individu $\mathbf{i}$ pour la variable $\mathbf{x}_j$ [27,28].

2-2-4 Recherche des composantes principales :

Soient les composantes : $C_1, C_2, \dots, C_k, \dots, C_q$

C_k = nouvelle variable = combinaison linéaire des variables d'origine $x_1, \dots, x_p$

$$C_k = a_{1k}x_1 + a_{2k}x_2 + \dots + a_{pk}x_p \dots \dots \dots (2.2)$$

Coefficients a_{jk} à déterminer telle que les C_k

Soient :

- deux à deux non corrélées,
 - de variance maximale,
 - d'importance décroissante: est la première composante principale qui doit être de variance maximale. Géométriquement,
- C_1 :détermine une nouvelle direction dans le nuage de points qui suit l'axe d'allongement maximal du nuage.
- C_2 :est la deuxième composante, orthogonale à C_1 et de variance maximale. Géométriquement,
- C_2 : détermine une droite perpendiculaire à C_1 , suivant un axe (perpendiculaire au premier)

2-2-5 Réalisation :

Le but est d'extraire le maximum de la variance projetée en trouvant les axes orthogonaux, Donc indépendants. La procédure s'énonce ainsi :

- déterminer la matrice des données centrées-réduites.
- déterminer la matrice des corrélations entre variables.
- déterminer la matrice des vecteurs propres. Ces derniers sont les coordonnées du point de norme 1 des nouveaux axes sur les anciennes variables.
- déterminer le vecteur des valeurs propres.
- calcul des coordonnées sur les nouveaux axes.
- calcul des corrélations des variables avec les nouveaux axes.

Si nous nous plaçons dans le cas instantané linéaire et où il y a autant de sources que de capteurs. Les sources sont notées $s_j(t)$ et les observations (mélanges) sont notés $x_i(t)$.

$$X = AS \dots\dots\dots (2.3)$$

où l'on a noté, pour le mélange :

$$S = \begin{bmatrix} x1(t1) & \dots & x1(tn) \\ \vdots & \ddots & \vdots \\ xn(t1) & \dots & xn(tn) \end{bmatrix} \dots\dots\dots (2.4)$$

et de même pour les sources :

$$S = \begin{bmatrix} s1(t1) & \dots & s1(tn) \\ \vdots & \ddots & \vdots \\ sn(t1) & \dots & sn(tn) \end{bmatrix} \dots\dots\dots (2.5)$$

On définit alors la matrice de covariance empirique des observations par :

$$C = \frac{1}{N} XX^T \dots\dots\dots (2.6)$$

La condition de décorrélation des sources implique que la matrice de covariance des sources soit diagonale comme vu précédemment. Il semble alors naturel de diagonaliser la matrice de covariance C et ainsi on peut écrire :

$$CU DU^T \dots\dots\dots (2.7)$$

$$\text{où } U \text{ est une matrice orthogonale} \quad U U^T = I \dots\dots\dots (2.8)$$

et D est une matrice diagonale. Une solution consiste alors à poser :

$$S = U^T X \dots\dots\dots (2.9)$$

$$\text{et on a :} \quad \frac{1}{N} \hat{S} \hat{S}^T = U^T C U = D \dots\dots\dots (2.10)$$

Les sources estimées $\hat{s}$ sont donc décorrélées, mais comme nous avons posé $A = U$, nous imposons à la matrice de mélange d'être orthogonale, ce qui n'a en fait aucune raison d'être dans la réalité. Nous pouvons d'autre part nous imposer la condition de sphéricité suivante :

$$\frac{1}{N} \hat{s} \hat{s}^T = I \dots \dots \dots (2.11)$$

qui est une condition a priori plus forte. En fait, les sources ne peuvent être ré-estimées qu'à un facteur d'amplitude près et aux permutations près.

2-3 Analyse en composantes indépendantes (ACI) :

Analyse en composantes indépendantes (ACI), dans un contexte linéaire, d'un vecteur aléatoire, consiste à trouver une transformation linéaire qui minimise la dépendance statistique entre ses composantes.

Soit $x_1, x_2, \dots, x_n$ des combinaisons linéaires des n variables aléatoires latentes, $s_1, s_2, \dots, s_n$, telles que :

$$X_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n, \quad i = 1, 2, \dots, n \dots \dots (2.12)$$

Où a_{ij} sont des coefficients réels inconnus.

Par définition, les variables aléatoires s_i sont mutuellement indépendantes.

C'est la base du concept ACI. Donc, la fonction densité de probabilité (pdf : probability density function) conjointe du vecteur $(s_1, s_2, \dots, s_n)^T$ est égale au produit des fonctions densité de probabilité marginales de chaque variable aléatoire s_i .

i.e

$$p_{s_1, s_2, \dots, s_n}(s_1, s_2, \dots, s_n) = \prod_i^n = 1 p_{s_i}(s_i) \dots \dots \dots (2.13)$$

$p_{s_1, s_2, \dots, s_n}(s_1, s_2, \dots, s_n)$ est la pdf conjointe du vecteur $(s_1, s_2, \dots, s_n)^T$

$p_{s_i}(s_i)$ est la pdf marginale de la variable aléatoire s_i .

