CONTRASTES POUR LA SÉPARATION AVEUGLE DE SOURCES

Dinh Tuan PHAM

Laboratoire de Modélisation et Calcul, IMAG, C.N.R.S.
Univ. de Grenoble, B.P. 53X, 38041 Grenoble Cedex, France
`Dinh-Tuan.Pham@imag.fr`

Résumé – Une méthode générale pour construire de contrastes pour la séparation aveugles de mélanges instantanés de sources est introduit. Elle est basée sur une fonctionnelle super-additive de classe II appliquées aux lois des sources reconstituées. Des exemples de telles fonctionnelles sont donnés. Notre approche permet d'exploiter la dépendance temporelle des sources en se servant d'une fonctionnelle sur la loi conjointe du processus sources dans un intervalle de temps. Cela fournit de nombreux nouveaux exemples et nous affranchit de la contrainte que les sources soient non gaussiennes. Le cas des contrastes basés sur les cumulants nécessitant la contrainte d'orthogonalité est également abordé.

Abstract – A general method to construct contrast functions for separating instantaneous mixtures of sources is introduced. It is based on a super-additive functional of class II applied to the distributions of the reconstructed sources. Examples of such functionals are given. Our approach permits exploiting the temporal dependence of the sources by using a functional on the joint distribution of the source process over a time interval. This yields many new examples and frees us from the constraint that the sources be non Gaussian. The case of contrasts functions based on cumulants requiring the orthogonality constraint is also considered.

1 Introduction

La séparation de source, dans sa formulation la plus simple, consiste à retrouver K sources $\{S_1(t)\}, \ldots, \{S_K(t)\}$ à partir de K enregistrements $\{X_1(t)\}, \ldots, \{X_K(t)\}$ qui sont leur combinaisons linéaires, soit $\mathbf{X}(t) = \mathbf{AS}(t)$ où

$$\mathbf{X}(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_K(t) \end{bmatrix}, \quad \mathbf{S}(t) = \begin{bmatrix} S_1(t) \\ \vdots \\ S_K(t) \end{bmatrix}$$

et $\mathbf{A}$ est une matrice carrée inversible. Naturellement, cela se fait par une transformation "inverse" $\mathbf{Y}(t) = \mathbf{BX}(t)$, mais dans le contexte aveugle, on n'a aucune information *a priori* ni sur $\mathbf{A}$ ni sur les lois des sources et doit s'appuyer uniquement sur l'indépendance de ces dernières pour réaliser la séparation. Ce qui conduit à la minimisation d'un critère d'indépendance entre les composantes $Y_1(\cdot), \ldots, Y_K(\cdot)$ de $\mathbf{Y}(\cdot)$.

Plus généralement on peut minimiser un contraste [2], définie comme une fonctionnelle des lois de $Y_1, \ldots, Y_K$ (et de $\mathbf{B}$) qui atteint son minimum quand $\mathbf{BA}$ est une matrice diagonale permutée. En effet, on ne peut, avec le seul critère d'indépendance, restituer les sources qu'à un facteur d'échelle et une permutation près. Suivant Comon [2], nous disons que le contraste est discriminant quand son minimum ne peut être atteint que si $\mathbf{BA}$ est une telle matrice.

Dans ce travail nous présentons une méthode générale pour la construction de contrastes discriminants. Elle est basée sur le concept de fonctionnelle super-additive de classe II, introduit par Huber [4]. Un nouveauté de ce travail est que nous ne nous restreindront pas aux fonctionnelles de la loi marginale de $Y_k(t)$ à un instant t donné, mais nous considérons aussi celles de la loi conjointe du processus $\{Y_k(t)\}$ sur un intervalle de temps. Cela permet d'exploiter la dépendance temporelle de sources pour mieux les séparer, en particulier, de construire de contrastes discriminants même pour les sources gaussiennes. À noter que la construction d'un contraste n'est qu'un premier pas vers un procédé de séparation. Nos contrastes sont des contrastes théoriques, dépendant des lois inconnues des sources restituées. Pour obtenir un contraste utilisable, ces lois ou plutôt certaine fonctionnelle de celles-ci doivent être estimées à partir des données. Ce problème ne sera pas considéré, ainsi que celui de la construction d'algorithme de minimisation du contraste ainsi obtenu. Ils seront considérés dans des travaux futurs pour des situations spécifiques, car notre approche générale peut conduire à des implémentations différentes, adaptées aux problèmes considérés.

