SEPARACIÓN CIEGA DE FUENTES DERIVANDO LOS CUMULANTES DE LAS SALIDAS.

Rubén Martín Clemente, José I. Acha

Área de Teoría de la Señal y Comunicaciones,
Escuela Superior de Ingenieros
Universidad de Sevilla
ruben@cica.es, acha@viento.us.es

RESUMEN

In this paper, we consider the Blind Separation of instantaneous mixtures of non-Gaussian sources. It is proven that the separation can be based on the cancellation of certain second-order partial derivatives of the output cross-cumulants.

1. INTRODUCCIÓN

La "Separación Ciega de Fuentes" es uno de los campos más activos en la moderna Teoría de la Señal. En este problema, se trata de estimar señales *no* observables (a las que se llama "fuentes") a partir de señales ("observaciones") que son combinación o "mezcla" de dichas fuentes, siendo estas observaciones *el único dato experimental disponible* (el adjetivo "ciega" enfatiza el hecho de que, por hipótesis, ni las señales fuente ni la estructura de su mezcla deben ser conocidas *a priori*). Las aplicaciones aparecen en varios ámbitos: p.ej., éste es el problema de separar las voces de personas que hablan simultáneamente (lo que, en inglés, se ha llamado "cocktail-party problem") o, en Obstetricia, de estimar el electrocardiograma fetal a partir de señales registradas sobre el vientre de la embarazada; de hecho, la "Separación de Fuentes" nació en el campo de la Ingeniería Biomédica, asociada al estudio de la codificación en el sistema nervioso del movimiento muscular.

La característica que hace posible la Separación es la *independencia estadística* de las fuentes. Así, la solución propuesta se basa en la transformación de las observaciones para hacerlas estadísticamente independientes entre sí.

2. DESCRIPCIÓN DEL PROBLEMA

El modelo de mezcla *lineal*, *invariante* en el tiempo y *sin memoria* viene dado por la ecuación:

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) \qquad (1)$$

donde $\mathbf{s}(t) = [s_1(t), ..., s_N(t)]^T$ es un vector de que contiene las N fuentes, $\mathbf{A} = (a_{ij})$ es una matriz *desconocida* de dimensiones N×N ("matriz de mezcla") que describe la propagación entre fuentes y electrodos y $\mathbf{x}(t)$ es el vector que recoge las observaciones, generalmente la medida de N sensores, siendo el *único* dato disponible. Por hipótesis, 1.) las señales fuente son realizaciones de procesos estocásticos estacionarios, de valor medio cero y varianza –potencia– unidad, 2.) las *señales fuente son estadísticamente independientes entre sí* –verosímil cuando tienen distinto origen–, 3.) a lo más *una* de las fuentes es de distribución estadística *gaussiana* y 4.) la *matriz de mezcla* **A** *es invertible* (lo que implica que debe haber tantos sensores como fuentes). Con estas suposiciones, es posible demostrar que la matriz de mezcla **A** *puede ser identificada, salvo factores de escala y/o permutaciones en sus filas* (corolario del Teorema de Darmois-Skitovich) [1].

La Separación se consigue determinando una matriz **B** ("matriz de separación") de dimensiones N×N tal que:

$$\mathbf{y}(t) = \mathbf{B}\,\mathbf{x}(t) = \mathbf{G}\,\mathbf{s}(t) \qquad (2)$$

es un estimador de $\mathbf{s}(t)$, esto es, la *matriz global de transferencia* $\mathbf{G} = \mathbf{B}\,\mathbf{A}$ tiene *un* y *sólo un* elemento *no nulo* por fila y columna. En este caso, se dice que **G** es una *matriz de permutación generalizada*.

3. DERIVANDO LOS CUMULANTES DE LAS ESTIMACIONES

Utilizando las propiedades de los cumulantes, el cumulante cruzado $cum_{31}(y_i(t),y_j(t))$ puede ser desarrollado como:

$$cum_{31}(y_i(t),y_j(t)) = \sum_{l=1}^{N} g_{il}^3\, g_{jl}\, \kappa_l \qquad (3)$$

donde κ_l denota la curtosis de la l-ésima fuente y g_{ij} es el elemento (i,j) de la matriz **G**. Igualando (3) a cero para todo $i \neq j$, obtenemos un sistema de ecuaciones resuelto cuando **G** es una matriz de permutación generalizada, aunque podrían existir soluciones no deseadas, dependiendo de los estadísticos de las fuentes. Sin embargo, que (3) se anule cuando **G** es una matriz de permutación generalizada se debe a que todos los sumandos de (3) se cancelan; pero no a que los coeficientes 'g_{ij}' que aparecen allí estén elevados al *cubo*. De hecho, podemos simplificar (3) derivándolo sin perder por ello ninguna propiedad fundamental. Sea:

$$\Delta_{ij} = \frac{1}{6}\frac{\partial^2}{\partial b_{ij}^2} \; cum_{31}(y_i(t),y_j(t)) = \sum_{l=1}^{N} g_{il} g_{jl} a_{jl}^2 \kappa_l \,, \quad (i \neq j) \qquad (4)$$

donde a_{ij} es el elemento (i,j) de la matriz **A**. Además, utilizando las propiedades de los cumulantes, (4) puede ser escrito como una función de los estadísticos de las observaciones, siendo, por tanto, medible:

$$\Delta_{ij} = \sum_{l=1}^{N} \sum_{m=1}^{N} b_{il} b_{jm} cum(x_l, x_m, x_j, x_j) \qquad (5)$$

Obsérvese que tanto (3) como (4) se cancelan cuando **G** es una matriz de permutación generalizada. Sin embargo, (4) es bastante más simple que (3), ya que sólo depende de potencias de orden *dos* de la matriz **B**. Pues bien, es *necesario y suficiente* que **B** satisfaga el sistema de ecuaciones $\Delta_{ij} = 0$ ($\forall\ i \neq j$) para ser una *matriz de separación*.

