EL ANALISIS DE COMPONENTES INDEPENDIENTES Y SUS APLICACIONES

1 Luz Noe Oliva Moreno
1 Miguel Ángel Alemán Arce
2 Miguel Olvera-Aldana

1.- Department of Electrical Engineering,
ESCOM-IPN, Mexico D.F., Mexico
2.- Department of basic formation, ESCOM-IPN,
Mexico D.F., Mexico
Department of Electrical Engineering,
Phone (55) 57296000 Ext 52022
E-mail: noetronix@hotmail.com
E-mail: maleman@hotmail.com
E-mail: molveraa@ipn.mx

0.1. Resumen

El Análisis de Componentes Independientes (**ICA** - Independent Component Analysis) es un método estadístico utilizado para el análisis de datos en espacios multidimensionales suponiendo que las señales de origen tienen una independencia estadística y son no-Gausianas. ICA se encuentra muy relacionado con el problema de Separación Ciega de Señales (**BSS** - Blind Signal Separation), el cual consiste en obtener las señales de las fuentes originales que intervienen en una mezcla, sin información alguna de las ponderaciones de la mezcla. Este es un problema clásico de procesamiento de señales con aplicaciones potenciales en sistemas de reconocimiento de voz, procesamiento de señales médicas, de telecomunicaciones, etc. En este artículo se muestra la Red-H-J e Infomax y sus arquitecturas se basan en Redes Neuronales para problemas en Tiempo-Real.

0.1.1. Introducción

El termino Análisis de Componentes Independientes (**ICA** - Independent Component Analysis), fue enunciado por primera vez en 1985 por J. Hérault y C. Jutten [JUT91], y generaliza la técnica de Análisis de Componentes Principales (**PCA** Principal Component Analysis), en que suele basarse. La tarea de ICA, consiste en encontrar los factores o componentes de datos estadísticos en espacios multidimensionales para obtener aquellos que pertenecen a una clase. Esta tarea es utilizada frecuentemente en el problema clásico de separación ciega de señales (**BSS** Blind Signal Separation), uno de los orígenes de la separación de señales se basó en encontrar una explicación al efecto “cocktail Party”.

Este efecto consiste en separar y estimar las múltiples señales de fuentes independientes (voz de la persona) de una mezcla, a través de un arreglo discreto de transductores (micrófonos), sin un previo conocimiento de las características de las fuentes originales y de los canales de transmisión. Asumiendo un número igual de fuentes desconocidas y de transductores, en donde cada transductor por sus características y su posición, reciben una combinación lineal diferente de cada una de las señales. En la figura 2 se muestra un esquema del arreglo de transductores y de fuentes originales de tamaño $n \times n$.

El sistema mostrado es descrito por la siguiente ecuación denominada modelo de ICA:

$$e_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t) \qquad (1)$$

$$\mathbf{e}(t) = \mathbf{A}\mathbf{s}(t)$$

Figura 1: Problema “cocktail Party”.

Figura 2: Arreglo de transductores y fuentes originales.

En donde: $e_i(t)$ es la señal continua en el tiempo, de la salida del i-ésimo transductor. $s_j(t)$ es la señal desconocida, emitida por la j-ésima fuente. a_{ij} es un número escalar de la combinación lineal.

Escrito en forma matricial, la ecuación 1 se expresa como:

$$\mathbf{e} = \mathbf{A}\mathbf{s} \qquad (2)$$

En donde $\mathbf{A}$ representa la matriz de mezcla de tamaño $n \times n$. El modelo de ICA es un modelo generativo y describe como los datos son generados por un proceso de mezcla de componentes $\mathbf{s}$. Un modelo general considera una transformación de datos del espacio $\mathbf{s}$ al espacio $\mathbf{e}$ y se le conoce como modelo no lineal [ALM99, PUN99]:

$$\mathbf{e} = \mathcal{F}(\mathbf{A}\mathbf{s}) \qquad (3)$$

Otros modelos de ICA suelen considerar los siguientes casos:

- **Modelo post-no lineal**: donde el medio es lineal y la no-linealidad es introducida en los sensores.
- **Modelo convolutivo**: el medio y los sensores se comportan como filtros lineales y estacionarios.
- **Mezcla lineal**: se considera que tanto el medio como los sensores se comportan de forma lineal (1)(2).