2 Contrastes basés sur la distribution marginale

Comme les sources sont restituées via une transformation instantanée, on peut espérer d'obtenir des contrastes basés seulement sur la distribution marginale du vecteur des sources restituées $\mathbf{Y}(t)$. Par stationnarité, cette distribution ne dépend pas de t et donc on supprime cet indice, pour simplifier.

Le concept de fonctionnelle super-additive de classe II, introduit par Huber [4] est à la base de notre méthode. Une fonctionnelle Q de la distribution d'une variable aléatoire Y, notée $Q(Y)$, est dite de classe II si $Q(Y + a) = Q(Y)$ et $Q(aY) = |a|Q(Y)$ pour tout nombre réelle a. Clairement si Q possède ces propriétés, il est en de même pour $|Q|$ et donc on peut supposer, sans perte de généralité que $Q \geq 0$. La fonctionnelle Q est dite super-additive si [4]:

$$Q^2(X+Y) \geq Q^2(X) + Q^2(Y) \tag{1}$$

pour toute paire de variables aléatoires indépendantes X et Y.

Proposition 2.1 *Soit Q une fonctionnelle super-additive de* *classe II, alors*

$$\sum_{k=1}^{K} \log Q(Y_k) - \log|\det \mathbf{B}| \tag{2}$$

est un contraste pour la séparation de mélanges instantanés de sources. Ce contraste est discriminant si $Q(S_k) > 0$ pour toute source S_k et si l'inégalité (1) est stricte pour toute paire de multiples non nuls de sources distinctes.

Exemples

- 1: Prenons $Q(Y) = e^{h(Y)}$, la racine carrée de la puissance d'entropie. Alors Q est de classe II et d'après un résultat de Blachman [1] (voir aussi [3]), elle est super-additive et de plus l'inégalité (1) peut être une égalité si et seulement si X et Y sont gaussiennes. Par suite (2) est un contraste discriminant s'il n'y a qu'au plus une source gaussienne. On peut voir que ce contraste n'est d'autre que celui, bien connu, basé sur l'information mutuelle [2, 8]. Notre résultat fournit donc une preuve qu'il est discriminant (une première preuve basée sur un théorème de Darmois a été obtenue par Comon [2]).
- 2: Prenons $Q(Y) = J(Y)^{-1/2}$ où $J(Y)$ est l'information de Fisher, définie par $J(Y) = \mathrm{E}\psi_Y^2(Y)$, ψ_Y étant l'opposé de la dérivée logarithmique de la densité de Y, appelée fonction score. Alors, Q est une fonctionnelle de classe II et, comme il est prouvé dans [4] ou [1], est super-additive avec l'inégalité (1) stricte sauf si X et Y sont gaussiennes. Par suite (2) est un contraste discriminant sous la même condition que dans l'exemple 1.
- 3: Supposons que les sources sont bornées et prenons $Q(Y) = R_Y$, l'étendue de la loi de Y. Alors, Q est de classe II et est super-additive avec l'inégalité (1) stricte, car $R_{X+Y} = R_X + R_Y$. Donc (2) est un contraste discriminant. Ce résultat est déjà obtenu dans [10].
- 4: Pour les variables aléatoires sous-gaussiennes (c'est-à-dire de cumulant d'ordre 4 est négatif ou nul), on peut vérifier que la fonctionnelle Q définie par $Q(Y) = [\mathrm{E}(Y - \mathrm{E}Y)^4]^{1/4}$ est super-additive, avec l'inégalité (1) stricte sauf si X et Y ont leur cumulants d'ordre 4 nuls. C'est bien sûr une fonctionnelle de classe II et donc (2) est un contraste discriminant si au plus une des sources peut avoir le cumulant d'ordre 4 nul.

