

Chapitre 3

Analyse en Composantes Indépendantes

Ce chapitre présente l'Analyse en Composantes Indépendantes (ACI). Nous adoptons dans un premier temps une démarche constructiviste en commençant par présenter le problème « historique » de séparation de source dans son contexte général (§3.1). Nous présentons ensuite des méthodes antérieures à l'Analyse en Composantes Indépendantes (§3.2) qui d'une part cherchent peu ou prou à résoudre les mêmes problèmes et d'autres part ont de forts liens avec elle. La suite du chapitre est construite de manière plus déductive. Partant de la définition la plus générale de l'ACI, nous en définissons les limites et indéterminations (§3.3) puis passons en revue les différentes approches mises en œuvre pour la réaliser (§3.4). Nous insistons à la fin de ce paragraphe sur les liens qui existent entre ces méthodes. Enfin nous présentons plusieurs applications ayant profité de manière significative de l'apport de l'ACI, ainsi que quelques utilisations prospectives de celle-ci (§3.5).

3.1 Représenter les données

3.1.1 Illustration : la soirée cocktail

Il est courant d'observer en milieu naturel des mélanges de signaux provenant de sources différentes. Le célèbre problème de la « soirée cocktail » (effet *cocktail party*) évoque le cas d'une soirée où les voix des convives se mélangent allègrement. Pourtant chacun a déjà constaté l'extraordinaire capacité de l'ouïe humaine à différencier l'une de ces voix en particulier, celle de leur interlocuteur par exemple. Cette capacité peut en effet être qualifiée d'extraordinaire lorsque l'on constate que l'ouïe humaine est capable d'effectuer cette discrimination dans des conditions extrêmes, que ce soit en présence de très nombreuses sources, ou encore lorsque le bruit ambiant est bien supérieur à la voix que l'on cherche à discerner. Et surtout, comme bien souvent, la nature réalise avec une facilité déconcertante cette tâche qui devient très ardue dès que l'on souhaite la réaliser artificiellement. Ce problème rentre dans le cadre plus général de la *séparation aveugle de sources* qui consiste à retrouver un certain nombre de sources à partir des observations d'un mélange de celles-ci. Le terme « aveugle » traduit simplement le fait que l'on ignore la façon dont les sources se mélangent, ainsi que le nombre de sources que l'on doit retrouver. Présenté ainsi dans son

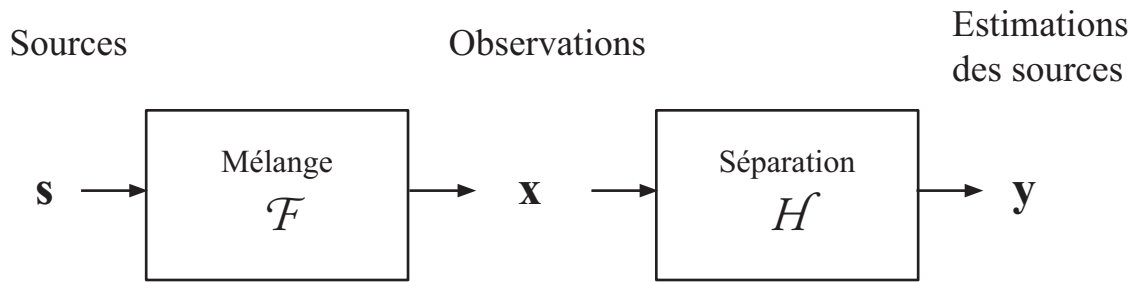


Figure 3.1: Modèle général de la représentation de données

contexte le plus général, le problème est très difficile à résoudre pour une machine. Pourtant ce formalisme permet de modéliser les difficultés rencontrées dans de nombreuses applications.

3.1.2 Formulation générale

Le problème de séparation de sources a initialement été formulé par Héroult, Jutten et Ans [HER85] pour séparer des signaux véhiculés par les fibres nerveuses. Le mélange résulte d'une part du fait que les champs récepteur de cellules voisines se recouvrent largement et d'autre part que les capteurs biologiques sont sensibles à plusieurs grandeurs simultanément. Pour résoudre le problème, ils proposèrent un algorithme utilisant une architecture non supervisée dont le fonctionnement est inspiré de celui de la cellule nerveuse. Indépendamment, Bar-Ness proposait une autre solution au problème appliqué aux communications par satellites [BAR82].

Si nous représentons les données observées par un vecteur aléatoire à p dimensions noté \mathbf{x} , le problème revient donc à trouver une fonction \mathcal{F} représentant le mélange d'un certain nombre n de « sources primitives » qui sont aussi considérées comme un vecteur aléatoire $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$, telles que :

$$\mathbf{x} = \mathcal{F}(\mathbf{s}) \quad (3.1)$$

Dans le cas du problème de la « soirée cocktail » par exemple, chacune des p dimensions représente un capteur (microphone par exemple) et les n sources sont les voix des convives et les autres bruits ambiants (musique de fond, bruit de l'extérieur...). \mathcal{F} est appelée *fonction de mélange*.

Formulé dans ce contexte très général, il s'agit de trouver la meilleure façon de représenter les données \mathbf{x} comme transformées des variables \mathbf{s} au moyen de la fonction \mathcal{F} . Cela revient donc à trouver un nouvel espace de représentation des données. La *meilleure* façon dépend bien entendu de la manière dont on veut comprendre les données, donc des hypothèses formulées dans un cadre applicatif déterminé. L'une des visions les plus anciennes de ce problème est l'Analyse en Composantes Principales (ACP), également appelée transformation de Kurhunen-Loève ou encore transformation de Hotelling. Dans ce cas, on cherche à exprimer les données observées comme résultant d'une transformation linéaire des sources permettant de trouver le plus petit sous-espace où l'erreur de reconstruction est minimale au sens des moindres carrés, ou de façon équivalente le sous-espace sur lequel les projections linéaires conservent le maximum de variance [HOT33]. Dans le cas de l'Analyse en Composantes Indépendantes, l'hypothèse sous-jacente permettant la *meilleure* représentation des données est que les sources sont

statistiquement indépendantes entre elles. C'est justement ce principe de « meilleure représentation » analogue à l'ACP qui a amené Hérault et Jutten à adopter le nom « Analyse en Composantes Indépendantes » [JUT88]. Elle sera cependant redéfinie plus précisément par Comon [COM94].

Quelles que soient les hypothèses formulées, nous nous plaçons dans un cadre statistique et sommes donc contraints à chercher une *estimation* des sources et de la transformation associée à partir des données. De plus, même si nous avons modélisé ces dernières par une variable aléatoire multidimensionnelle \mathbf{x} , nous ne disposons dans un cas réel que d'un nombre limité d'échantillons de cette variable. Formellement nous pouvons écrire:

$$y = H(\mathbf{x}) \tag{3.2}$$

Dans ce cas, y représente une estimation des sources et H est appelée *fonction de séparation*. C'est en réalité cette fonction de séparation que l'on cherche généralement à exprimer :

$$y = H(\mathcal{F}(s)) \tag{3.3}$$

Nous exprimerons la fonction \mathcal{F} de mélange comme l'inverse de la fonction de séparation H , si toutefois cet inverse existe. Si nous ne faisons aucune hypothèse sur la fonction de mélange nous ne savons pas résoudre ce problème. Cela nous amène donc à faire des hypothèses sur le canal de mélange, donc à contraindre la forme de celui-ci.

Comme dans bien des domaines scientifiques, la restriction au cas d'une transformation linéaire des sources est un cas particulier très important. Cela permet généralement de simplifier le problème à la fois d'un point de vue conceptuel et calculatoire. D'autre part de nombreuses méthodes ont été développées pour résoudre le cas linéaire, même si la plupart d'entre elles ont été étendues au cas non-linéaire ou à une restriction de ce dernier. Si les fonctions de mélange et de séparation sont des applications linéaires, elles s'expriment alors sous la forme de matrices et les équations précédentes s'expriment alors sous la forme:

$$y = W\mathbf{x} = W\mathbf{A}s \tag{3.4}$$

\mathbf{A} est la matrice de mélange et W la matrice de séparation. Dans ce cas linéaire, nous pouvons voir les sources comme les coordonnées des observations dans une base particulière. Dans le cas de l'ACP par exemple, cette base de représentation est composée des vecteurs de l'espace permettant le codage du maximum de variance.

3.1.3 Notations

Sauf mention contraire, nous adoptons les notations suivantes. Un vecteur aléatoire contenant n sources est noté \mathbf{s} et celui contenant p observations est noté \mathbf{x} (nous considérerons que $n = p$). Les composantes de ces vecteurs sont respectivement $(s_1, s_2, \dots, s_n)^T$ et $(x_1, x_2, \dots, x_n)^T$. Lorsque l'on considère des observations particulières de ces vecteurs aléatoires, nous adoptons une notation matricielle de la forme $X_T = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)]$ dans le cas de T échantillons :

$$X_T = \begin{bmatrix} x_1(1) & \cdots & x_1(T) \\ \vdots & & \vdots \\ x_n(1) & \cdots & x_n(T) \end{bmatrix} \quad (3.5)$$

La matrice de mélange est notée A et la matrice de séparation W . Dans les processus itératifs, W_t sera la notation prise pour désigner la matrice W à la t -ième itération (nous adopterons alors une notation semblable pour les vecteurs mais donnerons des précisions s'il y a un risque de confusion avec les composantes des vecteurs). Nous désignons la j -ième colonne de W par \mathbf{w}_j et la i -ième ligne de A par \mathbf{a}_i , adoptant la même notation pour les vecteurs déterministes et aléatoires.

Les estimations des sources à partir des observations sont notées \mathbf{y} . Pour désigner les estimateurs, nous utilisons la « notation chapeau », par exemple: $\hat{A} = W^{-1}$. Nous serons amenés à considérer l'ensemble du système « génération + séparation » noté G (donc $G = WA$ et $\mathbf{y} = G\mathbf{s}$).

3.2 Réduire la dimension des données

Représenter des données présuppose de contraindre le canal mélangeant les sources à une certaine forme. Les hypothèses faites sur ce dernier permettent d'exprimer les données dans un nouvel espace de représentation pour lequel un critère est optimisé. Dans ce paragraphe, nous allons d'une part étudier le cas de l'Analyse en Composantes Principales et d'autre part passer en revue une autre technique initialement développée pour observer des données en faible dimension, la Poursuite de Projection. Ces méthodes ont été développées dans le but de réduire la dimension de l'espace de représentation, mais permettent aussi de fournir une représentation pertinente des données.

3.2.1 Analyse en Composantes principales

L'Analyse en Composantes Principales d'un vecteur aléatoire réel \mathbf{x} de taille p et de matrice de covariance $V_x = E\{\mathbf{x}\mathbf{x}^T\}$ finie est définie dans [COM94] comme un couple de matrice $\{F, D\}$ tel que la matrice de variance/covariance se factorise sous la forme

$$V_x = F.D.F^T \quad (3.6)$$

D est une matrice diagonale réelle positive et F est une matrice de rang r et de taille $p \times r$ dont les colonnes sont orthogonales entre elles (c'est-à-dire que $F^T.F$ est une matrice diagonale).

Une méthode pratique pour réaliser une ACP est donc de diagonaliser la matrice de covariance des données¹ et de définir la matrice D comme une matrice diagonale contenant les valeurs propres non nulles de V_x rangées dans l'ordre décroissant et F telles que ses colonnes contiennent les vecteurs propres correspondants. Dans le cas d'une diagonalisation ou d'une décomposition en valeurs singulières de la matrice de covariance, les vecteurs propres ont

¹ Nous supposons que le processus stochastique \mathbf{x} est stationnaire. Voir [DON98] pour une présentation plus générale.

une norme unitaire, si bien que $F^T.F$ est égale à la matrice unité. Ainsi la projection des données sur le premier vecteur propre, appelée première composante principale, encode un maximum de variance puisque cela correspond au carré de la plus grande valeur propre des données originales. Si nous notons w_1 la direction de ce vecteur propre cela revient donc à l'estimer de façon à ce qu'il vérifie:

$$w_1 = \arg \max_{\|w\|=1} E \left\{ (w^T x)^2 \right\} \quad (3.7)$$

Les composantes principales suivantes sont déterminées de telle façon qu'elles encodent le maximum de la variance restante. Ainsi, si les $k-1$ premières composantes principales ont été définies, nous trouvons la direction de la k -ième par la formule:

$$w_k = \arg \max_{\|w\|=1} E \left\{ \left[w^T \left(x - \sum_{i=1}^{k-1} w_i w_i^T x \right) \right]^2 \right\} \quad (3.8)$$

Comme nous l'avons déjà évoqué, l'ACP revient à chercher un sous espace de projection des données dans lequel une l'approximation linéaire est optimale au sens des moindres carrés. Des modèles neuronaux ont aussi été proposés pour réaliser l'ACP, dont le principal initiateur a été Erkki Oja. Il a proposé un modèle de neurone à une seule sortie qui permet d'extraire la plus grande composante principale d'un ensemble de données. Si l'on note y la sortie du réseau, x_i les entrées et w_i les poids correspondants, la « règle de Oja » s'écrit:

$$\begin{aligned} y &= \sum_i w_i x_i \\ \Delta w_i &= \alpha (x_i y - y^2 w_i) \end{aligned} \quad (3.9)$$

Cette règle peut être vue comme une approximation de la règle d'apprentissage de Hebb classique, suivie d'une normalisation des poids (norme euclidienne unitaire) [FYF00]. Par suite, plusieurs modèles ont été développés afin d'extraire l'ensemble des composantes principales [OJA92]. Citons notamment l'*algorithme des sous espaces pondérés* développé par Oja [OJA91] et l'*algorithme de Hebb généralisé* (GHA) développé par Sanger [SAN89] qui permet de trouver les vrais vecteurs propres dans l'ordre des valeurs propres (estimation « au fil de l'eau »). Des extensions au cas non linéaire ont été faites, notamment par Karhunen et Joutsensalo [KAR94, KAR95]. Il s'avère que ces extensions aboutissent à une estimation des directions statistiquement indépendantes de l'espace d'entrée et effectuent donc une Analyse en Composantes Indépendantes [OJA97] sur laquelle nous reviendrons.

3.3.2 Blanchiment des données

Nous pouvons voir l'ACP comme un moyen de decorréler les données, donc à rendre leur matrice de covariance diagonale et même unitaire. Si on reprend la notation du paragraphe 3.1, on définit la matrice de séparation par :

$$W_{PCA} = D^{-\frac{1}{2}} F^T \quad (3.10)$$

La séparation des données à l'aide d'une telle matrice s'appelle un blanchiment spectral et correspond à une annulation des statistiques d'ordre 2 (variances). Il existe d'autres procédés pour effectuer cette opération, comme par exemple une solution symétrique [BEL97] :

$$W_{ZCA} = E\{xx^T\}^{-1/2} \quad (3.11)$$

La matrice de covariance $E\{yy^T\}$ des sorties $\mathbf{y}=W_{ZCA}\mathbf{x}$ est diagonale et les données sont donc décorréélées. De manière générale, multiplier à gauche une matrice de blanchiment par une matrice orthogonale, donne une nouvelle matrice orthogonale.

3.2.3 La poursuite de projection

La poursuite de projection est une méthode statistique d'analyse de données décrites en grande dimension cherchant à les projeter sur un espace de dimension faible de façon à faire apparaître des structures intéressantes. Comme précédemment, l'intérêt des projections en faible dimension dépend de l'application. La méthode est basée sur la définition d'un *indice* qui mesure les caractéristiques de la structure projetée. Par exemple, si cet indice est défini de façon à maximiser la variance des données projetées (sous contrainte de normalité des vecteurs de projection), la projection de poursuite revient à faire une ACP sur les données.

Friedman et Tukey [FRI74] ont défini un indice mesurant l'intérêt des structures projetées et permettant de rechercher les plus intéressantes. Le principe est d'éloigner les nuages de données les uns des autres, en se basant à la fois sur un critère de dispersion et de densité locale. Une alternative est de s'éloigner de la situation la plus « standard » en statistique, c'est-à-dire celle pour laquelle les données se projettent selon une distribution gaussienne [JON87, HOD56]. Pour cela, on définit des indices basés sur des mesures de non-gaussianité, notamment l'entropie différentielle [HUB85], ou une approximation de celle-ci par des moments ou des cumulants [JON87]. D'autres définitions d'indices sont revues en détail dans la thèse de Nason [NAS92] et des approximations de l'entropie différentielle (entropie de Shannon pour des variables continues) permettant des bonnes performances algorithmiques ont été établies par Hyvärinen [HYV98] pour l'estimation de l'ACI et de la poursuite de projections (voir 3.4.4).

