

2. Fundamentos básicos del Análisis de Componentes Independientes

2.1 Introducción

El objetivo fundamental del *Análisis de Componentes Independientes (ICA)* es el de proporcionar un método que permita encontrar una representación lineal de los datos no gaussianos de forma que las componentes sean estadísticamente independientes o lo más independiente posible. Una representación de este tipo permite obtener la estructura fundamental de los datos en muchas aplicaciones, incluidas la extracción de características y la separación de señales.

Seguidamente se va a describir el fundamento teórico en el que se basa dicho método, introduciendo a su vez los conceptos como el *Análisis de Componentes Principales (PCA)* o la *Separación Ciega de Fuentes (BSS)*.

2.2 Representación de datos

2.2.1 Consideraciones estadísticas iniciales

Un problema clásico en estadística y en diversas áreas relacionadas con la misma, es cómo encontrar una representación conveniente de los datos multidimensionales, entendiendo por ‘representación’ los medios a través de los cuales transformamos dichos datos de forma que su estructura esencial se haga lo más visible o accesible que podamos.

En computación neuronal, este problema pertenece al área del ‘Aprendizaje No Supervisado’, donde dicha representación debe aprenderse a partir del propio conjunto de datos, sin ninguna entrada externa previa, que es lo que constituye la diferencia esencial con el ‘Aprendizaje Supervisado’. En procesado de señales, podemos encontrar este mismo problema de la extracción de rasgos, como veremos en el problema de ‘Separación Ciega de Fuentes’ (BSS) más adelante.

Para ilustrar el problema inicial, asumamos que los datos consisten en un número determinado de variables que hemos observado juntas. Denotaremos dicho número de variables por m y el número de observaciones de las mismas por T . Podemos entonces denotar los datos (señales observadas) $x_i(t)$ donde el índice i y t toman valores $i = 1, \dots, m$ y $t = 1, \dots, T$. Las dimensiones m y T , pueden tomar valores bastante elevados.

Una formulación general del problema se puede realizar considerando que el objetivo es encontrar una función que lleve a cabo la transformación del conjunto original de datos en un espacio m -dimensional a otro espacio n -dimensional, de forma que las variables transformadas en este nuevo espacio nos den tanta información como sea posible sobre las componentes o rasgos que están ocultos en el conjunto de datos originales. En decir, buscamos que las variables transformadas sean las *componentes*

subyacentes que describan la estructura esencial de dichos datos. Es de esperar que dichas componentes correspondan a alguna causa física en la que estaba inmerso el proceso que generó los datos en primer lugar.

Además, nosotros vamos a considerar únicamente funciones de transformación lineales, ya que la interpretación de la representación de los datos que se lleva a cabo por medio de ellas es más simple, y también su programación. Consecuentemente, cada una de dichas componentes y_i , se puede formular como una combinación lineal de las variables observadas:

$$y_i(t) = \sum_j w_{ij} x_j(t), i = 1, \dots, n, j = 1, \dots, m \quad (2.1)$$

donde los w_{ij} se definen como coeficientes que permiten definir la representación. El problema por tanto se puede reformular como el del cálculo de la determinación de los coeficientes w_{ij} . Mediante el uso del Álgebra lineal es posible expresar la transformación lineal de la ecuación (2.1) como un producto de matrices. Así, agrupando los w_{ij} en una matriz \mathbf{W} , la ecuación anterior se convierte en:

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \mathbf{W} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \quad (2.2)$$

que en forma matricial será

$$\mathbf{y} = \mathbf{W} \cdot \mathbf{x} \quad (2.3)$$

Una aproximación estadística básica consiste en tratar a $x_j(t)$ como un conjunto de T realizaciones de m variables aleatorias. De este modo, cada muestra de la señal $x_j(t)$, $t = 1, \dots, T$, es a su vez una muestra de una variable aleatoria, la cual denotaremos por x_j . En este contexto, podríamos tratar de determinar la matriz \mathbf{W} de acuerdo a ciertas propiedades estadísticas de las componentes transformadas u_i . En las siguientes secciones discutiremos algunas propiedades de esta transformación, y a partir de una de ellas llegaremos al '*Análisis de las Componentes Independientes*' [Hyvärinen01].

2.2.2 Reducción de dimensión para la obtención de \mathbf{W}

Un principio estadístico clásico que podríamos emplear para la elección de la matriz \mathbf{W} es el de limitar el número de las componentes y_i a ser menor, en quizás una o dos unidades que el conjunto de datos originales, determinando entonces \mathbf{W} de forma que las y_i contengan tanta información sobre los datos originales como sea. Esto nos lleva a una familia de técnicas conocidas como '*Análisis de las Componentes Principales*'.

2.2.3 Uso de la *independencia* para la estimación de \mathbf{W}

Otro principio estadístico usado para determinar la matriz \mathbf{W} es la *independencia*, es decir, las componentes y_i son estadísticamente independientes unas de otras, lo que significa que el valor que tome cualquiera de ellas *no* da información alguna sobre el valor que pueda tomar el resto.

Esto resulta sencillo si los datos tienen una distribución Gaussiana, dado que es bastante simple encontrar componentes que sean independientes en este caso, atendiendo a que para datos Gaussianos, las componentes decorreladas son siempre independientes [Hyvärinen01].

Sin embargo, en la realidad los datos no suelen seguir una distribución Gaussiana, y la situación no es tan simple como estos métodos asumen. Muchos conjuntos de datos del mundo real tienen distribuciones supergaussianas, lo que significa que dichas variables aleatorias toman con mayor probabilidad valores que son cercanos al cero o valores muy grandes, en otras palabras, la *función densidad de probabilidad* ('pdf') de estos datos es puntiaguda en el cero y tiene las colas densas (debido a los valores grandes que toma lejos del cero), si la comparamos con la pdf de una variable Gaussiana de la misma varianza. Un ejemplo que ilustre esta situación lo encontramos en la *figura 2.1*.

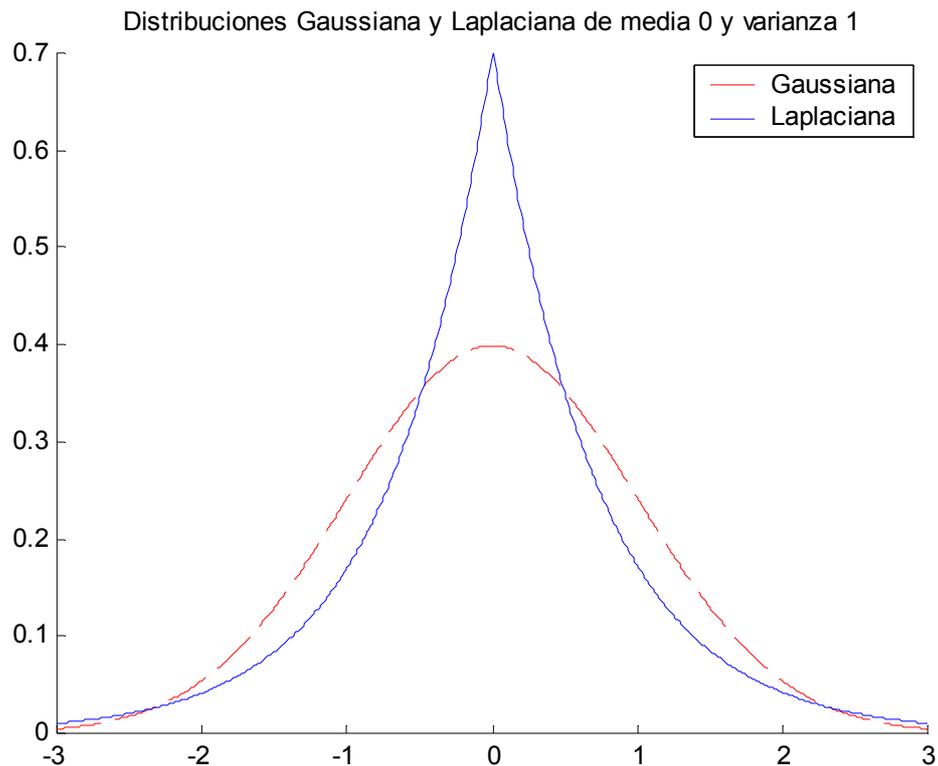


Figura 2.1 La función densidad de probabilidad de una distribución laplaciana es un ejemplo típico de distribución supergaussiana. En la figura se representa una distribución gaussiana (en línea discontinua roja) y una laplaciana (en línea continua azul), estando ambas normalizadas con media cero y varianza unidad.

2.3 Separación Ciega de Fuentes

El problema de la *Separación Ciega de Fuentes (BSS)*, representa una visión alternativa al de encontrar una representación adecuada de los datos. Además en él podemos observar gran parte del trasfondo de ICA, por lo que su estudio se antoja necesario antes de comenzar a analizar el *Análisis de Componentes Independientes* [Hyvärinen01].