3 Contrastes sous contrainte d'orthogonalité

Beaucoup des travaux antérieurs sur la séparation de sources s'appuient sur des contrastes basés sur les cumulants d'ordre supérieurs (que 2). Mais les cumulants ne correspondent pas aux fonctionnelles super-additives mais sous-additives et notre méthode ne s'applique pas. Toutefois on peut encore construire de contrastes basés sur de telles fonctionnelles si on impose *la contrainte que la matrice de séparation* $\mathbf{B}$ *soit orthogonale*. (En fait les contrastes de type cumulants que nous connaissons nécessitent tous une contrainte de ce genre.) Cette contrainte est justifiée quand les donnés sont préalablement blanchies de sorte que le vecteur d'observation $\mathbf{X}$ a pour matrice de covariance la matrice identité. Alors pour préserver cette propriété pour le vecteur des sources restituées, la matrice de séparation $\mathbf{B}$ doit être orthogonale. D'autre part, en absorbant un facteur d'échelle dans la matrice de mélange $\mathbf{A}$ pour avoir des sources de la variance unité, le blanchissage implique que $\mathbf{A}$ doit être aussi orthogonale.

Une fonctionnelle Q est appelée sous-additive si [4]

$$Q^2(X+Y) \le Q^2(X) + Q^2(Y)$$

pour toute paire de variables aléatoires indépendantes X and Y. Pour la généralité, nous étendons cette définition en appelant la fonctionnelle Q α-sous-additive ($\alpha \ge 0$) si

$$Q^\alpha(X+Y) \le Q^\alpha(X) + Q^\alpha(Y).$$

On peut montrer que α-sous-additivité implique β-sous-additivité pour tout $\beta \le \alpha$, donc en particulier la sous-additivité si $\alpha \ge 2$.

Proposition 3.1 *Soit Q une fonctionnelle de classe II, α-sous-additive pour un $\alpha \ge 2$ et supposons que la matrice de mélange $\mathbf{A}$ est orthogonale. Alors*

$$-\sum_{k=1}^{K} Q^\alpha(Y_k) \quad et \quad -\sum_{k=1}^{K} Q^{2\alpha}(Y_k)$$

sont des contrastes pour la séparation de mélanges instantanés de sources sous la contrainte que la matrice de séparation $\mathbf{B}$ soit orthogonale. Ces contrastes sont discriminants si $\alpha > 2$ et les $Q(S_k)$ ne sont pas nuls sauf peut être pour un seul indice k.

On vérifie facilement que la fonctionnelle Q définie par $Q(Y) = |\mathrm{cum}_r(Y)|^{1/r}$ où $\mathrm{cum}_r(Y)$ désigne le cumulant d'ordre r de Y, est r-sous-additive. Le résultat précédent permet de retrouver les contrastes obtenus par Comon [2] et Moreau et Macchi [5]. D'autres formes plus générales de ce type de contrastes sont données dans [13, 7, 6]

4 Contrastes basés sur la distribution conjointe

Pour exploiter la dépendance temporelle des sources afin de mieux les séparer, nous introduisons des contrastes faisant intervenir la distribution conjointe du processus sources sur un intervalle de temps (de longueur $m > 1$ pour être précis). Par stationnarité, il suffit de considérer la loi conjointe du vecteur aléatoire $[Y_k(1) \quad \cdots Y_k(m)]^{\mathrm{T}}$, que nous notons $Y_k(1:m)$. À noter que l'on peut également travailler avec le vecteur $[Y_k(q) \quad Y_k(2q) \quad \cdots \quad Y_k(mq)]^{\mathrm{T}}$ à la place de $Y_k(1:m)$ où q est un entier plus grand que 1. Cela est utile quand les observations proviennent d'un échantillonage trop fin d'un enregistrement en temps continu.

