



UNIVERSIDAD DE LA CORUÑA
Departamento de Electrónica y Sistemas

TESIS DOCTORAL

**ALGORITMOS ADAPTATIVOS
PARA SEPARACIÓN CIEGA DE FUENTES**

Autor: ADRIANA DAPENA JANEIRO

Director: LUIS CASTEDO RIBAS

Febrero 1999

Dr. Luis Castedo Ribas
Profesor Titular de Universidad
Departamento de Electrónica y Sistemas
Universidad de La Coruña.

CERTIFICA:

Que la memoria titulada “Algoritmos Adaptativos para Separación Ciega de Fuentes”, ha sido realizada por Dña. Adriana Dapena Janeiro bajo mi dirección en el Departamento de Electrónica y Sistemas de la Universidad de La Coruña y concluye la Tesis que presenta para optar al grado de Doctor.

La Coruña, 18 de Noviembre de 1998

Fdo: Dr. Luis Castedo Ribas
Director de la Tesis Doctoral

Tesis Doctoral: ALGORITMOS ADAPTATIVOS PARA
SEPARACIÓN CIEGA DE FUENTES

Autor: Dña. Adriana Dapena Janeiro

Director: D. Luis Castedo Ribas

Fecha:

Tribunal

Presidente:

Vocal 1:

Vocal 2:

Vocal 3:

Secretario:

A Miguel

Agradecimientos

En esta sección de agradecimientos merece un lugar privilegiado mi director de tesis, Dr. Luis Castedo Ribas, porque sin su ayuda hubiese sido imposible realizar este trabajo.

También me gustaría hacer llegar mi agradecimiento a los miembros del grupo GSSC (Carlos, Cris, Dani, Mónica y Xocas), a los del grupo SSR en Madrid y a todas aquellas personas que, de una u otra forma, me han ayudado a realizar esta tesis. Me resultaría imposible nombrarlas a todas y, por ello, me limitaré a decir *gracias compañeros y amigos*.

En el terreno de lo personal quisiera expresar mi más profunda gratitud a mis padres y a mi hermana en quienes siempre he encontrado apoyo y comprensión. Tampoco puedo olvidarme de la persona que durante todos estos años ha estado a mi lado escuchándome pacientemente. Me refiero, por supuesto, a Miguel a quien dedico esta tesis.

Por último, quisiera mencionar a los dos contribuyentes que, en gran medida, han subvencionado esta tesis: la Xunta de Galicia (proyecto XUGA 10502A96) y la CICYT (proyecto TIC 96-0500-C10-02).

Resumen

El problema de separación ciega de fuentes consiste en la recuperación de un conjunto de señales originales (fuentes) a partir de observaciones de mezclas lineales de ellas. Se trata de un problema de extraordinario interés en procesado de señal porque para resolverlo sólo es necesario que se cumplan un conjunto de hipótesis muy poco restrictivas, a saber, la invertibilidad del sistema de mezcla, la independencia mutua entre las fuentes y la distribución no gaussiana de las mismas. Por esta razón, las técnicas de separación ciega de fuentes encuentran numerosas aplicaciones en campos tan diversos como procesado en array, comunicaciones multiusuario y reconstrucción de voz o imágenes.

A grandes rasgos las mezclas lineales se pueden clasificar en dos grandes grupos: instantáneas o sin memoria y convolutivas o con memoria. En las mezclas instantáneas se supone que las observaciones en un determinado instante son sumas ponderadas de las fuentes en ese mismo instante. Por el contrario, el modelo de mezclas convolutivas es mucho más complejo porque las observaciones dependen de las fuentes en el instante presente y en instantes anteriores o posteriores lo cual plantea muchas más dificultades a la hora de separar las fuentes.

La separación ciega de fuentes es un campo en el que se ha venido investigando intensamente a lo largo de los últimos diez años. Son numerosos los trabajos en los que se proponen algoritmos adaptativos para su solución pero muy pocos se preocupan de analizar con detenimiento su estabilidad. Esta es una cuestión crucial para comprender las condiciones bajo las cuales estos algoritmos funcionarán adecuadamente y constituye el núcleo central de la presente tesis.