2.3.1 Observaciones de mezclas de señales desconocidas

Consideremos un problema donde una fuente emite un cierto número de señales, como por ejemplo, las señales eléctricas correspondientes a diferentes áreas del cerebro o las voces de gente hablando en una misma habitación. Asumamos además que existen una serie de sensores o receptores que van a permitir obtener una grabación de dichas señales mezcladas con unos ciertos coeficientes de ponderación.

Para fijar ideas, supongamos que tenemos tres señales procedentes de la fuente y además tres señales observadas, que han sido recogidas por los sensores. Llamemos $x_1(t)$, $x_2(t)$ y $x_3(t)$ a las señales observadas y $s_1(t)$, $s_2(t)$ y $s_3(t)$ a las señales originales. Se puede comprobar como la relación entre ellas vendrá dada por el siguiente sistema de ecuaciones:

$$\begin{aligned}x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)\end{aligned}\tag{2.4}$$

o de forma análoga:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s}\tag{2.5}$$

Los coeficientes a_{ij} son constantes que dan los pesos de la mezcla y que agruparemos en una matriz denotada por \mathbf{A} . Asumimos que son desconocidos, dado que no podemos conocer los valores de los coeficientes a_{ij} sin saber todas las propiedades físicas del sistema de mezclado (medio), tarea que puede llegar a ser, generalmente, bastante complicada. Supondremos además que las fuentes originales no son conocidas, de modo que lo único que conocemos son las mezclas.

Lo que nos gustaría hacer es encontrar las señales originales $s_1(t)$, $s_2(t)$ y $s_3(t)$, contando únicamente con las señales mezcla observadas en los sensores $x_1(t)$, $x_2(t)$ y $x_3(t)$. Ésta es el problema conocido como ‘Separación Ciega de Fuentes’ (BSS). El hecho de incluir el término ‘Ciega’ viene del hecho de que apenas tenemos una mínima información de las señales originales, aunque lo más normal es que no conozcamos nada de las mismas.

Asumiremos que los coeficientes de mezclado a_{ij} toman una serie de valores que permitan que la matriz \mathbf{A} sea invertible. De este modo, podemos decir que existe una matriz \mathbf{W} con coeficientes w_{ij} tal que nos permitirá separar las señales $s_i(t)$ como:

$$\begin{aligned}
s_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t) \\
s_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t) \\
s_3(t) &= w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t)
\end{aligned}
\tag{2.6}$$

o en forma matricial:

$$\mathbf{s} = \mathbf{W}\mathbf{x}
\tag{2.7}$$

Vemos entonces como podremos encontrar la matriz \mathbf{W} como la inversa de \mathbf{A} , la matriz que formaban los coeficientes a_{ij} , si conocemos dichos coeficientes de mezclado. Ahora observamos que este problema es matemáticamente similar al inicial, donde quisimos encontrar una buena representación de los datos aleatorios $x_j(t)$. Además es posible considerar cada señal $x_i(t)$, $t = 1, \dots, T$ como una muestra de una variable aleatoria x_i , de forma que el valor de dicha variable viene dado por las amplitudes de esa señal en cada instante de tiempo.

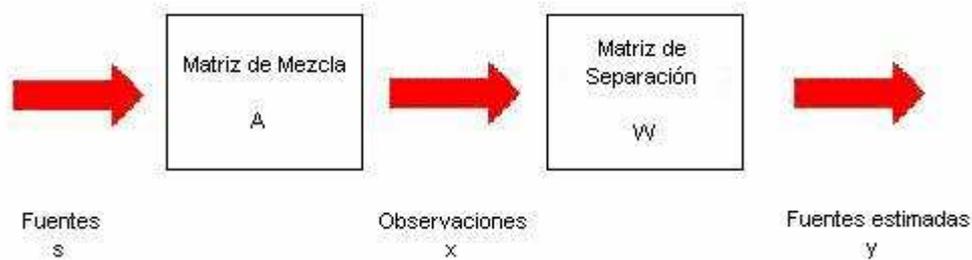


Figura 2.2 Esquema general que representa los procesos de generación de observaciones y estimación de fuentes en el problema de Separación Ciega de Fuentes.

2.3.2 Separación Ciega basada en la independencia

Ahora trataremos de responder a la pregunta de cómo estimar los coeficientes w_{ij} que introducimos en la ecuación (2.6). Para ello, buscaremos un método general que trabaje bajo diferentes circunstancias, y proporcione una respuesta al problema general con el que empezamos: encontrar una buena representación de los datos multidimensionales. Todo lo que tenemos son las señales mezcla, y queremos encontrar una matriz \mathbf{W} que nos permita obtener o estimar las señales originales de las fuentes lo mejor posible.

Una solución sorprendentemente simple al problema puede encontrarse suponiendo únicamente que existe independencia estadística entre las señales originales. De hecho, si las señales son no-Gaussianas, esto es suficiente para determinar los coeficientes w_{ij} , de modo que las señales:

$$\begin{aligned}
y_1(t) &= w_{11}s_1(t) + w_{12}s_2(t) + w_{13}s_3(t) \\
y_2(t) &= w_{21}s_1(t) + w_{22}s_2(t) + w_{23}s_3(t) \\
y_3(t) &= w_{31}s_1(t) + w_{32}s_2(t) + w_{33}s_3(t)
\end{aligned}
\tag{2.8}$$

o en forma matricial

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (2.9)$$

son *estadísticamente independientes*. Si las señales $y_1(t)$, $y_2(t)$ e $y_3(t)$ son realmente independientes entre sí, entonces serán iguales a las señales originales $s_1(t)$, $s_2(t)$ y $s_3(t)$ (salvo quizás un factor constante multiplicativo, hecho que no tiene mucha relevancia atendiendo al significado de lo que podemos conseguir).

Usando simplemente esta información sobre la independencia estadística de partida, podemos estimar la matriz \mathbf{W} a partir únicamente de las señales mezcla que observamos y conociéndola, podemos estimar las fuentes originales. Veamos a continuación un ejemplo en el que se ilustra este problema que nos servirá para comprender mejor estos conceptos.

Haciendo uso del algoritmo *FastICA*[©] (Gävert, Hurri, Särelä e Hyvärinen), que se ejecutará en el entorno de simulación matemática *MATLAB*[©], trataremos de extraer dos fuentes $s_1(t)$ y $s_2(t)$, a partir de dos observaciones procedentes de la mezcla lineal de ellas y que denotaremos como $x_1(t)$ y $x_2(t)$. Los resultados obtenidos se muestran en las siguientes figuras.