Comme dans la section 2, nous appelons une fonctionnelle Q, de la loi d'un *vecteur* aléatoire Y, de classe II, si (i) $Q(Y) = Q(Y+a)$ for tout vecteur réel a et (ii) $Q(aY) = |a|Q(Y)$ pour tout nombre réel a. Aussi cette fonctionnelle est dite super-additive si l'inégalité (1) est satisfaite pour toute paire de vecteurs aléatoires indépendants X and Y.

Proposition 4.1 *Soit Q une fonctionnelle super-additive de classe II sur les lois m-dimensionnelles, alors*

$$\sum_{k=1}^{K} \log Q[Y_k(1:m)] - \log|\det \mathbf{B}| \tag{3}$$

est un contraste pour la séparation de mélanges instantanés de sources, qui est discriminant si $Q[S_k(1:m)] > 0$ pour tout k et l'inégalité (1) est stricte pour toute paire de multiples non nuls de $S_j(1:m)$ et $S_k(1:m)$ avec $j \neq k$.

Exemples

– 1: Prenons $Q(Y) = e^{h(Y)/m}$, la racine carrée de la puissance entropie de la distribution du m-vecteur aléatoire $Y(1:m)$. Alors comme dans la section 2, Q une fonctionnelle super-additive de classe II, avec l'inégalité (1) stricte sauf si X et Y sont des vecteurs gaussiens de matrices de covariance proportionnelles. Donc

$$\frac{1}{m}\sum_{k=1}^{K} h[Y_k(1:m)] - \log|\det \mathbf{B}| \tag{4}$$

est un contraste, qui est discriminant dès qu'il n'existe pas de paire de sources gaussiennes ayant les autocovariances proportionnelles jusqu'au retard $m-1$. On peut aussi voir que ce contraste est équivalent à celui basé sur l'information mutuelle entre les vecteurs $Y_1(1:m), \ldots, Y_K(1:m)$].

– 2: Prenons $Q(Y) = [\det \mathbf{J}(Y)]^{-1/(2m)}$ où m est la dimension de Y and $\mathbf{J}(Y) = \mathrm{E}[\psi_Y(Y)\psi_Y' Y(Y)]$ est la matrice d'information de Fisher, ψ_Y désignant ici l'opposé du vecteur gradient du logarithme de la densité de Y). On montre que Q est une fonctionnelle super-additive de classe II, avec l'inégalité (1) stricte sauf si X and Y sont gaussiens de matrices de covariance proportionnelles. Par suite (4) avec $\frac{1}{2}\log\det J[Y_k(1:m)]$ à la place de $h[Y_k(1:m)]$ est un contraste discriminant sous les mêmes conditions que dans l'exemple 1 précédent.

– 3: Supposons que les sources sont bornées et prenons $Q(Y) = R_Y$ où pour une vecteur aléatoire Y de dimension m, R_Y est définie comme la racine m-ième du volume du support de sa loi. Alors, d'après l'inégalité de Brunn-Minkowski ([3], équation (16.98)), $R_{X+Y} \geq R_X + R_Y$ si X et Y sont indépendants. Par suite la fonctionnelle Q super-additive de classe II avec l'inégalité (1) stricte et donc $\sum_{k=1}^{K} R_{Y_k(1:m)} - \log\det \mathbf{B}$ est un contraste discriminant.

Exemples 1 – 3 sont des analogues des exemples 1 – 3 de la section 2. Mais dans ce cas vectoriel, il y a d'autres possibilités.

