

Capítulo 5. Características fundamentales del Análisis de Componentes Independientes.

5.1 Introducción

El problema del Análisis de Componentes Independientes (ICA) consiste en la proyección de un conjunto de componentes en otro conjunto de fuentes de manera que resulten lo más estadísticamente independientes que sea posible [Murillo01]. Esta técnica se aplica normalmente a la Separación Ciega de Fuentes (BSS) cuando la independencia estadística de las salidas asegura su separación. Este tipo de procesado asume un modelo de entrada y salida múltiples y ha sido desarrollado con buenos resultados en diferentes campos de aplicación, tales como, comunicaciones (por ejemplo en CDMA, donde permite la recuperación de los usuarios sin información acerca del código de los mismos), estudio de las componentes de señales biomédicas de tipo electroencefalográfico y electrocardiograma fetal (EEG y ECG), procesado de imagen (campo en el que se incluye el tema objeto del proyecto de marcado de imágenes), etcétera...

5.2 Representación lineal de datos de dimensión múltiple. **Contexto estadístico**

Un problema clásico en estadística consiste en la obtención de una representación de los datos multidimensionales que nos permita observar su estructura esencial de forma sencilla. En procesado de señales podemos encontrar este mismo problema referente a la extracción de rasgos esenciales en el caso del Análisis de Componentes Independientes.

Si consideramos que los datos consisten en un número n de variables que se han observado de forma conjunta, obteniéndose m observaciones, podemos denotar los datos, o componentes observadas, por $x_j(t)$, donde $j=1,2,\dots,m$ y $t=1,2,\dots,n$.

La formulación general del problema puede encontrarse [Hyvärinen01], [Hernández] y consiste en encontrar una función que lleve a cabo la transformación del conjunto original de datos en un espacio n -dimensional a otro espacio m -dimensional, de forma que las componentes resultado de la

transformación a este nuevo espacio nos proporcionen tanta información como sea posible acerca de los rasgos principales de los datos, ocultos en la representación original de los mismos. Consideramos únicamente transformaciones lineales puesto que en este caso se simplifica la interpretación de la representación de los datos. A partir de lo indicado, cada componente resultado de la transformación, que denotamos por $u_i(t)$, se expresa como una combinación lineal de las observaciones, es decir:

$$u_i(t) = \sum_{i,j} w_{ij} \cdot x_j(t), i=1..n, j=1..m \quad (5.1)$$

Donde w_{ij} son los coeficientes que definen la representación.

Expresado en forma matricial sería:

$$\begin{pmatrix} u_1(t) \\ u_2(t) \\ \dots \\ u_n(t) \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \dots & \dots & \dots & \dots \\ w_{n1} & \dots & \dots & w_{nm} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \dots \\ x_m(t) \end{pmatrix} \rightarrow \mathbf{u} = \mathbf{W} \cdot \mathbf{x} \quad (5.2)$$

Una aproximación estadística consiste en tratar a $x_j(t)$ como un conjunto de T realizaciones de m variables aleatorias, por lo que cada muestra de la señal $x_j(t), t=1..T$, es, a su vez, una muestra de una variable aleatoria. En este contexto se podría tratar de determinar la matriz \mathbf{W} conforme a ciertas propiedades estadísticas de las componentes transformadas u_i .

5.3 Planteamiento del problema ICA

5.3.1 Definición

Encontrar una solución para el problema de Análisis de Componentes Independientes consiste en buscar una representación lineal a partir de la que las componentes estimadas resulten estadísticamente independientes entre ellas. En situaciones reales, una representación en la que las componentes sean totalmente independientes será imposible de conseguir, pero sí una que estime componentes lo más independientes posible. Lo que nos permite formular el problema de ICA en los siguientes términos [Zeman00], [Hernández03].

Dado un conjunto de n observaciones $(x_1(t), x_2(t), \dots, x_n(t))$, se asume la existencia de n componentes independientes $(s_1(t), s_2(t), \dots, s_n(t))$, cuya relación

viene determinada por la denominada matriz de mezclas \mathbf{A} según la siguiente expresión en forma matricial:

$$\mathbf{x}_t = \mathbf{A} \cdot \mathbf{s}_t \quad (5.3)$$

El Análisis de Componentes Independientes busca estimar las componentes originales \mathbf{s}_t de la mejor forma posible a partir del conocimiento único de las componentes observadas \mathbf{x}_t , partiendo, lógicamente, de ciertas premisas que iremos viendo, entre las que destaca el requisito de independencia de las fuentes a estimar para lograr su obtención a partir de las observaciones (o mezclas de las mismas).