$(.)^T$: est la transposé d'un vecteur ou d'une matrice.

Sous forme matricielle (2.12) s'exprime :

$$X = As \dots \dots \dots (2.14)$$

Où $A = [a_{ij}]$ s'appelle la matrice de mélange (mixing matrix).

Dans ce cas, le but de l'ACI est d'estimer une matrice B , appelée matrice de séparation, en disposant de la relation (2.2) de sorte à ce que les composantes du vecteur $y = Bx$ soient mutuellement indépendantes.

$$\text{Ou} \quad y = (y_1, y_2, \dots, y_n)^T \quad (2.15)$$

2-3-1 Prétraitement pour l'ACI :

Avant l'application de n'importe quel algorithme d'ACI, il est impératif de procéder à un préalable prétraitement des observations, ce prétraitement consiste à centrer et à blanchir les variables aléatoires observées. L'intérêt d'un tel prétraitement est de permettre avantageusement de restreindre la recherche à l'espace vectoriel des matrices orthogonales.

2-3-1-1 Etape de centrage :

Le prétraitement de base est de centrer les observations, En retranchant du vecteur x sa valeur moyenne $m = E[x]$ afin de le rendre à valeur moyenne nulle.

$E[.]$ est l'opérateur de l'espérance mathématique.

$$\hat{x} = x - m \quad (2.16)$$

Après l'estimation de la matrice séparant B , nous pouvons rajouter le terme soustraite aux estimées centrées de s , qui est égal à : $A^{-1}m$

2-3-1-2 Etape de blanchiment :

Une deuxième étape du prétraitement consiste au blanchiment du vecteur d'observations. Donc, après avoir centré les données, transformer le vecteur résultant à un autre vecteur, $\tilde{x}$, dont les composantes sont non corrélées et de variances unité. Autrement dit, la matrice de covariance de $\tilde{x}$ est égale à la matrice identité

i.e

$$E[\tilde{x}\tilde{x}^T] = I_n \quad (2.17)$$

Une des méthodes de blanchiment les plus populaires, est la décomposition en vecteurs et valeurs propres (EVD Eigen-Value Décomposition) de la matrice de covariance.

$$E[\tilde{x}\tilde{x}^T] = EDE^T \dots\dots\dots(2.18)$$

Ou E est une matrice orthogonale des vecteurs propres

D est une matrice diagonale des valeurs propres u_i telle que :

$$D = \begin{pmatrix} u_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & u_n \end{pmatrix} = \text{diag}(u_1, u_2, \dots, u_n) \dots\dots\dots(2.19)$$

Et dans ce cas $\tilde{x}$ est obtenu par l'équation

$$\tilde{x} = D^{1/2} E^T \hat{x} \dots\dots\dots(2.20)$$

$$= D^{1/2} E^T A s = \tilde{A} s \dots\dots\dots(2.21)$$

En effet l'importance du blanchiment est d'avoir une matrice de mélange $\tilde{A}$ orthogonale :

$$E[\tilde{x}\tilde{x}^T] = \tilde{A} E [ss^T] \tilde{A}^T = \tilde{A} \tilde{A}^T = I \dots\dots\dots(2.22)$$

Il est clair que le blanchiment réduit le nombre de paramètres à estimer. Au lieu d'avoir n^2 paramètres de la matrice originale A, nous obtiendrons seulement $n(n-1)/2n$ paramètres de la matrice $\tilde{A}$.

2-3-2 Principales approches du problème ICA :

Des méthodes issues du maximum de vraisemblance et des contrastes basés sur l'information mutuelle ou autre mesures de divergence ont souvent été utilisées. Le concept de l'ICA est défini comme la maximisation du degré d'indépendance statistique entre sorties.

2-3-2-1 Non-gaussianité des variables

Pour que l'analyse en composantes indépendantes soit possible, les composantes indépendantes doivent être non-gaussiennes, ce qui représente une restriction fondamentale de l'ICA.

Pour voir cela, nous supposons que la matrice du mélange est orthogonale et les signaux si sont gaussiens, et x_1 et x_2 sont gaussiens, non corrélés, de variance unitaire. Leur densité conjointe est donnée par :

$$P(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \dots\dots\dots(2.23)$$

La distribution de toute transformation orthogonale de (x_1, x_2) gaussiens a exactement la même distribution que (x_1, x_2) , et X_1 et X_2 sont indépendants. Ainsi, dans le cas de variables gaussiennes, le modèle ICA ne peut être estimé qu'avec une

transformation orthogonale. En d'autres termes, la matrice A n'est pas identifiable dans le cas des composantes indépendantes gaussiennes. Par contre, le modèle ICA peut être estimé si seulement une des composantes indépendantes est gaussienne. La non gaussianité représente la clé de l'estimation du modèle ICA et l'analyse est impossible sans son estimation.