Autres exemples

– 4: Prenons pour Q la fonctionnelle définie par

$$Q[Y(1:m)] = e^{h[Y(m)|Y(1:m-1)]},$$

où $h[Y(m)|Y(1:m-1)]$ est l'entropie conditionnelle de $Y(m)$ sachant $Y(1:m-1)$. On montre que Q est une fonctionnelle super-additive de classe II avec l'inégalité (1) stricte sauf si les vecteurs aléatoires $X(1:m)$ et $Y(1:m)$ sont gaussiennes dont les matrices de covariance sont telles que les dernières colonnes de leur inverses soient proportionnelles. En se servant de la stationnarité des processus les sources, on voit que

$$\sum_{k=1}^{K} h[Y_k(m)|Y_k(1:m-1)] - \log \mathbf{B} \tag{5}$$

est un contraste discriminant sous les mêmes conditions que dans l'exemple 1 précédent.

– 5 Prenons pour Q la fonctionnelle définie par $Q(Y) = [\mathbf{J}(Y)_{mm}]^{-1/2}$ où $\mathbf{J}(Y)_{mm}$ est l'élément en bas à droite de $\mathbf{J}(Y)$, la matrice d'information de Fisher définie plus haut. On montre que Q est une fonctionnelle super-additive de classe II avec l'inégalité (1) stricte sauf si les vecteurs aléatoires $X(1:m)$ et $Y(1:m)$ sont comme dans l'exemple 4 précédent. On déduit alors que (5) avec $\frac{1}{2}\log\mathbf{J}(Y)_{mm}$ à la place de $h[Y_k(m)|Y_k(1:m-1)]$ est un contraste discriminant sous les mêmes conditions que dans l'exemple 1 précédent.

Il est intéressant de noter que les contrastes précédents, autres que celui de l'exemple 3, sont, contrairement à ceux dans la section 2, encore discriminant pour les sources gaussiennes, pourvu qu'une condition de "non proportionnalité" soit satisfait. Cela suggère la possibilité de séparer les sources à l'aide des statistiques du second ordre uniquement. Pour cela, nous associons, à une fonctionnelle Q, la fonctionnelle gaussienne Q_g définie par $Q_g(Y) = Q(\tilde{Y})$ où $\tilde{Y}$ est un vecteur gaussien de même matrice de covariance que celui de Y. Clairement, si Q est super-additive de classe II, il en est de même pour Q_g. Le problème est que l'inégalité (1) peut ne plus être stricte. En fait, dans le cas où Y est scalaire, on montre que c'est toujours une égalité et donc la fonctionnelle Q_g est sans intérêt. Mais dans le cas vectoriel, on obtient des contrastes intéressants.

Exemples de contrastes "gaussiens"

– $1'$: Prenons $Q_g(Y) = [\det \mathrm{cov}(Y)]^{1/(2m)}$ où $\mathrm{cov}(\cdot)$ désigne la matrice de covariance. C'est la fonctionnelle gaussienne associée à celles des exemples 1 et 2 précédents. Le contraste correspondant est discriminant s'il n'y a pas de paire de sources ayant leurs autocovariances proportionnelles jusqu'au retard $m-1$.

– $4'$: Prenons $Q_g(Y) = [\det \mathrm{cov}(Y)/\det \mathrm{cov}(Y)_-]^{1/2}$ où $\mathrm{cov}(Y)_-$ est la matrice déduite de $\mathrm{cov}(Y)$ par suppression de la dernière ligne et colonne. C'est la fonctionnelle gaussienne associée à celles des exemples 4 et 5 précédents. Le contraste correspondant est discriminant sous les même conditions que celle de l'exemple $1'$.

Filtrage et moyenne géométrique Dans la Proposition 4.1, la fonctionnelle Q est appliquée à un vecteur formé par les observations successives, mais ce n'est pas obligatoire. On a déjà vu que l'on peut prendre des observations espacées de $q > 1$ unités de temps. Mais il y d'autres possibilités. En fait, il est facile de voir que si $\tilde{Q}$ est une fonctionnelle super-additive de classe II sur les lois m-dimensionelles, et $\mathbf{T}$ est une application linéaire de $\mathbb{R}^m$ à $\mathbb{R}^p$, alors la fonctionnelle Q définie par $Q(Y) = \tilde{Q}(\mathbf{T}Y)$ est aussi super-additive de classe II. Plus généralement, on peut considérer une banque de m filtres spécifiés par les séquences de leur réponses impulsionelles $c_1(\cdot), \ldots, c_m(\cdot)$ et définit la fonctionnelle Q par:

$$Q[Y(\cdot)] = \tilde{Q}([(c_1 \star Y)(1) \ \cdots \ (c_m \star Y)(1)]^{\mathrm{T}}),$$

où $\star$ désigne la convolution et $^{\mathrm{T}}$ la transposition. Ici Q est une fonctionnelle sur les lois de processus et non plus de vecteur aléatoire, mais les notions de super-additivité et de classe II se généralisent immédiatement dans ce cas et les résultats de la Proposition 4.1 sont encore applicables. Clairement Q est super-additive de classe II si $\tilde{Q}$ est. Les conditions pour que le contraste correspondant est discriminant sont plus complexes mais peuvent être facilement obtenues cas par cas.

Une façon de construire de nouvelles fonctionnelles super-additives de classe II à partir d'autres fonctionnelles est d'effectuer une moyenne géométrique. Soient $\tilde{Q}_1, \ldots, \tilde{Q}_p$ des fonctionnelles super-additives de classe II, on montre que la fonctionnelle Q définie par

$$Q(Y) = \prod_{n=1}^{p} \tilde{Q}_n^{\alpha_n}(Y), \quad a_n > 0, \; \sum_{n=1}^{p} \alpha_n = 1,$$

l'est également. Les $\tilde{Q}_n$ peuvent être les mêmes mais opérant chacun sur une seule composante distincte de Y, par exemple.

En combinant les deux idées précédentes, on arrive à une fonctionnelle Q super-additive de classe II de la forme $Q[Y(\cdot)] = \prod_{i=1}^{p} \tilde{Q}_n^{\alpha_n}[c_n \star Y(1)]$, où c_n représentent des filtres linéaires, qui donne lieu au contraste

$$\sum_{k=1}^{K} \sum_{n=1}^{m} \alpha_n \log \tilde{Q}[(c_n \star Y_k)(1)] - \log|\det \mathbf{B}|.$$

Ce contraste en fait une combinaison convexe des contrastes simples basés sur une fonctionnelle super-additive de classe II d'une loi *uni-dimensionelle*. La condition pour qu'il soit discriminant dépend bien sur de cette fonctionnelle mais aussi du choix des filtres et peut être toujours obtenue cas par cas.

Si on prendre pour les filtres précédents des filtres passe-bande *étroites* centrés sur différentes fréquences, alors leur sorties tendent à être gaussiennes et donc on peut remplacer $\tilde{Q}$ par la fonctionnelle gaussienne associée sans trop perdre d'information. Or dans le cas uni-dimensionelle, une fonctionnelle gaussienne de classe II n'est d'autre qu'un multiple de la déviation standard. Par suite, on est amené au contraste

$$\sum_{k=1}^{K} \frac{1}{2} \sum_{n=1}^{m} \alpha_n \log \operatorname{var}[(c_n \star Y_k)(1)] - \log|\det \mathbf{B}|, \qquad (6)$$

où $\operatorname{var}(\cdot)$ désigne la variance. On montre que ce contraste est discriminant si pour chaque source S_k, $\operatorname{var}[(c_1 \star S_k)(1)], \ldots, \operatorname{var}[(c_m \star S_k)(1)]$ sont strictement positive et que le vecteur ayant ces variances comme composantes n'est pas proportionnel à tout autre vecteur défini de façon semblable correspondant à une autre source. Soit $\mathbf{f_Y}(n)$ la matrice de terme général $\operatorname{cov}[(c_n \star Y_i)(1), (c_n \star Y_j)(1)]$ où $\operatorname{cov}(\cdot)$ désigne la covariance et soit $\mathbf{f_X}(n)$ définie de la même façon avec Y_k remplacé par X_k, alors $\mathbf{f_Y}(n) = \mathbf{B}\mathbf{f_X}(n)\mathbf{B}^{\mathrm{T}}$ et s'interprète comme la matrice de densité spectrale lissée du processus $\{\mathbf{Y}(t)\}$ en la fréquence dominante du n-ième filtre. Le contraste (6) s'écrit alors, à une constante additive près