Es decir, aplicar ICA consiste en proyectar \mathbf{x}_t en un espacio de n componentes \mathbf{y}_t lo más estadísticamente independientes que sea posible. Esta proyección es representada por una matriz \mathbf{B} de dimensiones $n \times n$.

$$\mathbf{y}_t = \mathbf{B} \cdot \mathbf{x}_t \quad (5.4)$$

Puede demostrarse que si \mathbf{A} es una matriz no singular, \mathbf{x}_t es estacionario ergódico y como mucho una de las entradas de \mathbf{s}_t presenta una distribución gaussiana, entonces el vector resultado de aplicar ICA a las observaciones \mathbf{x}_t es una estima de las fuentes originales. Estos requerimientos son fundamentales y suponen la principal diferencia entre los análisis ICA y PCA (Análisis de Componentes Principales), ya que en éste último la restricción de no gaussianidad de los datos no es considerada.

Esquemáticamente el problema puede representarse de la forma ilustrada en la figura 5.1 [Murillo01].

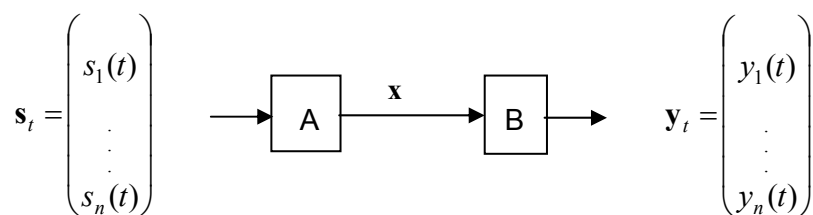


FIGURA 5.1. Proceso BSS.

Para lo que hemos elegido el caso más sencillo en que la matriz de mezcla es cuadrada. En un caso más general, si el número de observaciones es m y el de componentes independientes n , debe cumplirse que $m \geq n$.

5.3.2 **Ambigüedad del análisis ICA**

Hasta el momento se han indicado los requisitos básicos que deben cumplirse para que las estimaciones obtenidas a partir del Análisis de Componentes Independientes de un conjunto de vectores observados, sean similares a las fuentes originales. Pero, ¿es posible recuperar las señales originales exactamente?, pues bien, podría demostrarse que la matriz **B** obtenida representa la transformación inversa de **A** salvo, en general, un factor de escalado y una posible permutación de filas. Por lo que este análisis presenta ciertas ambigüedades:

- Imposibilidad de identificar la energía (varianza) de las fuentes, ya que un factor de escala cualquiera en la fuente i -ésima puede ser cancelado dividiendo la columna i -ésima de la matriz **A** por el mismo factor. Ambigüedad de poca importancia en la mayoría de las aplicaciones, por lo que normalmente se asume que $E[s_i^2]=1$. De tal forma que la componente que estimemos será igual a $\pm s_i$, ya que el modelo ICA no se ve afectado por el signo de las componentes.
- Tampoco es posible identificar la permutación acaecida en el orden de las fuentes, puesto que para cualquier permutación de las mismas, existe una permutación en las columnas de la matriz de mezcla que proporciona las mismas observaciones.

5.3.3 **Separabilidad e identificabilidad ciega. Unicidad**

Cuando se estudia la viabilidad de la solución del modelo ICA resulta imperativo analizar las restricciones que deben cumplir tanto las fuentes como la matriz de mezcla que aseguren la existencia de la misma y que ésta sea, a su vez, única, [Murillo01].

La matriz de mezcla debería ser regular para asegurar la separabilidad de las observaciones. Las condiciones de las fuentes para ser separables pueden obtenerse a partir los “principios de separabilidad” basados en el teorema de Darmois Skitovitch.

Teorema de Darmois Skitovitch. Sea $s = [s_1(t) \dots s_n(t)]$ (vector aleatorio n -dimensional $n > 2$) cuyas componentes son mutuamente independientes, y sean dos salidas:

$$\begin{aligned} y_i(t) &= b_{i,1}s_1(t) + b_{i,2}s_2(t) + \dots + b_{i,n}s_n(t) \\ y_j(t) &= b_{j,1}s_1(t) + b_{j,2}s_2(t) + \dots + b_{j,n}s_n(t) \end{aligned} \quad (5.5)$$

Suponiéndose ambas salidas estadísticamente independientes. Si para un índice k arbitrario $b_{k,i}$ y $b_{k,j}$ son distintos de cero, la fuente $s_k(t)$ presenta una función de distribución gaussiana. Si además se cumple el lema siguiente (propuesto por Cao).