Una primera parte de nuestro trabajo se ha dedicado al estudio de la estabilidad de algoritmos ya existentes obteniéndose condiciones suficientes para garantizar que los puntos donde se consigue una perfecta recuperación de las fuentes se correspondan con atractores de los algoritmos. La enorme dificultad de estos estudios nos ha obligado a considerar sólo el caso de sistemas separadores con dos entradas y dos salidas aunque conjeturamos que los resultados son extrapolables al caso más general.

En este trabajo también se propone una nueva familia de algoritmos de gradiente ascendente que permiten realizar la separación ciega de fuentes mediante la maximización de funciones de coste que utilizan estadísticos de segundo y cuarto orden. Esta aproximación se propone inicialmente para mezclas instantáneas y en un capítulo posterior se extiende al caso de mezclas convolutivas. Esta nueva familia de algoritmos surgen como extensión del criterio de Shalvi y Weinstein para igualación ciega. Aunque las funciones de coste resultan ser no cuadráticas, el hecho de que sólo utilicen estadísticos de orden dos y cuatro permite llevar a cabo un análisis completo de sus puntos estacionarios para el caso general de N fuentes complejas y mezclas instantáneas o convolutivas. El análisis ha revelado que los únicos atractores de los algoritmos se corresponden con la perfecta recuperación de las fuentes y que no existen atractores indeseados que puedan

dañar su convergencia. Hasta la fecha de ninguna otra familia de algoritmos adaptativos para separación ciega de fuentes se ha podido demostrar que está libre de atractores no deseados.

Abstract

The problem of blind source separation consists in the recovery of a set of original signals (sources) from observations of linear mixtures of them. It is a problem of paramount importance in signal processing because its solution only requires a set of little restrictive hypothesis, namely, the invertibility of the mixing system, the mutual independence of the sources and the non-gaussian distribution of them. For this reason, blind source separation techniques can be applied to numerous applications in diverse fields such as array processing, multiuser communications and image or voice restoration.

Linear mixtures can be classified in two large groups: instantaneous or memoryless and convolutive or with memory. In the instantaneous mixtures case it is assumed that the observations in a certain time are weighted sums of the sources in this same time. On the contrary, the convolutive mixtures model is more complex since observations depend on the sources at the present, past and future time which poses more difficulties when separating the sources.

Blind source separation has been an active area of research along the last ten years. There are a lot of works that propose adaptive algorithms for its solution but few of them carry out a detailed analysis of their stability. This is a crucial issue to understand the conditions under which these algorithms perform adequately and constitutes the core of the present work.

The first part of our work is devoted to the stability analysis of existing algorithms. Sufficient conditions to guarantee that the points where a perfect recovery or the sources correspond to the algorithms attractors are obtained. The extremely difficulty of these analysis has forced us to consider only the case of instantaneous mixtures and two inputs two outputs separating systems although we conjecture that the results are also valid in more general cases.

In this work we have also proposed a new family of gradient algorithms for blind source separation that maximize cost functions that use second and fourth order statistics. This approach is initially proposed for instantaneous mixtures and extended to convolutive mixtures in a subsequent chapter. This new family of algorithms arises as an extension of the Shalvi and Weinstein criterion for blind equalization. Although the proposed cost functions turn out to be non-quadratic, the fact that they only make use of second and fourth order statistics enables us to carry out a complete analysis of their stationary points for the general case of N complex sources and instantaneous or convolutive mixtures. The analysis has shown that the only algorithm attractors correspond to the perfect recovery of the sources and that there are no undesirable attractors that may impair the algorithms convergence. Up to date no other family of adaptive algorithms for blind source separation has been shown to be free of undesirable attractors.