En el presente artículo consideremos el modelo de mezcla lineal e instantánea, modeladas, con la expresión (2). Para estimar las señales de las fuentes independientes $\mathbf{s}$, buscamos una transformación lineal $(\mathbf{W})$ que las recupere, teniendo tan sólo acceso a las señales mezcladas $\mathbf{e}$, recibida por los transductores.

$$\mathbf{u} = \mathbf{W}\mathbf{e} \qquad (4)$$

Asumiremos dos señales de fuentes estadísticamente independientes con distribuciones de probabilidad no gaussianas. En la figura 3 se muestran dos señales, con distribuciones uniformes.

Figura 3: Dos señales de fuentes independientes en el tiempo, sus histogramas y en el plano bidimensional.

Figura 4: Mezcla de las señales independientes s_1 y s_2

Considerando la expresión 2 y asumiendo una matriz de mezcla $\mathbf{A}$. El resultado de la variable $e(t)$ se muestra en la figura 4.

Los algoritmos de ICA con enfoque estadístico, tratan de explotar la diversidad espacial producida por la mezcla percibida por los distintos censores, donde cada componente representará una señal independiente.

Para asegurar que los modelos ICA obtengan un buen estimado de cada una de las componentes, se deben hacer las siguientes consideraciones: Las componentes deben ser estadísticamente independientes y deben tener distribuciones no gaussianas.

En el modelo de ICA existen algunas indeterminaciones que no son identificables, como son: la amplitud o potencia y el orden de las fuentes. En los métodos de BSS se consideran validos que las señales recuperadas tengan sus amplitudes atenuadas o amplificadas con respecto a las señales originales, es decir, que pueden ser multiplicados o divididos por algún escalar. Otra indeterminación, concierne al orden de las señales recuperadas. Las componentes se pueden permutar, sin que por ello varíe el orden de las componentes de las observaciones $e_i(t)$.

El método de ICA desarrollado por Jutten y Hérault [JUT91] utilizó una técnica basada en descorrelación no lineal y la arquitectura de una red neuronal. La arquitectura se muestra en la figura 5.

Figura 5: Arquitectura neuronal de un filtro adaptable lineal recursivo

Cada círculo negro representa una operación de suma y cada triángulo representa la función de activación de la neurona. Esta arquitectura es similar a la red neuronal de Hopfield [HAY99]. El sistema es descrito por la siguiente ecuación:

$$e_i(t) = x_i(t) - \sum_{j \neq i} w_{ij} \varphi_j(x_j(t)) \qquad (5)$$

En donde $e_i(t)$ es la señal de salida del i-ésimo transductor, $x_i(t)$ es la i-ésima señal de salida de la red neuronal (señal estimada) y w_{ij} es el valor del peso sináptico. Expresando la ecuación 5 en notación matricial:

$$\mathbf{e} = \mathbf{x} - \mathbf{W}\varphi(\mathbf{x}) \qquad (6)$$

De esta ecuación se obtiene que:

$$\mathbf{x} = \mathbf{e} + \mathbf{W}\varphi(\mathbf{x}) \qquad (7)$$

Siendo $\mathbf{W}$ la matriz de los coeficientes de pesos inhibitorios de la red e $\mathbf{I}$ la matriz identidad.

Para el caso de dos fuentes se tiene:

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 0 & w_{12} \\ w_{21} & 0 \end{bmatrix} \begin{bmatrix} \varphi_1(x_1) \\ \varphi_2(x_2) \end{bmatrix} \qquad (8)$$

Jutten y Hérault desarrollaron su algoritmo matemático de la regla de aprendizaje basados en el método de descenso del gradiente.