3.3 Définition de l'Analyse en Composantes Indépendantes

3.3.1 Cadre pris en compte

Il existe plusieurs façons de définir l'ACI, ou ce qui revient au même, d'expliquer la manière dont on souhaite représenter les données. Heureusement, il a été établi des équivalences entre les différentes méthodes et toutes cherchent d'une manière ou d'une autre à retrouver des signaux sous la seule hypothèse d'indépendance statistique. Dans le cas le plus général cette hypothèse ne suffit pas à effectuer la séparation des signaux [DAR51]. Dans cette thèse nous nous restreignons d'une part au cas des mélanges linéaires des signaux, qui est de loin le cas le plus étudié et qui jusqu'à aujourd'hui a même souvent été pris comme point de départ pour la définition de l'ACI. L'intérêt est que dans ce cas, l'hypothèse d'indépendance statistique entre les signaux est suffisante pour effectuer la séparation¹. D'autre part, puisque nous nous intéressons à terme à l'utilisation de l'ACI pour des images, où les signaux

sont considérés comme variant dans l'espace, nous nous limitons également à l'étude de mélanges instantanés. Ainsi nous écartons l'ensemble des mélanges convolutifs qui intéressent plus particulièrement les chercheurs travaillant sur des signaux variant temporellement, notamment dans le domaine de la *déconvolution aveugle* (autrefois appelée *égalisation aveugle*), dont les applications directes concernent la séparation de signaux auditifs. On pourra se reporter à [HAY94] pour une présentation du problème et à [AMA98a] pour sa résolution par l'ACI.

Dans la suite, nous donnons la définition de l'ACI établie par Comon [COM94], qui est historiquement la première définition rigoureuse pour le cas des mélanges linéaires instantanés, mais aussi la plus générale. Nous indiquons ensuite les limitations qu'imposent les conditions d'identifications des signaux et les indéterminations que cela implique. Enfin, nous présentons un état de l'art de plusieurs approches possibles et développons certaines d'entre elles dans les paragraphes suivants.

3.3.2 Définition

L'Analyse en Composantes Indépendantes d'un vecteur aléatoire réel \mathbf{x} de taille p et de matrice de covariance $V_x = E\{\mathbf{x}\mathbf{x}^T\}$ finie est un couple de matrice $\{A, D\}$ tel que :

- (a) la matrice de variance/covariance se factorise sous la forme :

$$V_x = A.D^2.A^T \quad (3.12)$$

où D est une matrice diagonale réelle positive et A est une matrice de rang n et de taille pxn .

- (b) les observations peuvent être écrites sous la forme :

$$\mathbf{x} = A.s \quad (3.13)$$

où \mathbf{s} est un vecteur aléatoire de taille n dont D^2 est la matrice de covariance et dont les composantes $(s_1, s_2, \dots, s_n)^T$ sont les plus indépendantes possibles au sens de la maximisation d'une *fonction de contraste*.

Par soucis de clarification nous confondrons dans un premier temps la notion de fonction de contraste et de fonction mesurant l'indépendance. Nous renvoyons à [COM94] et au §3.3.4 pour la définition exacte des fonctions de contraste. Il est nécessaire de se donner une *fonction de coût* qui détermine les propriétés statistiques de l'ACI et un *algorithme d'optimisation* qui détermine ses propriétés calculatoires [HYV99b]. Ces deux concepts ne sont pas toujours indépendants l'un de l'autre. Une optimisation par gradient par exemple nécessite de pouvoir dériver la fonction de coût. Par contre, une même fonction pourra parfois être optimisée par différents algorithmes.

Une mesure d'indépendance apparaît immédiatement comme «naturelle». Nous pouvons en effet remarquer qu'un vecteur aléatoire réel $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ dont la densité de probabilité est notée $f_s(\mathbf{u}) = [f_{s_1}(u_1), f_{s_2}(u_2), \dots, f_{s_n}(u_n)]$ a par définition ses composantes (mutuellement) indépendantes si et seulement si :

¹ On parle de mélanges post-non-linéaires quand une non linéarité est appliquée à un mélange linéaire. Taleb et Jutten minimisent l'information mutuelle entre les sorties à l'aide de fonctions *score* (dérivée du logarithme de la densité des estimations \mathbf{y}) pour effectuer la séparation [TAL99]. Une revue des avancées dans le domaine de l'ACI non linéaire a été présentée lors de la conférence ICA2003 [JUT03].

$$f_s(u) = \prod_{i=1}^n f_{s_i}(u_i) \quad (3.14)$$

Ainsi, une mesure naturelle d'indépendance des composantes du vecteur \mathbf{s} est de comparer les deux membres de l'équation précédente au moyen d'une mesure appelée information de Kullback-Leibler dont nous rappelons en annexe A la définition et certaines propriétés. En l'absence de la propriété de symétrie, elle ne peut être rigoureusement considérée comme une distance, mais permet néanmoins de comparer des densités. Nous obtenons alors l'information mutuelle du vecteur \mathbf{s} , définie comme :

$$I(p_s) = \int f_s(u) \log \frac{f_s(u)}{\prod_{i=1}^n f_{s_i}(u_i)} du \quad (3.15)$$

Cette grandeur est toujours positive et s'annule uniquement si les composantes de \mathbf{s} sont mutuellement indépendantes. Malheureusement en pratique il est très difficile d'estimer directement l'information mutuelle, puisque cela nécessite une estimation de la densité conjointe multidimensionnelle, réputée difficile lorsque le nombre de composantes croît. Ce phénomène connu sous le nom de « démon de la dimensionalité » (*curse of dimensionality*) est expliqué par la diminution très rapide de la densité des échantillons dans l'espace probabiliste quand leur dimension augmente. Ainsi, même si l'information mutuelle est considérée comme une « référence » en ce qui concerne la mesure d'indépendance, elle l'est essentiellement au niveau théorique. En pratique d'autres mesures seront utilisées, pouvant éventuellement être des approximations directes de l'information mutuelle.

3.3.3 Reformulation et conditions d'identifiabilité

Si nous utilisons l'information mutuelle comme fonction de contraste particulière, il est montré dans [COM94] que la définition peut se simplifier à l'identification d'un modèle génératif non bruité, instantané et linéaire, ce qui constitue la définition adoptée par la grande majorité de la communauté s'intéressant au sujet [HYV99b]:

L'Analyse en Composantes Indépendantes d'un vecteur aléatoire $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ consiste à identifier le modèle génératif (non bruité) suivant:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (3.16)$$

où les composantes s_i du vecteur $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ sont supposées mutuellement indépendantes et la matrice \mathbf{A} est constante et de taille $p \times n$.

Néanmoins, les conditions d'identifiabilité [COM94, TON91] de ce modèle apportent quelques restrictions:

- Au plus une des sources (composantes de \mathbf{s}) peut suivre une distribution normale (gaussienne).
- Le rang de la matrice \mathbf{A} doit correspondre au nombre de sources.

La première condition vient du fait qu'une distribution gaussienne a tous ses moments et tous ses cumulants d'ordre supérieurs à deux nuls. Dans ce cas, l'indépendance est équivalente à une simple décorrélation telle que la réalise une Analyse en Composantes Principales et l'hypothèse d'indépendance statistique ne permet pas de

différencier les sources gaussiennes les unes des autres. Il faut cependant remarquer que si plus d'une source est gaussienne, il est toujours possible d'identifier les autres sources indépendantes non gaussiennes [HYV99b].

La seconde condition traduit le fait qu'il est nécessaire d'avoir plus de données observées que de sources à identifier. Il faut cependant noter que de récents travaux sur des « bases sur-complètes » (*overcomplete bases*) [OLS96, OLS97, LEW99, LEW00, HYV02] ont montré qu'il est possible d'extraire plus de sources que d'observations. La matrice de mélange n'est alors pas inversible, mais l'extraction des signaux est possible à l'aide d'une estimation bayésienne par exemple. Cela est particulièrement efficace dans le cas de signaux parcimonieux où la probabilité des sources *a posteriori* est modélisée par une distribution sur-gaussienne (*i.e.* ayant beaucoup de valeurs proches de zéro et des queues de distributions au dessus de la loi normale). Au contraire des travaux de Comon sur le modèle d'ACI standard, il n'existe à ce jour aucun résultat théorique assurant la convergence de tels modèles.

Réciproquement, dans le cas où le nombre d'observations est plus important que le nombre de sources que l'on souhaite identifier, nous pouvons réduire la dimension par l'une des techniques précédemment vues. Si les conditions d'identifiabilité sont respectées, nous pouvons donc toujours considérer que la matrice de mélange A est carrée.

Ces deux restrictions énoncées, il subsiste encore deux indéterminations dans le modèle d'ACI ainsi défini. D'une part, changer l'ordre des composantes indépendantes \mathbf{s} n'affecte pas leur indépendance mutuelle. D'autre part, l'indépendance statistique entre composantes est conservée si on les multiplie par une constante non nulle, ce qui revient à admettre une indétermination sur l'amplitude des sources. Ces deux indéterminations ne sont pas propres au modèle restreint présenté ici et existent dans le cas le plus général (§3.3.2). D'ailleurs, la définition des fonctions de contraste tient compte de ces indéterminations.

Dans le cas du modèle d'ACI non bruité, l'amplitude des sources est modélisée par la multiplication de la matrice de mélange A par une matrice diagonale, appelée « matrice d'échelle ». Nous pouvons aussi considérer que puisque l'ACI consiste à estimer simultanément la matrice de mélange A et les sources \mathbf{s} , toute multiplication d'une composante s_i par une constante non nulle revient à diviser la colonne de A correspondante par la même valeur. Le cas de la constante « -1 » montre en particulier l'indétermination sur le signe des signaux estimés.

L'incertitude sur l'ordre des sources dans le cas de l'ACI non bruitée peut être modélisée matriciellement par la multiplication des sources \mathbf{s} par une matrice de permutation P (matrice ayant exactement un seul « 1 » sur chaque ligne et colonne et des zéros sinon). De même que dans le cas précédent, changer l'ordre des sources est équivalent à une permutation des colonnes de la matrice de mélange A , ce qui revient à la multiplier à droite par P^{-1} .

3.3.4 Fonction de contraste

Nous sommes maintenant en mesure de donner la définition complète d'une fonction de contraste [COM94], appelée aussi plus simplement *contraste*. C'est une fonction Ψ à valeurs réelles qui, appliquée aux densités p_y des sorties doit vérifier les propriétés suivantes:

- Invariance par permutation : $\Psi(P.p_y) = \Psi(p_y)$ pour toute matrice de permutation P .

Chapitre 3

- Invariance à l'échelle : $\Psi(p_{\Delta y}) = \Psi(p_y)$ pour toute matrice diagonale Δ .
- Si les composantes y_i sont indépendantes entre elles, $\Psi(p_{My}) \geq \Psi(p_y)$ pour toute matrice M inversible.

On considère généralement des *contrastes discriminants*, c'est-à-dire des contrastes pour lesquels l'égalité est vérifiée uniquement pour des matrices de la forme $M = \Delta.P$. Ainsi avec de telles fonctions, l'indépendance des composantes est réalisée uniquement pour le minimum de la fonction de contraste.

L'information mutuelle est la fonction de contraste par excellence. Mais cette dernière étant difficile à calculer directement, on cherchera une approximation numérique de celle-ci, avec un développement en série d'Edgeworth ou de Gram-Charlier par exemple.

3.4 Etat de l'art

Comme indiqué au début du chapitre, le problème de séparation de sources ayant conduit à la formulation de l'Analyse en Composantes Indépendantes a été initialement défini par Héroult, Jutten et Ans [HER85], alors qu'ils s'intéressaient à des problèmes de neurophysiologie au début des années 80. Vingt ans plus tard, le concept intéresse des centaines de chercheurs dans le monde, du point de vue théorique et pratique. Depuis 1999 une conférence portant spécifiquement sur le sujet est organisée tous les 18 mois. La première a eu lieu à Aussois (France) et les suivantes à Espoo (Finlande), San Diego (Californie, Etat-Unis) et Nara (Japon). La prochaine aura lieu à Grenade (Espagne) au mois de septembre 2004.

L'objet de ce paragraphe est de passer en revue les principales approches de l'ACI effectuées au cours de cette période. Les « sources d'inspiration » sont essentiellement issues des domaines du traitement du signal dans une approche neuronale, de la théorie de l'information et des statistiques. Abordée et expliquée différemment dans chacun de ces domaines, l'ACI se trouve être un seul et même concept qui en retour permet de résoudre efficacement une multitude de problèmes et d'applications. Ce fait remarquable explique sans doute l'effervescence croissante qu'elle suscite chez les chercheurs depuis vingt ans. On trouvera une revue récente de l'ACI dans le livre de Hyvärinen, Karhunen et Oja [HYV01]. D'autres états de l'art sont présentés dans le livre de Lee [LEE98] et dans les articles [AMA98a, CAR98, HYV99b, LEE00]. Enfin signalons l'article de Jutten [JUT00] dans lequel il présente l'histoire de la genèse de l'ICA et de la séparation de sources.

3.4.1 Traitement du signal et statistiques

La première approche de la séparation de sources réalisée par Héroult et Jutten s'inspire du traitement du signal et plus particulièrement de l'approche neuronale ou, comme les auteurs l'appellent, l'approche neuromimétique [HER85], marquant ainsi clairement l'inspiration biologique initiale. L'algorithme « HJ » permettant la séparation [JUT91] est basé sur un réseau de neurones récurrents dont les poids sont les termes non diagonaux d'une matrice de séparation W (voir figure 3.2), les termes de la diagonale étant contraints à la nullité. Ainsi, l'algorithme calcule les estimations y des sources à partir des observations x :

$$y = (I+W)^{-1}x \quad (3.17)$$

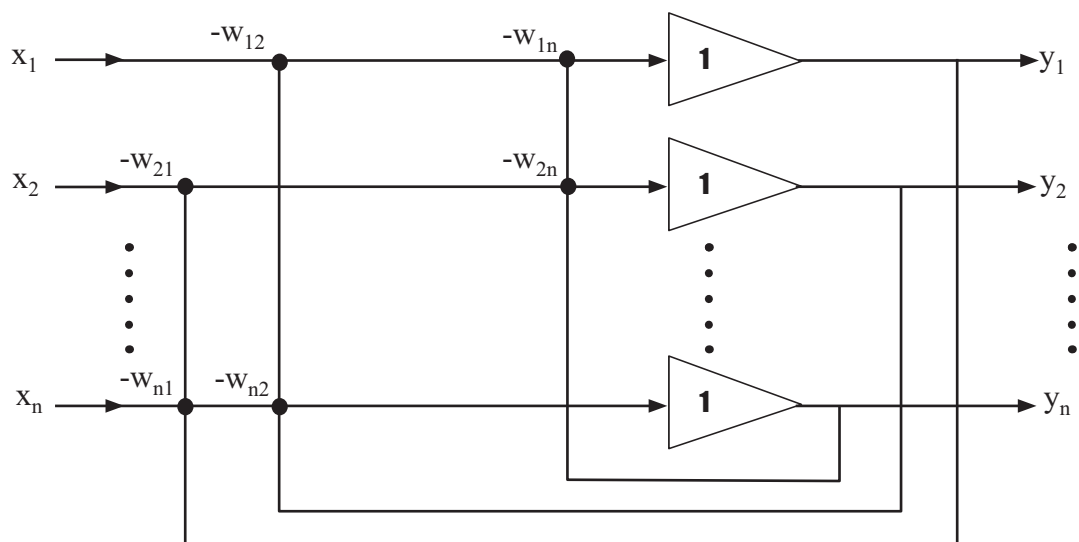


Figure 3.2: Architecture neuronale récursive de l'algorithme Héault-Jutten [JUT91]

avec la règle d'adaptation suivante pour les termes non diagonaux:

$$\Delta w_{ij} = f(y_i) \cdot g(y_j) \quad (3.18)$$

où f et g sont des fonctions non linéaires impaires différentes. Dans le papier original, les auteurs proposent la fonction « cube » pour f et la fonction « arctangente » pour g , en précisant que d'autres choix sont possibles (et souhaitables) en fonction de la forme des densités à estimer. Dans la seconde partie de l'article [COM91], des précisions sont apportées quand au choix de ces non linéarités. L'analyse mathématique de l'algorithme HJ [COM91] a aussi permis de préciser que la mesure d'indépendance sous-jacente est l'annulation des cumulants croisés d'ordre supérieur. C'est d'ailleurs dans la nécessité de recourir aux statistiques d'ordre supérieur pour identifier les sources que réside l'apport de l'ACI, comme cela sera montré dans [COM89] et [LAC92]. Pour une présentation des statistiques d'ordre supérieur, on pourra se reporter à l'ouvrage de Lacoume, Amblard et Comon [LAC97], ou à l'habilitation à diriger des recherches de ce dernier [COM95]. Ainsi, Ruiz et Lacoume proposent un algorithme annulant les cumulants d'ordre deux et quatre à l'aide d'un algorithme d'optimisation non linéaire sous contrainte revenant à annuler le carré des cumulants croisés [LAC92]. Mais en pratique cet algorithme présente une complexité calculatoire trop importante pour séparer plus de trois sources [COM95]. Dans [COM92] il est également proposé un algorithme basé sur le développement en séries d'Edgeworth des densités cherchant à annuler les cumulants d'ordre quatre. Dans [COM94], il est montré que cela revient à définir l'information mutuelle (ou son opposé plus exactement) comme une fonction de contraste que l'on cherche à minimiser. Expliquer l'ACI à l'aide des fonctions de contraste a permis leur étude mathématique rigoureuse et l'introduction de nouveaux algorithmes basés sur un apprentissage itératif de la matrice de séparation. Mais l'étude de la convergence de ces algorithmes a montré leur dépendance vis-à-vis de la matrice de mélange [MOR98]. Ce problème a été résolu par [CAR96] en utilisant des estimateurs *équivariants*, c'est-à-dire vérifiant la propriété suivante:

$$\hat{A}_{MX_T} = M\hat{A}_{X_T} \quad (3.19)$$

où M est une matrice de mélange inversible quelconque, \hat{A}_{X_T} est l'estimateur considéré (dans notre cas, l'inverse de la matrice de séparation W), estimé à partir de T échantillons des observations \mathbf{x} , rangés dans la matrice X_T et est noté \hat{A}_{MX_T} quand il est estimé à partir des mêmes échantillons multipliés par la matrice M. La recherche de tels estimateurs est justifiée dans le cas qui nous intéresse (3.17) puisque multiplier les observations par une matrice M est équivalent à multiplier le mélange par cette même matrice: $M(X_T) = M(AS_T) = (MA)S_T$. Or avec un estimateur équivariant de la matrice de mélange, nous pouvons constater que l'estimation des sources ne dépend plus du mélange A mais uniquement des sources:

$$\hat{\mathbf{s}}(t) = (\hat{A}_{X_T})^{-1}\mathbf{x}(t) = (\hat{A}_{AS_T})^{-1}A\mathbf{s}(t) \quad (3.20)$$

$$\hat{\mathbf{s}}(t) = (A\hat{A}_{S_T})^{-1}A\mathbf{s}(t) = (\hat{A}_{S_T})^{-1}\mathbf{s}(t) \quad (3.21)$$

Le passage de la première ligne à la seconde utilisant la propriété d'équivariance de \hat{A} .

