

Mesure de dépendance quadratique pour la séparation aveugle de source dans les mélanges post non linéaires.

Sophie ACHARD¹ Dinh Tuan PHAM¹ Christian JUTTEN^{2*}

¹Laboratoire de Modélisation et Calcul, IMAG, C.N.R.S.
Univ. de Grenoble, B.P. 53X, 38041 Grenoble Cedex, France

²Laboratoire des Images et des Signaux, INPG
46 avenue Félix Viallet, 38031 Grenoble Cedex, France

Sophie.Achard@imag.fr, Dinh-Tuan.Pham@imag.fr, Christian.Jutten@inpg.fr

Résumé – Dans le cadre de la séparation aveugle de sources, le problème requiert la définition d’une mesure de dépendance permettant de décider si des variables sont mutuellement indépendantes ou non. Jusqu’à présent, beaucoup de méthodes ont été développées en utilisant l’information mutuelle. Ici, nous proposons de construire une méthode basée sur une mesure de dépendance quadratique. Après avoir défini cette mesure, nous proposerons une estimation et la minimisation sera effectuée par une méthode de descente de gradient. Enfin nous illustrerons nos résultats en appliquant cette méthode dans le cas de mélanges post non linéaires.

Abstract – The problem of blind source separation requires the use of a dependence measure, which allows to decide whether the variables are mutually independent or not. Many developed methods are based on the use of mutual information. Here, we propose to construct a new method based on a quadratic dependence measure. After defining this measure, we will estimate it and minimize it through a gradient descent method. Finally, we will illustrate our results by applying the method in the case of post non linear mixtures.

1 Introduction

Durant ces dernières années, le problème de séparation aveugle de sources a suscité l’intérêt de diverses communautés de scientifiques, tant au niveau théorique, sur l’identifiabilité des mélanges [5] que au niveau algorithmique avec le développement de diverses méthodes concernant essentiellement le cas de mélanges linéaires instantanés [4], [3]. L’objectif de ces méthodes peut-être résumé comme suit:

On dispose d’observations X_1, \dots, X_K provenant d’un mélange de sources S_1, \dots, S_K indépendantes:

$$\mathbf{X} = (X_1, \dots, X_K)^T = f(\mathbf{S})^1$$

où $\mathbf{S} = (S_1, \dots, S_K)^T$ et f est une transformation inversible.

Le problème consiste donc à chercher une application g telle que

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_K)^T = g(\mathbf{X})$$

soit une estimation des sources.

Naturellement, comme l’ont fait remarquer de nombreux auteurs, il est indispensable de restreindre le type de mélange afin de pouvoir reconstituer les sources avec des indéterminations acceptables. De plus, si aucune hypothèse supplémentaire ne peut-être faite en ce qui concerne, par exemple, la densité des sources ou la non stationnarité des signaux, la seule hypothèse exploitable concerne l’indépendance des sources. C’est pourquoi, de nombreuses méthodes déjà construites sont basées sur la définition d’une mesure de dépendance, par exemple l’information mutuelle, qui permet d’entièrement caractériser

l’indépendance mutuelle de signaux. En effet, les méthodes basées sur le maximum de vraisemblance, [4], le principe d’Infomax, [3], ou bien les méthodes utilisant les statistiques d’ordre supérieur pour l’approximation de la negentropie, [5], [8], peuvent être reliées à l’information mutuelle. Actuellement, dans le cadre de mélanges non linéaires, les méthodes basées sur l’utilisation de l’information mutuelle sont assez compliquées à mettre en place à cause de la difficulté due à l’estimation de l’information mutuelle utilisée comme mesure de dépendance. Il est donc intéressant de proposer d’autres alternatives de mesure de dépendance. Murata [10], Eriksson *et al.* [6] ont développé une mesure de dépendance utilisant les fonctions caractéristiques, Bach et Jordan [2] utilisent une méthode basée sur la corrélation non linéaire calculée à partir de noyaux.

Dans un premier temps, nous nous attacherons à définir la mesure de dépendance quadratique utilisée pour la suite et à détailler certaines propriétés. Dans un deuxième temps, la mesure de dépendance quadratique sera appliquée au problème de séparation aveugle de sources dans le cadre de mélanges post non linéaires. Enfin, nous illustrerons notre méthode par une simulation.