SEÑALES ORIGINALES

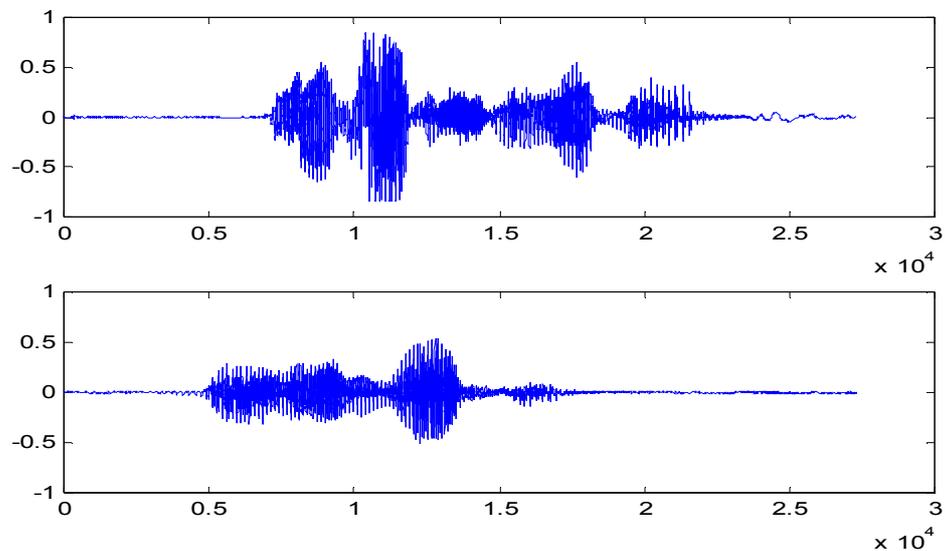


Figura 2.3 Señales originales

SEÑALES OBSERVADAS

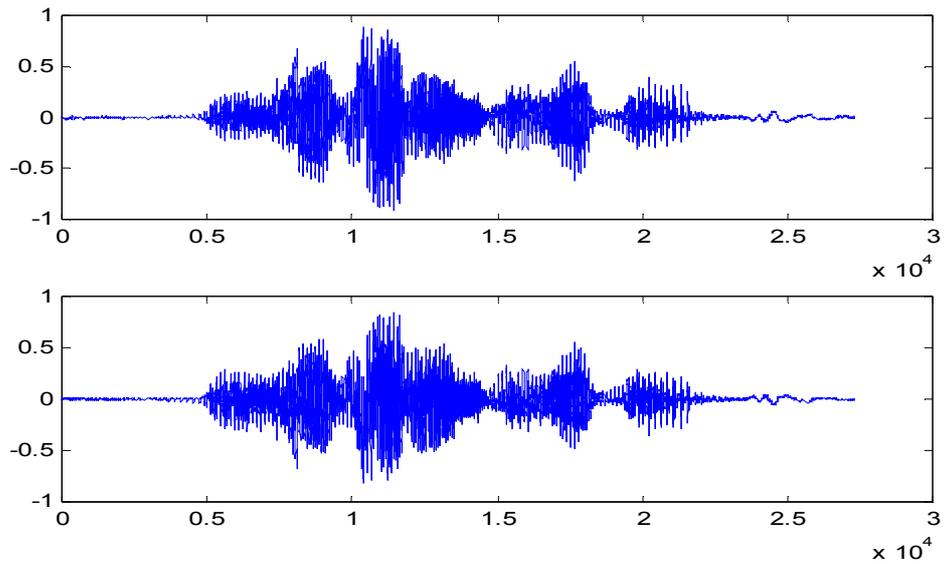


Figura 2.4 Señales observadas

SEÑALES ESTIMADAS

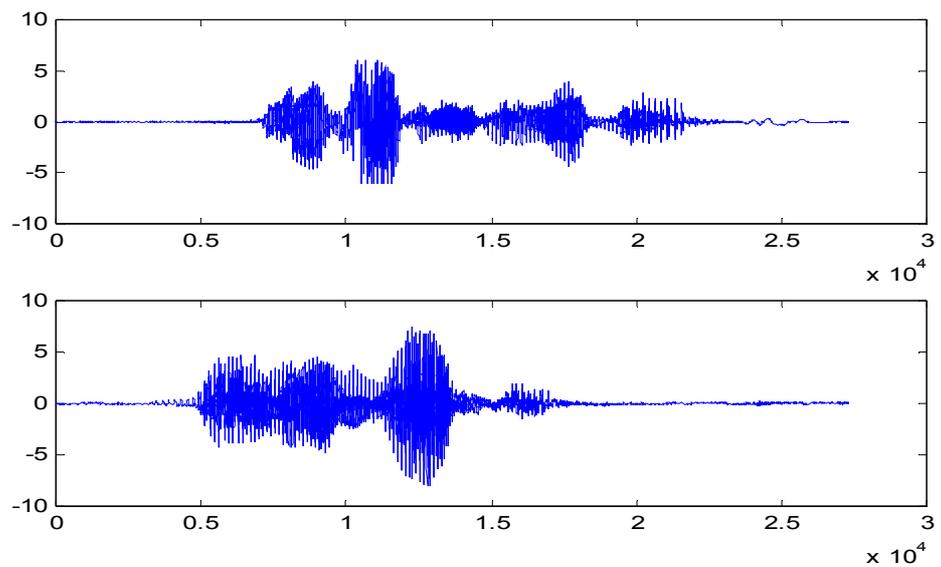


Figura 2.5 Señales estimadas

Podemos observar como las señales estimadas se asemejan a las fuentes originales salvo un factor de escala.

2.4 El Análisis de Componentes Principales

2.4.1 Introducción

El *Análisis de Componentes Principales* (*Principal Component Analysis, PCA*) es un método clásico de análisis de datos que tiene sus principales aplicaciones en el campo de la extracción de características y compresión de datos. Básicamente equivale

a la *transformada de Karhunen-Loeve* y está estrechamente relacionada con el *Análisis de Factores (Factor Analysis, FA)*. Ambas técnicas están basadas en estadísticos de segundo orden [Wong].

El *Análisis de Componentes Principales* tiene gran utilidad en aplicaciones de reconocimiento facial y de objetos como veremos sobre todo en la parte de este proyecto dedicada a la *Factorización No Negativa de Matrices*.

2.4.2 Blanqueado

Un vector de observaciones $\mathbf{z} = (z_1 \dots z_n)^T$ se dice que está *blanqueado* si los elementos z_i están **decorrelados** y tienen **varianza unidad**, es decir:

$$E\{z_i z_j\} = \delta_{ij} \quad (2.10)$$

que en términos de la matriz de covarianzas, implica que $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$, siendo \mathbf{I} la matriz identidad. La forma más clara de verlo es a partir del *ruido blanco*. Se dice que es *blanco* ya que todas las componentes espectrales son constantes para todo el dominio de la frecuencia [Hyvärinen01].

Además, dado que el blanqueado consiste básicamente en la decorrelación seguida del escalado, es posible usar *PCA*. El problema del blanqueado será ahora el de obtener una transformación \mathbf{V} tal que aplicada a las observaciones \mathbf{x} de lugar a los datos blanqueados, que denotaremos por \mathbf{z} , esto es

$$\mathbf{z} = \mathbf{V}\mathbf{x} \quad (2.11)$$

Una forma de verlo, consistiría en hacer uso de la conocida como *Descomposición en Valores Singulares (Singular Value Decomposition, SVD)*. Llamando $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^T\}$, a la matriz de covarianzas de \mathbf{x} , su descomposición en valores singulares dará lugar a a:

$$\mathbf{R}_x = \mathbf{U}\mathbf{D}\mathbf{U}^T \quad (2.12)$$

donde \mathbf{U} es la matriz de *autovectores*, \mathbf{D} la matriz diagonal cuyos elementos de la diagonal se corresponden con los *autovalores* de \mathbf{R}_x (en orden decreciente) y \mathbf{x} se supone de media cero [Jenssen00].

Considerando en este caso \mathbf{x} como una matriz real y simétrica, el conjunto de autovectores de la matriz de correlación \mathbf{R}_x forma un conjunto ortogonal, de forma que se cumple $\mathbf{U}\mathbf{U}^T = \mathbf{I}$, ya que \mathbf{U} resulta ser una matriz ortogonal para esas características de \mathbf{x} . De este modo, con la siguiente transformación las componentes de \mathbf{z} resultan estar decorreladas, dado que pueden ser vistas como una proyección de \mathbf{x} en el espacio PCA,

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} \quad (2.13)$$

y por consiguiente podremos calcular la matriz de covarianzas de \mathbf{z} de la siguiente manera:

$$\begin{aligned}\mathbf{R}_z &= E\{\mathbf{z}\mathbf{z}^T\} = E\{\mathbf{U}\mathbf{x}(\mathbf{U}\mathbf{x})^T\} = E\{\mathbf{U}\mathbf{x}\mathbf{x}^T\mathbf{U}^T\} = \\ &= \mathbf{U}\mathbf{R}_x\mathbf{U}^T = \mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T\mathbf{E}\mathbf{D}\mathbf{E}^T\mathbf{E}\mathbf{D}^{-\frac{1}{2}} = \mathbf{I}\end{aligned}\quad (2.14)$$

Por lo que queda demostrado que los nuevos datos están blanqueados ya que su matriz de covarianzas es la matriz identidad. En definitiva, para blanquear los datos existentes, tan sólo habrá que realizar la siguiente transformación:

$$\mathbf{z} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^T\mathbf{x}\quad (2.15)$$

ya que $\mathbf{z} = \mathbf{V}\mathbf{x}$, con $\mathbf{V} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^T$.

A continuación vamos a estudiar una matriz que cobra especial relevancia y es la que se conoce como *Matriz de Blanqueado*. Dicha matriz tiene la forma $\mathbf{E}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T$ y se puede verificar que es la raíz cuadrada inversa de la matriz de covarianzas de los datos \mathbf{R}_x , es decir:

$$\mathbf{E}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T = \mathbf{R}_x^{-\frac{1}{2}}\quad (2.16)$$

Por último, cabe destacar el hecho de que si las bases *PCA* vienen dadas por los autovectores de \mathbf{R}_x en orden decreciente de sus correspondientes autovalores, podemos descartar los autovectores correspondientes a los autovalores más pequeños de cara a obtener una dimensión menor, perdiendo la menor información posible en el sentido del error cuadrático medio.

Ejemplo

Para verlo de forma numérica, consideremos el siguiente ejemplo en el que tenemos dos muestras de voz, que denotaremos por x_1 y x_2 y que proceden de la mezcla a su vez de dos señales de voz s_1 y s_2 a partir de la matriz \mathbf{A} :

$$\mathbf{A} = \begin{pmatrix} 0.8 & -0.3 \\ 0.5 & 0.4 \end{pmatrix}\quad (2.17)$$

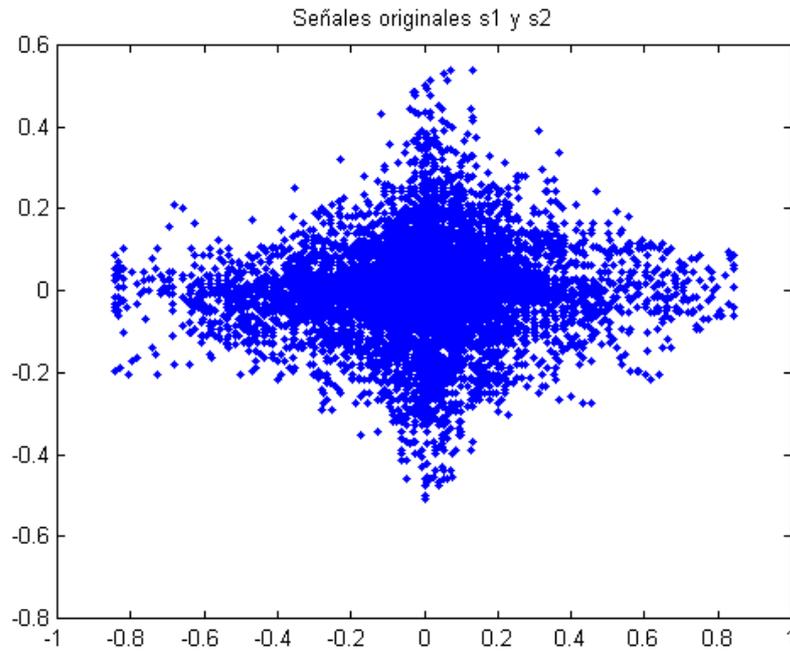


Figura 2.6 Señales originales s_1 (eje horizontal) y s_2 (eje vertical).