Contenido

Lista de figuras	xix
Lista de tablas	xxiii
1 Introducción	1
1.1 Modelo de separación de mezclas instantáneas	2
1.1.1 Conexión con Redes de Neuronas Artificiales	4
1.2 Modelo de separación de mezclas convolutivas	7
1.3 Aplicaciones de separación ciega de fuentes	9
1.3.1 Procesado en array	9
1.3.2 Sistemas de comunicación CDMA	11
1.4 Planteamiento y estructura de la tesis	14
2 Fundamentos teóricos de algoritmos adaptativos	17
2.1 Introducción	17
2.2 Sistemas adaptativos	18
2.2.1 Conexión con los sistemas dinámicos	19
2.2.2 Estabilidad de los algoritmos adaptativos	23
2.3 Algoritmos de gradiente	24
2.3.1 Algoritmos de gradiente relativo	25
2.3.2 Estabilidad de los algoritmos de gradiente	26
2.4 El criterio de Mínimo Error Cuadrático Medio	27
2.5 Conclusiones	30
3 Revisión y aportaciones a aproximaciones anteriores	33
3.1 Introducción	33
3.2 Algoritmos de decorrelación	34
3.3 El algoritmo de Héroult y Jutten	35
3.3.1 Estabilidad del algoritmo de Héroult y Jutten	36
3.4 El algoritmo de Cichocki y Undebauen	40
3.5 Funciones de contraste	42
3.5.1 Estabilidad de la función de contraste de Moreau y Macchi	45

3.6	El algoritmo EASI	48
3.6.1	Estabilidad del algoritmo EASI	51
3.7	Algoritmos de maximización de la transferencia de información	53
3.7.1	Estabilidad del algoritmo de Bell y Sejnowski	59
3.8	Maximización de la verosimilitud	63
3.9	Conclusiones	66
4	Nuevos criterios para separación ciega de fuentes	69
4.1	Introducción	69
4.2	Criterios para separación de mezclas instantáneas	70
4.3	Algoritmos adaptativos	73
4.3.1	Algoritmo I	75
4.3.2	Algoritmo II	75
4.3.3	Algoritmo III	77
4.3.4	Algoritmo IV	77
4.4	Análisis de estabilidad del algoritmo I	77
4.5	Resultados de simulaciones	82
4.5.1	Entornos sin ruido	82
4.5.2	Mezclas mal condicionadas	90
4.5.3	Entornos con ruido blanco gaussiano	91
4.5.4	Aproximación tipo bloque	93
4.6	Conclusiones	96
5	Criterios de separación para mezclas convolutivas	99
5.1	Introducción	99
5.2	Criterios para separación de mezclas convolutivas	100
5.3	Algoritmos adaptativos	102
5.3.1	Algoritmo I generalizado	104
5.3.2	Algoritmo II generalizado	105
5.4	Análisis de estabilidad del algoritmo I generalizado	106
5.5	Resultado de simulaciones	110
5.6	Conclusiones	116
6	Conclusiones y líneas futuras de trabajo	119
6.1	Conclusiones	119
6.1.1	Revisión y aportaciones a anteriores aproximaciones	120
6.1.2	Nuevos criterios de separación	121
6.2	Líneas futuras de investigación	122

A	Definiciones de derivadas de funciones	125
A.1	Definiciones de derivadas reales	125
A.2	Definiciones de derivadas complejas	127
B	Revisión de estadísticos de orden superior	129
B.1	Definiciones de HOS	129
B.2	Propiedades de los cumulantes	131
C	Revisión de Teoría de la Información	133
C.1	Minimización de la divergencia de Kullback-Leibler	135
C.1.1	Algoritmo de gradiente relativo	139
D	Jacobiano de los algoritmos del capítulo 3	141
D.1	Jacobiano del algoritmo de Hérault y Jutten	141
D.2	Puntos estacionarios del contraste de Moreau y Macchi	143
D.3	Jacobiano del algoritmo EASI	145
D.4	Jacobiano del algoritmo de Bell y Sejnowski	147
E	Puntos estacionarios de los criterios I y II	151
E.1	Puntos estacionarios del criterio I	151
E.2	Puntos estacionarios del criterio II	154
F	Puntos estacionarios del criterio I generalizado	165
G	Notación y abreviaturas empleadas	169
G.1	Notación	169
G.2	Abreviaturas	170
	Bibliografía	171