2 Définition de la mesure de dépendance quadratique

Rappelons que l’indépendance mutuelle de variables aléatoires peut-être caractérisée de plusieurs manières. Rosenblatt [11] a construit un test d’indépendance basé sur la comparaison entre la densité conjointe et le produit des densités marginales en prenant la norme L^2 de la différence

* Avec le soutien du projet européen BLISS.

1. \mathbf{x}^T désigne la transposée du vecteur \mathbf{x}

de ces deux quantités. L'information mutuelle est construite en prenant la distance de Kullback-Leibler entre la densité conjointe et le produit des densités marginales. Nous proposons d'utiliser ici une mesure de dépendance comparant la fonction caractéristique conjointe et le produit des fonctions caractéristiques marginales via la norme L^2 pondérée par une fonction de poids.

Déjà utilisée par Feuerverger [7] dans le cadre de méthodes de test d'indépendance, cette mesure de dépendance a aussi été étudiée par Kankainen [9] dont la définition est la suivante:

Définition 2.1 *Pour toutes variables aléatoires réelles, T_1, \dots, T_K , (on note $\mathbf{T} = (T_1, \dots, T_K)^T$ le vecteur aléatoire associé, et $\mathbf{t} = (t_1, \dots, t_K)^T$)*

$$\mathcal{D}(T_1, \dots, T_K) = \int |\psi_{\mathbf{T}}(\mathbf{t}) - \prod_{k=1}^K \psi_{T_k}(t_k)|^2 h(\mathbf{t}) d\mathbf{t}$$

où h est une fonction de R^N dans R intégrable, $\psi_{\mathbf{T}}$ est la fonction caractéristique associée à \mathbf{T} , c'est à dire, $\psi_{\mathbf{T}}(t_1, \dots, t_K) = E[\exp(i \sum_{k=1}^K t_k T_k)]$ et pour tout $k, 1 \leq k \leq K$, $\psi_{T_k}(t_k) = E[\exp(it_k T_k)]$.

Afin que cette mesure \mathcal{D} permette de caractériser l'indépendance mutuelle de variables aléatoires, c'est à dire que $\mathcal{D}(T_1, \dots, T_K) = 0$ si et seulement si T_1, \dots, T_K sont indépendantes, comme le note Kankainen [9], il est indispensable de faire des hypothèses sur la fonction h . En fait on peut distinguer deux classes de fonctions h qui entraînent cette propriété:

- soit on suppose h non nulle presque partout et positive,
- soit, si la fonction caractéristique conjointe est analytique, on suppose h nulle sauf sur un voisinage de 0 où celle-ci est positive.

Sous ces conditions, il est montré dans [9] que la mesure \mathcal{D} est bien une mesure de dépendance.

En partant de ces constatations, nous définissons h par,

$$g(t_1, \dots, t_K) = \prod_{k=1}^K \left| \frac{\sqrt{\sigma_{T_k}} \psi_{\mathcal{K}}(\sigma_{T_k} t_k)}{\sqrt{2\pi}} \right|^2$$

où \mathcal{K} est de carré intégrable tel que sa transformée de Fourier soit non nulle presque partout.

Par application de la formule de Parseval, on en déduit une nouvelle écriture pour \mathcal{D} que l'on appellera la mesure de dépendance quadratique, notée Q .

Définition 2.2 *Soit \mathcal{K} un noyau de carré intégrable tel que sa transformée de Fourier soit différente de zéro presque partout. On définit alors la mesure de dépendance quadratique de K variables aléatoires T_1, \dots, T_K telle que $\mathbf{T} = (T_1, T_2, \dots, T_K)^T$.*

$$Q(T_1, \dots, T_K) = \frac{1}{2} \int D_{\mathbf{T}}(t_1, \dots, t_K)^2 dt_1 \dots dt_K,$$

où pour tous t_1, \dots, t_K réels,

$$D_{\mathbf{T}}(t_1, \dots, t_K) = E \left[\prod_{k=1}^K \mathcal{K} \left(t_k - \frac{T_k}{\sigma_{T_k}} \right) \right] - \prod_{k=1}^K E \left[\mathcal{K} \left(t_k - \frac{T_k}{\sigma_{T_k}} \right) \right]$$

et σ_{T_k} est un coefficient de facteur d'échelle, c'est à dire une fonction positive dépendant seulement de la loi des T_k tel que $\sigma_{\lambda T_i} = |\lambda| \sigma_{T_i}$, pour toute constante réelle λ .