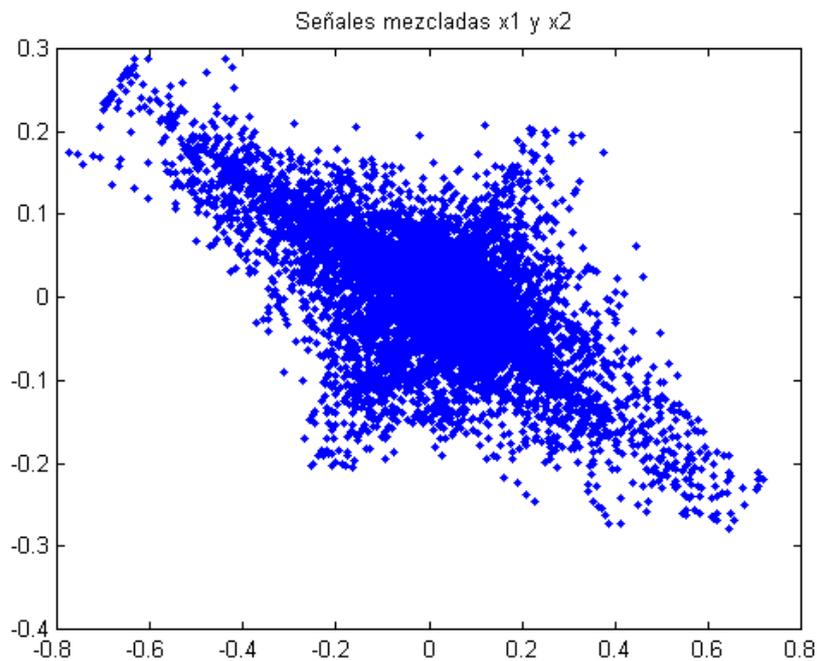


Figura 2.7 Señales resultantes de la mezcla mediante la matriz A , que denotaremos por x_1 (eje horizontal) y x_2 (eje vertical).

Si ahora aplicamos el algoritmo *PCA* a la matriz formada por los dos vectores x_1 y x_2 , obtendremos los siguientes valores para las matrices \mathbf{U} y \mathbf{D} :

$$\mathbf{U} = \begin{pmatrix} -0.9605 & -0.2782 \\ 0.2782 & -0.9605 \end{pmatrix} \quad (2.18)$$

$$\mathbf{D} = \begin{pmatrix} 0.0148 & 0 \\ 0 & 0.0013 \end{pmatrix}$$

por tanto, la matriz \mathbf{W} resultará:

$$\mathbf{W} = \mathbf{D}^{-\frac{1}{2}} \mathbf{U}^T = \begin{pmatrix} 0.9611 & -0.2763 \\ 0.2763 & 0.9611 \end{pmatrix} \quad (2.19)$$

y por último, podremos obtener las nuevas observaciones blanqueadas, que se calcularán como $\mathbf{y} = \mathbf{W}\mathbf{x}$:

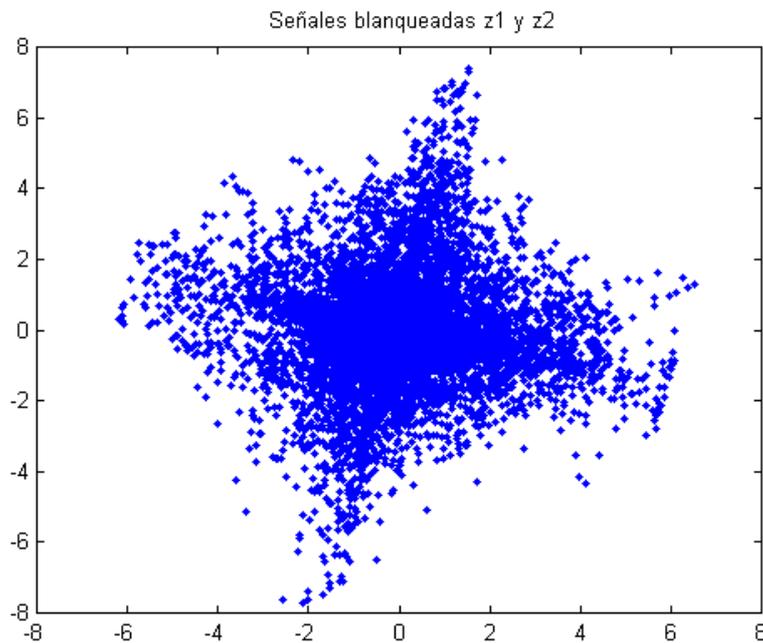


Figura 2.8 Señales obtenidas tras blanquear, z_1 (eje horizontal) y z_2 (eje vertical).

2.4.3 El Análisis de Factores y PCA

El *Análisis de Factores (Factor Analysis, FA)*, hace referencia a una técnica estadística de análisis de datos muy extendida. Su objetivo consiste en establecer las correlaciones existentes entre un conjunto de variables observadas en términos de un número más pequeño de *factores*. El *Análisis de Factores* se desarrolló originariamente en el ámbito de las ciencias sociales y psicología, donde su uso principal radicaba en el desarrollo de técnicas de medida de aspectos como la personalidad o la inteligencia [Baek].

Aunque *PCA* y *FA* persiguen un mismo objetivo consistente en expresar un conjunto de datos en función de una combinación lineal de un conjunto de datos de dimensión menor (*factores* en el caso de *FA* y *componentes* en el caso de *PCA*), veremos como existen una diferencia esencial entre ambas técnicas. En *FA* estudia tan

sólo la varianza compartida entre todos los datos, mientras que *PCA* analiza todas las varianzas existentes en ese conjunto de datos.

Por este motivo, en la mayoría de las aplicaciones *PCA*, presenta mejores prestaciones que *FA* en cuanto a la reducción de dimensiones de los datos estudiados [Dilip].

2.5 El Análisis de Componentes Independientes

2.5.1 Definición

Hemos visto como el problema de la ‘separación ciega de fuentes’ se basa en la obtención de una representación lineal en donde las componentes sean estadísticamente independientes. En la práctica, no es siempre posible encontrar una representación donde las componentes sean estadísticamente independientes, sin embargo si será viable la obtención de componentes que al menos sean lo más independientes que se pueda.

Esto nos lleva necesariamente a la definición del *Análisis de Componentes Independientes*, también conocido como *ICA* (Independent Component Analysis), que se puede hacer de la siguiente manera:

“Dado un conjunto de observaciones de variables aleatorias $\{x_1(t), x_2(t) \dots x_n(t)\}$, siendo t el tiempo o el índice de las muestras, asumimos que están generadas por una combinación lineal de componentes independientes:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \mathbf{A} \cdot \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{pmatrix} \quad (2.20)$$

o en forma matricial:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (2.21)$$

donde \mathbf{A} es una matriz de mezcla desconocida. El Análisis de Componentes Independientes consistirá ahora en estimar tanto la matriz \mathbf{A} como las fuentes $s_i(t)$ a partir de las observaciones $x_i(t)$ ” [Hyvärinen01]. Supondremos que el número de observaciones coincide con el de las fuentes originales, si bien esta simplificación no es completamente necesaria para resolver el problema.

De forma alternativa, podríamos definir *ICA*, como el problema de la obtención de la transformación lineal dada por la matriz \mathbf{W} como en (2.3), tal que las variables aleatorias estimadas $y_i(t)$ $i = 1, \dots, n$ sean tan independientes como sea posible. Este planteamiento no difiere en exceso del original ya que una vez obtenida la matriz \mathbf{A} , la matriz \mathbf{W} se obtiene invirtiéndola.

Se puede demostrar que el problema está completamente definido, es decir, el modelo planteado en (2.10) puede ser estimado si y sólo si las componentes $s_i(t)$ son *no*

gaussianas. Esta condición es esencial y servirá para explicar la diferencia principal que existe entre *ICA* y *PCA*, donde la no gaussianidad no se tiene en cuenta.