Lista de figuras

1.1	Modelo de separación de mezclas instantáneas.	2
1.2	Modelo de una neurona artificial	4
1.3	Ejemplos de funciones de activación.	5
1.4	Estructura de una red neuronal.	6
1.5	Modelo de separación de mezclas convolutivas.	7
1.6	Esquema de un conformador de banda estrecha.	9
1.7	Esquema de un conformador de banda ancha.	11
1.8	Esquema de un receptor CDMA síncrono.	12
2.1	Esquema de un sistema adaptativo para separación de señales.	18
2.2	Estabilidad de un sistema dinámico.	21
2.3	Clasificación de los puntos estacionarios.	22
2.4	Ajuste del sistema de separación utilizando el algoritmo LMS.	28
2.5	Funciones de error cuadrático medio.	29
2.6	Ajuste del sistema de separación utilizando el algoritmo p-vector.	30
3.1	Sistema de separación utilizado por Héroult y Jutten.	35
3.2	Condiciones de estabilidad del algoritmo de Héroult y Jutten.	39
3.3	Red neuronal directa utilizada en separación ciega.	41
3.4	Sistema de separación con AGCs.	44
3.5	Contraste J_c con $\alpha = 1$ (<i>izquierda</i>) y con $\alpha = 2.5$ (<i>derecha</i>).	47
3.6	Evolución de las ganancias del sistema separador ajustado con el algoritmo de Moreau y Macchi.	48
3.7	Sistema de separación de dos etapas.	49
3.8	Condiciones de estabilidad del algoritmo EASI para $f(y) = \tanh(y)$ (<i>izquierda</i>) y $f(y) = -\tanh(y)$ (<i>derecha</i>).	54
3.9	Evolución de las ganancias del sistema separador ajustado con el algoritmo EASI empleando $f(y) = -\tanh(y)$	54
3.10	Ejemplos de funciones de activación.	55
3.11	Condiciones de estabilidad del algoritmo de Bell y Sejnowski para $g(y) = 2\tanh(y)$	62

3.12	Evolución de las ganancias del sistema separador ajustado con el algoritmo de Bell y Sejnowski empleando $g(y) = 2 \tanh(y)$	62
3.13	Modelo del sistema de mezcla/separación para el criterio de máxima verosimilitud.	64
3.14	Comparación entre el criterio de máxima transferencia de información y el de máxima verosimilitud.	66
4.1	Sistema MIMO ajustado con el criterio de Shalvi y Weinstein.	72
4.2	Sistema MIMO adaptado considerando la dependencia entre sus salidas. . .	76
4.3	Clasificación de los puntos estacionarios indeseados del algoritmo I.	82
4.4	Separación de imágenes: fuentes y observaciones.	84
4.5	Separación de imágenes: señales recuperadas con el algoritmo I.	85
4.6	Separación de imágenes: señales recuperadas con el algoritmo II.	85
4.7	Separación de imágenes: señales recuperadas con el algoritmo III.	86
4.8	Separación de imágenes: evolución del índice $P(\mathbf{G})$	87
4.9	Separación de voz y música: fuentes y observaciones.	88
4.10	Separación de voz y música: señales recuperadas con los algoritmo propuestos. .	89
4.11	Separación de voz y música: evolución del índice $P(\mathbf{G})$	91
4.12	Separación de imágenes: evolución del índice $P(\mathbf{G})$ para el algoritmo IV y distintas matrices de mezcla.	92
4.13	Evolución de la SINR para los algoritmos I y II.	94
4.14	Evolución del índice $P(\mathbf{G})$ para separación de imágenes utilizando una aproximación tipo bloque.	97
4.15	Índice $P(\mathbf{G})$ en la iteración $S = 40$ para separación de señales de comunicación utilizando una aproximación tipo bloque.	98
5.1	Clasificación de los puntos estacionarios indeseados del algoritmo I generalizado.	110
5.2	Respuesta al impulso del sistema de mezcla.	111
5.3	Observaciones y salidas del primer experimento.	112
5.4	Evolución del índice $P(\mathbf{G})$ para el primer experimento.	113
5.5	Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo I generalizado en el primer experimento.	113
5.6	Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo II generalizado en el primer experimento.	114
5.7	Observaciones y salidas del segundo experimento.	115
5.8	Evolución del índice $P(\mathbf{G})$ para el segundo experimento.	116
5.9	Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo I generalizado en el segundo experimento.	117
5.10	Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo II generalizado en el segundo experimento.	118