Cette définition de la mesure de dépendance quadratique fait apparaître certaines propriétés importantes. Il est clair que, d'après ce qui précède, la mesure de dépendance quadratique permet aussi de caractériser l'indépendance de variables aléatoires dans le sens où, $Q(T_1, \dots, T_K) = 0$ si et seulement si, T_1, \dots, T_K sont indépendantes.

De plus, elle est invariante par translation et par changement d'échelle. D'autres propriétés sont détaillées dans [1].

Afin d'utiliser la mesure de dépendance quadratique Q pour résoudre le problème de séparation aveugle de sources, il est indispensable de pouvoir en donner une estimation. Présentons alors maintenant une nouvelle écriture de Q . En effet, on observe qu'en développant l'expression au carré sous l'intégrale, on a besoin d'évaluer la convolution de \mathcal{K} par la fonction symétrique associée ($u \mapsto \mathcal{K}(-u)$).

Définissons alors un nouveau noyau,

$$\mathcal{K}_2(u) = \int \mathcal{K}(u+v) \mathcal{K}(v) dv$$

On en déduit alors l'expression suivante de $Q(T_1, \dots, T_K) = \frac{1}{2} \left\{ E[\pi_{\mathbf{T}}(\mathbf{T})] + \prod_{k=1}^K E[\pi_{T_k}(T_k)] - 2E \left[\prod_{k=1}^K \pi_{T_k}(T_k) \right] \right\}$

où

$$\begin{aligned} \pi_{\mathbf{T}}(\mathbf{t}) &= E \left[\prod_{k=1}^K \mathcal{K}_2 \left(\frac{t_k - T_k}{\sigma_{T_k}} \right) \right] \\ \pi_{T_k}(t_k) &= E \left[\mathcal{K}_2 \left(\frac{t_k - T_k}{\sigma_{T_k}} \right) \right] \end{aligned}$$

D'après les propriétés vérifiées par \mathcal{K} en ce qui concerne l'existence de Q et ses propriétés, il est suffisant de choisir \mathcal{K}_2 pair, tel que sa transformée de Fourier soit sommable, et non nulle presque partout.

Nous avons alors à notre disposition une collection de mesures de dépendance dépendant du choix du noyau. De plus, on note que contrairement à l'estimation de la densité, le noyau n'est pas restreint aux noyaux adéquats à l'estimation de densité. Par contre, ces différents choix affectent les performances des différentes méthodes implémentées, ceux-ci doivent donc être choisis de manière *ad hoc*.

3 Séparation aveugle de sources via la mesure de dépendance quadratique

Comme nous l'avons évoqué dans l'introduction, il est indispensable de restreindre le type de mélange afin de pouvoir identifier les sources, à partir de la seule hypothèse d'indépendance de celles-ci et avec le moins d'indéterminations possibles. Dans la suite de notre travail, nous avons alors considéré une classe de mélanges non linéaires, les mélanges post non linéaires. Introduits par Taleb et Jutten, puis repris par Babaie-Zadeh *et al.*, cette classe de mélanges permet de modéliser certains mélanges non linéaires tout en gardant les propriétés d'identifiabilité des sources à un facteur d'échelle près et à une permutation près. Rappelons maintenant comment sont définis ces mélanges. On note $\mathbf{X} = (X_1, \dots, X_K)$ le vecteur des observations, et $\mathbf{S} = (S_1, \dots, S_K)$ le vecteur des sources, alors,

$$\mathbf{X} = \begin{bmatrix} f_1(\sum_{k=1}^K \mathbf{A}_{1k} S_k) \\ \vdots \\ f_K(\sum_{k=1}^K \mathbf{A}_{Kk} S_k) \end{bmatrix}$$

où \mathbf{A}_{ik} sont les éléments de la matrice inversible \mathbf{A} et f_1, \dots, f_K sont des fonctions monotones.