2.5.1.1 Definición de ICA basado en un modelo de variables ocultas

Para definir *ICA* de una forma rigurosa, es posible usar un modelo de variables ocultas [Hyvärinen01]. Se trata de observar n variables aleatorias x_1, \dots, x_n , que se modelan como una combinación lineal de las fuentes s_1, \dots, s_m :

$$x_i = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{in} \cdot s_n \quad \forall i = 1, \dots, n \quad (2.22)$$

donde los a_{ij} son coeficientes reales. Por definición los s_i son independientes entre si.

Este es el modelo básico ICA, que describe como las variables observadas son generadas por un proceso de mezcla de las fuentes s_j . Las componentes s_j son variables ocultas ya que no se pueden observar de forma directa. Además los a_{ij} pertenecientes a la matriz de mezcla se suponen también desconocidos. Las únicas variables que están ‘visibles’ serán las x_{ij} a partir de las cuales tendremos que estimar las fuentes s_j y la matriz de mezcla \mathbf{A} . Este problema se tendrá que resolver de la manera más general posible.

ICA guarda una estrecha relación con la ‘Separación Ciega de Fuentes’ (*BSS*). Una ‘fuente’ hace referencia en este contexto a una señal original, por ejemplo una persona hablando en el conocido como ‘cocktail-party effect’. El hecho de que sea ‘ciega’, significa que conocemos muy poco (o nada) de la matriz de mezcla, a la vez que se plantean hipótesis muy débiles acerca de las señales ‘fuente’ originales s_j . *ICA* es quizás el método más usado en la actualidad para resolver el problema de la separación ciega de fuentes.

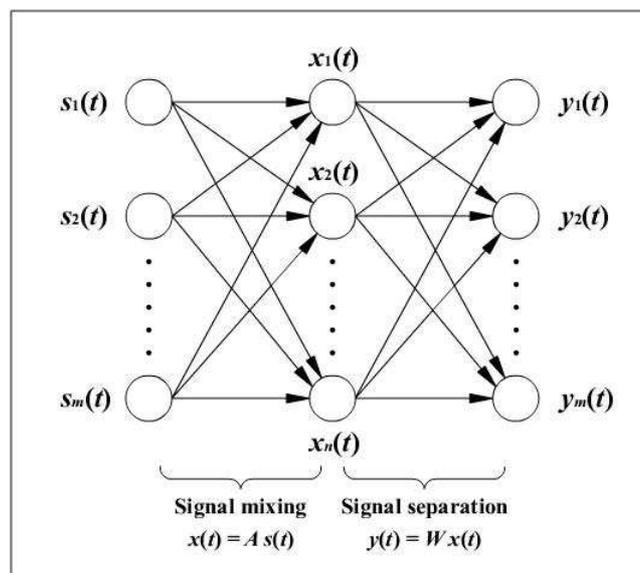


Figura 2.9 Cocktail-Party Effect. Las observaciones $x_i(t)$ con $i = 1, \dots, n$ proceden de la mezcla de las $s_j(t)$ con $j = 1, \dots, m$ a partir de la matriz \mathbf{A} . Las señales estimadas $y_j(t)$ con $j = 1, \dots, m$ se obtendrán a partir de la matriz de separación \mathbf{W} . Para que el problema

tenga solución se ha de verificar que el número de observaciones sea mayor o igual que el de señales a estimar, esto es, $n \geq m$.

2.5.2 Restricciones en ICA

Para poder asegurar que el modelo ICA que estamos planteando tiene solución, es necesario hacer una serie de suposiciones y restricciones que enumeraremos a continuación [Hyvärinen01].

1. *El número de observaciones ha de ser mayor o igual al número de componentes independientes a estimar.* En el caso de que sea mayor, es posible realizar una reducción de la dimensión usando el *Análisis de Componentes Principales (PCA)*, como se indica anteriormente en el apartado 2.2.2.
2. *Las componentes independientes se supondrán estadísticamente independientes.* En este principio descansa la base de ICA. Sorprendentemente no hace falta mucho más aparte de esta restricción para asegurar que el modelo va a tener solución. Es por esto por lo que ICA es un modelo tan potente y con tantas aplicaciones en diferentes campos.

Básicamente, las variables aleatorias y_1, y_2, \dots, y_n se dice que son independientes si la información en y_i no aporta ninguna información sobre el valor de y_j con $i \neq j$. Técnicamente, la independencia se puede definir a partir de las densidades de probabilidad. De esta forma, denotemos por $p(y_1, y_2, \dots, y_n)$ a la función densidad de probabilidad conjunta de y_i , y por $p(y_i)$ a la función densidad de probabilidad marginal de y_i . Entonces consideraremos que las y_i son independientes si y sólo si la función densidad de probabilidad conjunta se puede factorizar de la siguiente manera:

$$p(y_1, y_2, \dots, y_n) = p_1(y_1) \cdot p_2(y_2) \cdot \dots \cdot p_n(y_n) \quad (2.23)$$

3. *Las componentes independientes deben presentar distribuciones no gaussianas.* De forma intuitiva, se puede pensar que las distribuciones gaussianas son ‘muy simples’. Los cumulantes de orden superior son cero para distribuciones gaussianas, sin embargo dichos cumulantes son básicos en el análisis ICA. Por tanto, podemos deducir que el análisis ICA es imposible para distribuciones que sean gaussianas.

Por otro lado, cabe recordar que una combinación lineal de componentes gaussianas es a su vez gaussiana y además que para las estas distribuciones se cumple que la decorrelación entre dichas componentes va a implicar también independencia (propiedad que no se da en general para cualquier variable aleatoria en este sentido sino en el contrario, es decir, la independencia va a implicar existencia de decorrelación). Esto va a motivar que dada una combinación de componentes independientes gaussianas, sea imposible separarlas por los métodos que estamos describiendo.

4. *Consideraremos que la matriz de mezcla A es cuadrada.* Esto implicará que el número de componentes independientes a estimar es igual al número de

observaciones. A partir de esta suposición resulta fácil ver como una vez estimada la matriz de mezcla, al realizarle la inversa, podremos obtener la matriz **B**, que va a permitir estimar las fuentes originales, esto es:

$$\mathbf{s} = \mathbf{B}\mathbf{x} \quad (2.24)$$

Lógicamente asumiremos que la matriz de mezcla es *invertible* (si no lo fuera sería por la existencia de mezclas redundantes y llegaríamos a una situación en la que el número de componentes independientes y el de las observaciones es distinto y habría que recurrir a técnicas más complejas para llegar a una solución viable).

2.5.3 Ambigüedades del análisis ICA

En el modelo *ICA* dado por $\mathbf{x} = \mathbf{A}\mathbf{s}$, vemos una serie de ambigüedades que es necesario tener en cuenta antes de seguir desarrollando el modelo. Estas indeterminaciones se especifican en [Hyvärinen01] de la siguiente forma:

1. *No se puede determinar el orden de las componentes independientes.* Esto es debido a que tanto **s** como **A** son desconocidas a priori, de forma que si planteamos el modelo *ICA* como:

$$\mathbf{x} = \sum_{i=1}^n a_i s_i \quad (2.25)$$

siendo a_i cada una de las columnas de **A** y s_i las fuentes, entonces en todo momento sería posible cambiar el orden de los términos en el sumatorio y colocar cualquiera de las componentes independientes en primer lugar. De manera más rigurosa podemos decir que se podría insertar una *matriz de permutación P* y su inversa de forma que el modelo resultaría así:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{P}^{-1} \cdot \mathbf{P} \cdot \mathbf{s} \quad (2.26)$$

De esta manera, la matriz $\mathbf{A} \cdot \mathbf{P}^{-1}$ resultaría ser la nueva matriz de mezcla a estimar en el análisis *ICA*.

2. *No es posible determinar las energías de las componentes independientes.* Esto es debido de nuevo a que al ser **A** y **s** desconocidas, cualquier escalar λ_i que multiplicara a una fuente s_i se podría cancelar dividiendo la correspondiente columna a_i de **A**, esto es:

$$\mathbf{x} = \sum_{i=1}^n \left(\frac{1}{\lambda_i} a_i \right) (s_i \lambda_i) \quad (2.27)$$

La forma más habitual de proceder será suponer que las componentes independientes a estimar van a tener varianza unidad, es decir, $E[s_i^2] = 1$. De esta forma podremos hacer que la matriz **A** se calcule teniendo en cuenta esta restricción en nuestra solución *ICA*. También hay que destacar que sigue

habiendo problemas con el signo ya que por lo visto anteriormente podríamos multiplicar una fuente por '-1' sin que ello afectara al modelo. Sin embargo este caso es poco relevante en la mayoría de las aplicaciones.

2.5.4 Búsqueda de las componentes independientes

A priori puede resultar sorprendente que las componentes independientes puedan ser estimadas a partir de las combinaciones lineales, sin haber impuesto ninguna restricción además de su independencia. Para poder explicar este hecho vamos a ver una serie de aspectos que quizás puedan aclarar esta cuestión.