Afin d'utiliser cette propriété pour estimer la matrice de séparation, Cardoso et Laheld ont introduit le *gradient relatif* qui remplace l'itération additive habituelle d'un gradient par une itération multiplicative :

$$W_{t+1} = W_t - \lambda_t \nabla J_\psi(y_t).W_t = (I - \lambda_t \nabla J_\psi(y_t)).W_t \quad (3.22)$$

où $\nabla J_\psi(y_t)$ désigne le gradient d'une fonction de coût dépendant d'une fonction de contraste ψ calculée à partir des estimées y_t . Ainsi l'itération multiplicative (autrement appelée « mise à jour en série » pour la traduction de *serial update*) permet à l'estimateur global des sources $G = W.A$ de vérifier la propriété d'équivariance :

$$\begin{aligned} y_t &= W_t A s = G_t s \\ G_{t+1} &= W_{t+1} A \\ G_{t+1} &= (I - \lambda_t \nabla J_\psi(G_t s)).G_t \end{aligned} \quad (3.23)$$

Ainsi l'estimation globale des sources n'est pas dépendante du mélange. Par suite dans [CAR96] un algorithme baptisé EASI (la signification n'est pas donnée dans [CAR96], mais le premier auteur étant français il peut s'agir de *Estimation Adaptative de Sources Indépendantes*) est dérivé de ces règles générales en faisant les choix suivants:

$$\begin{aligned} \psi(y) &= \sum_{i=1}^n |y_i|^4 \\ J_\psi(y) &= E[\psi(y)] \end{aligned} \quad (3.24)$$

Il est ainsi montré que la règle d'adaptation de EASI pour la matrice de séparation devient :

$$W_{t+1} = W_t - [y_t y_t^T - I + g(y_t) y_t^T - y_t g(y_t)^T].W_t \quad (3.25)$$

Amari est parvenu à un algorithme semblable [AMA96, AMA98b] en exprimant l'information mutuelle comme un développement en série de Gram-Charlier et l'a appelé *gradient naturel*. L'algorithme du gradient naturel a aussi été proposé et mis en oeuvre dans [CIC96]. L'approche est justifiée par le fait que cela permet de faire tendre

la matrice des corrélations des sorties vers l'identité.

Une autre classe de méthodes basées sur la diagonalisation tensorielle a été introduite pour rechercher une optimisation des contrastes. L'algorithme le plus connu est JADE (*Joint Approximate Diagonalisation of Eigenmatrices*), développé par Souloumiac et Cardoso [CAR93], qui fait suite à FOBI (*Fourth Order Blind Identification*) [CAR89]. Leur popularité est en partie due au fait qu'ils furent parmi les premiers algorithmes à permettre une réalisation pratique de l'ACI. Un *tenseur de cumulants* (à l'ordre quatre) est une matrice en quatre dimensions contenant tous les cumulants croisés d'ordre quatre. Pour un vecteur aléatoire \mathbf{x} de taille n chaque élément de son tenseur (d'ordre quatre) est $Cum(x_i, x_j, x_k, x_l)$ avec $1 \leq i, j, k, l \leq n$; cela peut être vu comme la généralisation d'une matrice de covariance au delà de l'ordre deux. Nous pouvons surtout le voir comme une application linéaire d'un espace de matrice $n \times n$ dans un autre espace de matrice $n \times n$ et le représenter par la matrice bloc en trois dimensions \mathcal{N}_x^4 contenant tous les cumulants d'ordre quatre de \mathbf{x} , comme représenté à la gauche de la figure 3.3. Comme toute application linéaire, celle-ci peut être diagonalisée et, sous contrainte de blanchiment des signaux d'entrée, il a été montré dans [TON93] que toutes les « tranches » de la matrice \mathcal{N}_x^4 pouvaient être diagonalisées à l'aide d'une même matrice unitaire U , qui permet d'effectuer la séparation dans le cas où toutes les valeurs propres sont différentes. Dans le cas contraire [TON93] propose d'utiliser une combinaison linéaire de « matrices tranches » et de retenir la combinaison offrant le spectre (au sens « ensemble des valeurs propres ») le plus large. Cette méthode a le désavantage de négliger l'information des cumulants non pris en compte dans la combinaison choisie. Dans [CAR93], le choix de la matrice unitaire parmi toutes celles possible se fait par diagonalisation directe de l'application linéaire associée au tenseur d'ordre quatre, en mesurant la « diagonalité » de la matrice par la somme du carré des éléments diagonaux. Puisque l'on est sous contrainte de normalité, rendre minimale la somme du carré des éléments « hors diagonale » est équivalent à rendre maximal la somme des carrés des éléments diagonaux. Par suite, il est prouvé qu'une telle opération revient à optimiser la fonction de contraste :

$$c(\mathbf{e}) = \sum_{i,k,l} |Cum(e_i, e_i^*, e_k, e_l^*)|^2 \quad (3.26)$$

où \mathbf{e} est le vecteur d'entrée blanchi. En pratique c'est la diagonalisation de la matrice \mathcal{N}_x^4 dépliée (figure 3.3 droite) de taille $n^2 \times n^2$ qui permet d'identifier la matrice unitaire appropriée. Le problème essentiel de cette approche est qu'elle utilise tous les cumulants d'ordre 4, ce qui conduit à des calculs d'une complexité d'ordre n^4 . Ainsi elle ne pourra être utilisée en pratique que pour de faibles dimensions.

Une troisième classe de méthode a été développée dans l'approche « traitement du signal statistique » de l'Analyse en Composantes Indépendantes avec l'estimateur du maximum de vraisemblance (MV). La première proposition a été formulée par [GAE90] puis dans [HAR96] en approchant la log-vraisemblance des sources par un développement en série de Gram-Charlier basé sur leurs cumulants jusqu'à l'ordre quatre. La mise en œuvre a plutôt été faite par [PHA97] qui tient compte de l'ensemble des statistiques. Pour le modèle considéré, la vraisemblance des observations conditionnées par la matrice de mélange s'exprime comme :

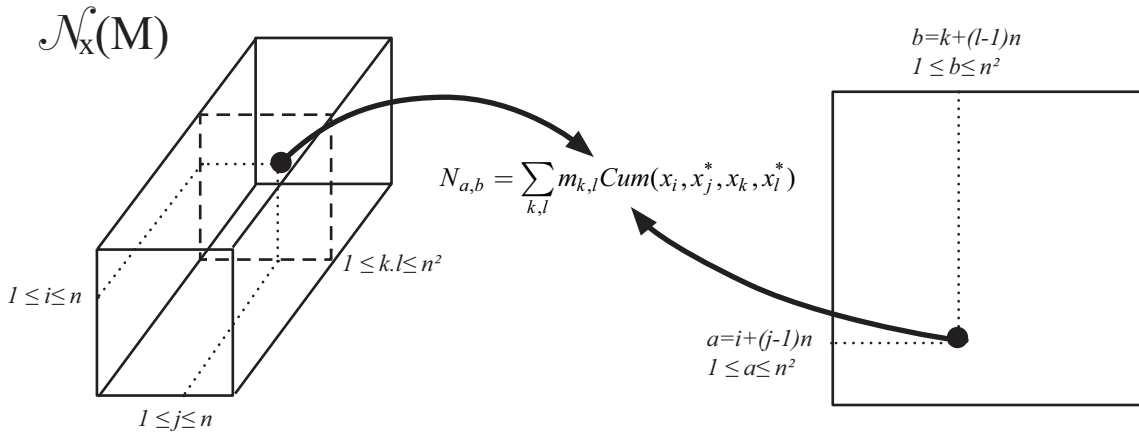


Figure 3.3: Matrice de cumulants pour l'algorithme JADE.

$$p_{x|A}(y) = \int p_s(A^{-1}u) |\det(A)|^{-1} du \quad (3.27)$$

En notant \hat{E} la moyenne temporelle sur T échantillons considérés comme indépendants, e_i un vecteur ayant un « 1 » à la i -ième position et des zéros ailleurs et en posant $\Phi_i = [\log(p_{s_i})]'$ (où le signe ' marque la dérivée), l'estimateur du maximum de vraisemblance est obtenu en résolvant :

$$\hat{E}[\Phi_i(e_i^T A^{-1}x)e_j^T A^{-1}x] = 0 \quad \forall i \neq j \quad (3.28)$$

Et en notant $\hat{s}_i = e_i^T A^{-1}x$ l'estimation des sources, on obtient:

$$\hat{E}[\Phi_i(\hat{s}_i)\hat{s}_j] = 0 \quad \forall i \neq j \quad (3.29)$$

Ce résultat justifie la forme de la règle d'apprentissage de l'algorithme HJ et donne la forme de la fonction non linéaire impaire qui doit être choisie au sens du maximum de vraisemblance. Dans [PHA97], la solution de cette équation est obtenue par le biais d'une optimisation itérative à l'aide de l'algorithme de Newton-Raphson. Dans [CHO01], c'est le gradient naturel développé par Amari qui est utilisé pour effectuer l'optimisation. Enfin, [PEA96] dérive deux gradients à partir de la formulation de la vraisemblance, l'un servant à l'estimation de la matrice de séparation et l'autre à l'estimation des densités de chaque sortie y_i conditionnée par la colonne w_i correspondante. L'une de leur règle du gradient étant identique à celle de [BEL95], les auteurs en déduisent l'équivalence entre la méthode d'estimation par maximum de vraisemblance et l'approche *Infomax* qui sera développée ultérieurement. Cette équivalence a été démontré différemment par Cardoso [CAR97].

3.4.2 Approche ACP non linéaire

Une autre façon d'aborder l'Analyse en Composantes Indépendantes est de la considérer comme une extension non linéaire de l'Analyse en Composantes Principales. Le point de départ est la règle de Oja généralisée à plusieurs unités [OJA92] qui s'exprime linéairement :

$$W_{t+1} = W_t + \lambda_t [I - W_t W_t^T] x_t x_t^T W_t \quad (3.30)$$

Il a été proposé dans [OJA91] d'appliquer des non linéarités à un ou plusieurs des produits $W_t^T x_t$ ou $x_t^T W_t$. Karhunen et Joutsensalo [KAR94] dérivent un algorithme à partir d'un critère non linéaire permettant de minimiser l'erreur de représentation, pouvant toujours se mettre sous la forme :

$$J_1(w_i) = E \left\{ f_1 \left(x - W f_2(W^T x) \right) \right\} \quad (3.31)$$

où $f_1(\cdot)$ et $f_2(\cdot)$ sont deux fonctions non linéaires s'appliquant à chaque composante de leur argument vectoriel. Ils en dérivèrent alors une règle d'adaptation pour un apprentissage par réseau de neurones :

$$W_{t+1} = W_t + \lambda_t \left[x_t g_1(e_t^T) W_t G_2(x_t^T W_t) + g_1(e_t) f_2(x_t^T W_t) \right] \quad (3.32)$$

où $g_1(\cdot)$ et $g_2(\cdot)$ sont respectivement les dérivées de $f_1(\cdot)$ et $f_2(\cdot)$. e_t est l'erreur de reconstruction :

$$e_t = x_t - W_t g_2(W_t^T x_t) \quad (3.33)$$

et $G_2(\cdot)$ est la matrice diagonale :

$$G_2(x_t^T W_t) = \text{diag} \left[g_2(x_t^T w_t(1)), \dots, g_2(x_t^T w_t(n)) \right] \quad (3.34)$$

Notons que le choix $f_1(t) = t^2/2$ permet de retrouver le critère de minimisation de l'erreur quadratique habituel pour l'Analyse en Composantes Principales. D'autres choix sont possibles, mais pour des raisons de stabilité, il est nécessaire que sa dérivée $g_1(\cdot)$ soit une fonction impaire croissante. Les choix courants pour ces fonctions sont représentés sur la figure 3.4. Si f_1 est choisie quadratique et f_2 est choisie linéaire, nous retrouvons l'ACP standard. Notons par ailleurs qu'après une période d'apprentissage, l'erreur de reconstruction devient suffisamment petite pour que le premier terme dans les crochets de (3.31) soit négligé devant le second. La règle d'adaptation apparaît comme une approximation de gradient stochastique permettant de minimiser le critère $J_1(W)$. Un autre critère d'optimisation a été introduit dans [KAR94] et étudié plus particulièrement dans [KAR95]. Plusieurs formes proches ont été proposées, la plus significative s'exprimant pour chaque neurone $w(i)$ ($i=1, \dots, n$) :

$$J_2(w_i) = E \left\{ f \left(x^T w_i \right) \right\} + \sum_{j=1}^{I(i)} \lambda_{ij} \left[w_i^T w_j - \delta_{ij} \right] \quad (3.35)$$

où $\lambda_{ij} = \lambda_{ji}$ sont les multiplicateurs de Lagrange, δ_{ij} est la notation habituelle pour le produit de Kröneckner permettant d'imposer l'orthonormalité des vecteurs w et $I(i)$ indique le nombre de neurones sur lequel est fait la sommation. Lorsque $I(i) = n$, cela donne une généralisation de l'algorithme des sous espaces pondérés et pour $I(i) = i$, nous obtenons une généralisation de l'algorithme de Hebb généralisé (GHA) de Sanger [SAN89]. En notant $g(\cdot)$ la dérivée de la fonction $f(\cdot)$ précédente, la règle d'apprentissage est :

$$w_{t+1}(i) = w_t(i) + \lambda_t \left[I - \sum_{j=1}^{I(i)} w_t(j) w_t(j)^T \right] x_t g \left[x_t^T w_t(i) \right] \quad (3.36)$$

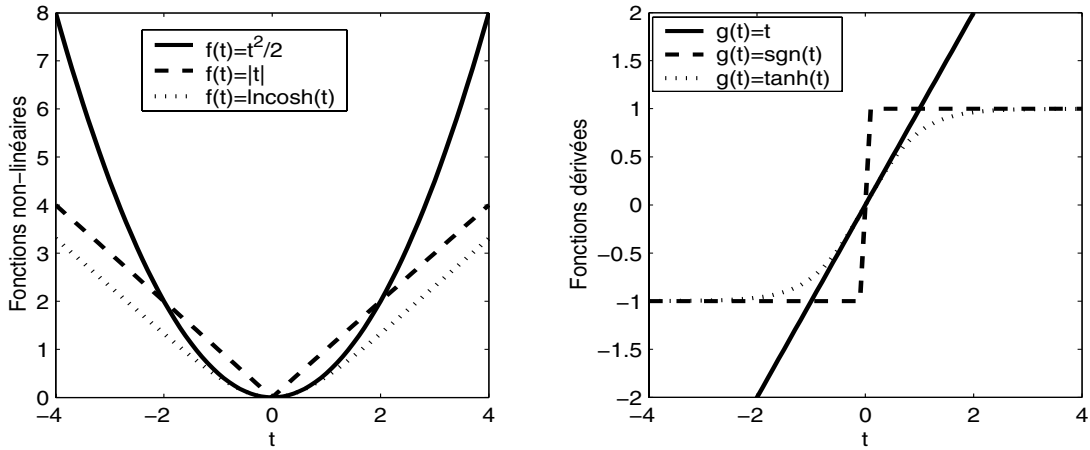


Figure 3.4: Choix typiques de fonctions non linéaires (gauche) et leurs dérivées (droite) pour la PCA non linéaire [KAR94].

