Analyse en composantes indépendantes et séparation de sources : approches basées sur la théorie de l'information et recherche de conditions minimales

Jean-Pierre Nadal(1) **et Nestor Parga**(2)

(1)Laboratoire de Physique Statistique de l'E.N.S.
Laboratoire associé au C.N.R.S. (U.R.A. 1306) et aux Universités Paris VI et Paris VII
Ecole Normale Supérieure
24, rue Lhomond, 75231 Paris Cedex 05, France

(2)Departamento de Física Teórica
Universidad Autónoma de Madrid
Cantoblanco, 28049 Madrid, Espagne

Résumé

Nous passons en revue des travaux récents sur l'analyse en composantes indépendante (ACI), montrant que différentes approches, basées sur la théorie de l'information et de l'estimation statistique, sont essentiellement équivalentes.
Nous considérons le problème plus particulier de la séparation aveugle de source (SAS), pour lequel on peut chercher à déterminer des conditions minimales sur les sorties d'un filtre qui, si elles sont vérifiées, garantissent que la SAS a été réalisée.

Abstract

We revue recent results on independent component analysis (ICA), showing that several approaches, based on information theory and statistical estimation concepts, are essentially equivalent.
We consider more specifically the problem of blind source separation (BSS), for which one may try to find minimal conditions on a filter output which, if satisfied, guarantee that BSS has been achieved.

1 Introduction

Depuis quelques années un nombre important de nouveaux résultats ont été obtenus sur le (vieux) problème de la *séparation aveugle de sources* (SAS), et plus généralement sur l'*analyse en composantes indépendantes* (ACI). Ces travaux se situent à l'interface entre le traitement du signal vu par les ingénieurs, et les neurosciences computationnelles, plus précisément la modélisation des systèmes sensoriels. De fait, deux courants principaux sont à l'origine de la plupart des recherches actuelles. D'un coté, on trouve des travaux s'appuyant sur une réflexion théorique développée dès les années 50[4, 6] sur le traitement effectué par les systèmes sensoriels, et en particulier sur la nature du codage sensoriel. Ainsi, selon le biologiste H. Barlow[6], un codage (ou représentation neuronale) *a priori* utile pour le système nerveux doit permettre une interprétation rapide de l'information, avec la meilleure utilisation possible des ressources disponibles (nombre de neurones, temps de réponse d'une cellule nerveuse...). Barlow montre que ces aspects d'efficacité et de lisibilité conduisent au concept clef de *réduction de redondance*, qui implique qu'un codage optimal est *factoriel* : chaque neurone doit coder pour une caractéristique du signal statistiquement indépendante de celles codées par les autres neurones. Autrement dit, dans le langage actuel propre au Traitement du Signal, le système sensoriel doit réaliser une ACI des stimuli auxquels il est soumis. Il faut noter que dès le début[4, 6] ces considérations qualitatives ont été formalisées à l'aide des outils de la Théorie de l'Information[11]. Ce n'est cependant que depuis la fin des années 80 que des études analytiques systématiques, utilisant les concepts issus de la Théorie de l'Information, ont été développées pour explorer les conséquences de ces idées pour la modélisation des systèmes sensoriels (voir par exemple [5, 3]), tandis que l'application au Traitement du Signal est toute récente [27, 9].

D'un autre coté, et tout à fait indépendamment, tout un domaine s'est créé il y a une quinzaine d'années en Traitement du Signal autour de la séparation de source à partir d'un algorithme inspiré de données empiriques en... neurosciences ! C'est d'ailleurs à l'occasion d'un colloque GRETSI, en 1985, que ce travail de J. Hérault, C. Jutten et B. Ans a été présenté pour la première fois. Bien qu'il existe des précurseurs [8], on doit à Hérault et Jutten[20] la popularisation de ce sujet grâce à une heuristique aux performances qu'on pourrait qualifier d'insolentes vue la simplicité de l'algorithme. Mais ce fut aussi l'occasion de comprendre que ce problème de SAS peut, ou doit pouvoir, s'étudier mathématiquement, et de très nombreux travaux théoriques ou algorithmiques ont vu le jour (voir par exemple [19, 15, 16, 24]). Sur le plan théorique, un progrès notable est l'identification de l'ensemble des propriétés qui, si elles sont vérifiées, font d'une quantité une bonne fonction de coût à optimiser pour réaliser une SAS ; une telle quantité est alors appelée un *contraste*[15]. Un contraste dépend des propriétés statistiques des sorties du réseau de neurone (ou du filtre) utilisé pour analyser le signal. Les contrastes les plus fréquemment considérés impliquent des cumulants d'ordre élevés[15, 12, 25]. Un autre, s'exprime comme une quantité d'information[15], laquelle s'identifie en fait à la redondance *à la Barlow*.