6.1	Sistema separador de mezclas convolutivas en el dominio de la frecuencia. .	123
B.1	Comparación de distribuciones con la gaussiana	130
C.1	Relación entre la entropía de dos variables aleatorias y la información mutua.	134
E.1	Clasificación de los puntos estacionarios.	151

Lista de tablas

3.1	Distribución de probabilidad discreta empleada para verificar las condiciones de estabilidad del algoritmo de Héroult y Jutten.	39
3.2	Estadísticos de las fuentes empleadas para verificar las condiciones de estabilidad del algoritmo de Héroult y Jutten.	39
3.3	Distribución de probabilidad discreta empleada para verificar las condiciones de estabilidad del algoritmo EASI.	52
3.4	Estadísticos de las fuentes empleadas para verificar las condiciones de estabilidad del algoritmo EASI para $f(y) = \tanh(y)$	53
3.5	Distribución de probabilidad discreta utilizada para verificar las condiciones de estabilidad del algoritmo de Bell y Sejnowski.	61
3.6	Parámetros de estabilidad del algoritmo de Bell y Sejnowski para $g(y) = 2 \tanh(y)$	61
4.1	Separación de imágenes: parámetros de los algoritmos.	87
4.2	Separación de voz y música: parámetros de los algoritmos.	90
4.3	Parámetros de las fuentes utilizadas en las simulaciones del entorno con ruido.	93
4.4	Parámetros de los algoritmos empleados en las simulaciones del entorno con ruido.	93
4.5	Entornos con ruido: <i>SINRs</i> obtenidas con el algoritmo I.	95
4.6	Entornos con ruido: <i>SINRs</i> obtenidas con el algoritmo II.	96
5.1	Parámetros de los algoritmos empleados en las simulaciones.	114
C.1	Entropías e información mutua para variables discretas y continuas.	136

Capítulo 1

Introducción

La presente tesis doctoral aborda el problema de recuperar un conjunto de señales originales (fuentes) a partir de observaciones de mezclas lineales de ellas. Este problema se plantea en numerosas aplicaciones prácticas en las que, por ejemplo, se persigue recuperar las señales que se propagan por un medio abierto que exhiba un comportamiento lineal (como es el caso del aire o del agua) a partir de las señales suministradas por una colección o *array* de sensores.

La forma en que se lleva a cabo esta mezcla lineal puede ser de naturaleza muy diversa y depende tanto de las propiedades del medio como de las características físicas y eléctricas de las fuentes. El modelo más sencillo considera que las mezclas son instantáneas o sin memoria porque en cada instante de tiempo se tiene la suma de las fuentes en ese instante ponderadas por ciertos parámetros que describen el entorno. Este modelo, sin embargo, es insuficiente para describir muchas situaciones prácticas en las que nos encontramos con mezclas convolutivas o con memoria. En este caso las observaciones en un instante no son sólo combinaciones lineales de las fuentes en ese instante sino también de instantes anteriores y posteriores.