$$\frac{dw_{ij}}{dt} = h \cdot \varphi_i(x_i) \cdot \psi_j(x_j) \qquad (9)$$

Donde h es la razón de aprendizaje y define la velocidad y estabilidad de la red. El algoritmo matemático ajusta en forma continua los pesos sinápticos ($\mathbf{W}$) de la red neuronal de acuerdo con la siguiente expresión, conocida como la regla de Hebb [HAY99] o regla de adaptación:

$$\Delta w_{ij} = h \cdot \varphi_i(x_i) \cdot \psi_j(x_j) \qquad (10)$$

Para lograr la independencia estadística, Jutten y Hérault propusieron el uso de dos funciones no lineales e impares. Expresando estas dos funciones no lineales (f y g) en series de Taylor:

$$f(x) = x - \frac{1}{3}x^3, \quad g(x) = x - \frac{1}{3}x^3 \qquad (11)$$

El n-ésimo momento de una señal estacionaria se define como:

$$m_n = \int_{-\infty}^{\infty} x^n p(x)\, dx \qquad (12)$$

En donde m_n es el promedio en un intervalo de tiempo Δt. De la ecuación (11) y (9), se define la convergencia del algoritmo con la siguiente condición:

$$m_{1,3}(x_1, x_2) = \mathbb{E}\{x_1 x_2^3\} - 3\mathbb{E}\{x_1 x_2\}\mathbb{E}\{x_2^2\} = 0 \qquad (13)$$

La convergencia se alcanza cuando los momentos de orden superior $m_{1,3}$ son cero, esto implica $\mathbb{E}\{x_1 x_2^3\} = \mathbb{E}\{x_1 x_2\} = 0$, logrando independencia estadística entre x_1 y x_2. Un gran número de funciones impares pueden ser utilizadas para la separación.

Experimentalmente fueron determinadas las siguientes restricciones en f y g asegurando una convergencia rápida y estable de los pesos.

- Para una convergencia estable, f debe tener una curvatura positiva y g una curvatura negativa.
- Para una convergencia rápida, f y g deben ser ortogonales en el origen.

Las funciones f y g propuestas por Jutten y Hérault son:

Par 1: $f(x) = x$, $g(x) = x^3$ (14)

Par 2: $f(x) = \tanh(x)$, $g(x) = \tanh(10x)$ (15)

A pesar de que $f(x) = x$ es una función lineal, la combinación de ambas funciones consiguen la separación de fuentes independientes. Un estudio más detallado del trabajo de Hérault y Jutten puede encontrarse en [JUT91].

Un sistema de separación de señales puede implementarse de forma analógica funcionando en forma continua [COH92]. Lo usual, sin embargo, es implementarlo en un sistema digital. En este caso, la adquisición de las señales, $\mathbf{e}(t)$, se efectúa con convertidores A/D situados después de los censores. De esta forma las señales a procesar están discretizadas en el tiempo, $e(m)$, con $t = mT_s$ y $m = 0, 1, \ldots$, siendo T_s el periodo de muestreo.

Otro método desarrollado se basó en el principio de maximización de la información mutua [WON00] y se encuentra muy relacionado al de máxima verosimilitud [CAR97b]. Este método consiste en la medida de la entropía relativa de dos o más señales de entrada a fin de obtener la maximización de la información (Infomax). En la figura 6 se muestra la arquitectura de la red neuronal utilizada para efectuar la maximización de la transferencia de información y minimiza la información mutua entre sus salidas.

Figura 6: Arquitectura de la red neuronal.

La entropía se define como:

$$H(X) = -\int p(x) \ln p(x)\, dx \qquad (16)$$

Donde X es el conjunto de variables aleatorias de x. La regla de aprendizaje de esta red, que encuentra a la matriz $\mathbf{W}$, es derivada partiendo de la maximización de la entropía ($H(X)$) de salida de la red. Como fue propuesto por Bell y Sejnowski [BEL95].

La entropía conjunta de la salida de la red, está dada por:

$$H(\mathbf{y}) = \sum_i H(y_i) - I(\mathbf{y}) \qquad (17)$$

Donde $H(y_i)$ son las entropías marginales de la i-ésima salida, e $I(\mathbf{y})$ es la información mutua entre las salidas. Maximizar la entropía de juntura $H(\mathbf{y})$ de la salida de la red, equivale a maximizar las entropías marginales $H(y_i)$ y minimizar la información mutua ($I(\mathbf{y})$).