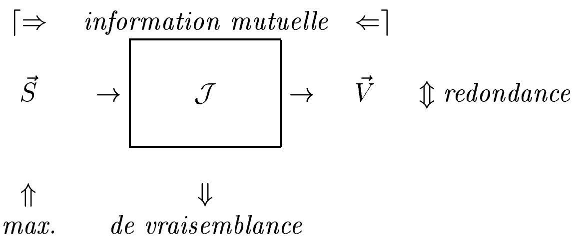
L'objet de notre contribution à ce colloque est en particulier de montrer comment, précisément, les différentes approches ont finie par converger, aboutissant à une clarification théorique remarquable. Ainsi au moins trois points de vue, basés sur différents aspects de la théorie de l'information, se sont révélés fondamentalement équivalents. On peut faire l'historique de ces développements récents de la façon suivante. En 1994 nous avons démontré[27] que la maximisation de l'*information mutuelle* entre l'entrée et la sortie du filtre (du réseau de neurones) conduit à une ACI. Il en résulte d'abord l'équivalence entre la maximisation de l'information et la minimisation de la *redondance* évoquée plus haut. Ensuite il résulte que la SAS peut être obtenue par maximisation de cette quantité d'information avec un réseau de neurone d'architecture simple mais avec fonctions de transfert non linéaires[27]. La mise en œuvre algorithmique, due à Bell et Sejnowski[9], a conduit à réaliser que la fonction de coût ainsi obtenue, pour ce cas particulier de la SAS, est en fait la même que celle introduite antérieurement par Pham *et col* à partir du critère statistique de maximisation de la vraisemblance[29, 18]. La démonstration de l'équivalence des deux approches se révèle elle-même très simple[14].

Dans la section 2 ci-après nous résumons ces équivalences. Ensuite, dans la section 3, nous montrons que de nouvelles directions émergent si on tient compte de la spécificité de la SAS. En particulier nous donnons des conditions sur des cumulants des sorties du système de traitement du signal (filtre, réseau de neurones) qui, si elles sont satisfaites, garantissent que la séparation de source a été réalisée.

2 Vers "la pensée unique" ?

2.1 Cadre général

La figure ci-dessous illustrera notre propos.



Nous considérons un signal multidimensionnel $\vec{S} = \{S_j, j = 1, ..., N\}$, traité par un système (e.g. un réseau de neurones) charactérisé par un ensemble de paramètres ajustables $\mathcal{J}$. La sortie du système est dénotée $\vec{V}$. Pour simplifier la discussion, nous nous limitons au cas où le nombre de sorties est égal au nombre N d'entrées, $\vec{V} = \{V_i, i = 1, ..., N\}$, et nous supposons que chaque sortie peut s'écrire :

$$V_i = f_i(h_i) \tag{1}$$

où la "fonction de transfert" f_i est une fonction déterministe, inversible - par exemple monotone strictement croissante entre 0 et 1 - d'une variable réelle, et le scalaire h_i une fonction déterministe du signal d'entrée, $\vec{V}$. Pour une généralisation au cas de sorties stochastiques, voir [26]. Dans le cas particulier d'intérêt pour la SAS,