Dans ce cadre-là, on recherche un système de séparation défini comme ci-dessous, de telle sorte que les composantes Y_k de \mathbf{Y} soient le plus indépendantes possibles dans le sens de la mesure de dépendance quadratique,

$$\mathbf{Y} = \begin{bmatrix} \sum_{k=1}^K \mathbf{B}_{1k} g_k(X_k) \\ \vdots \\ \sum_{k=1}^K \mathbf{B}_{Kk} g_k(X_k) \end{bmatrix}$$

où \mathbf{B}_{ik} représentent les éléments de la matrice inversible \mathbf{B} et g_1, \dots, g_K sont des fonctions monotones.

D'après la définition de la mesure de dépendance quadratique, on remarque que celle-ci ne peut pas être évaluée directement. En effet, elle dépend de la distribution des observations qui est inconnue. Il est donc indispensable d'envisager l'estimation de la mesure de dépendance quadratique. Ceci nous permettra ensuite de proposer une méthode d'optimisation conduisant à la solution du problème.

Dans la suite, pour $1 \leq k \leq K$, l'échantillon associé à la composante X_k du vecteur des observations sera noté $X_k(n)$ pour $n = 1, \dots, N$ où N désigne la taille de l'échantillon. De même, l'échantillon correspondant à la composante Y_k de \mathbf{Y} sera noté $Y_k(n)$ pour $n = 1, \dots, N$.

3.1 Estimation de Q

D'après la définition de Q , son estimation nécessite seulement l'utilisation de la moyenne empirique, $\widehat{E}\Phi(\mathbf{X}) = \sum_{n=1}^N \Phi(\mathbf{X}(n))/N$, où Φ est une fonction quelconque des données.

Une estimation de Q sera alors: $\widehat{Q}(Y_1, \dots, Y_K) =$

$$\frac{1}{2} \left\{ \widehat{E}\widehat{\pi}_{\mathbf{Y}}(\mathbf{Y}) + \prod_{k=1}^K \widehat{E}\widehat{\pi}_{Y_k}(Y_k) - 2\widehat{E} \prod_{k=1}^K \widehat{\pi}_{Y_k}(Y_k) \right\}$$

où

$$\begin{aligned} \widehat{\pi}_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{N} \sum_{n=1}^N \prod_{i=1}^K \mathcal{K}_2 \left(\frac{y_i - Y_i(n)}{\widehat{\sigma}_{Y_i}} \right) \\ \widehat{\pi}_{Y_k}(y_k) &= \frac{1}{N} \sum_{n=1}^N \mathcal{K}_2 \left(\frac{y_k - Y_k(n)}{\widehat{\sigma}_{Y_k}} \right). \end{aligned}$$

D'après les contraintes imposées sur \mathcal{K}_2 , le choix de ce dernier doit se limiter à des noyaux \mathcal{K}_2 pair, de transformée de Fourier intégrable et différentiable de zéro presque partout. Voici trois exemples de noyaux utilisés,

1. Le noyau Gaussien :

$$\mathcal{K}_2(x) = e^{-x^2}, \psi_{\mathcal{K}_2}(t) = \sqrt{\pi} e^{-t^2/4}$$

2. Le noyau de Cauchy carré:

$$\mathcal{K}_2(x) = 1/(1+x^2)^2, \psi_{\mathcal{K}_2}(t) = \pi(|t|+1)e^{-|t|}$$

3. L'opposé de la dérivée seconde du noyau de Cauchy carré: $\mathcal{K}_2(x) = -(20x^2 - 4)/(1+x^2)^4$,

$$\psi_{\mathcal{K}_2}(t) = 4t^2\pi^3(|t|+1)e^{-|t|}.$$

3.2 Gradient du critère empirique

Ici, nous choisissons de nous limiter au cas des mélanges post non linéaires pour le développement des calculs du gradient. Cependant, dans [1], les calculs sont détaillés dans un cadre général identique à celui donné en introduction.

Dans la suite, nous allons détailler plus particulièrement une méthode semi-paramétrique consistant en l'approximation des transformations non linéaires par des fonctions linéaires par morceaux. Puis nous expliciterons une méthode utilisant une approximation des transformations non linéaires basée sur les fonctions quantiles. Pour plus de détail sur une méthode non paramétrique, on peut consulter [1]. Dans la suite, on note \widehat{G}_k , la k -ième composante du gradient relatif de la mesure de dépendance quadratique. Son expression étant assez complexe, nous renvoyons à [1] pour le détail de celle-ci.