1. La decorrelación no es suficiente

Lo primero que tenemos que ver es que la independencia es una propiedad mucho más fuerte que la propia decorrelación. Considerando el problema de la separación ciega de fuentes, vemos que como es posible encontrar multitud de configuraciones de señales decorreladas que por el contrario no serían independientes y por tanto no extraerían las fuentes. Es decir, *la decorrelación de por sí, no garantiza una correcta separación de las componentes*. Este es el motivo principal por el que *PCA* no es válido para separar las fuentes ya que exclusivamente permite obtener componentes que son decorreladas, pero poco más.

2. La decorrelación no lineal es la base de ICA

Una forma de ver que la independencia es una condición más fuerte que la decorrelación consiste en establecer que la independencia implica *decorrelación no lineal*. Si s_1 y s_2 son independientes, entonces las transformaciones no lineales $g(s_1)$ y $h(s_2)$ son decorreladas (su covarianza es cero). Por el contrario, para dos variables aleatorias en general que son decorreladas (pero no independientes), dichas transformaciones no tienen covarianza cero en general.

De esta manera, podríamos intentar abordar nuestro problema ICA usando una forma más potente de decorrelación, es decir, buscando una cierta representación donde las y_i son decorreladas incluso después de aplicarle alguna transformación no lineal. Esto nos llevará a un primer principio de estimación de la matriz de separación \mathbf{W} :

- *Principio de decorrelación no lineal*: Consiste en encontrar la matriz \mathbf{W} tal que para todo $i \neq j$ las componentes y_i e y_j están decorreladas, y las correspondientes transformaciones $g(y_i)$ y $h(y_j)$ están decorreladas, siendo g y h funciones no lineales apropiadas.

Si escogemos las no linealidades de una forma apropiada para nuestro fin, ésta resulta una aproximación válida para que el método que buscamos halle las componentes independientes.

Sin embargo, llegados a este punto cabría preguntarse la forma de elegir las transformaciones no lineales g y h . Para ello debemos basarnos en principios de teoría de la estimación y teoría de la información, que proporcionarán métodos tales como el de la *máxima verosimilitud* e *información mutua* respectivamente.

3. Las componentes independientes son las componentes máximamente gaussianas

Otro principio que requiere de nuestra atención dentro del estudio del ‘Análisis de Componentes Independientes’, no es otro que el de la *estimación de la máxima no gaussianidad*. Este concepto se fundamenta en el conocido como ‘*Teorema Central del Límite*’, que establece que la suma de variables no gaussianas dará lugar a una nueva variable que será ‘más gaussiana’ que las originales.

Esto va a motivar que dada una combinación lineal de las variables observadas $y = \sum_i b_i x_i$, ésta será máximamente no gaussiana si es igual a alguna de las componentes independientes que conforman la mezcla. Esto es debido a que si se tratara de una mezcla real de dos o más componentes, estaría más cerca de parecerse a una distribución gaussiana en virtud del Teorema Central del Límite.

Esto nos llevará a formular un segundo principio de la siguiente manera:

- *Principio de máxima no gaussianidad*: Se trata de encontrar el máximo local de no gaussianidad de la combinación lineal $y = \sum_i b_i x_i$, sujeto a la condición de que la varianza de y ha de ser constante. Cada máximo local implicará que estamos antes una componente independiente.

La kurtosis y su importancia en la medida de la no gaussianidad

En la práctica, para medir la no gaussianidad haremos uso de la *kurtosis*. La kurtosis se define a partir de los momentos de tercer y cuarto orden, y para el caso en el que la media sea cero, tendrá la siguiente forma:

$$kurt(y) = E[y^4] - 3\{E[y^2]\}^2 \quad (2.28)$$

o de forma normalizada:

$$kurt(y) = \frac{E[y^4]}{\{E[y^2]\}^2} - 3 \quad (2.29)$$

A partir de esta definición resulta sencillo establecer medidas de la gaussianidad de una cierta variable aleatoria. Se puede demostrar que si y viene dada por una distribución gaussiana, entonces se verifica que su kurtosis es cero. Por eso, la kurtosis será una buena medida de la no gaussianidad.

De esta forma podremos hacer una clasificación de cualquier distribución, atendiendo al valor que toma su kurtosis y llegaremos a las siguientes conclusiones [Hyvärinen01]:

- Una distribución cuya kurtosis vale cero, se denomina *gaussiana*.

- Si la kurtosis es negativa, dicha variable se denomina *subgaussiana*. Las distribuciones subgaussianas son en general más ‘planas’ que las gaussianas.
- Se llaman variables *supergaussianas* a aquellas cuya kurtosis es positiva. Las distribuciones supergaussianas suelen tener ‘picos’ muy acusados y colas más largas que las gaussianas.

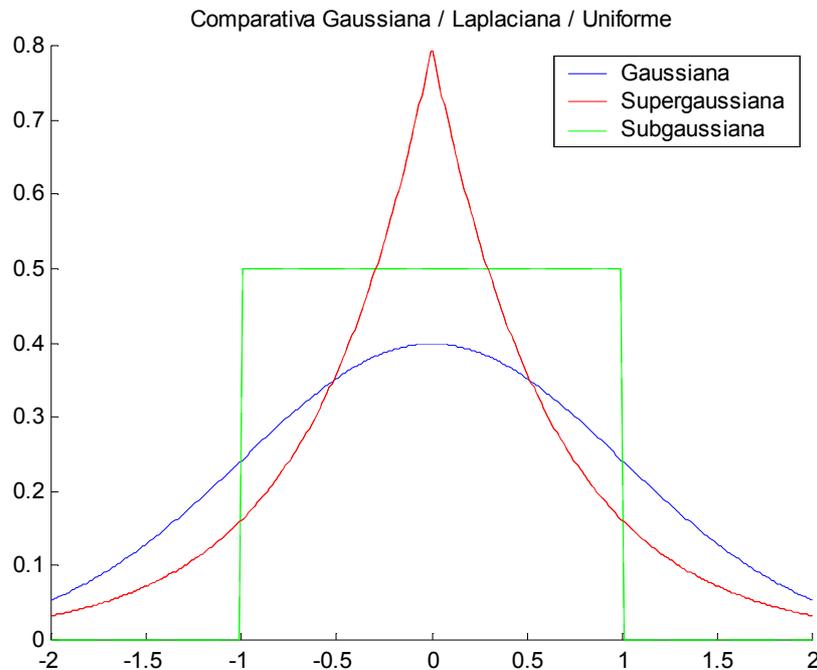


Figura 2.10 Comparativa entre distribuciones gaussiana (en azul), laplaciana (en rojo) y uniforme (en verde). Estas distribuciones representan ejemplos típicos de distribuciones supergaussianas ($kurtosis > 0$), gaussianas ($kurtosis = 0$) y subgaussianas ($kurtosis < 0$), en función del valor que tome la kurtosis.

Sin embargo es preciso tomar una serie de precauciones a la hora de tomar la kurtosis como medida de la no gaussianidad. El motivo es que la kurtosis de una variable supergaussiana puede tomar un valor excesivamente elevado, pero en el caso de las variables subgaussianas existe un valor mínimo que puede tomar, que será ‘-2’ (cuando la varianza está normalizada a la unidad). Esto implica que comparar la no gaussianidad de una variable supergaussiana y otra supergaussiana, no sería del todo correcto.

4. La importancia de los estadísticos de orden superior

En la práctica existen multitud de métodos para realizar la estimación *ICA*, si bien todos ellos se caracterizan por usar estadísticos que no aparecen en la matriz de covarianzas, son los llamados estadísticos de orden superior. Mediante la matriz de covarianzas podemos decorrelar de forma lineal pero no más allá. Es por eso por lo que

se antoja necesario el uso de dichos estadísticos, como la kurtosis y las correlaciones no lineales, aunque es posible emplear una amplia variedad para llevar a cabo el análisis *ICA*.

5. Los métodos numéricos

Debido a la enorme carga computacional que puede presentar el problema de estimación *ICA*, es necesario disponer de herramientas matemáticas potentes y algoritmos que permitan subsanar las necesidades de cálculo existentes. Sin embargo estos serán no lineales, por lo que tendremos que utilizar métodos numéricos para compensar las carencias que presenta el álgebra lineal, que no será válida.

Algunos algoritmos utilizados son el *Método del Gradiente*, o los algoritmos *FastICA* y *ThinICA*, desarrollados para explotar las características especiales que presenta *ICA*.

2.5.5 Teoría de la Información: una forma alternativa de estimar la transformación *ICA*

2.5.5.1 Entropía e Información Mutua

La *entropía diferencial* H de una variable aleatoria y con densidad $p(y)$ se define como [Hyvärinen01]:

$$H(y) = -\int p(y) \log p(y) dy \quad (2.30)$$

La entropía está estrechamente relacionada con la longitud del código de la variable aleatoria. Una versión normalizada de la entropía, viene dada por lo que se conoce como *entropía negativa* (o *negentropy*), que se define de la siguiente manera:

$$J(y) = H(y_{gauss}) - H(y) \quad (2.31)$$

siendo y_{gauss} una variable aleatoria gaussiana con la misma matriz de covarianzas que y . La entropía no negativa es siempre positiva y cero en el caso de las variables aleatorias gaussianas.