$$h_i = \sum_j J_{i,j}\ S_j \tag{2}$$

où les $J_{i,j}$ sont les coefficients du filtre, ou poids du réseau de neurones. Ce qui suit s'applique cependant au cas plus général de h_i qui peuvent être des fonctions non linéaires du signal. Nous dénoterons dans tous les cas par $\mathbf{J}$ l'ensemble des paramètres définissant ces fonctions, de sorte que l'ensemble (potentiellement) ajustable $\mathcal{J}$ est constitué de $\mathbf{J}$ et des fonctions de transfert $\{f_i(.), i = 1, ..., N\}$. Nous nous limitons encore en ne considérant que le cas où les signaux $\vec{S}(t), t = 1, ...$ aux instants $t = 1, ...$ sont des variables aléatoires indépendantes, de même loi, $P(\vec{S})$.

2.2 Information mutuelle et redondance

Dans cette section nous résumons les résultats principaux publiés dans [27]. L'*information mutuelle* I entre les variables aléatoires $\vec{V}$ (la sortie) et $\vec{S}$ (l'entrée) est une mesure de la dépendance statistique entre ces deux variables.

Avec les hypothèses énoncées ci-dessus, la sortie est une fonction déterministe de l'entrée, l'information mutuelle est infinie. En fait, en introduisant un bruit aussi petit qu'on veut sur la sortie (qui peut représenter la précision numérique finie sur la lecture des sorties), on obtient[27] que I est une constante (d'autant plus grande que le bruit est petit), plus un terme qui n'est rien d'autre que l'entropie de la sortie :

$$I = cst. + H[Q], \tag{3}$$

$$H(Q) = -\int d^N V\ Q(\vec{V})\ \log Q(\vec{V}), \tag{4}$$

$Q(\vec{V})$ étant la loi de probabilité de la sortie $\vec{V}$ induite par la loi des entrées et les paramètres du réseau, $\mathcal{J}$. Ainsi, plus la sortie parait "désordonnée" (plus l'entropie est grande), plus elle contient d'information sur le signal.

Puisque chaque V_i est une fonction déterministe de h_i, nous pouvons faire le changement de variable $\vec{V} \rightarrow \vec{h}$ dans l'expression (4) de $H(Q)$. Compte tenu de $dV_i = f_i'(h_i)dh_i$

pour chaque i, et en notant par $\Psi(\vec{h})$ la loi de probabilité de $\vec{h}$, on obtient

$$H(Q) = -\int d^N h \Psi(\vec{h}) \, \ln \frac{\Psi(\vec{h})}{\prod_{i=1}^N f_i'(h_i)} \tag{5}$$

Noter que Ψ dépend des poids $\mathbf{J}$ mais pas des fonctions de transfert f_i. Si $\Psi_i(h_i)$ est la distribution marginale de h_i, il est commode de réécrire l'opposé de l'entropie (5) comme une somme de termes, tous positifs ou nuls :

$$-H(Q) = \mathcal{R} + \sum_i \mathcal{D}_i \tag{6}$$

où $\mathcal{R}$ est la *redondance* de $\vec{h}$ (la sortie avant d'appliquer les fonctions de transfert), c'est-à dire l'information mutuelle entre les N variables aléatoires h_i :

$$\mathcal{R} \equiv \int d^N h \Psi(\vec{h}) \, \ln \frac{\Psi(\vec{h})}{\prod_{i=1}^N \Psi_i(h_i)} \tag{7}$$

Chaque $\mathcal{D}_i$ est la divergence de Kullback entre la loi marginale Ψ_i et la loi définie par la dérivée de la fonction de transfert, f_i' :

$$\mathcal{D}_i \equiv \int dh_i \Psi_i(h_i) \ln \frac{\Psi_i(h_i)}{f_i'(h_i)}. \tag{8}$$

Maximiser l'information mutuelle c'est donc, de façon équivalente, :

- minimiser la redondance $\mathcal{R}$, c'est-à dire trouver les $\mathbf{J}$, s'ils existent, qui réalisent un code factoriel,

$$\Psi(\vec{h}) = \prod_{i=1}^N \Psi_i(h_i), \tag{9}$$

  autrement dit qui réalisent une ACI. Cette redondance, que H. Barlow préconise de minimiser, est aussi le contraste considéré par Comon[15] (sous le nom d'*information mutuelle* entre les h_i, à ne pas confondre avec l'information mutuelle entre l'entrée et la sortie du réseau que nous avons considérée ci-dessus).