De plus, on note $Z_k(n) = g_k(X_k(n))$ pour tout $k = 1, \dots, K$ et $n = 1, \dots, N$. Alors, $Y_k(n) = \sum_{i=1}^K B_{ki} Z_i(n)$. On note aussi, $Z_j(1:N) \leq \dots \leq Z_j(N:N)$ les statistiques d'ordre associées à l'échantillon de Z_j .

Dans le cas semi-paramétrique ou paramétrique avec la fonction quantile, le gradient relatif s'écrit, pour ϵ une petite variation de \mathbf{B} et $\delta_1, \dots, \delta_K$ des petites variations de g_1, \dots, g_K respectivement:

$$\begin{aligned} (\epsilon, \delta_1, \dots, \delta_K) &\rightarrow \sum_{1 \leq j \neq k \leq K} (\widehat{\Gamma}_{kj} - \widehat{\Gamma}_{kk} \widehat{\Sigma}_{kj} / \widehat{\Sigma}_{kk}) \epsilon_{kj} \\ &+ \frac{1}{N} \sum_{n=2}^N \left\{ \sum_{m=n}^N \sum_{k=1}^K \widehat{G}_k^*(\mathbf{Y}(o_{j,m})) \mathbf{B}_{kj} \right\} \widetilde{g}_{j,\theta_j}(n) \end{aligned}$$

où $\widehat{\Gamma}_{kj} = \widehat{E}[\widehat{G}_k(\mathbf{Y})Y_j]$, $\widehat{\Sigma}_{kj} = \widehat{E}[(Y_k - \widehat{E}(Y_k))(Y_j - \widehat{E}(Y_j))]$ et $\widehat{G}_k^*(\mathbf{y}) = \widehat{G}_k(\mathbf{y}) - (\widehat{\Gamma}_{kk}/\widehat{\Sigma}_{kk})(y_k - \widehat{E}(Y_k))$. De plus, $\widetilde{g}_{j,\theta_j}(n) = \delta_j(Z_j(n:N)) - \delta_j(Z_j(n-1:N))$ avec $\delta_j = \dot{g}_{j,\theta_j} := g'_{j,\theta_j} \circ g_{j,\theta_j}^{-1}$, g'_{j,θ_j} le gradient de g_{j,θ_j} par rapport à θ_j .

3.2.1 Méthode semi-paramétrique

Cette méthode consiste simplement à approcher les transformations non linéaires par des fonctions linéaires par morceaux. C'est à dire que, par rapport aux notations précédentes, le vecteur θ_k des paramètres correspondant à la fonction g_k désigne les coefficients directeurs de chaque morceau de droite. En effet, le fait de rajouter une constante à chaque fonction g_k ne modifie pas la mesure de dépendance quadratique. Cette dernière ne dépend donc que des dérivées g'_k des transformations non linéaires. En paramétrant les dérivées g'_k par des fonctions très simples, constantes par morceaux, ceci revient à approcher les transformations g_k par des fonctions linéaires par morceaux. Cette méthode permet d'obtenir une approximation des fonctions g_k assez facile à mettre en place et de diminuer le nombre de degré de liberté du problème par rapport au cas non paramétrique.

3.2.2 Méthode paramétrique par la fonction quantile

Détaillons à présent cette nouvelle approche. Elle consiste simplement à faire la remarque suivante, dans le cadre d'un mélange post non linéaire, chaque transformation non linéaire g_k vérifie l'égalité suivante, $g_k = Q_{Z_k} \circ F_{X_k}$ où Q_{Z_k} est la fonction quantile de $Z_k = g_k(X_k)$, et F_{X_k} est la fonction de répartition de X_k . Pour les mêmes raisons que précédemment, il est intéressant de travailler avec seulement les dérivées de Q_{Z_k} . En outre, les fonctions Q_{Z_k} ne sont pas bornées et donc difficilement paramétrisables. Pour toutes ces considérations, la paramétrisation protera sur les fonctions appelées densité-quantile, $1/Q'_{Z_k} = p_{Z_k} \circ Q_{Z_k}$.

D'après l'écriture du gradient ci-dessus, seules les quantités $z_{k,i} = Q_{Z_k}((i - 0.5)/N)$ sont nécessaires. En revenant aux notations utilisées dans l'écriture du gradient, on a

$$\tilde{g}_{j,\theta_j}(i) = \frac{1}{N(p_{k,i+1} - p_{k,i})} \log \left(\frac{p_{k,i+1}}{p_{k,i}} \right), \quad i = 1, \dots, N-1$$

où

$$p_{k,i} = \sum_{j=1}^M \tilde{s}_j \left(\frac{i - 0.5}{N} \right) \theta_{kj}$$

et $\tilde{s}_1, \dots, \tilde{s}_M$ est une base de B-splines d'ordre 2.