Lema: Si en una matriz no singular \mathbf{U} de dimensiones $n \times n$ la primera fila es ortogonal al resto de las $n-1$ filas, entonces $b_{1,k} = 0$ para $k = 2, \dots, n$, si y sólo si $b_{k,1} = 0$ para $k = 2, \dots, n$.

A partir del teorema y lema propuestos se puede concluir lo siguiente:

“Si todas las fuentes menos una son no gaussianas, la propiedad de independencia dos a dos de las salidas garantiza la separación. Sin embargo, en presencia de más de una fuente gaussiana la propiedad de independencia dos a dos de las salidas solo garantiza la separación de las señales no gaussianas.”

Veamos una ilustración que muestra la razón por la que no valen las funciones de distribución de probabilidad gaussianas para las fuentes independientes. Sean dos fuentes con distribuciones gaussianas, como las que se muestran en la figura 5.2.

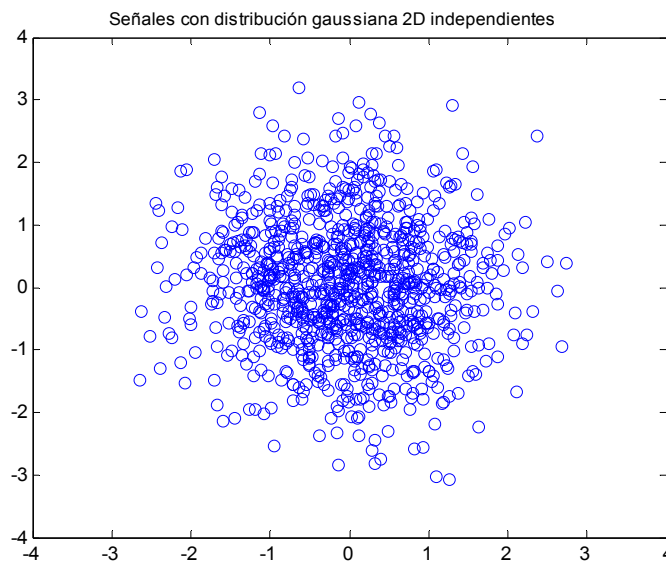


FIGURA 5.2 Componentes independientes con distribución gaussiana.

Si se mezclan según la matriz $\mathbf{A} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ se obtienen las observaciones ilustradas en la figura 5.3.

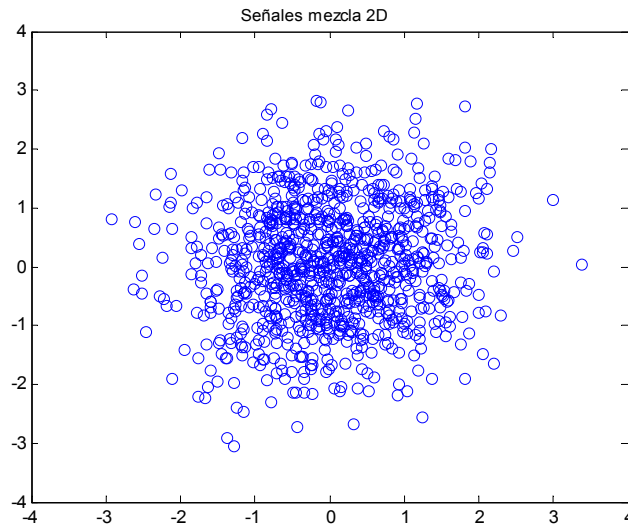


FIGURA 5.3 Mezclas incorreladas de la componentes.

En las figuras anteriores se observa que no se podría determinar la rotación sufrida por las componentes originales, de modo que la identificación de la matriz de mezcla o su inversa no resulta clara. Cualquier matriz de mezcla ortogonal rota la función de densidad de probabilidad dejándola igual, por lo que no se puede identificar, de manera que el resultado del análisis ICA no tendría efecto. De ahí la razón para exigir la no gaussianidad de las fuentes.

En conclusión, para conseguir la separación de fuentes las condiciones que deben cumplirse son:

- La matriz de mezcla debe ser regular.
- No más de una fuente gaussiana.