La separación de mezclas lineales de señales es un campo en el que se ha venido investigando a lo largo de los últimos años y en el que se han propuesto numerosas soluciones [12]. Las técnicas clásicas se basan fundamentalmente en el criterio de Mínimo Error Cuadrático Medio (MMSE, del inglés Minimum Mean Square Error) y requieren conocer o bien las fuentes o bien los parámetros de la mezcla. Esta información, sin embargo, no está disponible en numerosas aplicaciones prácticas (en comunicaciones, por ejemplo) por lo que se han desarrollado un conjunto de técnicas que son capaces de llevar a cabo la separación a partir de las observaciones exclusivamente. Estas técnicas se denominan ciegas o no supervisadas y poseen un ámbito de aplicación mucho más extenso que las aproximaciones convencionales. La clave de su funcionamiento reside en el aprovechamiento de determinadas propiedades estadísticas de las fuentes. En particular veremos que la independencia estadística y la naturaleza no gaussiana de las fuentes son suficientes para poder llevar a cabo la separación de forma ciega.

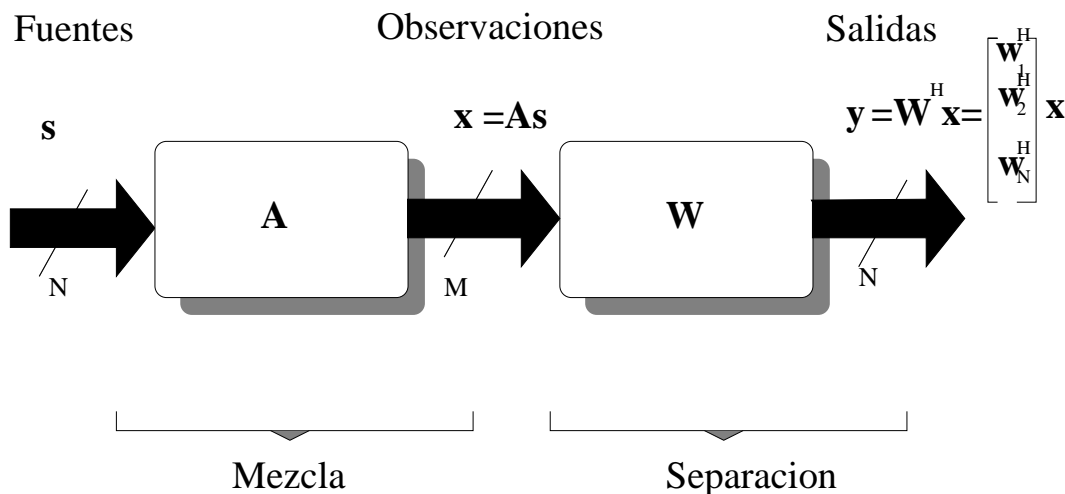


Figura 1.1: Modelo de separación de mezclas instantáneas.

El presente capítulo se organiza de la siguiente manera. En la sección 1.1 se describe el problema de separación ciega de fuentes para mezclas instantáneas y su conexión con las Redes Neuronales Artificiales. La sección 1.2 presenta el modelo de señal para el caso de mezclas convolutivas. La sección 1.3 presenta varios ejemplos de aplicaciones prácticas en las que se debe separar mezclas lineales de señales. Finalmente, la sección 1.4 expone los objetivos de la tesis y su organización.

1.1 Modelo de separación de mezclas instantáneas

El problema básico de separación de fuentes consiste en recuperar señales a partir de mezclas lineales e instantáneas de ellas. El modelo que se asume es el presentado en la figura 1.1 donde $\mathbf{s} = [s_1, \dots, s_N]^T$ es un vector de dimensión $N \times 1$ formado por las fuentes que modelaremos como un conjunto de procesos estocásticos estacionarios de media cero y varianza unidad. En general asumiremos que pueden tomar valores complejos para considerar las señales que se manejan en el ámbito de las comunicaciones. En situaciones prácticas las fuentes se corresponden con señales que tienen un origen físico diferente y, por lo tanto, supondremos que son estadísticamente independientes entre sí.