La *información mutua* I entre ‘ m ’ variables aleatorias y_i , $i = 1 \dots m$, se define así:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(y) \quad (2.32)$$

2.5.5.2 La información mutua como medida de la independencia

La información mutua entre variables aleatorias tiene en cuenta la dependencia completa que existe entre dichas variables y no sólo la covarianza como ocurre en el caso del *Análisis de Componentes Principales (PCA)*.

Por eso, es posible usar la información mutua de cara a obtener la representación *ICA*. Para ello definimos el modelo *ICA* de una variable aleatoria \mathbf{x} como una transformación invertible del tipo:

$$\mathbf{y} = \mathbf{B}\mathbf{x} \quad (2.33)$$

donde la matriz \mathbf{B} se calcula de forma que la información mutua de las componentes y_i sea mínima. Además se puede demostrar que la minimización de la información mutua va a llevar a maximizar la independencia de las componentes.

2.5.5.3 Entropía de una transformación

Consideremos una transformación *invertible* de una variable aleatoria \mathbf{x} que denotaremos por $\mathbf{y} = \mathbf{f}(\mathbf{x})$. Llamando $\mathcal{J}\mathbf{f}(\boldsymbol{\zeta})$ a la matriz jacobiana de la función \mathbf{f} , la relación entre las funciones densidad de probabilidad de \mathbf{y} y de \mathbf{x} , que llamaremos p_y y p_x respectivamente se define como:

$$p_y(\boldsymbol{\eta}) = p_x(\mathbf{f}^{-1}(\boldsymbol{\eta})) \left| \det \mathcal{J}\mathbf{f}(\mathbf{f}^{-1}(\boldsymbol{\eta})) \right|^{-1} \quad (2.34)$$

Expresando ahora la entropía a partir del valor esperado:

$$H(\mathbf{y}) = -E \left\{ \log p_y(\mathbf{y}) \right\} \quad (2.35)$$

e introduciendo el valor de p_y de la ecuación anterior, llegaremos a la ecuación que permite obtener la entropía de una transformación:

$$H(\mathbf{y}) = H(\mathbf{x}) + E \left\{ \log \left| \det \mathcal{J}\mathbf{f}(\mathbf{x}) \right| \right\} \quad (2.36)$$

2.5.5.4 Información mutua y no gaussianidad

A partir de la ecuación (2.36), vamos a construir el razonamiento que nos va a permitir llegar a un resultado para la información mutua. Sea una transformación lineal e invertible $\mathbf{y} = \mathbf{B}\mathbf{x}$, de forma que su información mutua se podrá escribir como:

$$I(y_1, y_2, \dots, y_n) = \sum_i H(y_i) - H(\mathbf{x}) - \log \left| \det \mathbf{B} \right| \quad (2.37)$$

Ahora veamos lo que ocurre si tomamos las y_i como incorreladas entre si y con varianza unidad. Esto significa que $E \left\{ \mathbf{y}\mathbf{y}^T \right\} = \mathbf{B}E \left\{ \mathbf{x}\mathbf{x}^T \right\} \mathbf{B}^T = \mathbf{I}$, lo cual implica que al tomar determinantes en ambos miembros de la expresión anterior:

$$\det(\mathbf{B}E \left\{ \mathbf{x}\mathbf{x}^T \right\} \mathbf{B}^T) = \det(\mathbf{B}) \det(E \left\{ \mathbf{x}\mathbf{x}^T \right\}) \det(\mathbf{B}^T) = \det(\mathbf{I}) = 1 \quad (2.38)$$

lo cual lleva a que $\det(\mathbf{B})$ ha de ser constante ya que $\det(E\{\mathbf{xx}^T\})$ no depende de \mathbf{B} . Además para las y_i , que tienen varianza unidad, la entropía y la entropía negativa difieren sólo en una constante y en el signo como vemos en la ecuación (2.31). De esta forma, llegaremos a que:

$$I(y_1, y_2, \dots, y_n) = \text{const.} - \sum_i J(y_i) \quad (2.39)$$

donde el término *constante* no depende de \mathbf{B} . Esta ecuación muestra la relación existente entre la información mutua y la entropía no negativa.

Hemos visto en la ecuación (2.39) que encontrar una transformación lineal e invertible que minimice la información mutua es equivalente a encontrar las direcciones en las que entropía negativa se maximiza. Además se sabe que la entropía negativa es una medida de la no gaussianidad por lo que podemos afirmar que *la estimación de la transformación ICA mediante la minimización de la información mutua es equivalente a maximizar la suma de las no gaussianidades de las componentes independientes estimadas.*

2.5.6 Un ejemplo práctico sobre ICA

A continuación vamos a ver un ejemplo numérico sobre el modelo ICA desde el punto de vista estadístico. Consideremos dos componentes independientes que presentan las siguientes distribuciones uniformes:

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}}, & |s_i| \leq \sqrt{3} \\ 0 & \text{en otro caso} \end{cases} \quad (2.40)$$

donde se verifica que la media va a ser cero y la varianza unidad. La función densidad de probabilidad conjunta de s_1 y s_2 va a ser uniforme en un cuadrado. Esto se debe a que la función densidad de probabilidad conjunta de dos variables independientes es el producto de las marginales como vimos en la ecuación (2.13). La función densidad de probabilidad conjunta se muestra en la siguiente figura:

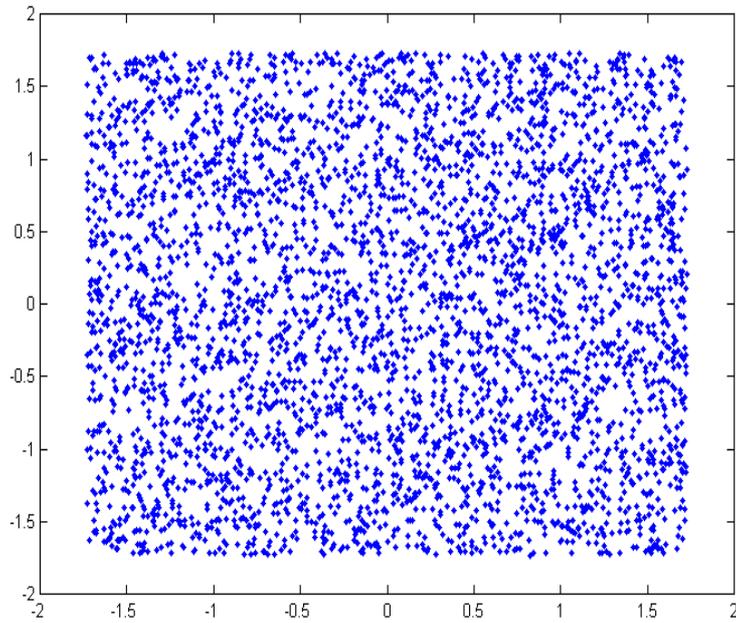


Figura 2.11 *Densidad de probabilidad conjunta de la distribución (2.40). El eje horizontal representa 's₁' mientras que el vertical representa 's₂'.*

A continuación, vamos a proceder a realizar una mezcla de las componentes independientes usando para ello la siguiente matriz de mezcla:

$$\mathbf{A}_0 = \begin{pmatrix} 3 & 5 \\ 9 & 6 \end{pmatrix} \quad (2.41)$$

y tras realizar las operaciones pertinentes, mostramos en la siguiente figura el resultado obtenido:

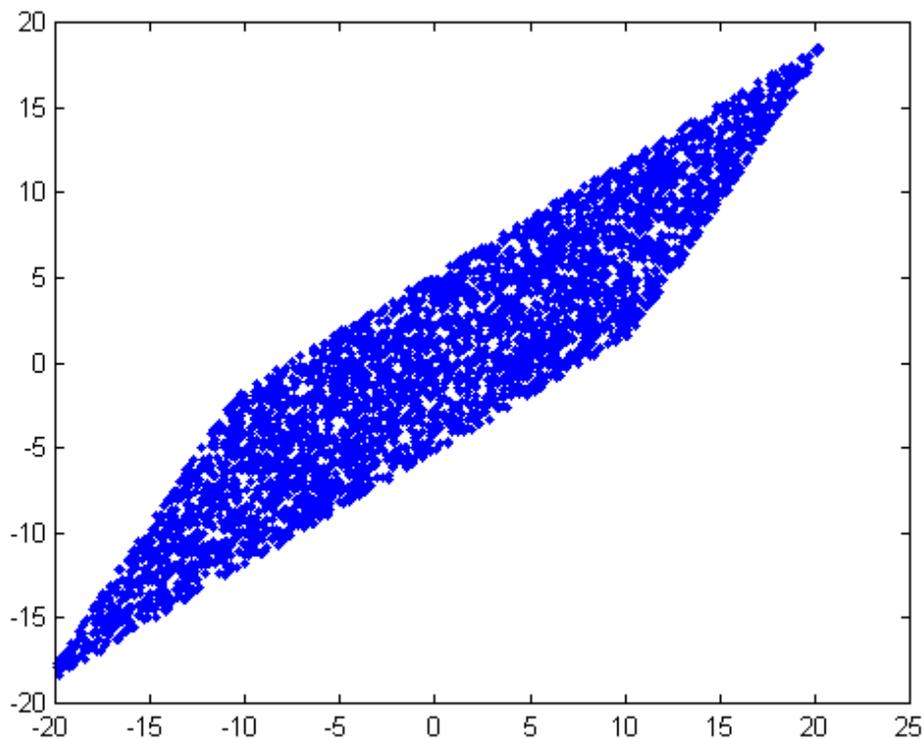


Figura 2.12 Densidad de probabilidad conjunta de la distribución (2.40) tras aplicarle la rotación dada por la matriz A_0 . El eje horizontal representa ' x_1 ' mientras que el vertical representa ' x_2 '.