Pero, ¿es única la solución? Bajo la hipótesis de que como mucho una de las fuentes sea gaussiana unida a la condición de independencia estadística entre ellas queda asegurada la obtención de las fuentes a partir de sus mezclas, pero resulta imposible recuperar el orden original ni el escalado de las mismas, sin embargo, la propiedad de independencia de las salidas se preserva para cualquier escalado de estas y el orden de recuperación de las fuentes no influye en la independencia entre las mismas.

5.3.4 Información de segundo orden. Blanqueado y rotación

La información de segundo orden (decorrelación) permite expresar el espacio de mezclas en un sistema ortogonal (blanqueado), pero ello no implica la independencia estadística de estas proyecciones. La teoría de análisis de componentes principales (PCA) resuelve el blanqueado, pero deja indeterminada la rotación del mismo que asegura la independencia [Murillo01], [Zeman00].

Resulta sencillo eliminar la componente de continua de las mezclas observadas y, a partir de lo indicado acerca de que, en general, el escalado obtenido para las fuentes es distinto al original, podemos suponer que las fuentes tienen varianza unidad. Por todo lo dicho, resulta una matriz identidad como matriz de covarianza $E[s \cdot s^T] = 1$, y el vector s se considera espacialmente blanco.

Si W es una matriz de blanqueado para x , esto es, $z = Wx$ es espacialmente blanca, el resto de la transformación es, por fuerza, una rotación, puesto que ésta relaciona dos vectores blancos s y z . Por lo que blanquear los datos reduce la mezcla a una rotación, de tal forma que la matriz de separación se puede expresar como el producto de dos matrices, una de blanqueado y otra de rotación:

$$B = W \cdot U \quad (5.6)$$

Los vectores estimados serán:

$$y = B \cdot x = U \cdot W \cdot x = U \cdot z = U \cdot W \cdot A \cdot s \quad (5.7)$$

Un conjunto de mezclas de señales Gaussianas no puede separarse con sólo imponer independencia estadística, para ellas, los estadísticos de orden superior a dos son nulos. De manera que dicha mezcla podría blanquearse, pero la rotación quedaría indefinida, debido a la simetría circular de su función de densidad de probabilidad conjunta.

Los autovalores de la matriz de correlación pueden usarse para determinar la matriz de fuentes.

5.3.5 El Análisis de Componentes Principales

El planteamiento del problema PCA puede formularse en los términos siguientes.

Sea un vector aleatorio de observaciones de dimensión n $x_1(t), x_2(t), \dots, x_n(t)$ con media nula $E[x] = 0$ y función de autocorrelación $R_x = E[xx^T]$. El objetivo del análisis en componentes principales es encontrar una transformación lineal \mathbf{W} a un espacio de dimensión menor $m < n$, tal que la reconstrucción de la señal tenga un error cuadrático medio mínimo, es decir:

Proyección en un espacio de dimensión menor:

$$\mathbf{y}_{m \times 1} = \mathbf{W}_{m \times n} \cdot \mathbf{x}_{n \times 1} \quad (5.8)$$

Reconstrucción en el espacio original:

$$\hat{\mathbf{x}}_{n \times 1} = \mathbf{F}_{n \times m} \mathbf{y}_{m \times 1} \quad (5.9)$$

Objetivo: Error de reconstrucción $J_e = E[|\mathbf{x} - \hat{\mathbf{x}}|^2]$ mínimo (5.10)

El vector \mathbf{y} se denomina, a veces, “vector de características”: es el vector de dimensión m que mejor representa (de manera lineal) al vector original \mathbf{x} .

5.3.5.1 Solución

La matriz de autocorrelación puede descomponerse en autovalores y autovectores:

$$\mathbf{R}_x = \sum_{i=1}^m \lambda_i \mathbf{u}_i \mathbf{u}_i^T = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (5.11)$$

Siendo $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$ y $\mathbf{\Lambda}$ una matriz diagonal con los autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

La matriz óptima de proyección la forman los m autovectores correspondientes a los mayores autovalores, es decir, aquellas direcciones donde la varianza resulta máxima $\mathbf{W}_{m \times n} = [\mathbf{u}_1 \dots \mathbf{u}_m]^T$ y la matriz de reconstrucción es

$$\mathbf{F}_{m \times n} = [\mathbf{u}_1 \dots \mathbf{u}_m] \quad (5.12)$$

Es la transformada de Karhunen-Loeve, que es óptima desde un punto de vista de compactación de la energía. En definitiva el PCA se reduce a encontrar aquellos vectores en los que las proyecciones den máxima varianza.