- et adapter les fonctions de transfert f_i suivant

$$f_i'(h_i) = \Psi_i(h_i), \quad i = 1, ..., N. \tag{10}$$

Dans le contexte de la modélisation en neurosciences, de nombreuses études ont porté sur l'adaptation d'un réseau de neurones, soit selon le critère de maximisation de l'information (le principe "infomax" de Linsker[23]) [21, 23, 32] soit selon le principe proposé par Barlow de minimisation de la redondance [3, 7, 30, 22]. La grande similitude des prédictions des différents modèles, en ce qui concerne la partie linéaire du traitement, est ainsi en grande partie expliquée par notre résultat qui montre que la maximisation de l'information implique la minimisation de la redondance - dans le cas d'un traitement peu ou pas bruité (pour l'analyse d'un système bruité, voir [3, 17]).

2.3 SAS : information mutuelle et maximum de vraisemblance

Comme nous l'avions exposé dans [27], l'une des conséquences de ce résultat est que pour réaliser l'ACI , et en particulier la SAS, on peut utiliser l'information mutuelle comme fonction de coût, à condition d'adapter aussi les fonctions de transfert.

Dans le cas particulier de la SAS, c'est à dire pour des données $\vec{S}$ qui sont un mélange linéaire de sources indépendantes, les $\vec{h}$ sont linéaires, selon (2). Les paramètres $\mathbf{J}$ sont alors la matrice des poids $\mathbf{J} = \{J_{i,j}, i = 1, ..., N, j = 1, ..., N\}$. L'information mutuelle prend dans ce cas une expression simple en fonction de cette matrice des poids et des fonctions de transfert, comme montré dans [9] :

$$I = \text{constante} + \ln \mid \mathbf{J} \mid + \sum_i \int d^N S P(\vec{S}) \, \log f_i'(h_i[\vec{S}]) \tag{11}$$

où $\mid \mathbf{J} \mid$ est le déterminant de la matrice $\mathbf{J}$.

Or, il se trouve que cette quantité à maximiser (11) est essentiellement identique au critère de maximum de vraisemblance introduit par Pham et al [29] quelques années auparavant. Cette approche est *a priori* complètement différente. On part d'un modèle explicite de génération des données, et on choisit les paramètres du modèle (ici la matrice de mélange et les probabilités de distribution des sources) qui maximisent la probabilité de générer les données $\vec{S}$ effectivement observées. Le lien entre les deux approches est facile à comprendre. Choisir un réseau de neurones comme ci-dessus c'est en fait faire implicitement une hypothèse sur la structure des données. Si on suppose que la matrice de mélange est $\mathbf{J}^{-1}$ et que les probabilités de distribution des sources sont $\Psi_i \equiv f_i'$, alors h_i donné par (2) est la valeur de la source numero i ; la probabilité d'observer $\vec{S}$ (étant donné $\mathbf{J}^{-1}$ et les Ψ_i), est donc, les sources étant indépendantes,

$$P(\vec{S}) = \mid \mathbf{J} \mid \prod_i \Psi_i(h_i) \tag{12}$$

En prenant le logarithm de $P(\vec{S})$, et en moyennant sur $\vec{S}$, on obtient (11) (à la constante près), avec, rappelons le, l'identification entre f_i' et la loi de la source i (pour une dérivation plus formelle voir [14]).