L'utilisation de fonctions non linéaires dans des réseaux de neurones du type ACP permet l'introduction de statistiques d'ordre supérieur et peut donc se ramener à une ACI. Par exemple dans [HYV01], Oja remarque qu'en choisissant un critère quadratique pour J_1 et en notant les sorties $\mathbf{y}=\mathbf{W}\mathbf{x}$ et sous contrainte d'orthogonalité pour la matrice de séparation ($\mathbf{W}\mathbf{W}^T=\mathbf{W}^T\mathbf{W}=\mathbf{I}$), on peut écrire :

$$\begin{aligned} \|\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})\|^2 &= [\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})] \mathbf{W}^T \mathbf{W} [\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})] \\ \|\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})\|^2 &= \|\mathbf{W}\mathbf{x} - \mathbf{W}\mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})\|^2 \\ \|\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})\|^2 &= \|\mathbf{t} - \mathbf{g}(\mathbf{y})\|^2 = \sum_{i=1}^n [y_i - g(y_i)]^2 \end{aligned}$$

et si on choisit la fonction non linéaire comme :

$$g_i(y) = \begin{cases} y^2 + y & \text{si } y \geq 0 \\ -y^2 + y & \text{si } y < 0 \end{cases}$$

alors, le critère J_1 revient à :

$$J_{kurt}(W) = \sum_{i=1}^n E \left\{ (y_i - y_i \pm y_i^2)^2 \right\} = \sum_{i=1}^n E \{ y_i^4 \}$$

où l'on reconnaît une fonction de contraste introduite dans [COM94].

3.4.3 Théorie de l'information

L'approche *Infomax* de l'Analyse en Composantes Indépendantes est souvent assimilée à l'approche par le maximum de vraisemblance car une équivalence a été établie entre les deux méthodes [CAR97]. Néanmoins, il nous semble important de lui réserver une place à part dans cette thèse, puisque d'une part elle a été formulée à partir de principes de la théorie de l'information et que d'autre part c'est cette approche qui permet de voir que l'ACI réalise un processus pouvant expliquer le codage de l'information visuelle dans le cortex des vertébrés et plus particulièrement des primates.

Nous avons vu au chapitre 2 que Nadal et Parga [NAD94] ont montré l'équivalence entre le principe de réduc-

tion de redondance formulé par Barlow [BAR61] et le principe Infomax de Linsker [LIN88]. Bell et Sejnowsky ont exploité ce résultat [BEL95] :

$$\frac{\partial}{\partial w} I(y, x) = \frac{\partial}{\partial w} H(y) \quad (3.37)$$

où $I(y, x)$ est l'information mutuelle entre les sorties \mathbf{y} et les entrées \mathbf{x} d'un réseau de neurone, $H(y)$ est l'entropie des sorties et w les paramètres du réseau. La relation ci-dessus exprime donc exactement que rendre maximum l'information mutuelle des sorties du réseau est équivalent à rendre maximale l'information qui « passe » à travers le réseau. De la relation liant les densités de probabilités des entrées et des sorties, ils dérivent une règle d'apprentissage des paramètres du réseau qui permet d'obtenir un code factoriel et d'avoir une représentation en composantes indépendantes des entrées. Dans le cas général cette règle s'écrit:

$$\Delta W = [W^T]^{-1} + \frac{\partial}{\partial W} \ln \prod_i |y_i'| \quad (3.38)$$

où y_i' est la dérivée de chaque sortie. Celle-ci dépend donc des non linéarités (sigmoïdes) qui sont choisies pour chaque unité du réseau. L'hypothèse sous jacente est que la fonction de répartition des données suit la non linéarité. On constate heuristiquement que les distributions sous-gaussiennes ne sont pas toujours séparées [BEL95]. Cet inconvénient est résolu et la vitesse de convergence améliorée, en utilisant une règle du type « gradient relatif » [CAR96] (ou « gradient naturel » [AMA98b]):

$$\Delta W = [I - K \tanh(y)y^T - yy^T] W \quad (3.39)$$

K est une matrice diagonale dont les éléments valent «1» si la source est sur-gaussienne et «-1» si elle est sous-gaussienne [LEE99]. Le paramètre est estimé à chaque pas d'itération pour assurer la stabilité [CAR98].

3.4.4 Eloignement à la gaussianité

L'Analyse en Composante Indépendantes peut être abordée, par la recherche de distributions les plus éloignées possibles de la distribution normale. La justification essentielle de ce point de vue est le théorème central limite qui stipule que la somme de variables indépendantes tend asymptotiquement vers une distribution normale. Or selon le modèle d'ACI pris en compte, toutes les estimations y_i en sortie de la matrice de séparation sont la somme de variables indépendantes ($\mathbf{y} = G\mathbf{s}$), donc elles tendent à se rapprocher d'une distribution gaussienne. En cherchant à les en éloigner, elles tendent à évaluer une seule des variables s_i et à réaliser ainsi l'estimation souhaitée (à une permutation et un facteur d'échelle près). C'est l'approche généralement adoptée par Hyvärinen pour présenter l'Analyse en Composantes Indépendantes [HYV01]. Le problème revient à trouver une « mesure de non-gaussianité » qui est appliquée aux estimations des sources puis rendue maximale par une méthode itérative. La méthode a initialement été appliquée pour résoudre des problèmes de déconvolution aveugle, mais a été appliquée dans le cadre de l'ACI par Delfosse et Loubaton [DEL95] en utilisant des grandeurs dérivées du moment et du cumulants d'ordre quatre des sorties pour mesurer la non-gaussianité. Cependant, l'apport principal de ce travail est l'introduction d'une pro-

Chapitre 3

cédure de déflation pour estimer les sources. Cette procédure exploite l'existence de points fixes pour un processus itératif, lui assurant non seulement la garantie de converger, mais permet aussi une convergence beaucoup plus rapide qu'avec une descente de gradient ordinaire. C'est cette propriété qui a permis à Hyvärinen et Oja de baptiser leur algorithme « FastICA ». Dans la première version de l'algorithme [HYV97], la mesure de non-gaussianité est la valeur absolue du kurtosis. Mais cette mesure étant insuffisamment robuste, la seconde version de l'algorithme [HYV99c] utilise une autre mesure, la néguentropie qui est définie par :

$$J(y) = H(y_{\text{gauss}}) - H(y), \quad H(y) = - \int p_y(u) \log(p_y(u)) du \quad (3.40)$$

où $H(\cdot)$ indique l'entropie différentielle (entropie de Shannon pour des variables continues) et y_{gauss} est une variable gaussienne de même moyenne et covariance que la variable aléatoire y mesurée. Cette mesure est toujours positive, invariante par une transformation linéaire et ne s'annule que pour une variable gaussienne. Elle a été introduite dans [COM94], pour exprimer l'information mutuelle comme une fonction de contraste et en dériver un algorithme. En dérivant des approximations de (3.39) on aboutit à l'algorithme « FastICA ». En première approximation cependant, la néguentropie est équivalente au carré du kurtosis pour des distributions symétriques *i.e.* ayant leur cumulant d'ordre trois (aplatissement ou *skewness* en anglais) nul. Afin d'obtenir des estimateurs plus robustes, la néguentropie est approchée par :

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \quad (3.41)$$

où $G(\cdot)$ est une fonction non quadratique, typiquement de la forme :

$$\begin{aligned} G_1(y) &= \log \cosh(y) \\ G_2(y) &= -\exp(-y^2/2) \end{aligned} \quad (3.42)$$

Comme pour beaucoup d'algorithmes, les données sont contraintes à être centrées et décorréelées. Par suite, la dérivation de l'algorithme se fait à partir de (3.38). Dans le cas où l'on considère toutes les colonnes à la fois, il faut ajouter une contrainte de blanchiment global à chaque itération. Le point clé est que sous contraintes de blanchiment, la décorrélation est équivalente à une orthogonalisation. Cette orthogonalisation évite que les différentes colonnes convergent vers la même source. Deux stratégies peuvent alors être choisies pour contraindre à l'orthogonalité. La première consiste à effectuer le schéma de la table 3.1 pour chaque colonne itérativement en ajoutant simplement une orthogonalisation de Gram-Schmidt avec les autres colonnes avant l'étape de normalisation. L'inconvénient de cette méthode, appelée « approche par déflation », est que une erreur d'estimation sur une composante se répercute sur les suivantes du fait de cette orthogonalisation. L'autre méthode, appelée « approche symétrique », orthogonalise simultanément toutes les colonnes. Elle réclame donc plus de mémoire, mais n'est pas pour autant moins rapide [HYV01, chap 14].

Voir l'ACI comme un éloignement à la gaussianité permet d'établir de forts liens avec la poursuite de projections, où le but est de chercher des directions de projection où les données sont les moins gaussiennes possibles. Ainsi, les mesures de non-gaussianité présentées dans ce paragraphe peuvent être utilisées en poursuite de projection [HYV98].

| <u>Déflation</u> | <u>Symétrique</u> |
|--|--|
| $W = \emptyset$ {Matrice initiale vide} <u>TANT QUE</u> $i \leq N_{ica}$ {Pour toutes les sources...} $w = \text{rand}(\cdot)$ $w = w - WW^T w$ {orthogonalisation} $w = w / \ w\ $ $t=0$ <u>TANT QUE</u> $t < t_{\max}$ { t_{\max} itérations max} $w_{i0} = w$ $w = E\{zg(w^T z)\} - E\{g'(w^T z)\} w^T$ $w = w - WW^T w$ $w = w / \ w\ $ <u>SI</u> $\ w - w_{i0}\ < \varepsilon$ <u>OU</u> $\ w + w_{i0}\ < \varepsilon$ $i = i + 1$ $W = [W \ w]$ <u>BREAK</u> {Source suivante} <u>FIN</u> $t = t + 1;$ <u>FIN</u> | $W = \text{rand}()$ {Matrice initiale aléatoire} $W = (WW^T)^{-1/2} W$ {Orthogonalisation} <u>POUR</u> $i : 1 \rightarrow N_{ica}$ $w_i = w_i / \ w_i\ $ <u>FIN</u> $t=0$ <u>TANT QUE</u> $t < t_{\max}$ { t_{\max} itérations maximum} $W_0 = W$ <u>POUR</u> $i : 1 \rightarrow N_{ica}$ $w_i = E\{zg(w_i^T z)\} - E\{g'(w_i^T z)\} w_i^T$ <u>FIN</u> $W = (WW^T)^{-1/2} W$ <u>SI</u> $1 - \min(\text{diag}(W * W_0)) < \varepsilon$ <u>RETOUR</u> <u>FIN</u> $t = t + 1;$ <u>FIN</u> |

Table 3.1 : Les deux versions de l’algorithme du point fixe [HYV97, HY99c, HYV01]. (a) La version par déflation orthogonalise les filtres itérativement. (b) La version symétrique fait une orthogonalisation globale. Les non linéarités testées sont indiquées dans la table 3.2

3.4.5 Liens entre les méthodes

Toutes les méthodes précédemment décrites ont bien entendu des liens entre elles, au delà du fait qu’elles résolvent toutes le problème posé par l’ACI et fort heureusement plusieurs de ces liens ont été mis en évidence. Rappelons que chacune des méthodes précédentes réunit en fait deux aspects : une « méthode statistique » d’une part permettant de mettre en évidence et de mesurer la propriété d’indépendance recherchée et une « méthode algorithmique » d’autre part permettant d’optimiser la fonction précédente. Ce sont bien entendu les liens entre les diverses « méthodes statistiques » que nous allons mettre en évidence dans ce paragraphe, puisque les différences entre algorithmes n’influent que sur l’aspect purement calculatoire (temps de convergence, mémoire requise...).

L’équivalence des approches « Infomax » et « maximum de vraisemblance » a été énoncée dans [PEA96] après que l’auteur ait montré que l’on pouvait dériver une règle d’adaptation semblable à celle de Bell et Sejnowski [BEL95] à partir de la vraisemblance. Cette démonstration est reprise dans [LEE00]. Une autre démonstration a été

| $G(t)$ | $g(t)$ | $g'(t)$ |
|----------------------------|-----------------------------------|--|
| $G_1(t) = \log \cosh(t)$ | $g_1(t) = \tanh(t)$ | $g_1'(t) = 1 - \tanh^2(t)$ |
| $G_2(t) = -\exp(-t^2 / 2)$ | $g_2(t) = t \cdot \exp(-t^2 / 2)$ | $g_2'(t) = (1 - t^2) \cdot \exp(-t^2 / 2)$ |
| $G_3(t) = t^4 / 4$ | $g_3(t) = t^3$ | $g_3'(t) = 3t^2$ |

Table 3.2 : $g(t)$ et sa dérivée $g'(t)$ sont les non linéarités utilisées dans l’algorithme du point fixe. $G(t)$ fait référence à la fonction correspondante dans la définition du contraste associé (eq. 3.39)

proposée dans [CAR97] qui a montré que les fonctions de contraste des deux approches coïncident. Plus précisément, ces deux contrastes correspondent à la divergence de Kullback-Leibler entre la distribution des estimations en sortie de la matrice W et de la distribution supposée des sources réelles \mathbf{s} . Ainsi c'est aussi le contraste associé à l'information mutuelle [CAR99] comme cela a été défini dans [COM94]. Dans cet article, ce même contraste a été mis en relation avec la négentropie, ce qui établit un lien avec les méthodes basées sur une approximation de l'information mutuelle (annulation des cumulants croisés) mais aussi celles calculées à partir d'approximations de la négentropie (éloignement à la gaussianité). La relation entre l'ACP non linéaire et d'autres critères a été étudiée dans [KAR98]. Il a aussi été montré que la règle d'apprentissage développée dans [KAR94] est équivalente à celle que Girolamy et Fyfe obtiennent avec une approche « poursuite de projection » [GIR97].

3.5 Utilisations de l'analyse en composantes indépendante

Pour toutes les méthodes présentées précédemment, les auteurs ont bien entendu appliqué leur algorithme à un cas plus ou moins concret afin de démontrer ses capacités à séparer des sources. Ces applications consistaient donc à générer quelques signaux, puis à les mélanger artificiellement avant d'utiliser l'algorithme pour retrouver avec succès les signaux originaux. Dans ce paragraphe, nous allons plutôt nous intéresser à l'utilisation de l'ACI avec des données issues du monde réel.

3.5.1 Séparation de signaux de parole

Une première application est la séparation de signaux de parole, telle que présentée dans le « problème de la soirée cocktail ». Malheureusement le modèle d'ACI présenté dans notre cadre (mixture linéaire instantanée) n'est pas très adapté pour le résoudre, d'une part parce que les signaux ont tendance à être convolués et surtout parce qu'il faut prendre en compte les délais temporels entre chaque micro comme cela est fait dans [TRK96]. De plus, dans un contexte réel, nous connaissons mal le modèle de mélange des voix, ce qui rend la séparation d'enregistrements réels difficile [NGU95]. On pourra se reporter à [TRK99] pour une revue de l'ensemble des méthodes applicables au problème convolutif.

3.5.2 Imagerie médicale

Une classe importante de problèmes résolus par le modèle instantané linéaire d'ACI concerne les applications en imagerie médicale, en particulier la détermination de l'activité cérébrale [JUN01]. Celle-ci est étudiée à l'aide de deux types d'images : les images encéphalographiques d'une part et les images obtenue par résonance magnétique d'autre part.

L'activité électrique du cerveau peut être détectée à l'aide d'enregistrement électroencéphalographiques (EEG) ou magnétoencéphalographiques (MEG) puisque toute activité électrique induit aussi bien un champ électrique que magnétique. Les ERPs (*Event-Related Potentials*) sont des EEG enregistrées sur des patients qui réagissent plusieurs fois à un même stimuli et qui ont été moyennées en vue d'augmenter leur rapport signal sur bruit. La

boîte crânienne agit comme un filtre passe-bas sur les signaux provenant du cerveau [MAK00] mais l'hypothèse de superposition linéaire des signaux reste néanmoins valide. Si on suppose d'autre part que les activations cervicales sont temporellement indépendantes, il n'en n'est pas de même spatialement puisque plusieurs lieux peuvent être actifs simultanément. Cette technique ne permet donc pas d'effectuer la localisation spatiale des sources, mais plutôt une localisation temporelle d'un ensemble d'activités. Notons néanmoins que puisque la somme d'activités indépendantes tend vers une distribution gaussienne, l'ACI peut théoriquement avoir quelques difficultés pour faire la séparation. En pratique, l'utilisation de l'algorithme de Bell & Sejnoski [BEL97, LEE99] permet de détecter des variations faibles par rapport à la distribution normale. Vigário et ses collègues ont quand à eux appliqué l'algorithme « FastICA » à des données EEG et MEG [VIG00].

L'imagerie par résonance magnétique fonctionnelle (IRMf ou fMRI : *Functional Magnetic Resonance Imaging*) est une technique permettant de détecter les zones actives du cerveau lors de l'exécution de tâches spécifiques. C'est une technique récente qui contrairement à celle qui était utilisée précédemment pour cette tâche (TEP : tomographie par émission de positrons) ne nécessite pas de traceur radioactif et peut donc être pratiquée plus souvent sur un patient. Elle utilise au contraire un marqueur naturel très commun dans l'organisme : l'oxygène. Plus précisément, l'hémoglobine perd son oxygène après être passée dans les « zones actives » du cerveau et la « désoxy-hémoglobine » résultante possède des propriétés para-magnétiques qui peuvent être détectées par des aimants puissants (0,5T à 3T). C'est donc l'effet de l'activité neuronale sur la désoxygénation sanguine qui est détectée. L'avantage immédiat par rapport aux images encéphalographiques est la possibilité de repérer spatialement les sources. L'ACI permettra donc de rechercher des zones du cerveau spatialement indépendantes pour un intervalle de temps donné, pouvant correspondre à des zones fonctionnelles [BEC03].