Si quisiésemos caracterizar (codificar) de forma óptima el vector bidimensional original mediante un escalar deberíamos usar la proyección sobre el autovector correspondiente al autovalor máximo.

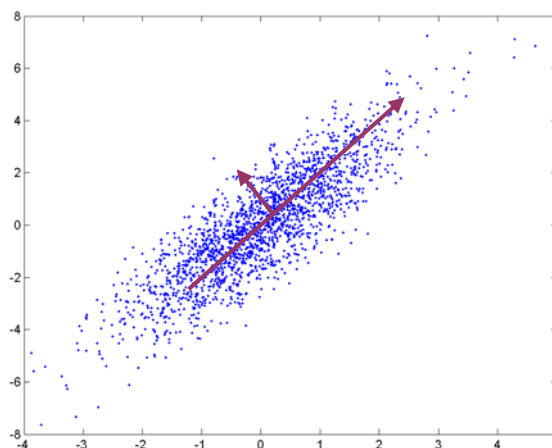


FIGURA 5.4. Dirección de autovectores del vector bidimensional representado.

En conclusión, el análisis en componentes principales (PCA) es una herramienta fundamental del tratamiento estadístico de señales que permite la extracción de las características más relevantes en los datos, proporcionando las direcciones de máxima varianza de los mismos. La dirección principal coincide con el autovector correspondiente al autovalor máximo de la matriz de autocorrelación.

5.3.6 Búsqueda de las componentes independientes

En esta sección se explicará de forma breve cómo es posible recuperar las señales fuente a partir de observar mezclas lineales de las mismas, con la única suposición de independencias estadística de las señales fuente.

5.3.6.1 Decorrelación y no gaussianidad

Es sabido que la independencia estadística es una restricción más severa que la decorrelación, por lo que ésta última no es suficiente. Si consideramos el problema de Análisis de Componentes Independientes podemos encontrar muchas representaciones incorreladas, pero no independientes, de manera que no representan a las señales originales. El Análisis de Componentes Principales, sin embargo, puede realizarse a partir de esa suposición más relajada.

Para ilustrar lo dicho supongamos que tenemos dos componentes independientes s_1 y s_2 con distribuciones uniformes en el intervalo $[-0.5, 0.5]$, tal como se representa en la figura 5.5.

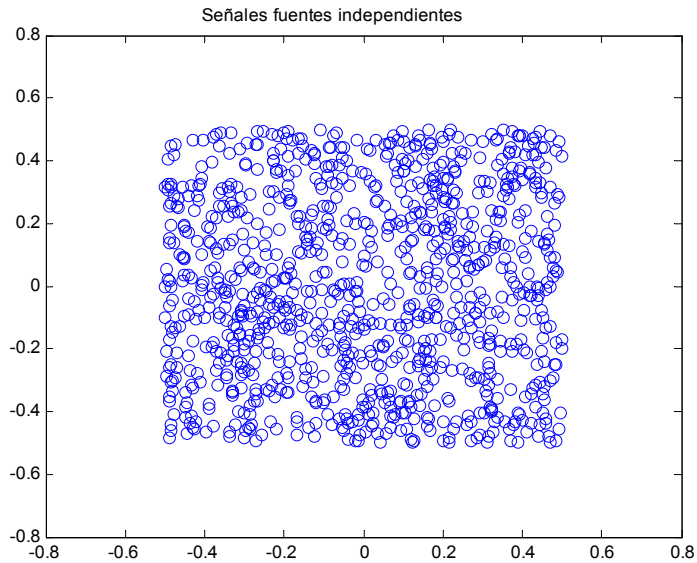


FIGURA 5.5. Señales fuente independientes con distribución uniforme en el intervalo $[-0.5, 0.5]$