Por lo tanto, el valor máximo que se puede lograr alcanzar, es cuando no existe información mutua entre sus salidas, es decir que no contienen información adicional acerca de otra salida. En consecuencia, el término de $I(\mathbf{y})$ será igual a cero y su entropía marginal será igual a su entropía de juntura (entropía máxima).

Bell y Sejnowski eligieron una función no lineal $g(u)$ (función logística), de la cual su derivada es una función de distribución de probabilidad similar a una función super-gaussiana (gaussiana con distribución más concentrada en la media y más dispersa en los extremos), asumiendo de antemano que las fuentes originales tienen distribuciones super-gaussianas, como se verá más adelante.

Un parámetro de ajuste es la matriz $\mathbf{W}$. Esta matriz obtiene su valor óptimo a través de maximizar la entropía $H(\mathbf{y})$ con relación a $\mathbf{W}$. Por lo tanto, al derivar la ecuación 17 con respecto a $\mathbf{W}$, se obtiene la igualdad con la divergencia de Kullback-Leibler entre el estimado de la distribución de la fuente $\hat{p}(\mathbf{s})$ y la fuente $p(\mathbf{s})$.

$$\frac{\partial}{\partial \mathbf{W}} H(\mathbf{y}) = \frac{\partial}{\partial \mathbf{W}} \left[\ln |\det \mathbf{W}| + \sum_i \ln \frac{dg_i}{du_i} \right] \qquad (18)$$

Considerando el caso del valor $\mathbf{W}$ óptimo cuando la entropía es máxima (y la información mutua es igual a cero), el estimado de la distribución de la fuente $\hat{p}(\mathbf{s})$ y la fuente $p(\mathbf{s})$ serán iguales. Puesto que, la función no lineal $g_i(u_i)$ a la salida de la red (ver Fig. 6) no introduce dependencia entre las señales, la información mutua seguirá siendo igual a cero ($I(\mathbf{y}) = 0$).

Figura 7: Método de máxima entropía para la separación de fuentes.

En donde cada entrada $e_i(t)$ está formada por la mezcla de las señales y cada salida $u_i(t)$ es la señal estimada de la fuente independiente. El objetivo es encontrar una matriz $\mathbf{W}$ inversa a la matriz de mezcla $\mathbf{A}$, que recupere las señales de las fuentes independientes.

De acuerdo a la figura 7, la función de transferencia entre u_i, u_i se denota como:

$$u_i = \varphi_i(e_i) = \frac{1}{1 + e^{-e_i}} \qquad (19)$$

La relación entre e_i, u_i y la función de transferencia no lineal es (PAP91). Para una distribución uniforme de e_i, se tiene que:

$$p(u_i) = \frac{d\varphi_i^{-1}(u_i)}{du_i} \qquad (20)$$

Esto asume que u_i es una variable independiente con una distribución de la forma de la derivada de la no linealidad. En este caso la función empleada es la función logística (Ver figura 8).

Figura 8: Función logística y su derivada.

La función logística (o curva logística, modeliza a una función sigmoide) es ampliamente utilizada en redes neuronales para estimar distribuciones. Esta función de activación fue propuesta por Bell y Sejnowski, demostrando su convergencia para señales de audio como la voz y música [WON99].

La función no lineal ($g_i(u_i)$) es empleada principalmente para minimizar la información mutua y así obtener salidas independientes. Otra interpretación para el uso de la no-linealidad es la minimización de la correlación de orden superior (por su expansión en series de Taylor).

Como se mencionó anteriormente de la ecuación 17, la independencia estadística se obtiene cuando la información mutua de la red es cero y la entropía de juntura es igual a la suma de sus entropías marginales, donde la expresión de la entropía marginal se representa como:

$$H(u_i) = -\int p(u_i) \ln p(u_i)\, du_i \qquad (21)$$

La no linealidad realiza una transformación de la variable u_i a y_i. Por lo tanto la ecuación 29 se puede expresar como:

$$H(y_i) = H(u_i) + \mathbb{E}\left[\ln\left|\frac{du_i}{dy_i}\right|\right] \qquad (22)$$