Bien que l'utilisation de l'ACI en imagerie médicale semble prometteuse, quelques limitations subsistent. Le modèle supposé est généralement non bruité et suppose la présence d'autant de sources que de capteurs. Dans le cas de l'EEG/MEG, cela reste donc limité par le nombre d'électrodes (une vingtaine pour des schémas standards). Par ailleurs, l'hypothèse d'indépendance temporelle peut être remise en cause quand les enregistrements sont courts, ou lorsque des événements spatialement séparés surviennent simultanément. Pour le moment, ces limitations sont surmontées à l'aide de post-traitements statistiques ou d'une interprétation humaine des résultats [JUN01, BEC03]. Il semble aussi prometteur de combiner des enregistrements encéphalographiques fournissant une bonne résolution temporelle et des enregistrements provenant de l'IRMf qui ont grande résolution spatiale.

3.5.3 Données financières

Une première application de l'ACI à des données financières a été réalisée par [BAC97]. Cette étude, quelque peu prospective, utilise comme données d'entrées le cours des actions des 28 plus grosses entreprises cotées à la bourse de Tokyo entre 1986 et 1989¹. Afin d'avoir des signaux stationnaires, ils s'intéressent en fait au « retour d'action » qui est la différence entre deux valeurs successives du cours. En appliquant l'algorithme JADE sur de telles données, ils espèrent trouver des facteurs indépendants dont l'interprétation expliquerait les structures sous-

¹ Un « crack boursier » mondial a eu lieu durant l'été 1987...

jacente des marchés d'actions. Les résultats restent néanmoins très qualitatifs. Ils montrent en particulier que l'ACI permet une mise évidence de phénomènes plus intéressants que l'ACP. De plus l'utilisation des quatre composantes indépendantes les plus dominantes (définies à partir de l'amplitude maximale) permet de retrouver l'essentiel de la variation du cours de la principale banque japonaise.

Dans [KIV98], les données utilisées sont les flux de liquidité de 40 magasins appartenant à une même chaîne sur une période de trois ans. L'algorithme « FastICA » est utilisé pour extraire cinq composantes indépendantes (la réduction de dimension est effectuée par ACP). Dans ce cas, l'interprétation de certaines de ces composantes se fait très aisément et révèle les pics de vente de Noël ou bien les baisses pendant la saison estivale. D'autres composantes peuvent avoir une interprétation plus délicate (mais d'autant plus intéressante) concernant par exemple la place relative que peut avoir la chaîne de magasin par rapport à ses concurrents.

Dans [MAL99], l'ACI est utilisée pour transformer des séries temporelles de façon à construire un prédicteur. Le processus est testé sur des données simulées et des données réelles et testé avec un prédicteur auto-régressif. Pour les deux jeux de données, le prétraitement par l'ACI permet une meilleure prédiction des séries temporelles.

Ces premières application de l'ACI aux données financières sont assez prometteuses. Néanmoins, l'ACI pré-suppose un modèle linéaire et ne prend en compte qu'un nombre restreint de composantes. Etant donné que l'évolution de telles données dépend non seulement d'indicateurs économiques mais surtout de facteurs psychologiques, il semble assez difficile d'obtenir de bonnes prédictions dans tous cas ! L'ACI semble tout de même révéler des structures intéressantes pour de tels problèmes et être un prétraitement efficace pour les méthodes existantes.

3.5.4 Caractéristiques fondamentales des images et des séquences naturelles

Selon les idées formulées par Attneave [ATT54], Barlow [BAR61] et Watanabe [WAT60], le but du système sensoriel et particulièrement le système visuel des vertébrés, est de réduire la redondance des données d'entrée afin d'en avoir une représentation interne la plus efficace possible. Dans ce cas, l'information est codée selon un code factoriel et a une structure parcimonieuse (voir chapitre 2 et le paragraphe 3.4.3). En construisant un réseau de neurone cherchant à reconstruire au mieux les images (au sens des moindres carrés) sous contrainte de rendre maximale la structure parcimonieuse des codes générés, Olshausen et Fields [OLS96] ont obtenu des unités de codage localisées et orientées. Harpur et Prager [HAP96] ont indépendamment développé un modèle semblable. Par ailleurs il a été démontré [NAD94] que le principe *infomax* [LIN88] était équivalent à l'hypothèse de réduction de redondance formulée par Barlow. L'algorithme [BEL95] étant basé sur ce principe, leurs auteurs eurent l'idée de l'appliquer à des images naturelles [BEL97] et obtinrent des filtres semblables à ceux de Olshausen et Field. Le modèle d'image supposé est que toute partie d'une image est la superposition linéaire de fonctions de bases activées par des « causes » indépendantes sous jacentes (figure 3.5 et chapitre 5). Les fonctions de base estimées à partir d'images naturelles (figure 3.6) ressemblent en première approximation à des filtres de Gabor à différentes orientations et échelles fréquentielles. Cette structure est cohérente avec les mesures effectuées par [HUB68] sur le cortex des macaques ayant révélé une organisation en colonnes par orientation et par résolution. La comparai-

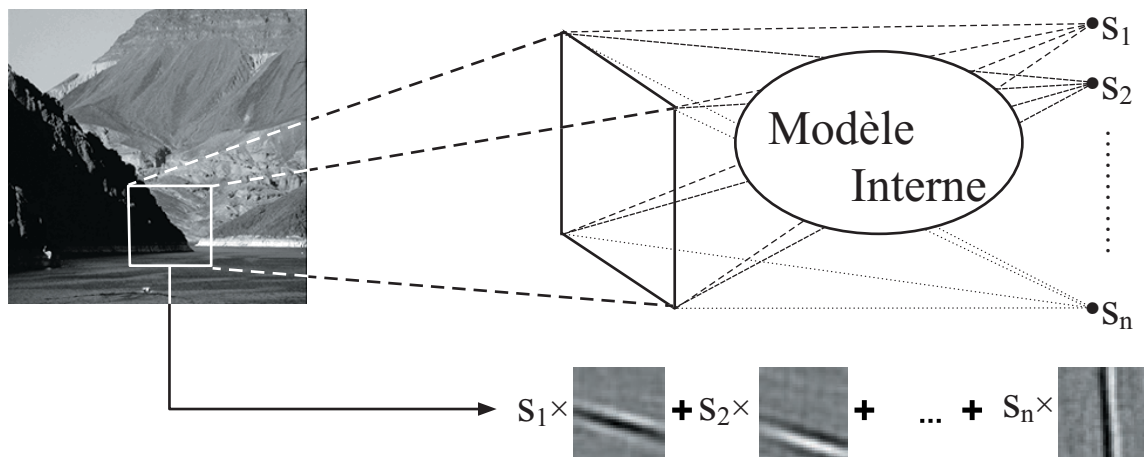


Figure 3.5: Modèle d'image permettant d'appliquer l'ACI aux images naturelles d'après [OLS96].

son entre les propriétés de ces filtres et les données physiologiques sur le cortex visuel des macaques [DEL82a, DEL82b] confirma la ressemblance [HAT98a].

Dans [HOY00], l'application de l'algorithme « FastICA » à des images en couleur fit émerger des filtres spatio-chromatiques codant la couleur selon le même schéma que le système visuel humain (*i.e* selon une opposition rouge/vert d'une part et bleu/jaune d'autre part). Dans le même article, l'algorithme a aussi été appliqué à des images binoculaires conduisant alors à des paires de filtres semblables aux cellules simples du cortex visuel.

Enfin, la même ressemblance a été constatée dans [HAT98b] lorsque des séquences d'images naturelles sont utilisées. En plus de la localisation spatiale des filtres, on observe une localisation temporelle (figure 3.7).

Ces similitudes entre les unités codantes résultant de l'ACI et les cellules du cortex visuel ont été exploitées par Hoyer pour développer divers modèles de vision biologiquement valides. De nombreux raffinements de l'ACI et d'autres avancées majeures sont présentées dans sa thèse [HOY02] et les articles qui la complètent. En dehors de la modélisation des cellules simples, il présente une modélisation des cellules complexes qui utilise les dépendances rémanentes des cellules simples conduisant à une organisation topographique des filtres (TICA), ou encore un modèle de codage neuronal spécifique des contours dans les images. Ces modèles étendus de l'ACI sont destinés à modéliser le comportement visuel des humains en respectant une architecture neuronale plausible.

3.5.5 Classification et reconnaissance d'images

Appliquée à des images naturelles, l'ACI permet de faire émerger les structures fondamentales de celles-ci (les «bords» [BEL97]). Cette capacité d'adaptation aux données a naturellement été utilisée pour des applications de reconnaissance et de discrimination d'image.

[BAR98] a appliqué l'algorithme [BEL95] avec le modèle d'image précédent sur des images de visage. Les fonctions de bases obtenues ressemblant alors à des « visages propres » (*eigenfaces*) telles que celles qui résultent de l'application de l'ACP. Ils implantèrent aussi une seconde architecture revenant à appliquer l'ACI sur la trans-

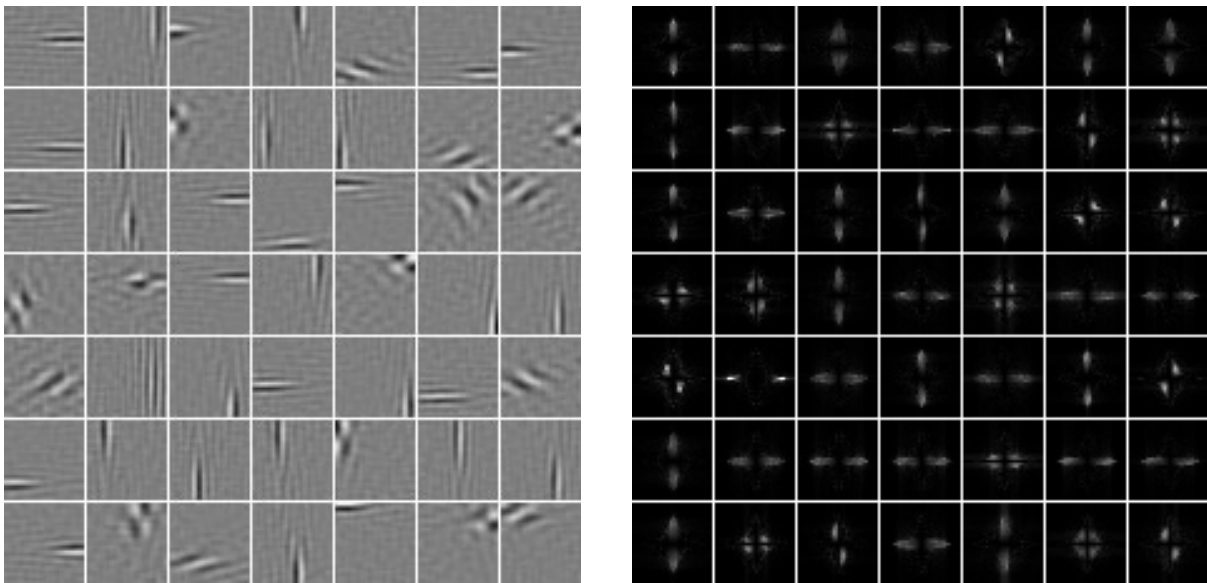


Figure 3.6: Exemple de fonctions de bases extraites d'images naturelles par ACI (droite) et du module de leur transformée de Fourier (gauche)

posée de la matrice de données telle que présentée dans le modèle précédent. Les sources et leurs estimées sont alors des imagerie et les visages sont caractérisés par les coefficients de la matrice de séparation. Il est alors remarquable de constater que les sources indépendantes représentent des morceaux de visages caractéristiques: lèvre supérieure, lèvre inférieure, yeux, sourcils... Les deux protocoles ont été testés avec une base d'images de visages où chaque sujet avait posé avec une expression neutre pour l'apprentissage et une autre expression (joie, colère...) pour le test. La même séance de photo avait été reproduite deux ans plus tard, fournissant ainsi deux autres ensembles d'images de test. Les images sont caractérisées par le code indiqué précédemment et la distance entre deux images est égale à l'angle entre leurs vecteurs caractéristiques. Les performances de discrimination sont évaluées avec un classifieur aux K plus proches voisins. Pour les trois bases de test, les deux protocoles d'ACI permettent une meilleure reconnaissance des visages que l'ACP, mais sont à peu près équivalentes entre elles.

La discrimination d'objets a été abordée dans [LAB99a] et a suscité un vif intérêt chez de nombreux chercheurs [ASH02]. Tout comme dans l'expérience précédente, le principe est d'appliquer le modèle d'image de Olshausen et Field à des images d'objets, généralement représentés par une collection de photos prises sous différents angles de vue. Une partie des images sert à l'apprentissage et le test est réalisé sur les images restantes. L'objet est caractérisé par la collection des réponses énergétiques moyennes des filtres ainsi générés aux images. Une sélection ou une pondération des filtres est faite en fonction de leur pouvoir discriminant évaluant sa capacité à séparer deux objets sur la base d'apprentissage. Lors de la phase de test, les distances entre les objets-test et les prototypes calculés lors de la phase d'apprentissage sont évaluées par la norme euclidienne pondérée par le pouvoir discriminant des filtres et la plus petite d'entre elles permet d'attribuer l'objet à la classe correspondante. Là encore les tests montrent que l'ACI donne de meilleures performances que l'ACP.

La reconnaissance de scènes naturelles au moyen de l'ACI est traitée en détail au chapitre 6. Les méthodes

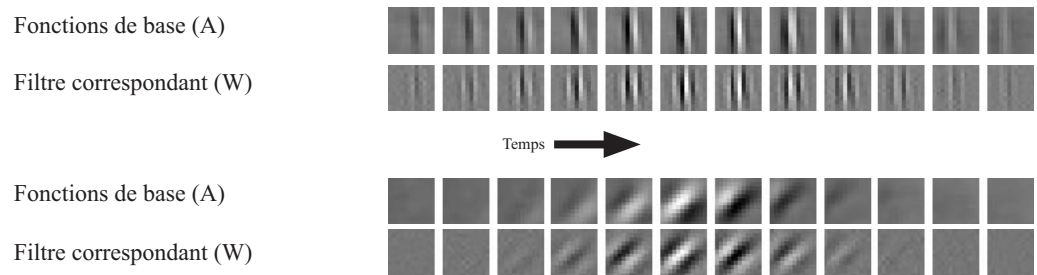


Figure 3.7: Deux exemple de “séquences indépendantes” extraites de séquences naturelles [HAT98b].

existantes [LAB99b, LAB99c, LAB01] exploitent la capacité de l’ACI à s’adapter aux statistiques des données concernées, les images naturelles en particulier. [LAB99b, LAB01] aborde le problème d’une façon semblable à la reconnaissance d’objets, tandis que [LAB99c] propose une méthode pour caractériser les images qui repose sur l’identification du filtre ayant donné la réponse la plus forte et non sur la réponse elle-même. [LEN01] analyse des images hyperspectrales à l’aide de paquets d’ondelettes, puis applique l’ACI à ces coefficients afin de retrouver les fréquences pures présentes dans la scène.

Dans le contexte de la modélisation de données multimédias, l’ACI présente un potentiel suscitant un intérêt croissant [LAR03]. L’intérêt de l’ACI est que la fusion des caractéristiques très hétérogènes, établie par extraction des sources multimédia combinant les informations de nature différente, semble correspondre plus fidèlement à celle qui serait pratiquée par un humain que d’autres techniques [LAR03, KOL02]. Par exemple, [KOL02] combine du texte, caractérisé par l’histogramme d’occurrence des termes le composant [SAL89] et des images caractérisée par des attributs bas niveau de couleur (histogramme de la description HSV) et de texture (banque de filtres de Gabor). Les trois caractéristiques sont centrées, normalisées (variance unitaire) et réduites par ACP, puis les sources multimédias extraites par ACI. Le classifieur de Bayes conduit à la discrimination de trois catégories («sport», «aviation» et «paintbal») combinant le texte et les images de pages web. Le texte (seul) permet une meilleure classification que les caractéristiques d’images, mais la combinaison des trois réduit l’erreur significativement par rapport au taux de classification du texte seul.

Dans le domaine de la fusion audio-visuelle en particulier, la cohérence entre le mouvement des lèvres d’un locuteur et les caractéristiques spectrales de son discours est exploitée par Sodoyer et ses collègues [SOD02]. Les performances de séparation sont pour le moment équivalentes à d’autres algorithmes de séparation de sources, mais cette approche est très prometteuse, en particulier quand le nombre de micros est inférieur au nombre de sources. Dans ce cas, l’information visuelle pourrait permettre une extraction plus performante de l’une des sources.

3.5.6 Autres applications de l’ACI

Dans ce paragraphe nous recensons quelques autres applications utilisant l’ACI que nous avons jugé remarquables, étonnantes ou attrayantes, mais qui sont encore assez prospectives, au sens où elles commencent juste à être explorées et que parfois les résultats ne montrent pas toujours d’améliorations majeures par rapport à d’autres techniques.