4 Résultats expérimentaux

Nous présentons sur la figure 1, les résultats d'une simulation de la méthode paramétrique avec les fonctions quantiles. Le mélange est défini par, $\mathbf{A} = \begin{pmatrix} 1 & 0.6 \\ 0.7 & 1 \end{pmatrix}$ et $f_1(x) = f_2(x) = \tanh(4x) + 0.1x$.

En pratique, nous utilisons aussi le résultat de Solé *et al.* [12], on applique une procédure de gaussianisation afin d'obtenir une valeur initiale pour l'algorithme de descente du gradient, ce qui augmente très sensiblement la vitesse de convergence.

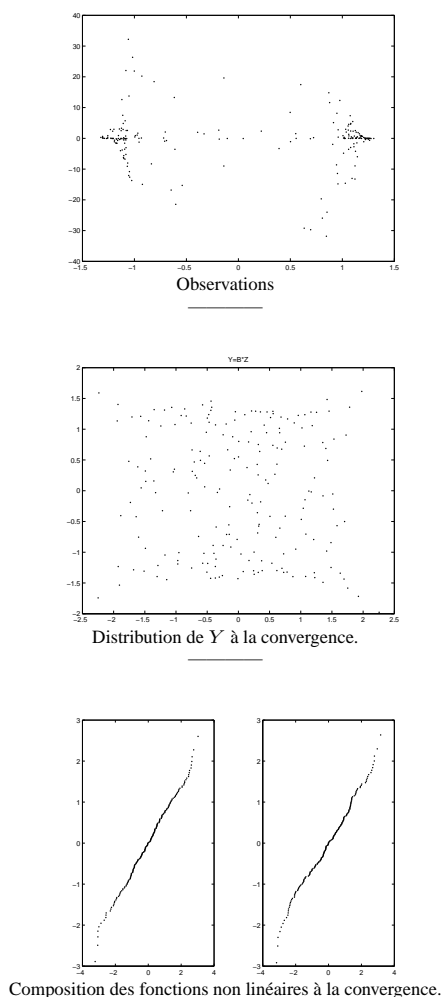


FIG. 1: Simulation

5 Conclusion

Utilisant une propriété d'indépendance basée sur les fonctions caractéristiques, nous avons développé une nouvelle méthode permettant de résoudre le problème de séparation de sources. Il a été montré aussi comment envisager différentes paramétrisations du modèle de mélange post non linéaire, en particulier à l'aide de fonctions quantiles. Pour finir, on présente le résultat d'une simulation. Les choix du noyau et de la taille de fenêtre étant très importants, nous envisageons d'étudier plus précisément leur influence dans les algorithmes.

Références

- [1] S. Achard, D.T. Pham, and C. Jutten. Quadratic dependence measure for non linear blind sources separation. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation, ICA2003*, pages 263–268, Nara, Japan, Apr. 2003.
- [2] F.R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, Jul. 2002.
- [3] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- [4] J.F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11:157–192, 1999.
- [5] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 3(36):287–314, Apr. 1994.
- [6] J. Eriksson, A. Kankainen, and V. Koivunen. Novel characteristic function based criteria for ICA. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation, ICA2001*, pages 108–113, San Diego, California, Dec. 2001.
- [7] A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3):419–433, 1993.
- [8] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. New York: John Wiley & Sons, 2001.
- [9] A. Kankainen. *Consistent testing of total independence based on empirical characteristic functions*. PhD thesis, University of Jyväskylä, 1995.
- [10] N. Murata. Properties of the empirical characteristic function and its application to testing for independence. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation, ICA2001*, pages 19–24, San Diego, California, Dec. 2001.
- [11] M. Rosenblatt. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics*, 3(1):1–14, 1975.
- [12] J. Solé, C. Jutten, and D.T. Pham. Improving algorithm speed in PNL mixture separation and wiener system inversion. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation, ICA2003*, pages 639–644, Nara, Japan, Apr. 2003.