3 A la recherche de conditions minimales

Nous avons donc vu que trois points de vues, la maximisation de l'information, la minimisation de la redondance, et le maximum de vraisemblance, chacun pouvant prétendre avoir un caractère fondamental, conduisent non seulement au même traitement du signal (une ACI ou SAS dans le cas d'un mélange linéaire), mais aussi plus précisément à la même fonction de coût (le même contraste). Sur le plan qualitatif, cela n'est finalement pas surprenant, et après tout rassurant : le meilleur réseau est celui qui rend le plus compte de la structure des données, la plus grande quantité d'information est obtenue en reproduisant cette structure. Cette clarification théorique a été obtenue grace à l'apport de chercheurs de disciplines différentes. Maintenant que tout le monde est d'accord, sur ce plan théorique -et hormis les aspects purement algorithmiques - que pouvons nous attendre de plus ? En fait, si on se limite au problème de la séparation de sources, l'objectif recherché n'est pas identique à celui considéré par exemple en neurosciences pour la modélisation du codage sensoriel. En effet, pour la SAS, l'information mutuelle (entre l'entrée et la sortie) contient trop... d'information, par rapport à ce qu'on cherche : il ne devrait pas être nécessaire de connaître tous les détails des lois de probabilité des sources pour arriver à les séparer.

A partir de l'expression (6), on voit que si on part d'un choix de fonctions de transfert raisonnable, c'est-à-dire si leur dérivées sont assez proches des lois des sources, on peut alors espérer réaliser la SAS par maximisation de l'information mutuelle sans avoir à adapter davantage les fonctions de transfert. C'est le pari fait par Bell et Sejnowski[9], qui obtiennent des résultats prometteurs sur plusieurs problèmes de SAS. On peut alors se demander quelles sont les conditions minimales permettant d'effectuer la SAS.

Une façon d'aborder le problème est d'approximer les lois de probabilités par leur développement en cumulants[15, 28, 2]. Un problème important est que, en général, le critère à maximiser découlant de ce type d'approximation va avoir des maxima locaux en plus du maximum correspondant à la solution cherchée. De tels développements sont cependant une source d'inspiration pour dégager de nouveaux contrastes[15, 28]. Un autre point de vue est alors de chercher les conditions, si possible minimales, suffisantes pour garantir que la SAS a été effectivement réalisée, si ces conditions sont satisfaites. Plus précisément, avec un traitement linéaire du signal comme en (2), les conditions portent sur l'annulation de certains cumulants des h_i. Nous avons ainsi obtenu plusieurs familles de conditions, chacune liée à des cumulants d'un ordre donné[28]. Certaines de ces conditions sont en fait équivalentes à la possibilité d'obtenir des solutions algébriques [12, 31, 13, 10], d'autres sont entièrement nouvelles, ou impliquent un nombre de cumulants des h_i plus petit que ce qui était connu auparavant.

A titre illustratif, nous donnons ici, sans la preuve, deux des principales propositions démontrées dans [28]. Dans les deux énoncés ci-dessous, pour chaque i, h_i est défini comme en (2), et nous dénotons par $< . >_c$ le cumulant de la quantité entre crochets, et $\delta_{i,i'}$ est le symbol de Kronecker (= 1 si $i = i'$ et 0 sinon).

Théorème 1 *Soit k un entier* impair *au moins égal à 3. On suppose que parmi les N sources au plus une a son cumulant d'ordre k nul. Alors :*

J *est l'inverse de la matrice de mélange* **M** *(à une permutation près, et modulo des signes et facteurs d'échelle arbitraires, voir[28]), si et seulement si*

pour toute paire i, i',

$$\begin{cases} < h_i h_{i'} >_c = \delta_{i,i'} \\ < h_i^{(k-1)} h_{i'} >_c = 0 \text{ pour } i \neq i'. \end{cases} \tag{13}$$

Théorème 2 *Soit k et m deux entiers, avec $m \geq 2$ et $k > m$. On suppose que parmi les N sources au plus une a son cumulant d'ordre k nul. Alors :*

J *est l'inverse de la matrice de mélange* **M** *(à une permutation près, et modulo des signes et facteurs d'échelle arbitraires, voir[28]), si et seulement si*

pour chaque triplet i, i', i'',

$$\begin{cases} < h_i h_{i'} >_c = \delta_{i,i'} \\ < h_i^{(k-m)} h_{i'}^{m-1} h_{i''} >_c = 0 \\ \text{pour au moins deux indices non identiques.} \end{cases} \tag{14}$$

Remarque : plus généralement les mêmes conditions impliquent la séparation du sous ensemble de sources pour lesquelles le cumulant d'ordre k est non nul, voir [28].