Chapitre 3

C'est le cas pour la compression d'image qui est abordée dans [FEI03]. Les auteurs synthétisent des « fonctions de base » conformément à la méthode précédemment décrite en utilisant l'algorithme *FastICA* et les orthogonalisent au moyen d'une transformation de Karhunen-Loève ou d'une procédure de Gram-Schmidt. Cet algorithme est appliqué à quatre types de bases d'images : des images naturelles, des images d'empreintes digitales, des visages et enfin des images synthétiques. Les images sont divisées en blocs et ceux-ci sont caractérisés par leur code après projection sur les fonctions de bases ICA (éventuellement un nombre restreint), puis quantifiés par un quantificateur de Lloyd dont l'apprentissage a été fait hors ligne, suivi d'un codage entropique. La valeur moyenne de chaque bloc est codée séparément étant donné que l'ACI travaille sur des données centrées. Après décodage des mesures quantitatives et qualitatives sont faites en comparaison des algorithmes de référence dans le domaine de la compression d'image: JPEG, JPEG2000 et l'encodeur utilisé par la police fédérale des Etats-Unis (WSQ) pour les images d'empreintes digitales. Les résultats sont souvent meilleurs que pour le JPEG (surtout avec les visages), mais restent inférieurs à ceux de JPEG2000. Pour les images d'empreintes digitales, l'ACI a des performances proches de WSQ, tout deux surpassant JPEG mais restant inférieurs à JPEG2000.

[HYVO1a] a développé une méthode de débruitage des images naturelles utilisant l'ACI lorsque le bruit est additif et gaussien. La méthode utilise une matrice de séparation W qui est estimée à partir d'images naturelles selon la méthode décrite dans le paragraphe précédent, puis qui est orthogonalisée globalement (méthode symétrique de la table 3.1). Appliquée à des données bruitées $\mathbf{z} = \mathbf{x} + \mathbf{n}$ où \mathbf{n} est un bruit additif gaussien cela donne alors la somme d'une estimation des sources indépendantes et de $W\mathbf{n}$ qui est aussi gaussien. En supposant une forme très sur-gaussienne pour les sources, les auteurs dérivent plusieurs classes de fonctions modélisant ces densités qui appliquées au mélange permet d'effectuer la séparation. Des tests sont effectués sur des images naturelles et sont commentés qualitativement (appréciation visuelle). Les résultats sont visiblement meilleurs qu'un débruitage par filtre de Wiener mais ne sont pas comparés à d'autres méthodes.

[FAR99] utilise aussi un modèle d'image différent de [OLS96] dans le but d'étudier les transparences. Ils prennent en compte le cas d'une transparence additive, typiquement celle qui peut être observée lorsqu'un personnage regarde un tableau ou un paysage à travers une vitre. Le problème est alors de séparer l'image du tableau ou du paysage et le reflet de l'observateur. Afin de réaliser une telle tâche les auteurs ont besoin d'au moins deux prises de vue différentes de la scène et supposent ensuite que les deux objets à séparer sont indépendants et se mélangent additivement. L'ACI s'applique alors parfaitement au problème et les résultats sont assez convaincants. Néanmoins cette méthode ne peut pas s'appliquer à tous les types de transparence car l'hypothèse d'indépendance statistique et surtout de mélange linéaire n'est pas toujours valide, ou bien n'est pas valide partout dans l'image et peut dépendre de l'angle de prise de vue. On trouvera dans [PIN03] une présentation des problèmes liés à l'étude des transparences et des méthodes existantes pour les résoudre.

Bibliographie

- [ABR00] Abramovich F., Benjamini Y., Donoho D., Johnstone I. "Adapting to unknown sparsity by controlling the false discovery rate". Rapport technique N° 2000-19, Stanford univ., dept. stat, 2000.
- [ACH01] Achard S., Pham D.T., "Blind source separation in post nonlinear mixtures". Actes *ICA 2001*, San Diego, CA, USA, 9-13 décembre 2001.
- [ALL99] Alleyson D. "Le traitement du signal chromatique dans la rétine: un modèle de base pour la perception humaine des couleurs". *Manuscrit de thèse*, UJF, Grenoble, France, 3 Mars 1999.
- [AMA96] Amari A., Cichocki A., Yang H.H., "A new learning algorithm for blind signal separation". Dans: *advances in neural information processing systems*, vol 8, editors D. Touretzky, M. Mozer, and M. Hasselmo, pp 757-763, MIT press, Cambridge MA, 1996.
- [AMA98a] Amari S.I., Cichocki A. "Adaptative Blind Signal Processing - Neural Network Approaches". *Proceedings of the IEEE*, vol 86, N° 10, Octobre 1998.
- [AMA98b] Amari S.-I., "Natural Gradient works efficiently in learning", *Neural computation*, 10, pp 251-276, 1998.
- [AMA03] Amato U., Antoniadis A., Grégoire G., "Independent Component Discriminant Analysis". *International Mathematical Journal*, vol 3, N° 7, pp 735-753, 2003.
- [ASH02] Ashutosh G., Agarwal S., Huang T.S., " Fusion of Global and Local Information for Object Detection". Actes *ICPR 2002*, Québec City, Canada, 2002.
- [ATI92] Atick J.J., "Could information theory provide an ecological theory of sensory coding ?". *Network: computation in neural systems*, N° 3, pp 213-251, 1992.
- [ATI92a] Atick J.J., Redlich A.N., "What does the retina know about natural scenes?". *Neural computation*, 4, 196-210, 1992.
- [ATI93] Atick J.J., Redlich A.N., "Convergent Algorithm for sensory receptive field development", *Neural Computation*, 5, pp 45-60, 1993.
- [ATT54] Attneave F., "Some informational aspects of visual perception". *Psychological Reviews*, 61:183-93, 1954.
- [BAC97] Back A.D., Weigend A.S. "A first application of independent component analysis to extracting structure from stock returns". *International journal of neural systems*, vol 8, N° 5, octobre 1997.
- [BAR61] Barlow HB., "Possible principles underlying the transformation of sensory messages". *Sensory Communication*, ed. WA Rosenblith, pp. 217-34. Cambridge, MA: MIT Press, 1961.

Bibliographie

- [BAR98] Barlett M., Lades H.M., Sejnowski T.J. "Independent component representation for face recognition", Actes du *SPIE symposium on electronic imaging: science and technology, conference on human vision and electronic imaging III*, San Jose, Californie, janvier 1998.
- [BAR01a] Barlow H., "Redundancy reduction revisited". *Network : computation in neural systems*, 12, 241-253, 2001.
- [BAR01b] Barlow, H., "The Exploitation of Regularities in the Environment by the Brain", *Behavioral and Brain Sciences*, 24, <http://www.bbsonline.org/documents/a/00/00/04/25/>, 2001.
- [BAR82] Bar-Ness Y., Carlin J.W., and Steinberg M.L., "Bootstrapping Adaptive Cross Pol Cancelers for Satellite Communication". Actes *The International Conference on Communication*, N° 4F.5, Philadelphie, PA, Etats-Unis, juin 13-17, 1982.
- [BAS96] Baseville M., "Information: entropies, divergences et moyennes". *Publication interne* N° 1020, INRIA, Mai 1996.
- [BEC03] Beckmann C.F., Smith S.M., "probabilistic independent component analysis for functional magnetic resonance imaging", *fMRI Technical Report TR02CB1*, accepté à IEEE TMI, 2003.
- [BEL95] Bell A.J., Sejnowski T.J., "An information-maximisation approach to blind separation and blind deconvolution". *Neural computation*, vol 7, pp 1129-1159, 1995
- [BEL97] Bell A.J., Sejnowski T.J., "The Independent Component of Natural Scenes are Edge Filter". *Vision Research*, vol 37, n° 23, pp 3327-3338, 1997.
- [BIE82] Biederman I., Mezzanotte R.J., Rabinowitz J.C., "Scene perception: detecting and judging objects undergoing relational violations". *Cognitive psychology*, vol 14, pp 143-177, 1982.
- [BIE87] Biederman I., "Recognition-by-components: a theory of human understanding". *Psychological review*, 94:115-47, 1987
- [BIE88] Biederman I., "Aspect and extensions of a theory of human image understanding". Dans *Computational processes in human vision: an interdisciplinary perspective*, editeur Pylyshyn Z., pp 370-428. Norwood, NJ: Ablex, 1988.
- [BIE01] Biederman I., "Recognizing Depth-Rotated Objects: a review of recent research and theory". *Spatial Vision*, vol 13, pp 241-253, 2001/
- [BOD00] Bodt E. de, Cottrell M., "Bootstrapping self-organising maps to assess the statistical significance of local proximity". Actes *European symposium on artificial neural networks (ESANN'00)*, Bruges (Belgique), 26-28 Avril 2000.
- [BOS00] Bosch H., "Object segmentation and recognition using temporal coding and independent component analysis". Université de Genève, 31 mars 2000.
- [BOV90] Bovik, A. C., Clark, M. and Geisler, W.S. "Multichannel Texture Analysis Using Localized Spatial Filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, pp. 55-73, 1990
- [BRA99] Brandt S., "Use of shape in content-based image retrieval". *Thèse de doctorat*, Espoo, Finlande, 1999.
- [BUN00] Bunke H., "Recent developments in graph matching". Actes *International Conference on Pattern Recognition*, vol 2, pp 117-124, Barcelone, Espagne, 2000.
- [BUR83] Burt P.J., Adelson E.H., "The laplacian pyramid as a compact image code". *IEEE transaction on communication*, vol COM-31, pp 532-540, avril 1983.

- [BUR89] Burman, P. "A comparative study of ordinary cross-validation, v-fold cross validation and the repeated learning testing methods". *Biometrika*, 76(3), 503 - 514, 1989.
- [CAN98] Candès E., "Ridgelets: theory and application". *Manuscrit de thèse*, Université de Stanford, 1998.
- [CAN00] Candès E., Donoho D.L., "Curvelets: optimally sparse representation of objects with edges". Dans *Curve and surface fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, L.L. Schumaker (eds), Vanderbilt university press, Nashville, TN. ISBN 0-8265-1357-3, 2000.
- [CAR89] Cardoso J.-F., "Source separation using higher order moments". Actes *IEEE ICASSP*, pp 2109-2112, Glasgow, Ecosse, UK, 1989.
- [CAR93] Cardoso J.-F., Souloumiac A. "Blind beamforming for non gaussian signals". *IEE-proceedings-F*, vol 140, N°6, pp 362-370, décembre 1993.
- [CAR97] Cardoso J.-F. "Infomax and maximum likelihood for blind source separation". *IEEE signal processing letters*, vol 4, N° 4, pp 112-115, avril 1997.
- [CAR98] Cardoso J.F., "Blind Signal Separation: Statistical Principles". *Proceedings of the IEEE*, vol 86, N° 10, Octobre 1998.
- [CAR99] Cardoso J.-F., "High-order contrasts for independent component analysis". *Neural computation*, vol 11, pp 157-192, 1999.
- [CHA02] Chauvin A., Héroult J., Marendaz C., Peyrin C., "Natural scene perception: visual attractors and image neural computation and psychology". Dans W. Lowe et J. Bullinaria (Eds.), *Connexionist Models of Cognition and Perception*, World scientific press, 2002.
- [CHE92] Chéhikian A., "Algorithmes optimaux pour la génération de pyramides d'images passe-bas et laplaciennes". *Traitement du signal*, vol 9, N°4, pp 297-307, 1992.
- [CHO01] Choi S., Cichocki A., Zhang L., Amari S.-I. "Approximate maximum likelihood source separation using the natural gradient". *Third IEEE signal processing advances in wireless communication*, Taiwan, 20-23 mars 2001.
- [CIC96] Cichocki A., Unbehauen R., "Robust neural network with on-line learning for blind identification and blind separation of sources". *IEEE transaction on circuits and systems I: fundamental theory and application*, 43(11):894-906, 1996.
- [COL94] Coleman T.F., Li Y., "On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to biund". *Mathematical programming*, vol 67, N° 2, pp 189-224, 1994.
- [COL96] Coleman T.F., Li Y., "An interior trust region approach for nonlinear minimization subject to bounds". *SIAM journal on optimization*, vol 6, pp 418-445, 1996.
- [COM89] Comon P., "Separation of sources using high-order cumulants". *SPIE conference on advanced algorithms and architectures for signal processing*, vol. Real-time signal processing XII, pp 170-181, San Diego, California, 8-10 août 1989.
- [COM91] Comon P., Jutten C., Héroult J., "Blind separation of sources, Part II: problem statement". *Signal Processing*, vol 24, N° 1, pp 11-20, juillet 1991.
- [COM92] Comon P. "Independent Component Analysis". *International signal processing workshop on high-order statistics*, Chamrousse, France, 10-12 juillet 1991, pp 111-120; republié dans J.L Lacoume, ed., *High order statistics*, Elsevier, Amsterdam, 1992, pp 29-38.
- [COM94] Comon P., "Independent Component Analysis, A new concept?". *Signal Processing*, vol. 36, N° 3, pp 287-314, 1994.

Bibliographie

- [COM95] Comon P., "Quelques développements récents en traitement du signal". *Habilitation à diriger des recherches*, université de Nice Sophia-Antipolis, 18 septembre 1995.
- [COX00] Cox I.J., Miller M.L., Minka T.P., Papathomas T.V., Yianilos P.N., "The bayesian image retrieval system, PicHunter: theory, implementation, and psychological experiments". *IEEE transaction on Image processing*, vol 9, N° 1, janvier 2000.
- [DAU85] Daugman J. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized two-dimensional visual cortical filters". *Journal Optical Soc. Am.*, 2:1160- 1168, 1985.
- [DEB78] De Boor C., "A practical guide to splines". *Springer-Verlag*, New York, 1978.
- [DEB 97] Del Bimbo A., Pala. P., "Visual image retrieval by elastic matching of user sketches". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):121--132, 1997.
- [DEB99] Del Bimbo A., "Visual Information Retrival". *Morgan Kaufmann Publishers*, San Francisco, 1999.
- [DEL82a] De Valois R.L., Yund E.W., Hepler N., "The orientation and direction selectivity of cells in macaque visual cortex". *Vision research*, vol 22, pp 531-544, 1982
- [DEL82b] De Valois R.L., Albrecht D.G., Thorell L.G., "Spatial frequency selectivity of cells in macaque visual cortex". *Vision research*, vol 22, pp 545-559, 1982.
- [DEL95] Delfosse N., Loubaton P., "Adaptive blind separation of independent sources: a deflation approach". *Signal processing*, vol 45, pp 59-83, 1995.
- [DEL97] Delorme, A, Makeig, S, et al. "EEGLAB: Matlab Toolbox for Electrophysiological Research". WWW Site, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, www.sccn.ucsd.edu/eeglab [World Wide Web Publication], 1997.
- [DEM94] Demartines P., "Analyse de données par réseau de neurones auto-organisés". *Manuscrit de thèse*, INPG, 1995.
- [DEM97] Demartines P., Héroult J., "Curvilinear Component Analysis: a self-organising neural network for non-linear mapping of data sets". *IEEE transaction on neural networks*, 8(1):148-154, 1997
- [DEN02] Denquive N., Tarroux P. "Multi-resolution codes for scene categorization". *Actes European symposium on artificial neural networks (ESANN02)*, d-side publi., ISBN 2-930307, pp 281-287, Bruges, Belgique, 24-26 avril 2002.
- [DON98] Donoho D.L., Vertelli M., DeVore R.A., Daubechie I., "Data compression and harmonic anlysis", *IEEE transaction on information theory*, vol 6, pp 2435-2476, 1998.
- [DON99] Donato G., Barlett M.S., Hager J.C., Ekman P., Sejnowski T.J., "Classifying facial actions". *IEEE transaction on pattern analysis and machine intelligence*, vol 21, N° 10, pp 974-989, 1999.
- [DON00] Donoho D.L., "Orthonormal ridgelet and linear singularities", *SIAM J. Math Anal.*, 31, pp 1062-1099, 2000
- [DON01] Donoho D.L., Flesia A.G., "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?". *Network: computation in neural systems*, vol 12, pp 371-393, 2001.
- [DON02] Donoho D.L., Flesia A.G., "Digital Ridgelet Transform based on true Ridge Functions". Rapport technique, université de Stanford, 22 janvier 2002.
- [DOV00a] Do M. N., Vetterli M., "Image denoising using orthonormal finite ridgelet transform". *Actes SPIE on wavelet applications in signal and image processing VIII*, San Diego, Californie, Etats-Unis, 2000.
- [DOV00b] Do M. N., Vetterli M., "Orthonormal finite ridgelet transform for image compression". *Actes IEEE International Conference on Image Processing (ICIP)*, Vancouver, Canada, September 2000