Sustituyendo la ecuación 21 y 22 en la ecuación 17, se obtiene:

$$H(\mathbf{y}) = \sum_i H(u_i) + \sum_i \mathbb{E}\left[\ln\left|\frac{du_i}{dy_i}\right|\right] - I(\mathbf{u}) \qquad (23)$$

$$H(\mathbf{y}) = \sum_i H(u_i) + \mathbb{E}\left[\ln\left|\det\frac{\partial \mathbf{u}}{\partial \mathbf{y}}\right|\right] - I(\mathbf{u}) \qquad (24)$$

Realizando la derivada de la ecuación anterior respecto a $\mathbf{W}$ se obtiene:

$$\frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} = \frac{\partial}{\partial \mathbf{W}}\left[\ln\left|\det \mathbf{W}\right| + \mathbb{E}\left[\sum_i \ln\left|\frac{du_i}{dy_i}\right|\right] - I(\mathbf{u})\right] \qquad (25)$$

En esta ecuación se observa la relación de la entropía de juntura con la información mutua. La minimización directa de la información mutua genera la maximización de la entropía de juntura. Dicho de otra forma, la información mutua será minimizada ($I(\mathbf{u})$) cuando la no linealidad $\frac{du_i}{dy_i}$ es aproximada a la función de densidad acumulativa (f.d.a) de la fuente estimada (u_i).

En el caso en el que la fuente estimada y la función no lineal resulten en un valor $I(\mathbf{y})$ diferente de cero, existirá un error. Por lo tanto, en la ecuación 25 se plantea que el proceso de minimización de $I(\mathbf{y})$ depende de las fuentes estimadas y de la función no lineal empleada. No obstante, el término de error para las aplicaciones propuestas (empleando una función logística y señales con distribución super-gausiana) se puede considerar como despreciable. En este caso, el término del error desaparece y el máximo de la entropía de juntura se obtiene derivando la entropía $H(\mathbf{y})$ con respecto a $\mathbf{W}$. Esto es calculando el gradiente de $H(\mathbf{y})$.

$$\frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} = \left(\mathbf{W}^{-1}\right)^T + \mathbb{E}\left[\varphi(\mathbf{y})\mathbf{e}^T\right] \qquad (26)$$

Esta ecuación es resultado del gradiente de la función de la entropía. Una forma de maximizar la entropía es por medio del gradiente “natural”, el cual se logra multiplicando por $\mathbf{W}^T\mathbf{W}$. Por lo tanto, la ecuación 26 se expresa como:

$$\Delta \mathbf{W} \propto \left(\mathbf{I} + \mathbb{E}\left[\varphi(\mathbf{y})\mathbf{y}^T\right]\right)\mathbf{W} \qquad (27)$$

Donde $\mathbf{I}$ es la matriz identidad y el segundo factor denota la no linealidad.

$$\varphi(y_i) = \frac{\partial}{\partial y_i} \ln \frac{dg_i}{dy_i} = \frac{g_i''(y_i)}{g_i'(y_i)} \qquad (28)$$

La ecuación (27) resultante es:

$$\Delta \mathbf{W} \propto \left(\mathbf{I} - \mathbb{E}\left[\varphi(\mathbf{y})\mathbf{y}^T\right]\right)\mathbf{W} \qquad (29)$$

En donde el término $\varphi(y_i)$ es la función no lineal y es determinante para la separación de fuentes sub y súper-gaussianas. Algunas funciones no lineales fueron propuestas por Bell y Sejnowski [BEL95][WON00]. La ecuación (29) se conoce con el nombre de regla de aprendizaje y se emplea en un modelo neuronal (regla de Hebb) recursivo para el ajuste de los pesos sinápticos en donde en cada iteración se realizará una actualización hasta encontrar su valor óptimo.

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta \Delta \mathbf{W}_n \qquad (30)$$

Una forma alternativa de generar la regla de aprendizaje antes mencionada, se deriva de la estimación de máxima verosimilitud (“Maximum Likelihood Estimator” MLE) propuesta por Pearlmutter, Parra y Cardoso [CAR97]. Otros métodos han sido desarrollados [HYV03] utilizando la máxima no-gaussianidad por Curtosis, por Negentropía [HYV00], por búsqueda de proyección, redes de predicción, métodos geométricos, entre otros.