PRUEBA. Soluciones triviales, en las que alguna de las señales de salida se cancela, pueden ocurrir si la matriz de separación no es de rango completo. Para evitarlo, supondremos en adelante que tanto **A** como **B** y **G** son matrices *ortogonales*. Esto no implica pérdida de generalidad, pues siempre puede ser conseguido decorrelando las observaciones [1].

De (4), es evidente que $\Delta_{ij} = 0$ ($i \neq j$) si **G** es una matriz de permutación generalizada. Ahora bien, la proposición recíproca también es cierta: sea $\boldsymbol{L}_j$ la matriz diagonal cuya entrada (p,p) es $a_{jp}^2 \kappa_p$. Podemos suponer que todos los autovalores de la matriz $\boldsymbol{L}_j$ son distintos: de no ser así, siempre podemos encontrar una matriz ortogonal **U** tal que la nueva matriz de mezcla **U A** hace que se verifique esta condición. Entonces, (4) se puede escribir como:

$$\Delta_{ij} = \boldsymbol{g}_i \boldsymbol{L}_j \boldsymbol{g}_j \qquad (6)$$

siendo $\boldsymbol{g}_p = [g_{p1}(t), ..., g_{pN}]^T$. El vector $\boldsymbol{L}_j \boldsymbol{g}_j$ puede ser siempre obtenido a partir de la combinación lineal de los vectores $\boldsymbol{g}_1, ..., \boldsymbol{g}_N$, al ser estos una base ortogonal del espacio. Así:

$$\boldsymbol{L}_j \boldsymbol{g}_j = \Delta_{jj}\, \boldsymbol{g}_j + \Sigma_{\,i \neq j}\, \Delta_{ij}\, \boldsymbol{g}_i \qquad (7)$$

Supongamos que $\Delta_{ij} = 0$ ($i \neq j$). Entonces, de (7):

$$\boldsymbol{L}_j \boldsymbol{g}_j = \Delta_{jj}\, \boldsymbol{g}_j \qquad (8)$$

Esta última relación implica que $\boldsymbol{g}_j$ es un autovector de una matriz diagonal, es decir, un *vector canónico*. Con ello queda demostrado que **G** es una matriz de permutación generalizada.

4. ALGORITMO DE SEPARACIÓN

En vista del resultado anterior, proponemos optimizar la siguiente función de coste:

$$f(\mathbf{B}) = \Sigma_{\,i \neq j}\, \Delta_{ij}^2 \qquad (9)$$

que debe ser minimizada con la restricción:

$$\mathbf{B}\mathbf{B}^T = \mathbf{I} \qquad (10)$$

Para este fin, proponemos el algoritmo:

$$\Delta\mathbf{B} = -\mu(\nabla f(\mathbf{B}) - \mathbf{B}\, \nabla f(\mathbf{B})^T\, \mathbf{B}) \qquad (11)$$

siendo $\nabla f(\mathbf{B})$ la matriz cuya componente (i,j) es la derivada parcial de $f(\mathbf{B})$ respecto a b_{ij} y $\mu > 0$. La regla de adaptación (11) actualiza **B** en la dirección de *máximo descenso* de la función de coste, sujeta a la restricción (10) [2].

5. EXPERIMENTOS

La validez del algoritmo (11) ha sido corroborada por medio de numerosos experimentos. La Figura 1 presenta la evolución de la relación señal a ruido para una mezcla de tres fuentes de distribución uniforme. Los estadísticos de las observaciones fueron estimados utilizando 1000 muestras. La gráfica es el promedio de diez experimentos independientes.

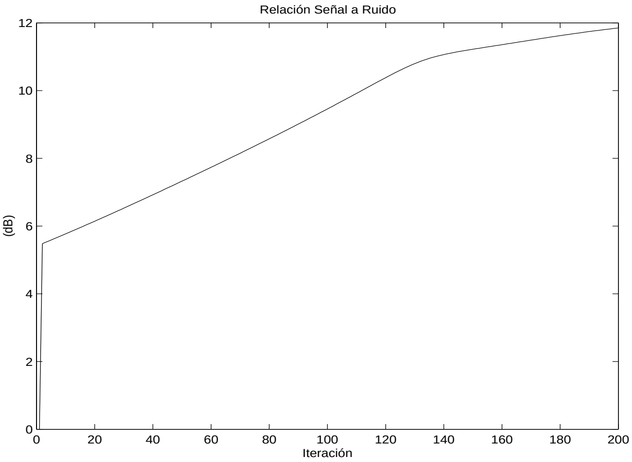

Figura 1. Separación de tres fuentes.

6. CONCLUSIONES

En el presente artículo se obtiene un conjunto de ecuaciones que es *necesario* y *suficiente* para garantizar la Separación de las Fuentes. Es destacable el hecho de que no se presupone ningún conocimiento *a priori* sobre las funciones de distribución de probabilidad de las fuentes.

7. REFERENCIAS

[1] P. Comon, "Indep. Component Analysis-A new Concept?", *Signal Processing*, 36(3), 1994.

[2] J. Cardoso, B. Laheld, "Equivariant Adaptive Source Separation", *IEEE Trans. on SP*, 45(2), 1996.