- [DOV02] Do M.N., Vetterli M., "Wavelet-based texture retrieval using generalised gaussian density and Kulback-Leibler distance". *IEEE transaction on image processing*, vol 11, N° 2, février 2002.
- [DRE02] Dréo J., Siarry P., "Un nouvel algorithme de colonie de fourmis exploitant le concept d'hétérarchie pour l'optimisation en variables continues". NSI'2002, La Londe les Maures, France, 15-18 septembre 2002.
- [DUC03] Duchêne C., « Traitement de données multidimensionnelles par Analyse en Composantes Curvili-gnes ». Rapport de DEA, université de Cergy-Pontoise, 2003.
- [EFR93] Efron, B, Tibschirani, R.J., "An introduction to the Bootstrap". *Monographs on statistics and Applied Probability*. Chapman & Hall, New York, 1993.
- [FAR99] Farid H., Adelson E.H., "Separating Reflections from Images using independent component analysis". *Journal of the optical society of america*, 16(9):2136-2145, 1999.
- [FEI03] Feirreira A, Figueiredo M.A.T "Image compression using orthogonalised independent component bases". *IEEE workshop on Neural Network for Signal Processing*, Toulouse, France, 17-19 septembre 2003.
- [FIE87] Field D.J., "Relations between the statistics of natural images and the response properties of cortical cells". *Journal of the Optical Society of America*, vol 4, N° 12, pp 2379-2393, 1987.
- [FOL90] Földiak P., (1990), Forming sparse representation by local anti-Hebbian learning, *Biological Cybernetics*, vol 64, pp. 165-170, 1990.
- [FRE91] Freeman, W.T., Adelson, E.H., "The design and use of steerable filters". *IEEE transaction on Pattern Analysis and Machine intelligence*, 13 (9), pp 891-906, 1991.
- [FRI74] Friedman J.H., Tukey J.W., "A projection pursuit algorithm for exploratory data analysis". *IEEE transaction on computers*, c-23(9):881-890, 1974.
- [FYF00] Fyfe C., "Artificial Neural Networks and Information Theory". Cours, Université de Paisley, 2000.
- [GAB46] Gabor D., "Theory of communication", *Journal of IEEE*, 93:429-457, 1946.
- [GAE90] Gaeta M., Lacoume J.L., "Source separation without prior knowledge: the maximum likelihood solution". *Dans Actes EUSIPCO'90 - Signal Processing V: Theories and Applications*, L. Torres, E. Masgrau et M.A. Lagunas (eds), pp 621-624, Barcelone, Espagne, 1990.
- [GAR01] Garrard P., Lambon Ralph M.A., Hodges J.R., Patterson K., "Prototypicallity, distinctiveness, and intercorrelation: analyse of the semantic attributes of living and nonliving concepts". *Cognitive neuropsychology*, vol 18, N° 2, pp 125-174, 2001.
- [GAR02] Garg A., Agarwal S. and Huang T.S., "Fusion of local and global information for Object detection," *Actes International conference on Pattern Recognition (ICPR02)*, 2002.
- [GIB66] Gibson J.J, "The perception of the visual world". *Houghton Mifflin*, Boston, 1966.
- [GIR97] Girolami M., Fyfe C., "An extended exploratory projection pursuit network with linear and nonlinear anti-hebbian lateral connections applied to the cocktail party problem". *Neural networks*, vol 10, N° 9, pp 1607-1618, 1997.
- [GOR94] Gokani M.M., Picard R.W., "Texture orientation for sorting photos "at a glance"". *IEEE conference on pattern recognition*, vol 1, pp 459-464, Jérusalem, Israël, Octobre 1994.
- [GRO84] Grossmann A., Morlet J. "Decomposition of Hardy functions into square integrable wavelets of constant shape". *SIAM Journal of Math. Anal.*, 15(4) : 723-736, juillet 1984.

Bibliographie

- [GUE00] Guérin-Dugué A., Oliva A., "Classification of scene photographs from local orientations features". *Pattern Recognition Letters*, 21, pp 1135-1140, 2000.
- [GUY01] Guyader N, Hérault J., "Représentation espace-fréquence pour la catégorisation d'images". Actes *GRETSI 2001*, Toulouse, France, 2001.
- [GUY03] Guyon I., Elisseeff A., "An introduction to variable and feature selection". *Journal of machine learning research*, 3, pp 1157-1182, 2003.
- [HAP96] Harpur G.F., Prager R.W. "Development of low entropy coding in a recurrent network". *Network: computation in neural systems*, 7, pp 277-284, 1996.
- [HAR96] Harroy F., Lacoume J.-L., "Maximum likelihood estimators and Cramer-Rao bounds in source separation", *Signal processing*, vol 55, pp 167-177, 1996.
- [HAT98a] Hateren J.H. van, Schaaf A. van der, "Independent component filters of natural images compared with simple cells in primary visual cortex". *Proceedings of the Royal Society Series B*, 265, pp 359-366, 1998
- [HAT98b] Hateren J.H. van, Ruderman D.L., "Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex". *Proceedings of the Royal Society Series B*, 265, pp 2315-2320, 1998.
- [HAY94] Haykin, Ed., "Blind deconvolution". Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [HEN94] Henery, R.J., "Methods for comparison". Dans: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds), *Machine learning, Neural and Statistical Classification*. Ellis Horwood, 1994.
- [HEN99] Henderson J.M., Hollingworth A., "High-level scene perception". *Annual review of Psychology*, vol 50, pp 243-271, 1999.
- [HER85] Hérault J., Jutten C. et Ans B., "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé". Actes *du Xième colloque GRETSI*, Nice, France, volume2, pages1017-1022, Mai1985.
- [HER97] Hérault J, Oliva A., Guérin-Dugué A., "Scene categorisation by curvilinear component analysis of low frequency spectra". Actes *ESANN'97*, pp 91-96, Bruges, 16-18 avril 1997.
- [HER01] Hérault J., "De la rétine biologique aux circuits neuromorphiques". Dans "Les système de vision", chap 3, J.M. Jolion (Ed.), IC2 col, *Hermès*, 2001.
- [HER02] Hérault J., Guérin-Dugué A., Villemain P., "Searching for the embedded manifolds in high-dimensional data, problems and unsolved questions". Actes *ESANN'96*, Bruges, Belgique, 2002.
- [HOD56] Hodges J.L., Lehman E.L., "The efficiency of some non-parametric competitors on the t-test". *Annals of the Mathematical Statistics*, 27:324-335, 1956.
- [HOP82] Hopfield J.J., "Neural networks and physical systems with emergent collective computational abilities,". *Proc. Nat. Acad. Sci.*, vol. 79, pp 2554-2558, Apr. 1982.
- [HOT33] Hotelling H., "Analysis of a complex of statistical variables into principal components ". *Journal of Educational Psychology*, 24, pp 417-441, 1933.
- [HOY00] Hoyer P.O., Hyvärinen A., "Independent Component Analysis Applied to Feature Extraction from Colour and Stereo Images". *Network: Computation in Neural Systems*, 11(3):191-210, 2000.
- [HOY02] Hoyer P.O., "Probabilistic models of early vision". *Manuscrit de thèse*, Espoo, Finlande, 2002.
- [HUA99] Huang J., Mumford D., "Statistics of Natural Images and Models". Actes *IEEE Conference Computer Vision and Pattern Recognition*, Fort Collins (Colorado), Etats-Unis, pp 541-547, 1999.

- [HUB68] Hubel D.H., Wiesel T.N., "Receptive fields and functional architecture of monkey striate cortex". *Journal of physiology*, 195, pp 215-244, 1968.
- [HUB85] Huber P.J., "Projection pursuit". *The Annals of Statistics*, 13(2):435-475, 1985.
- [HUM00] Hummel J.E., "Where view-based theories break down: the role of structure in shape perception and object recognition". Dans E. Dietrich & A. Markman (Eds). *Cognitive Dynamics: conceptual change in humans and machines*, pp 157-185, Hillsdale, NJ: Erlbaum, 2000.
- [HUR97] Hurri J., "Independent component analysis of image data". *Master's thesis*, Espoo, Finlande, 1997.
- [HYV97] Hyvärinen A., Oja E., "A fast fixed-point algorithm for independent component analysis", *Neural computation*, vol 9, N° 7, pp 1483-1492, 1997
- [HYV98] Hyvärinen A., "New approximations of differential entropy for independent component analysis and projection pursuit". Dans *Advances in Neural Information Processing Systems 10*, pages 273-279. MIT Press, 1998.
- [HYV99a] Hyvärinen A., Pajunen P., "Nonlinear Independent Component Analysis: Existence and Uniqueness Results". *Neural Networks*, vol 12, N° 3, pp 429--439, 1999
- [HYV99b] Hyvärinen A., "Survey on Independent Component Analysis", *Neural Computing Surveys*, vol 2, pp 94-128, 1999.
- [HYV99c] Hyvärinen A., "Fast and robust fixed-point algorithms for independent component analysis". *IEEE transaction on neural networks*, vol 10, N°3, 626-634, 1999.
- [HYV01] Hyvärinen A., Karhunen J., Oja E., "Independent Component Analysis". *John Wiley & Sons*, 2001.
- [HYV01a] Hyvärinen A., Hoyer P.O., Oja E. "Image Denoising by Sparse Code Shrinkage". Dans *S. Haykin and B. Kosko (eds), Intelligent Signal Processing*, IEEE Press, 2001
- [HYV01b] Hyvärinen A., Hoyer P., "A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images". *Vision research*, 41, pp 2413-2423, 2001.
- [HYV02] Hyvärinen A., Inki M., "Estimating overcomplete independent component bases for image windows.". *Journal of Mathematical Imaging and Vision*, 17:139-152, 2002
- [ITT98] Itti L., Koch C., Niebur E., "A model of saliency-based visual attention for rapid scene analysis". *IEEE transaction on pattern analysis and machine intelligence*, vol 20, pp 1254-1259, 1998.
- [IZE91] Izenman, A.J., "Recent developments in non parametric density estimation". *Journal of the American Statistical Association*, 86 (413), 204-224, 1991.
- [JAI97] Jain A.K., Ratha N, Lakshmanan S, "Object detection using Gabor filters" *Pattern Recognition*, 30, 295-309, 1997.
- [JAI00] Jain A.K., Duin R.P.W., Mao J., "Statistical pattern recognition: a review". *IEEE transaction on pattern analysis and machine intelligence*, vol 1, N°22, janvier 2000.
- [JOH02] Johansson B., "A survey on : Content Based Search in Image Databases". <http://www.isy.liu.se/cvl/Projects/VISIT-bjojo/survey/surveyonCBIR/index.html>, 2002.
- [JON87] Jones M.C, Sibson R., "What is projection pursuit?". *Journal of the Royal Statistical Society, serie A*, 150:1-36, 1987.
- [JPE00] JPEG2000 part 1 final committee draft version 1.0. Technical report, ISO/IEC FCD15444-1, March 2000.
- [JUN01] Jung T.-P., Makeig S., McKeown M.J., Bell A.J., Lee T.-W., Sejnowski T.J., "Imaging brain dynamics using independent component analysis". *Proceedings of the IEEE*, vol 89, N° 7, juillet 2001.

Bibliographie

- [JUT88] Jutten C., Héroult J., "ICA versus PCA". *Dans Actes EUSIPCO 88- Signal Processing IV: Theories and Applications*, J.L Lacoume, A. Chehikian, N. Martin, J. Malbos (Eds), pages 643-646, Grenoble, France, 1988.
- [JUT91] Jutten C., Héroult J., "Blind separation of sources, Part I: An adaptative algorithm based on neuro-mimetic architecture", *Signal Processing*, vol 24, N° 1, pp 1-10, juillet 1991.
- [JUT00] Jutten, C., Taleb, A., "Source separation: From dusk till dawn". Actes *ICA 2000*, pages 15-26 (papier invité), Helsinki, Finland, June 2000.
- [JUT03] Jutten, C., Karhunen J., "Advances in Nonlinear Blind Source Separation". Actes *ICA2003*, pp 245-256, Nara, Japon, 2003.
- [KAR94] Karhunen J., Joutsensalo J., "Representation and separation of signals using nonlinear PCA type learning". *Neural Networks*, 7(1):113-127, 1994.
- [KAR95] Karhunen J., Joutsensalo J., "Generalizations of principal component analysis, optimization problems, and neural networks". *Neural Networks*, 8(4):549-562, 1995.
- [KAR98] Karhunen J., Pajunen P., Oja E., "The nonlinear PCA criterion in blind source separation: relations with other approaches". *Neurocomputing*, vol 22, pp 5-20, 1998.
- [KIV98] Kiviluoto K., Oja E. "Independent component analysis for parallel financial time series". Actes *ICONIP98*, Kitakyushu, Japon. 'S Usui et T. Omori, eds), vol 2, (Tockyo, Japon), pp 895-898, APNNA, JNNS., Ohmsha, Octobre 1998.
- [KIR01] Kirkpatrick, K., "Object recognition". In R. G. Cook (Ed.), *Avian visual cognition* [En ligne à : www.pigeon.psy.tufts.edu/avc/kirkpatrick/], 2001
- [KOC85] Koch C., Ullman S. "Shifts in selective visual attention : towards the underlying neural circuitry", *Human Neurobiology*, vol 4 : pp219-227, 1985.
- [KOF35] Koffka K., "Principles of Gestalt Psychology". *Lund Humphries*, Londres, 1935.
- [KOH84] Kohonen T. "Self-organization and associative memory", *Springer-Verlag*, 1984.
- [KOH95] Kohonen T. "Self-organizing maps", *Springer*, 1995.
- [KOL02] Kolenda T., Hansen L.K., Larsen J., Winther O. "Independent component analysis for understanding multimedia content". *Actes du workshop IEEE Neural Network for Signal Processing XII*, pp 757-766, Martigny, alais, Suisse, 4-6 septembre 2002
- [KOO92] Kooperberg C., Stone C.J., "Log spline density estimation for censored data". *J. Comput. Graph. Stat.*, 1, 301-328, 1992.
- [KUN93] Kunt M., Granlund G., Kocher M., "Traitement numérique des images". *Presses polytechniques et universitaires romandes et CNET-ENST*, Lausanne, 1993.
- [KUNT00] Kunt M., Coray G., Granlund G., Haton J-P., Ingold R., Kocher M., "Reconnaissance des formes et analyse de scènes". *Presses polytechniques et universitaires romandes et CNET-France Télécom*, Lausanne, 2000.
- [LAA00] Laaksonen J., Koskela M., Laakso S., Oja E., "PicSOM - content-based image retrieval with self-organizing maps". *Pattern recognition letters*, 21, pp 1199-1207, 2000.
- [LAB99a] Labbi A., Bosch H., Pellegrini C., Gerstner W. "Viewpoint-Invariant object recognition using independent component analysis". Actes *NOLTA 99*, Hawaï, Etats-Unis, 28 nov-3 dec 1999.

- [LAB99b] Labbi A., Bosch H., Pellegrini C., "Image categorization using independent component analysis". *ACAI workshop on biologically inspired machine learning (BIML'99)*, conférencier invité, 14 juillet, Crete, Grèce.
- [LAB99c] Labbi, A., "Sparse-Distributed Codes for Image Categorization". Résumé de projet sur l'ACI et le codage des images, 1999.
- [LAB01] Labbi A., Bosch H., Pellegrini C., "High order statistics for image classification". *International Journal of Neural Systems*, vol 11, N° 4, pp 371-377, 2001.
- [LAC92] Lacoume J.-L., Ruiz P., "Separation of independent sources from correlated inputs". *IEEE transaction on signal processing*, 40(12):3074-3078, 1992
- [LAC97] Lacoume J.-L., Amblard P.-O., Comon P., "Statistiques d'ordre supérieurs pour le traitement du signal". *Masson*, 1997.
- [LAR03] Larsen J., Hansen L.K., Kolenda T., Nielsen F.A., "Independent Component Analysis in Multimedia Modeling". conférencier invité *ICA2003*, Nara, Japan, 1-4 Avril, pp. 687-696, 2003.
- [LAU81] Laughlin, S. "A simple coding procedure enhances a neuron's information capacity", *Z. Naturforsch*, c 36, 910-2, 1981.
- [LEE96] Lee T.S., "Image representation using 2D gabor wavelets". *IEEE transaction on pattern analysis and machine intelligence*, vol 18, N°10, 1996.
- [LEE98] Lee T.W., "Independent Component Analysis, theory and applications". *Kluwer Academic Publishers*, Boston, 1998.
- [LEE99] Lee T.-W., Girolami M., Sejnowski T.J., "Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources". *Neural computation*, vol 11, N°2, pp 409-433, 1999.
- [LEN01] Lennon, M., Mercier G., Mouchot M.C., Hubert-Moy L., "Spectral unmixing of hyperspectral images with the Independent Component Analysis and wavelet packets". *Actes International Geoscience and remote sensing symposium*, Sydney, Australie, Juillet 2001.
- [LEE00] Lee T.W., Girolami M., Bell A.J., Sejnowski T.J., "A unifying information-theoretic framework for independent component analysis", *Computer & mathematics with application*, 39(11):1-21, 2000.
- [LEW99] Lewicki M.S., Olshausen B.A., "A probabilistic framework for the adaptation and comparison of image codes". *Journal of the Optical Society of America*, A 16:1587-1601, 1999.
- [LEW00] Lewicki S., Sejnowski T.J., "Learning overcomplete representation", *Neural computation*, vol 12, N°2, pp 337-365, 2000.
- [LIN88] Linsker, R. "Self-organization in a perceptual network". *IEEE Computer*, 21:105-117, 1988.
- [LIU03] Liu C., Wechsler H., "Independent Component Analysis of gabor features for face recognition". *IEEE transaction on neural networks*, vol 14, N° 4, pp 919-928, 2003.
- [LOG95] Logothetis N.K., Pauls J., Poggio T., "Shape representation in the inferior temporal cortex of monkeys". *Current Biology*, vol 5, N° 5, pp 552-563, 1995.
- [MAK00] Makeig, S *et al.* "Frequently Asked Questions about ICA applied to EEG and MEG data". WWW Site, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, www.sccn.ucsd.edu/eeglab www.sccn.ucsd.edu/~scott/icafaq.html [World Wide Web Publication], 2000
- [MAL99] Mălăroiu S., Kiviluoto K., Oja E. "Time series prediction with independent component analysis". *Actes AIT'99 (Advances Investment Technologies)*, Gold coast, Australie, 20-21 décembre 1999.