Observando las figuras anteriores, resulta inmediato ver que las nuevas variables aleatorias generadas x_1 y x_2 ya *no son independientes*. Una manera sencilla de verlo, consiste en comprobar si es posible determinar el valor de una de ellas a partir de la otra. A partir de la *figura 2.9* es posible determinar que si la variable x_1 (representada en el eje horizontal) toma su valor máximo, esto es, en la esquina superior derecha de la distribución, entonces el valor de x_2 quedaría perfectamente determinado. Por el contrario, si se realiza el mismo experimento para la distribución de la *figura 2.8*, vemos como en este caso, no sería posible determinarlo.

El problema de la estima del modelo *ICA* va a consistir por tanto en la determinación de la matriz de mezclas A usando tan sólo la información contenida en las variables aleatorias x_1 y x_2 . De forma intuitiva, en el ejemplo que estamos desarrollando se podría estimar esta matriz A teniendo en cuenta que los ejes del paralelogramo de la *figura 2.9* están según las direcciones de las columnas de dicha matriz.

Esto significa que en principio, sería posible estimar el modelo *ICA* determinando en primer lugar la densidad de probabilidad conjunta de la distribución formada por x_1 y x_2 y posteriormente localizar los bordes que permitirían obtener la matriz A .

Por otro lado, podríamos considerar la mezcla de dos componentes independientes que no fueran uniformes (y por tanto subgaussianas), por ejemplo una

distribución supergaussiana como es el caso de la laplaciana. Para verlo, consideraremos la distribución laplaciana dada por la siguiente función densidad de probabilidad, donde tomaremos $\lambda = 1.6$:

$$p(s_i) = \frac{\lambda}{2} \exp(-\lambda |s_i|) \quad (2.42)$$

cuya representación en el plano formado por s_1 y s_2 será de esta forma:

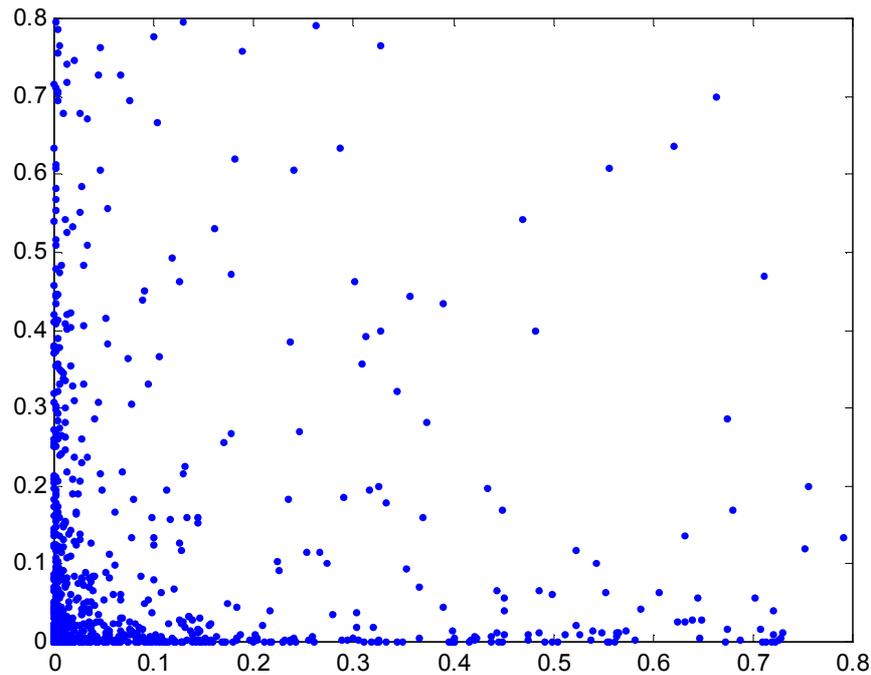


Figura 2.13 *Función densidad de probabilidad conjunta de las componentes s_1 y s_2 representadas por distribuciones laplacianas (supergaussianas). En el eje horizontal está s_1 y el vertical s_2 .*

Si procedemos de igual forma que para la distribución uniforme, realizaremos una rotación usando para ello la misma matriz \mathbf{A} dada por la ecuación (2.41) y obtendremos la siguiente figura:

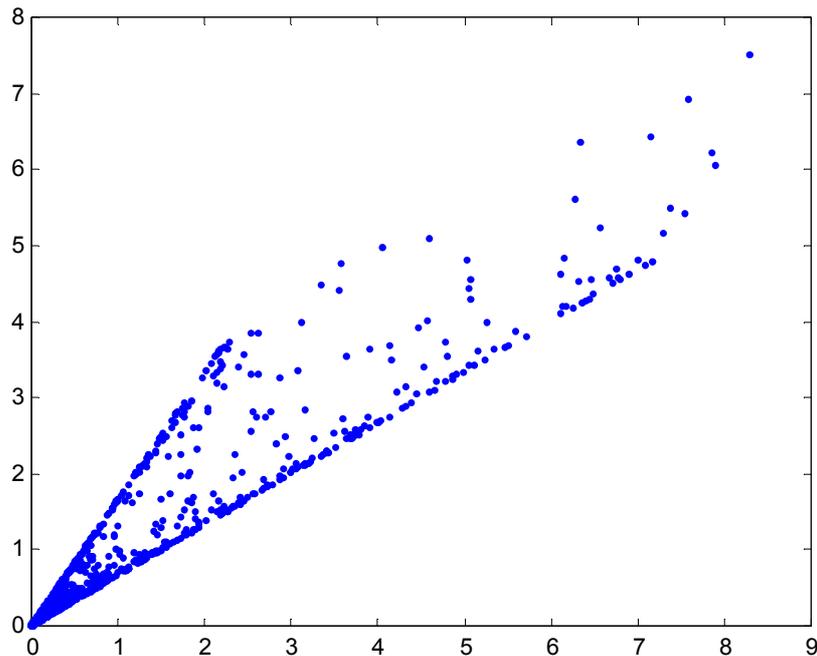


Figura 2.14 Rotación de la distribución de la figura 2.10 usando la matriz \mathbf{A}_0 . El eje horizontal representa ' x_1 ' mientras que el vertical representa ' x_2 '.

En este ejemplo vemos como ya resulta más difícil localizar los bordes que permitirían discernir el valor de los coeficientes que formarían la matriz \mathbf{A}_0 . En la práctica veremos como usar este método puede resultar computacionalmente complicado e inviable. Lo que realmente necesitaremos será un método que funcione para cualquier distribución de componentes independientes y que lo haga de forma rápida.

2.6 Conclusiones

En este primer capítulo dedicado a estudiar los fundamentos del *Análisis de Componentes Independientes* hemos visto la importancia que tiene el problema de la *Separación Ciega de Fuentes* de cara a comprender el funcionamiento de *ICA*.

Luego comprobamos como *PCA* es una herramienta estadística de vital importancia de cara a la reducción de las dimensiones del problema y a la extracción de características de los datos. En este mismo apartado, vimos el funcionamiento del proceso de *blanqueado*, que será de gran utilidad como preprocesado antes de realizar el análisis *ICA* y que permitirá que las observaciones resulten incorreladas y con varianza unidad.

Por último realizamos un estudio inicial del *Análisis de Componentes Independientes*, en el que se definió el modelo básico y se realizaron consideraciones referentes a las restricciones que se han de plantear y a las ambigüedades que presenta dicho modelo. Seguidamente vimos que la decorrelación no es suficiente para poder realizar la separación de las componentes independientes, por lo que se antoja necesario recurrir a criterios de búsqueda basados en la *no gaussianidad*. Mediante el estudio de

los estadísticos de orden superior (*kurtosis*) llegamos a un criterio que permitía establecer una clasificación de las variables aleatorias según su gaussianidad.

Por último destacamos como era posible estimar la matriz *ICA* mediante la conocida como ‘Teoría de la Información’, cuyo fundamento consiste en minimizar la *información mutua* entre las salidas de dicha matriz.

Con todo, pudimos extraer las siguientes conclusiones:

2.6.5.1 La decorrelación no va a garantizar el éxito en el proceso de estima de las componentes independientes.

2.6.5.2 Las componentes independientes son aquellas que son *máximamente gaussianas*.

2.6.5.3 Una forma alternativa de calcular la matriz de separación **W** se puede fundamentar en la utilización de técnicas basadas en la *Teoría de la Información*.