0.1.2. Aplicaciones del análisis de componentes independientes

En los últimos años ha surgido un gran interés en los algoritmos de ICA debido a sus resultados obtenidos bajo las condiciones de no obtener o contar con información detallada de las fuentes originales. Ya que ICA puede ser aplicado a cualquier sistema que emita señales que se propaguen en un medio y que se desee recuperar mediante la percepción de las señales observadas. La aplicación clásica es el problema “cocktail party” y se enfoca al problema de separación de señales EEG (Electroencephalography) y MEG (Magnetoencephalography).

Hari y Oja [OJA95] desarrollaron el método para separar actividad de artefactos utilizando ICA. La suposición esta basada en el hecho de que la actividad y los artefactos son anatómicamente y fisiológicamente procesos independientes, y esta separación se refleja en la independencia estadística entre las señales magnéticas generadas por dichos procesos. Los experimentos con señales EEG fueron reportados por Bell y Sejnowski en 1996 y consisten en separar las señales eléctricas generadas en la corteza cerebral, revelando información importante de la actividad cerebral. Otras aplicaciones se pueden encontrar, por ejemplo, todas aquellas en las que interese eliminar un ruido, la biomedicina, sistemas de audio, radiocomunicaciones, etc.

Existen otros campos de aplicación, como la eliminación de ruido en imágenes [HYV99][JEN03] y procesamiento, en general [SAH96], análisis de datos sísmicos [THI96], identificación de componentes químicos en una mezcla [TRI03], análisis de datos financieros [BAC98], etc.

Conclusiones

Se realizó una descripción y el planteamiento general del problema, y se justificaron las indeterminaciones e hipótesis que habitualmente se realizan en los algoritmos de ICA. Una vez establecido este marco general, se resumen los métodos basados en la minimización de la dependencia estadística y principio de máxima entropía.

Desde que en Abril de 1986 Jeanny Hérault y Christian Jutten, en una reunión sobre Computación con Redes Neuronales celebrada en Snowbird (Utah), enunciaron por primera vez el problema de separación de fuentes y presentaron un modelo de red neuronal recurrente con una regla de aprendizaje hebbiana capaz de separar mezclas lineales de señales independientes, no han dejado de surgir aplicaciones reales de este tipo de sistemas. Se trató en este artículo de enumerar algunas de las de mayor interés.

Los dos métodos abordados en este trabajo han sido desarrollados mediante modelos neuronales y su aplicación se enfoca a sistemas que operen en tiempo real, ya sea mediante el diseño de un Circuito Integrado de Aplicación Específica (ASIC - Application-Specific Integrated Circuits) [COH92][OLI06b], Dispositivos Programables (FPGA - Field Programmable Gate Array) [MIN03][KUO08][OLI08] o Procesadores de Señales (DSP - Digital Signal Processor) [OLI06].

0.2. Referencias

[ALM99] L.B.Almeida, G.C.Marques, “Nonlinear Blind Source Separation by Pattern Repulsion”. Lecture Notes in Computer Science, vol. 1607, pp.674-682, Springer-Verlag, 1999.

[BAC98] A. D. Back, A. S. Weigend, “A first application of independent component analysis to extracting structure from stock returns”. International Journal of Neural Systems, vol. 9. Special Issue on Data Mining in Finance, 1998.

[BEL95] A. Bell, T. Sejnowski, “An information-maximisation approach to blind separation and blind deconvolution”. Neural Computation, vol. 7, pp. 1129-1159, 1995.

[CAR97] J. F. Cardoso, “Statistical principles of source separation”. Proceedings of the SYSID’97, 11th IFAC symposium on system identification, pp. 1837-1844, 1997.

[CAR97b] J. F. Cardoso, “Infomax and Maximum Likelihood for Blind Source Separation”. IEEE Signal Processing Letters, Vol. 4; Issue 4, pp. 112-114, April 1997.

[COH92] M. H. Cohen, A. G. Andreou, “Current-mode subthreshold MOS implementation of the Hérault-Jutten autoadaptive network”. IEEE journal of solid-state circuits, vol. 27, n. 5, May 1992.

[HAY99] S. Haykin, “Neural Networks”, Prentice Hall, 1999.