2-3-2-2 Mesure de non-gaussianité utilisant le kurtosis :

La mesure classique de la non-gaussianité est le kurtosis ou le cumulants d'ordre 4.

- pour trouver l'inverse de la matrice de mélange afin d'identifier les sources, on procède source par source.

- On cherche un vecteur w tel que :

$$s = w^T x \dots \dots \dots (2.24)$$

ait le kurtosis le plus différent de 0 possible.

* On part d'un vecteur w pris au hasard

* On calcule la direction où le kurtosis croît le plus vite s'il est positif ou décroît le plus vite s'il est négatif. Par des techniques de type «gradient», on peut trouver un optimum local. On a alors identifié une source.

* On recommence. . .

Minimisation de l'information mutuelle

Une autre approche pour l'estimation de l'ICA, inspiré par la théorie de l'information, est la minimisation de l'information mutuelle. Utilisant le concept de la dérivée de l'entropie, on définit l'information mutuelle I entre N variables aléatoires si avec $i = 1 \dots N$ comme suit :

$$I(s_1, s_2, \dots, s_n) = \sum_i H(s_i) - H(s) \dots \dots \dots (2.25)$$

Une propriété importante de l'information mutuelle est que nous avons pour une transformation linéaire inversible :

$$I(s_1, s_2, \dots, s_n) = \sum_i H(s_i) - H(s) - \log |\det w| \dots \dots \dots (2.26)$$

Avec w la matrice de séparation. Si s_i décorrélés et de variance unitaire :

$$E \{ss^T\} = w E \{xx^T\} w^T = I \dots \dots \dots (2.27)$$

ce qui implique :

$$\det I = 1 = (\det w E \{xx^T\} w^T) = (\det w) (\det \{xx^T\}) (\det w^T) \dots \dots (2.28)$$

et donc $\det w$ doit être constant. De plus, pour si de variance unitaire, l'entropie et la néguentropie diffèrent seulement d'une constante et de signe. Ainsi nous avons :

$$I(s_1, s_2, \dots, s_n) = C - \sum_i j(s_i) \dots \dots \dots (2.29)$$

où C est une constante ne dépendant pas de w . Ceci montre la relation entre la néguentropie et l'information mutuelle.

2-4 Différences et similarités entre PCA et ICA :

- ✓ L'analyse en composantes indépendantes ne peut rien apporter à l'analyse de données gaussiennes puisque l'indépendance se réduit alors à la décorrélation. On pourrait dire que l'ICA est l'art d'intégrer l'information non gaussienne dans la recherche de composantes.
- ✓ La PCA est une décomposition unique grâce à la combinaison de deux contraintes : $C(A^{-1}X) = 0$ et $A^T A = 0$. Cette dernière contrainte d'orthogonalité n'a rien de physique : si des composantes existent (ont une réalité physique), il n'y a en général pas de raison de supposer leur orthogonalité géométrique.
- ✓ On peut voir la PCA comme une décomposition hiérarchique, qui extrait des composantes (décorrélées, orthogonales) par ordre décroissant d'énergie. Par contraste, L'ICA est complètement invariante par le changement d'échelle.
- ✓ L'objectif implicite de l'ICA est souvent de trouver des composantes «physiquement significatives». C'est à la fois une force et une faiblesse : les objectifs de la PCA sont plus modestes, mais ils sont toujours atteints : la PCA ne suppose pas, ne cherche pas, ne trouve pas (en général) de composantes ayant une réalité physique.
- ✓ L'algorithmique de la PCA est simple à comprendre et à mettre en œuvre. Par contre, il n'y a pas d'algorithme d'ICA « universel » car il n'est pas direct, ne définit pas des versions empiriques du critère $I(S)$ ni les optimise.
- ✓ La PCA est une méthode utilisant les statistiques d'ordre deux qui permettent de décorréler les données et réduire la dimension du problème, mais ne permet pas de réaliser l'indépendance totale. Par contre l'ICA a la capacité de réaliser l'indépendance totale des signaux en utilisant les statistiques d'ordre supérieur.

2-5 CONCLUSIONS :

Dans ce chapitre nous avons présenté le principe des méthodes de décomposition en appliquées à la séparation de sources. Ces méthodes sont : PCA et ICA. Nous avons tout d'abord, présenté le principe de l'analyse en composantes principales linéaires. L'idée de base du PCA est de réduire la dimension de la matrice des données, en retenant le plus possible les variations présentes dans le jeu de données de départ. Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et ont des coefficients de

corrélation entre elles non nuls. L'ICA est un concept général avec un très grand nombre d'applications en calcul neuronal, traitement du signal, et des statistiques. L'ICA donne une représentation, ou une transformation des données multidimensionnelles qui semble être bien adaptée pour le traitement de l'information. Les composantes de la représentation doivent être indépendantes entre elles, et en même temps elles doivent être «non-gaussiennes».