Si se mezclan estas fuentes usando la matriz de mezcla \mathbf{A} siguiente se obtiene la figura 5.6:

$$\mathbf{A} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (5.13)$$

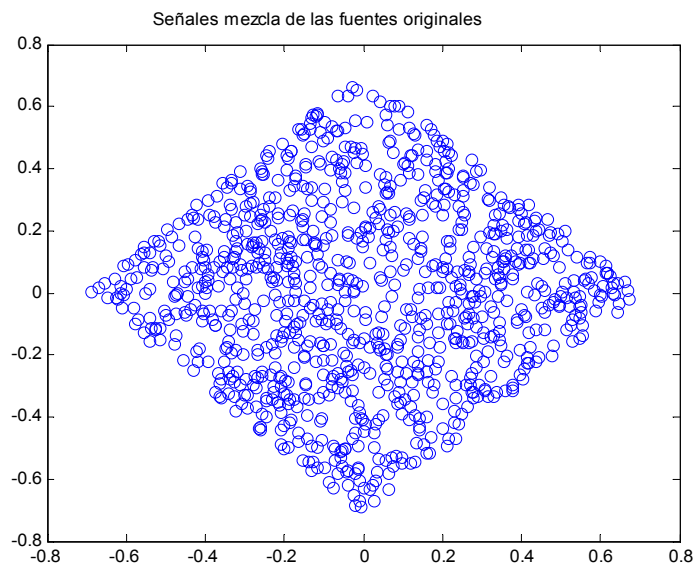


FIGURA 5.6. Mezclas incorreladas de las componentes independientes de la figura 5.5.

Por observación de la figura 5.6 se puede ver que para obtener las fuentes independientes a partir de sus mezclas solo es necesario realizar una

rotación de los ejes mostrados, lo que se consigue por medio de la transformación inversa \mathbf{A}^{-1} .

La independencia implica decorrelación no lineal, en el sentido de que, las componentes obtenidas tras aplicar cualquier transformación no lineal a las variables independientes deben estar incorreladas entre sí (en el sentido de que su covarianza sea nula). Si las componentes iniciales sólo estuvieran incorreladas, después de una transformación no lineal no tendrían covarianza nula, en general. Si se escogen las no linealidades de forma apropiada, ésta resulta una aproximación válida para que el método que busquemos determine las componentes independientes. La respuesta a la cuestión referente a qué no linealidades elegir puede hallarse en la “Teoría estadística” o bien en la “Teoría de la Información”, como más tarde se verá.

Otro principio importante consiste en considerar que las componentes independientes son las componentes máximamente no gaussianas. De acuerdo al Teorema Central del Límite sabemos que la suma de una serie de variables aleatorias no gaussianas presenta una distribución más cercana a una gaussiana que sus componentes de partida. Por tanto, si tomamos una combinación lineal de las componentes observadas, que a su vez son una combinación lineal de las fuentes originales independientes, será máximamente no gaussiana si corresponde a una componente independiente. Lo que nos lleva a considerar el principio de no Gaussianidad como el segundo principio de estimación de ICA.

“Se busca un máximo local de la no gaussianidad de una combinación lineal $u = \sum b_i x_i$ bajo la condición de que la varianza de u sea constante. Cada máximo local da una componente independiente”.

5.3.6.1.1 La estimación de las componentes independientes necesita estadísticos de orden superior

Para realizar el análisis ICA, partir de la información única proporcionada por la matriz de covarianzas (que proporciona el valor de la covarianza entre cada dos variables) es insuficiente, ya que con ella solo podemos conseguir la recorrelación de las componentes de la transformación en el sentido lineal. De este modo, el Análisis de Componentes Independientes requerirá la información de estadísticos de orden superior que expresen información no contenida en la matriz de covarianzas, como puede ser la kurtosis.

5.3.6.2 Medidas de la no-Gaussianidad y la independencia

En este apartado vamos a ver la kurtosis como una medida de la no-Gaussianidad y la Información mutua como una medida de la dependencia entre variables, [GTAS], [Hernández].

5.3.6.2.1 La kurtosis

La combinación lineal de variables aleatorias independientes y no Gaussianas tiende a “Gaussianizarse” de acuerdo al Teorema de Límite Central, como se ha indicado anteriormente. De manera que una posibilidad para encontrar la matriz de separación **B** puede ser maximizar una cantidad que mida la no Gaussianidad de los datos, como es la kurtosis.

La kurtosis de una variable aleatoria **s** de media nula se define como:

$$K[s] = E[s^4] - 3(E[s^2])^2 \quad (5.14)$$

Definida de esta forma, la kurtosis de una Gaussiana es cero. Además, variables aleatorias con “colas” más largas que la Gaussiana (como la laplaciana) tienen una kurtosis positiva y se las conoce como variables “super-gaussiana”, mientras que, aquellas que son más planas que la Gaussiana (como la distribución uniforme) tienen una kurtosis negativa por lo que se denominan “sub-Gaussianas”.

A partir de estos resultados puede verse que se pueden actualizar las columnas de la matriz de separación con el objetivo de maximizar (o minimizar) la kurtosis. Sin embargo la estimación de la kurtosis es muy sensible a espureos.