[HYV00] A. Hyvarinen, E. Oja, “Independent Component Analysis: algorithm and applications”. Neural Networks, vol 13 , March 28, pp. 411-430, Elsevier Science, 2000.

[HYV03] A. Hyvärinen, J. Karhunen, E. Oja. “Independent Component Analysis”. John Wiley & Sons, Inc. 2003.

[HYV99] A. Hyvarinen, P. Hoyer, E. Oja, “Denoising of nongaussian data by independent component analysis and sparse coding”. Proc of ICA’99, pp. 485-490, January 11-15, Aussois (France), 1999.

[JEN03] R. Jenssen and T. Eltoft, “ICA FilterBank for Segmentation of Textured Images”. 4th International Symposium on Independent Component Analysis and Blind Signal Separation, Proc of ICA’2003, April 2003, Nara (Japan), 2003.

[JUT91] C. Jutten, and J. Hérault, ”Blind Separation of sources, part I, II, III”. Signal Processing, vol.24, pp.1-29, 1991.

[KUO08] Kuo-Kai Shyu, Ming-Huan Lee and Po-Lei Lee, “Implementation of Pipelined FastICA on FPGA for Real-Time Blind Source Separation”. IEEE transactions on Neural Networks, pp 958-970, Vol. 19, No. 6, June 2008.

[MIN03] C. Min Kim, H. Min Park, T. Kim, Y. Kyung Choi, S. Young, “FPGA Implementation of ICA Algorithm for Blind Signal Separation and Adaptive Noise Canceling”. IEEE transactions on Neural Networks, pp. 1038-1046, Vol. 14, No. 5, September 2003.

[OJA95] E. Oja, J. Karhunen, L. Wang, and R. Vigario, ”Principal and independent components in neural-networks - recent developments”, In Proc. Italian Workshop on Neural Networks, WIRN’95, Vietri, Italy, 1995.

[OLI06] L. N. Oliva, J.A. Moreno, L. M. Flores and F. Gomez, “DSP Implementation of Extended Infomax ICA Algorithm for Blind Source Separation”, 3rd International Conference on Electrical and Electronics Engineering (ICEEE 2006), Veracruz, Mexico, September 6-8, 2006.

[OLI06b] L. N. Oliva, A. Oliverio, J.A. Moreno, L. M. Flores y F. Gómez , “Implementación de un Algoritmo ICA con Circuitos CMOS Analógicos para el Problema de Separación Ciega de Fuentes”, XII Workshop Iberchip, San José, Costa Rica, 22 a 24 de Marzo de 2006.

[OLI08] L. N. Oliva, M. A. Aleman, J. García, J. A. Moreno, “Implementation of Blind Source Separation for Multi-input”, Proceedings of the 5th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE 2008), pp. 470-474. ISBN: CFP08827-CDR. Mexico City, Mexico, Nov. 12-14, 2008.

[PUN99] C.G. Puntonet; M. R. Alvarez; A. Prieto; B. Prieto. ”Separation of Speech Signals for Nonlinear Mixtures”. Lecture Notes in Computer Science, vol. 1607, pp.665-673, Springer-Verlag, 1999.

[SAH96] H. Sahlin, H. Broman, “Blind separation of images”, Proceedings of the 30th ASILOMAR conference on signals, systems and computers, 1996.

[THI96] N. Thirion, J. Mars, F. Glangeaud, "Séparation d'ondes dans le cas des signaux de prospection sismique", Proceedings of Ecole des Techniques Avancées en Signal Image Parole, pp. 357-366, Grenoble, France, Sept. 2-6, 1996.

[TRI03] S. Triadaphillou, A.J.Morris and E.B. Martin, "Application of Independent Analysis to Chemical Reactions". 4th International Symposium on Independent Component Analysis and Blind Signal Separation, Proc of ICA'2003, April 2003, Nara (Japan), 2003.

[WON00] Te-Won Lee, "Independent Component Analysis Theory and applications". Kluwer Academic Publishers, 2000.

[WON99] H-C. Wu, J. C. Principe, "Simultaneous diagonalization in the frequency domain (SDIF) for source separation". Proc of ICA'99, pp. 245-250, January 11-15, Aussois (France), 1999.