On peut noter que le théorême 1 pour le cas particulier $k = 3$ implique que l'algorithme de Hérault et Jutten[20] contient *toujours*, parmi ses points fixes, les bonnes solutions (dans la mesure où les sources ont des cumulants d'ordre 3 non nuls).

Ce genre de résultat est interessant sur le plan qualitatif pour la raison suivante. Adapter les fonctions de transfert, autrement dit approximer les probabilités de distribution des sources, demande d'explorer un espace de functions. En se limitant à la détermination de cumulants, on se ramène à un espace de petite dimension. On peut alors poser simplement le problème en terme de *nombre d'équations, nombre d'inconnues*. On s'aperçoit alors du fait suivant. Si on cherche à avoir exactement le même nombre d'équations que d'inconnues, on trouve[28] des conditions liées à l'existence de solutions algébriques simples (blanchiement + diagonalisation d'une autre matrice de corrélation). L'inconvénient est que ces solutions ne permettent pas de séparer les sources lorsque certains cumulants sont identiques. Pour lever cette dégénérescence, il faut plus d'équations que d'inconnues, et on obtient des conditions telles que celles des théorêmes présentés ci-dessus. Ceci n'exclut pas la possibilité d'avoir des solutions algébriques. Ainsi, le cas particulier $m = 2$ de (14) est relié à la méthode de diagonalisation jointe proposée par Cardoso et Souloumiac[13]. Il faut noter aussi que les conditions (13) et (14) n'impliquent qu'un petit sous ensemble des cumulants d'ordre k des h_i.

Bien entendu on peut imaginer d'autres approches cherchant à effectuer une SAS selon un critère qui ne demande pas la détermination complète des lois des sources. Ainsi un travail récent de Amari et Cardoso[1] formule ce problème dans le cadre de l'estimation statistique, dans lequel les lois des sources apparaissent comme des "paramètres de nuisance" : ils ne sont pas connus et on ne souhaite pas les connaître...

4 Conclusion

Nous avons montré que des recherches issues de la modélisation en neurosciences d'une part et du traitement du signal d'autre part ont conduit à une clarification théorique montrant l'équivalence entre différentes approches de l'analyse en composantes indépendante et de la séparation aveugle de sources.

Cependant la spécificité du problème de la séparation de source permet d'envisager d'aller plus loin, à la recherche de conditions minimales permettant de séparer les sources sans avoir à connaître leur lois de probabilités de façon précise. Nous avons en particulier exposé des résultats montrant qu'un nombre limité de conditions sur des cumulants suffisent.

Bien que nous ayons principalement discuté le cas particulier de la séparation de sources, nous insistons sur la portée plus générale de la plupart des résultats cités. Comme mentionné dans la section 2, la maximisation de l'information mutuelle conduit à la minimisation de la redondance de façon générique[27]. Ainsi, on peut par exemple utiliser un réseau de neurones avec couches cachées, pour traiter un mélange non linéaire de sources. Dans l'hypothèse où ce mélange est inversible, les théorêmes généraux sur l'approximation de fonctions par un réseau de neurones garantissent la possibilité d'approximer cette transformation inverse ; par ailleurs, pour un réseau multicouches avec le même nombre de neurones dans chaque couches, l'information mutuelle prend une forme simple, généralisation directe de l'expression (11) (voir [28] pour plus de détails). Enfin, les conditions sur les cumulants s'appliquent aussi à la dernière couche d'un réseau de neurones multicouches.

Références

[1] S.-I. Amari & J.-F. Cardoso, Blind Source Separation - Semiparametric Statistical Approach, sumitted to *IEEE Tr. on Signal Processing*

[2] S. I. Amari, A. Cichocki, & H. H. Yang. A new learning algorithm for blind signal separation. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*. MIT Press, 1996, in press.

[3] J. J. Atick. Could information theory provide an ecological theory of sensory processing. *NETWORK*, 3:213–251, 1992.

[4] F. Attneave. Informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.