El algoritmo Fast-ICA es un ejemplo de procesado que usa la información de la kurtosis por defecto, aunque permite utilizar otras funciones no lineales, para realizar la estimación de las componentes independientes.

5.3.6.2.2 Información y entropía

Retomando algunos conceptos básicos de la teoría de la Información [Haykin97], dado un experimento que puede tomar un conjunto discreto de posibilidades, al observar el suceso $x = x_k$ que aparece con probabilidad p_k , la información obtenida se define de la siguiente manera:

$$I(x_k) = -\ln(p_k) \quad (5.15)$$

Cumpléndose que:

$$I(x_k) = 0 \text{ si } p_k = 1 \quad (5.16)$$

$$I(x_k) < I(x_j) \text{ si } p_k > p_j \quad (5.17)$$

$$I(x_k) \geq 0 \text{ para } 0 < p_k \leq 1 \quad (5.18)$$

La entropía de Shannon se define como el valor medio de la información que obtenemos, o como el grado de desinformación previo a la observación del suceso.

Para el caso discreto se puede formular de la siguiente manera:

$$H(X) = E[I(x_k)] = - \sum_k p_k \cdot \ln(p_k) \quad (5.19)$$

Siendo su expresión en el caso de que la variable sea continua:

$$H(X) = E[I(x)] = - \int_{-\infty}^{\infty} f(x) \cdot \ln(f(x)) dx \quad (5.20)$$

Para una varianza fija, la variable aleatoria cuya entropía es máxima es la gaussiana si el rango de la misma no está acotado y la uniforme si el rango está acotado.

Una medida de la información disponible acerca de una variable aleatoria S a partir de la observación de otra variable X puede representarse por la llamada entropía condicional:

$$H(S|X) = H(S, X) - H(X) \quad (5.21)$$

La información mutua entre ambas variables se puede calcular como:

$$I(S, X) = H(X) - H(X|S) \quad (5.22)$$

Donde $H(X)$ representa la entropía de X , interpretable como el desconocimiento asociado a X o la cantidad de información que puede aportar la observación de la variable X en media. La información mutua cumple lo siguiente:

$$I(S, X) \geq 0 \text{ si } I(S, X) = 0 \rightarrow f(s, x) = f(s) \cdot f(x) \text{ Independencia} \quad (5.23)$$

Donde $f(s, x)$, $f(s)$ y $f(x)$ son las funciones densidad de probabilidad conjunta y marginales respectivamente de las variables aleatorias S y X , caso de

que consideremos variables aleatorias continuas. La propiedad dada en (5.18) se debe a la propia definición de la información mutua (o divergencia de *Kullback-Leibler*) entre variables continuas:

$$I(S, X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s, x) \cdot \ln \left(\frac{f(s, x)}{f(s) \cdot f(x)} \right) ds dx \quad (5.24)$$

A partir de lo visto puede concluirse que un posible criterio para realizar ICA puede ser minimizar la Información mutua entre las salidas de la aplicación matriz de separación a los vectores observados. Se puede demostrar, además, que esta idea lleva también (de una forma más rigurosa) a buscar las direcciones de menos gaussianidad.

5.3.6.2.2.1 ICA mediante minimización de la Información mutua

Un esquema básico de la implementación de ICA puede ser el mostrado en la figura 5.7.