$$\frac{1}{2} \sum_{n=1}^{m} \alpha_n [\log\det\operatorname{diag}\mathbf{f_Y}(n) - \log\det\mathbf{f_Y}(n)], \qquad (7)$$

où diag est l'opérateur qui met à zéro les termes non diagonale de son argument. L'expression entre les crochets précédents est en fait une mesure de déviation à la diagonalité de la matrice $\mathbf{f_Y}(n)$, car d'après l'inégalité d'Hadamard ([3], p. 233 ou 502) $\det\operatorname{diag}\mathbf{M} \le \det\mathbf{M}$ pour toute matrice définie positive $\mathbf{M}$, avec égalité si et seulement si cette matrice est diagonale. Par suite (7) est un critère de diagonalisation jointe des matrices de densité spectrales croisées lissées des sources restituées.

Les contrastes (4) et (5) ont été introduits dans Pham [9]. Leur analogies gaussiennes et le contraste (6) ou (7) ont été introduits dans Pham [11]. Pham [11, 12] a aussi donné un algorithme efficace trouver la solution au problème de diagonalisation jointe approchée de plusieurs matrices définies positives.

5 Conclusion

Nous avons introduit une méthode très générale pour construire des contrastes qui *nécessitent pas de blanchissage au préalable* et peuvent exploiter l'information temporelle dans les différentes sources.

Références

[1] Blachman, N. M. "The convolution inequality for entropy powers". *IEEE Trans. Inform. Theory*, **11**, 267–271, 1968.

[2] Comon, P. "Independent components analysis, a new concept". *Signal Processing*, **36**, 3, 287–314, 1994.

[3] Cover, Th. and Thomas, J. A. *Elements of Information Theory*. New-York: Wiley, 1991

[4] Huber, P. J. "Projection pursuit". *Ann. Statist.* **13**, 2, 435–475, 1985.

[5] Moreau, E. and Macchi, 0. "High Order Contrasts for Self-Adaptive Source Separation", *International Journal of Adaptive Control and Signal Processing*, **10**, no. 1, 19–46, 1996.

[6] Moreau, E. "A Generalization of Joint-Diagonalization Criteria for Source Separation". *IEEE Trans. Signal Processing*, To appear 2001.

[7] Moreau, E and Thirion-Moreau, N. "Non Symmetrical Contrasts for Sources Separation"; *IEEE Trans. Signal Processing*, **47**, 8, 2241–2252, 1999.

[8] D. T. Pham, "Blind Separation of Instantaneous Mixture of Sources via an Independent Component Analysis." *IEEE Trans. Signal Processing*, **44**, 11, 2768–2779, 1996.

[9] Pham, D. T. "Mutual Information Approach to Blind Separation of Sources. In *Proceeding of ICA'99 workshop"*, 215 – 220, Aussois, France, January 1999.

[10] Pham, D. T. "Blind Separation of Instantaneous Mixture of Sources via Order Statistics. *IEEE Trans. Signal Processing*, **48**, 2, 363–375, 2000.

[11] Pham, D. T. "Blind Separation of Instantaneous Mixture of Sources via the Gaussian Mutual Information Criterion". In *Signal Processing X*, 3–6, Proceeding of EUSIPCO 2000, Tampere, Finland, Eds. Moncef Gabbouj and Pauli Kuosmanen. Tampere: Finland, September 2000.

[12] Pham, D. T. "Joint approximate diagonalization of positive definite Hermitian matrices". *SIAM J. on Matrix Analysis and Applications*, To appear, 2001.

[13] Stoll, B. and Moreau, E. "A Generalized ICA Algorithm", *IEEE Signal Processing Letters*, **7**, 4, 90–92, 2000.