[5] W. Bialek, F. Rieke, R. de Ruyter van Steveninck, & D. Warl, Reading a neural code, *Science*, 252:1854–57, 1991.

[6] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W. Rosenblith, editor, *Sensory Communication*, page 217. M.I.T. Press, Cambridge MA, 1961.

[7] H. B. Barlow, T. P. Kaushal, & G. J. Mitchison. Finding minimum entropy codes. *Neural Comp.*, 1:412–423, 1989.

[8] Y. Bar-Ness. Bootstrapping adaptive interference cancelers: some practical limitations. In *The Globecom Conf.*, pages 1251–1255, paper F3.7, Miami, November 1982.

[9] A. Bell & T. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Comp.*, 7:1129–1159, 1995.

[10] A. Belouchrani & K. A. M. Séparation aveugle au second ordre de sources corrélées. In *GRETSI'93*, Juan-Les-Pins, 1993.

[11] R. E. Blahut. *Principles and Practice of Information Theory.* Addison-Wesley, Cambridge MA, 1988.

[12] J.-F. Cardoso. Source separation using higher-order moments. In *Proc. Internat. Conf. Acoust. Speech Signal Process.-89*, pages 2109–2112, Glasgow, 1989.

[13] J.-F. Cardoso & A. Souloumiac Blind beamforming for non Gaussian signals *IEE Proceedings-F*, 140(6):362–370, 1993.

[14] J-F Cardoso. Infomax and maximum likelihood for blind separation. *to appear in IEEE Signal Processing Letters*, 1997.

[15] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36:287–314, 1994.

[16] N. Delfosse & Ph. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45:59–83, 1995.

[17] P. Del Giudice, A. Campa, N. Parga, & J.-P. Nadal. Maximization of mutual information in a linear noisy network: a detailed study. *NETWORK*, 6:449–468, 1995.

[18] M. Gaeta & J.L. Lacoume. Source separation without apriori knowledge: the maximum likelihood approach *Signal Processing V*, proceedings of EUSIPCO 90, L. Tores, E. MasGrau & M.A. Lagunas eds, pp 621-624, 1990

[19] J. Herault & C. Jutten, *Réseaux neuronaux et traitement du signal* Hermes, Paris, 1994

[20] C. Jutten & J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[21] S. B. Laughlin, A simple coding procedure enhances a neuron's information capacity, *Z. Naturf.*, C36:910-2, 1981.

[22] Z. Li & J. J. Atick. Efficient stereo coding in the multiscale representation. *Network: Computation in Neural Systems*, 5:1–18, 1994.

[23] R. Linsker. Self-organization in a perceptual network. *Computer*, 21:105–17, 1988.

[24] Z. Malouche & O. Macchi, Adaptive unsupervised extraction of one component of a linear mixture with a single neuron, *soumis à IEEE Neural Networks*, preprint 1996

[25] E. Moreau & O. Macchi, High order contrasts for self-adaptive source separation, *Int. Journ. of Adaptive Control and Signal Proc.*, 10:19–46, 1995.

[26] J.-P. Nadal, N. Brunel & N. Parga. Nonlinear feed-forward networks with stochastic ouputs: infomax implies redundancy reduction. *Submitted to NETWORK.*

[27] J.-P. Nadal & N. Parga. Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer. *NETWORK*, 5:565–581, 1994.

[28] J.-P. Nadal & N. Parga. Redundancy reduction and independent component analysis: Algebraic and adaptive approaches. To appear in *Neural Computation*, preprint 1996.

[29] D.-T. Pham, Ph. Garrat & Ch. Jutten Separation of a mixture of independent sources through a maximum likelihood approach. in *Proc. EUSIPCO*, pp 771–774, 1992.

[30] A. N. Redlich. Redundancy reduction as a strategy for unsupervised learning. *Neural Comp.*, 5:289–304, 1993.

[31] B. Shraiman. *communication privée, et: technical memomorandum: AT&T Bell Labs TM-11111-930811-36*, 1993.

[32] J.H. van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly lmcs, and experimental validation. *J. Comp. Physiology A*, 171:157–170, 1992.