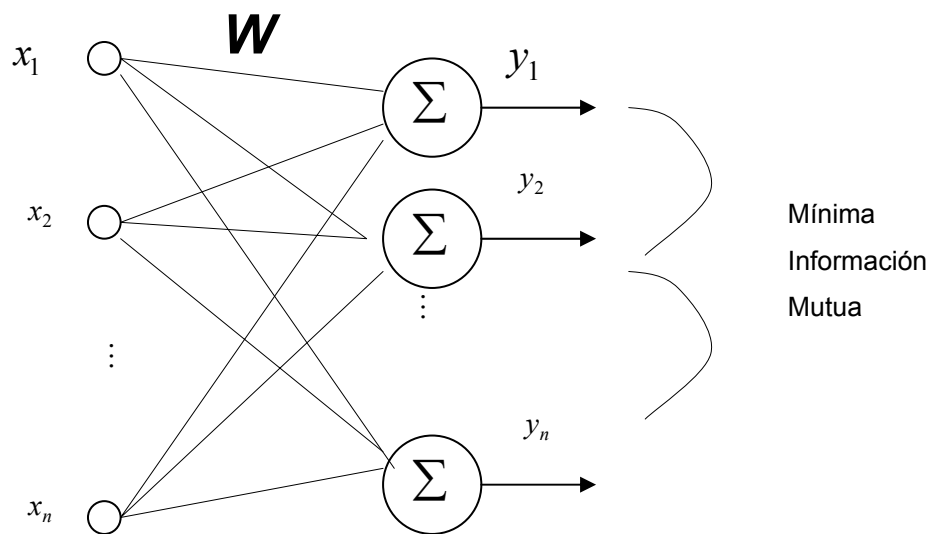


FIGURA 5.7. ICA basado en minimización de la Información Mutua

Donde:

$$I(y_1, y_2, \dots, y_n) = \sum_{i=1}^n H(y_i) - H(\mathbf{y}) = \sum_{i=1}^n H(y_i) - (H(\mathbf{x}) + \ln|\det \mathbf{W}|) \quad (5.25)$$

$$I(y_1, y_2, \dots, y_n) = -\sum_{i=1}^n E[\ln(f(y_i))] - \ln|\det \mathbf{W}| \quad (5.26)$$

Una estima instantánea de la Información Mutua genera una función de coste:

$$J(\mathbf{W}) = -\ln \left(\det \mathbf{W} \prod_{i=1}^n f(y_i(k)) \right) \quad (5.27)$$

Que puede minimizarse mediante un algoritmo de gradiente natural aplicado a (5.28) resultando la expresión dada en (5.30)

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu [\mathbf{W}^{-T}(k) - f(\mathbf{y}(k))\mathbf{x}^T(k)] \quad (5.28)$$

$$f(y) = -\frac{\partial \ln f_s(y)}{\partial y} \quad \text{i.i.d. fuentes} \quad (5.29)$$

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mu [\mathbf{W}^{-T}(k) - f(\mathbf{y}(k))\mathbf{x}^T(k)] \mathbf{W}^T(k) \mathbf{W}(k) = \\ \mathbf{W}(k+1) &= \mathbf{W}(k) + \mu [I - f(\mathbf{y}(k))\mathbf{y}^T(k)] \mathbf{W}(k) \end{aligned} \quad (5.30)$$

Aunque el algoritmo anterior requiere, en principio, conocer la f.d.p. de las fuentes, funciones sencillas como $f(y)=y^3$ (para sub-Gaussianas) o $f(y)=\text{sgn}(y)$ (para super-Gaussianas) dan buenos resultados.

Un criterio de aprendizaje no supervisado propuesto en el campo de las Redes Neuronales consiste en la maximización de la información mutua entrada/salida. El criterio, denominado INFOMAX, se basa en que la salida debe ser tal que contenga la máxima información mutua acerca de su entrada y fue propuesto por Linsker (1989) para una red lineal y extendido por Bell & Sejnowsky (1995) para una red no lineal.

Bajo ciertas restricciones se puede demostrar que maximizar la información mutua entrada/salida es equivalente a minimizar la información mutua entre las salidas, de manera que INFOMAX es equivalente a la estimación ML del modelo ICA.

5.4 Conclusiones

En definitiva, se ha visto que la suposición de no gaussianidad e independencia de los datos es suficiente para separar una mezcla instantánea de señales. El Análisis de Componentes Principales emplea estadísticos de segundo orden mientras que el Análisis de Componentes Independientes requiere el uso de estadísticos de orden superior. Para llevar a cabo la

estimación ICA se puede, o bien forzar la no Gaussianidad de los datos de salida o bien buscar su independencia minimizando la Información Mutua.

El método ICA empleado en los procesos de recuperación de marcas de agua desarrollados en los diferentes algoritmos realizados consiste en el método **SICA** (Análisis de Componentes Independientes Sinusoidal), que se corresponde con un método ICA en el que la función objetivo queda reducida a una sinusode (puesto que *“la función objetivo basada en momentos de cuarto orden que minimiza la información mutua entre dos mezclas lineales de dos variables aleatorias, es equivalente, bajo el supuesto de decorrelación a una sinusode”* [Murillo01]), de manera que el problema de minimización se reduce a calcular la fase de dicha sinusode. La justificación del empleo del método señalado radica en las buenas prestaciones que presenta en para alcanzar la independencia de las salidas, así como en la reducida carga computacional que presenta lo que lo hace adecuado para tratamiento digital de imágenes.