Bibliographie

- [MAL00] Mallat S., “Une exploration des signaux en ondelettes”, *Les éditions de l'école polytechnique*, Palaiseau, 2000.
- [MAN96] Manjunath B.S., Ma W.Y., “Texture features for browsing and retrieval of image data”. *IEEE pattern analysis and machine intelligence*, vol 18, pp 837-842, août 1996.
- [MAO92] Mao J., Jain A.K., “Texture classification and segmentation using multiresolution simultaneous autoregressive models”, *Pattern recognition*, vol 25, N° 2, pp 173-188, 1992.
- [MAR78] Marr D., Nishihara H.K., “Representation and recognition of the spatial organization of tree-dimensional shapes”. *Proceeding of the Royal Society of London*, B, 200, pp 269-294, 1978.
- [MAR82] Marr D., “Vision: a computational investigation into the human representation and processing of visual information”. *Freeman*, San Francisco, 1982.
- [MAS03] Massot C., Héroult J., “Extraction d'indices d'orientation et de forme dans les scènes naturelles par modèles corticaux”, Actes *GRETSI03*, toulouse, France, 2003.
- [MIN75] Minsky M., “A framework for representing knowledge”. In Patrick Henry Winston (Eds.), *The Psychology of Computer Vision*, McGraw-Hill, New York, USA, 1975
- [MOJ01] Mojsilovic A., Rogowitz B. “Capturing image semantic with low-level descriptors”. Actes *International conference on image processing*, vol 1, pp 18-21, Thessaloniki, Grèce, 7-10 octobre 2001.
- [MOR98] Moreau E., Macchi O., “Self-adaptative source separation, part II: comparison of the direct, feedback, and mixed linear network”. *IEEE transaction on signal processing*, vol 46, N° 1, pp 39-50, janvier 1998.
- [NAD94] Nadal J.-P., Parga N., “Non linear neurons in the low noise limit: a factorial code maximizes information transfer”. *Network: computation in neural systems*, 5:565-581, 1994.
- [NAS92] Nason G.P., “Design and choice of projection indices”. *Thèse de doctorat*, université de Bath, 1992.
- [NEI67] Neisser U., “Cognitive Psychology”. *New-York: Appleton-Century-Crofts*, 1967.
- [NGU95] Nguyen Thi H.-L., Jutten C., “Blind source separation for convolutive mixtures”. *Signal processing*, vol 45, N° 2, pp 209-229, 1995.
- [OJA82] Oja E., “A simplified neuron model as a principal component analyser”. *Journal of Mathematical Biology*, vol 15, pp 267-273, 1982.
- [OJA91] Oja E., Ogawa H., Wangviwattana J., “Learning in non-linear constrained Hebbian networks”. Dans *T. Kohonen et al. (Eds.), Artificial neural networks*, pp 385-390, Amsterdam, Pays Bas, 1991.
- [OJA92] Oja E., “Principal Components, Minor Analysis, and Linear Neural Networks”. *Neural Networks*, 5(6):927-935, 1992.
- [OJA97] Oja E., “The nonlinear PCA learning rule in independent component analysis”. *Neurocomputing*, 17(1):25-46, 1997.
- [OLI97] Oliva A., Schyns P., “Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli”. *Cognitive psychology*, vol 34, pp 72-107, 1997.
- [OLI99] Oliva A., Torralba A., Guérin-Dugué A., Héroult J. “Global semantic classification of scenes using power spectrum templates”. Actes *Challenge of Image Retrieval. Elect. work. in Computing series*, springer-Verlag, Newcastle, 1999.
- [OLI01] Oliva O., Torralba A., “Modeling the shape of the scene: a holistic representation of the spatial envelope”. *International journal of computer vision*, 42(3):145-175, 2001.

- [OLI03] Oliva, A., Torralba, A., Castelhana, M. S., and Henderson, J. M. "Top-Down control of visual attention in object detection". *Actes IEEE International Conference on Image Processing*, 14-17 septembre, Barcelone, Espagne, 2003.
- [OLS96] Olshausen B.A, Fields D.J., "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature*, 381:607-609, 1996.
- [OLS97] Olshausen B.A, Fields D.J., "Sparse coding with an overcomplete basis set: a strategy employed by V1?". *Vision research*, vol 37, N° 23, pp 3311-3325, 1997.
- [PAR02] Park S.-J., Shin J.-K., Lee M; "Biologically inspired saliency map model for bottom-up visual attention". *Second workshop on biologically motivated computer vision*, Tübingen, Allemagne, pp 418-426, Springer-verlag, Berlin, Heidelberg, 22-24 novembre 2002.
- [PAL99] Palmer, S. E., "Vision science: From Photons to Phenomenology". *Cambridge, MA: Bradford Books/MIT Press*, 1999.
- [PEA96] Pearlmutter B.A., Parra L.C., "A context-sensitive generalisation of ICA". *Actes ICONIP'96*, pp 151-157, Hong-Kong, 1996.
- [PES01] Pesquet-popescu B., Pesquet J-C., "Ondelettes et applications". *Techniques de l'ingénieur*, 2001.
- [PHA92] Pham D.T., Garat P., Jutten C., "Separation of a mixture of independent sources through a maximum likelihood approach". *Actes EUSIPCO*, pp 771-774, 1992.
- [PHA97] Pham D.T., Garat P., "Blind separation of mixture of independent sources through a quasimaximum likelihood approach". *IEEE transactions on signal processing*, 45(7):1712-1725, 1997.
- [PIN03] Pingault M., "Estimation du mouvement d'objets transparents". *Manuscrit de thèse*, Université Joseph Fourier, Grenoble, France, 2003.
- [POG90] Poggio T., Edelman S., "A network that learns to recognition three-dimensional objects", *Nature*, vol 343, pp 263-266, 1990.
- [POL83] Pollen, D. and Ronner, S "Visual cortical neurons as localized spatial frequency filters". *IEEE Transaction. on Systems, Man, and Cybernetics*, 13:907--916, 1983.
- [POS84] Posner, M.I., Cohen, Y., "Components of Visual Orienting". H. Bouma and D.G. Bouwhuis, eds., *Attention and Performance*, vol. 10, pp. 531-556. Hilldale, N.J.: Erlbaum, 1984.
- [POT76] Potter M., "Short-term conceptual memory for pictures". *Journal of experimental psychology: human learning and memory*, vol 2, pp 509-522, 1976.
- [PUZ99] Puzicha J., Rubner Y., Tomasi C., Buhmann J.M., "Empirical evaluation of dissimilarity measures for color and texture". *Acte International Conference on Computer Vision*, Kerkyra, Corfu, Grèce, pp 1165-1173, 1999.
- [RAN99] Randen T., Håkon Husøy J., "Filtering for texture classification: a comparative study". *IEEE transaction on pattern analysis and machine intelligence*, vol 21, N° 4, avril 1999.
- [RIJ79] Rijsbergen C.J. van, "Information retrieval" (2nd ed.), *Butterworths*, Londres, 1979.
- [RIP02] Ripley B., Kooperberg C., Logspline density estimation package, version 1.0-7, disponible à <http://www.cran.r-project.org/>, 28 août 2002.
- [ROG98] Rogowitz B.E., Frese T., Smith J.R, Bouman C.A., Kalin E., "Perceptual image similarity experiment". *ISAT/SPIE Symposium on Electronic Imaging: Science and Technology, Conference on Human Vision and Electronic Imaging III*, pp. 576-590, 1998

Bibliographie

- [ROS75] Rosch, E., "Cognitive representations of semantic categories". *Journal of Experimental Psychology*, General 104, pp. 192-233, 1975.
- [RUD94] Ruderman D.L., "The statistics of natural images". *Network: computation in neural systems*, vol 5, pp 517-548, 1994.
- [RUI97] Rui Y., Huang T.S., Chang S-F., "Image retrieval: past, present, and future". *Actes International Symposium on Multimedia Information Processing*, Taiwan, décembre 1997.
- [SAL89] Salton G., "Automatic text processing: the transformation, analysis, and retrieval of information by computer", *Adison-Wesley*, 1989.
- [SAM69] Sammon J.W., A nonlinear mapping algorithm for data structure analysis. *IEEE transaction on Computers*, C-18(5):401-409, 1969.
- [SAN89] Sanger T.D., "Optimal unsupervised learning in a single-layer linear feedforwrd network". *Neural Netwoks*, 2(6), 459-473, 1989.
- [SAN99] Santini S, Jain R., " Similarity measures ". *IEEE transaction on pattern analysis and machine intelligence* , vol 21, N° 9, pp 871-883, 1999.
- [SAN01] Santini S., "Exploratory image databases : content-based retrieval". *Academic press*, Londres, 2001.
- [SAN02] Sanfeliu A., Alquézar R., Andrade J., Climent J., Serratos F., Vergés J., "Graph-based representations and techniques for image processing and image analysis". *Pattern recognition* 35, N°3, pp 639-650, mars 2002.
- [SAP90] Saporta G., "Probabilités, analyse des données et statistiques". Editions technip, paris, 1990.
- [SCH94] Schyns P., Oliva A., "From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition". *Psychological Science*, vol 5, pp 195-200, 1994.
- [SCH96] Schaaf van der A., Hateren van J.H., "Modelling the power spectra of natural images: statistics and information". *Vision research*, 36, pp 2759-2770, 1996.
- [SCH97] Schmid C., Mohr R., "Mise en correspondance par invariants locaux". *Traitement du signal*, vol 13, N° 6, pp 591-618, 1997.
- [SHA49] Shannon, C.E. & Weaver, W. (Ed.). "The mathematical theory of communication". Urbana: Univ. Illinois Press, 1949.
- [SHA76] Shafer G., "A mathematical theory of evidence". Princeton university press, 1976.
- [SHE72] Shepard R.N., Romney K., Nerlove S.B., "Multidimensional scaling: Theory and Application in the behavioral sciences (volume 1: theory), *Seminar press*, New York, 1972.
- [SIL86] Silverman B.W., "Density estimation for statistics and data analysis", *Chapman & Hall*, Londres, 1986
- [SIM01] Simoncelli E.P., Olshausen B.A., "Natural image statistics and neural representation". *Annual review of neuroscience*, 24:1193-216, 2001.
- [SME00] Smeulders A.W.M., Worring M., Santini S., Gupta A., Jain R., "Content-based image retrieval at the end of the early years", *IEEE transaction on pattern analysis and machine intelligence*, vol 22, N° 12, décembre 2000.
- [SOD02] Sodoyer D., Schwartz J.-L., Girin L., Klinskisch J., Jutten C., "Separation of audio-visual speech sources: a new approach exploiting the audio-visual coherence of speech stimuli". *EURASIP journal on applied signal processing*, vol 1, pp 1165-1173, 2002.

- [STA02] Starck J.-L., Candès E.J., Donoho D.L., “The curvelet transform for image denoising”. *IEEE transaction on image processing*, vol 11, N°6, juin 2002.
- [STA03] Starck J.-L., Murtagh F., Candès E.J., Donoho D.L., “Gray and color image contrast enhancement by the curvelet transform”. *IEEE transaction on image processing*, vol 12, N°6, juin 2003.
- [STE00] Stetter M., Obermayer K., “Biology and theory of early vision in mammals“. *Brains and Biological Networks*, H. H. Szu (ed), chapter . INNS press, 2000.
- [STR95] Stricker M., Orengo M., “Similarity of color images”. Actes *SPIE 95*, San jose, CA, USA, 1995.
- [SWA91] Swain M.J., Ballard D.H., “Color indexing”. *International journal of computer vision*, vol 7, N° 1, pp 11-32, 1991.
- [SZU98] Szummer M., Picard R.W., “Indoor-outdoor image classification”. *IEEE international workshop on content-based access of image and video databases*, Bombay, Inde, janvier 1998.
- [UNS95] Unser M., “Texture classification and segmentation using wavelet frames”. *IEEE transaction on image processing*, vol 4, pp 1549-1560, novembre 1995.
- [TAL99] Taleb A., Jutten C., “Source Separation in Post Non Linear Mixtures”. *IEEE Transaction on Signal Processing*, Vol. 47, n° 10, pp. 2807-20, Octobre 1999.
- [TAR95] Tarr M.J., Bülthoff H.H., “Is human object recognition better described by geon structural descriptions or by multiple views?” *Journal of experimental psychology: human perception and performance*, vo 21, pp 1494-1505, 1995.[TAR98] Tarr, MJ, Bülthoff H.H., “Image-based object recognition in man, monkey and machine”. *Cognition* 67, 1-20, 1998.
- [TAR00] Tarr M.J., “Visual pattern recognition”. *Encyclopedia of psychology*, A.E. Kazdin (Ed.), Washington, DC: American Psychological Association, 2000.
- [THI97] Thiria S., Lechevallier Y., Gascuel O., Canu S., “Statistique et méthodes neuronales”. *Dunod*, Paris, 1997
- [TON91] Tong L. Liu R.-W., Soon V.C., Huang Y.-F., “Indeterminacy and identifiability of blind identification”. *IEEE Transaction on Signal Processing*, Vol. 38, n° 5, pp. 499-509, mai 1991.
- [TON93] Tong L., Inouye Y., Liu R.W., “Waveform-Preserving Blind estimation of multiple independent sources”. *IEEE transaction on signal processing*, 41(7):2461-2470, 1993.
- [TOR52] Torgerson W.S., « Multidimensional scaling, part I : theory and method ». *Psychometrika*, vol 17, pp 401-419, 1952.
- [TOR99] Torralba A., Oliva O., “Semantic organization of scenes using discriminant structural templates”. Actes *international conference on computer vision*, pp 1253-1258, Korfu, Grèce, septembre 1999.
- [TOR02] Torralba A., Oliva A., “Depth estimation from image structure”. *IEEE transaction on pattern analysis and machine intelligence*, vol 24, N° 9, pp 1226-1238, septembre 2002
- [TOR03a] Torralba A., “Contextual priming for object detection”. *International Journal of Computer Vision*, vol 53, N° 2, pp 157-167, juillet 2003.
- [TOR03b] Torralba A, Oliva A., “Statistics of Natural image categories”. *Network: computation in neural systems*, vol14, pp 391-412, 2003.
- [TRE80] Treisman A., Gelade G., “A feature integration theory of attention”. *Cognitive psychology*, vol 12, pp 97-136, 1980.
- [TRE88] Treisman A., “Preattentive processing in vision”. Dans *computational processes in human vision: an interdisciplinary perspective*, Zelon Pylyshyn (Eds), pp 341-369, 1988.

Bibliographie

- [TRK96] Torkkola K., "Blind deparation of delayed sources based on information maximization". Actes *ICASSP*, Atlanta, GA, Etats-Unis, 7-10 mai 1996.
- [TRK99] Torkkola K., "Blind separationfor audio signals - are we there yet?". Actes *ICA99*, pp 239-244, Aussois, France, janvier 1999.
- [ULL96] Ullman S. "High level vision: object recognition and visual cognition". *Cambridge MA: MIT press*, 1996.
- [VAI98] Vailaya A., A. Jain, A., Zhang H.J., "On Image Classification: City vs. Landscape", *Pattern recognitions*, vol 31, N° 12, pp 1921-1935, 1998.
- [VAI01] Vailaya A., Figueiredo M.A.T., Jain A.K., Zhang H.J., "Image classification for content-based indexing". *IEEE transaction on image processing*, vol 10, N° 1, janvier 2001.
- [VER01] Verpeaux B., "Analyse et amélioration d'une chaîne de catégorisation d'images par ACI". Rapport de stage de seconde année, ENSERG, 2001.
- [VIG00] Vigário R., Oja E., "Independence: a new criterion for the analysis of the electromagnetic fields in the global brain". *Neural Netwoks*, 13, pp891-907, 2000.
- [WAT60] Watanabe, S., "Information-theoretical aspects of inductive and deductie inference". *IBM journal of research and development*, 4, pp 208-231, 1960.
- [WIL00] Willmore B., Watters P. A., Tolhurst D.V., "A comparison of natural-image-based models of simple-cell coding", *Perception*, vol 29, pp 1017-1040.
- [WOL89] Wolfe J. M., Cave K. R., Franzel S. L., "Guided search: an alternative to the feature integration model for visual search". *Journal of experimental psychology: human perception & performance*, 15, pp 419-433, 1989.
- [YAN97] Yang H.H., Amari S.-I., "Adative online learning algorithms for blind separation: maximum entropy and minimum mutual information". *Neural computation*, vol 9, N° 7, pp 1457-1482, 1997.
- [ZAD78] Zadeh L.A., "Fuzzy sets as a basis for a theory of possibility", *Fuzzy sets and systems*, vol 1, N°1, pp 3-28, 1978.
- [ZHU03] Zhu S-C., "Statistical modeling and conceptualization of visual patterns". *IEEE transaction on pattern analysis and machine intelligence*, vol 25, N°6, pp 691-712, juin 2003.