Las fuentes son mezcladas por un sistema que se modela como una matriz \mathbf{A} de dimensión $M \times N$. Al igual que las fuentes, supondremos que \mathbf{A} es desconocida aunque siempre garantizaremos que las columnas son linealmente independientes entre sí. El resultado de esta mezcla es un vector de observaciones $\mathbf{x} = [x_1, \dots, x_M]^T$ que se puede escribir como sigue

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1.1)$$

Las observaciones \mathbf{x} son combinaciones lineales de las fuentes, lo cual da cabida a la utilización de un sistema lineal para invertir el proceso de mezcla. En la figura 1.1 se

ha representado la etapa de separación como un sistema MIMO (Multiple Input Multiple Output) formado por N combinadores lineales cuyas salidas vienen dadas por la siguiente expresión

$$y_i = \mathbf{w}_i^H \mathbf{x} \quad i = 1, \dots, N \quad (1.2)$$

donde \mathbf{w}_i es un vector complejo de dimensión $M \times 1$ que contiene los coeficientes del i -ésimo combinador (el superíndice H denota conjugado traspuesto). Agrupando los vectores de coeficientes en una matriz $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N]$ de dimensión $M \times N$ podemos escribir las salidas de una forma mas compacta

$$\mathbf{y} = \mathbf{W}^H \mathbf{x} \quad (1.3)$$

donde $\mathbf{y} = [y_1, \dots, y_N]^T$ representa al vector de salidas. Adicionalmente, combinado las expresiones (1.1) y (1.3) se obtiene que

$$\mathbf{y} = \mathbf{G} \mathbf{s} \quad (1.4)$$

donde $\mathbf{G} = \mathbf{W}^H \mathbf{A}$ es una matriz $N \times N$ que agrupa las etapas de mezcla y separación.

Observando (1.4) es inmediato deducir que el proceso de mezcla será invertido cuando la matriz \mathbf{G} sea la identidad. Sin embargo, también se consideran validas aquellas soluciones donde las fuentes son recuperadas en otro orden o con una cierta ganancia $g_{ij} \neq 1$. Como consecuencia, la separación de las fuentes se produce en aquellas situaciones en las que la matriz de ganancias \mathbf{G} adquiere la siguiente forma

$$\mathbf{G} = \mathbf{\Lambda} \mathbf{P} \quad (1.5)$$

donde $\mathbf{\Lambda}$ es una matriz diagonal invertible y \mathbf{P} es una matriz de permutación. Algunos de los métodos de separación que abordaremos en el presente trabajo fijan *a priori* las ganancias con las que serán extraídas las fuentes (generalmente, $\mathbf{\Lambda} = \mathbf{I}$). Sin embargo, no establecen ningún control sobre el orden en que serán recuperadas.

Es interesante hacer notar que el problema de separación ciega de fuentes se puede resolver imponiendo únicamente la condición de que las fuentes sean no-gaussianas y estadísticamente independientes entre sí. Esto es una consecuencia del siguiente teorema.

Teorema 1.1 *Teorema de Darmois-Skitovich [9, 19]:* si $\mathbf{s} = [s_1, \dots, s_N]^T$ es un vector de componentes estadísticamente independientes entre sí con a lo sumo una componente gaussiana y $\mathbf{y} = \mathbf{G} \mathbf{s}$ es un vector cuyos componentes y_i y y_j son independientes, entonces todas las variables s_k para la cual $g_{ik}g_{jk} \neq 0$ son gaussianas.

Así pues, si el vector de fuentes \mathbf{s} tiene a lo sumo una componente con distribución no gaussiana, las salidas serán estadísticamente independientes entre sí si y sólo si la matriz de ganancias tiene la forma (1.5) y, por tanto, cuando se consiga la separación de las fuentes. Esto quiere decir que si las fuentes no son gaussianas el criterio de seleccionar el

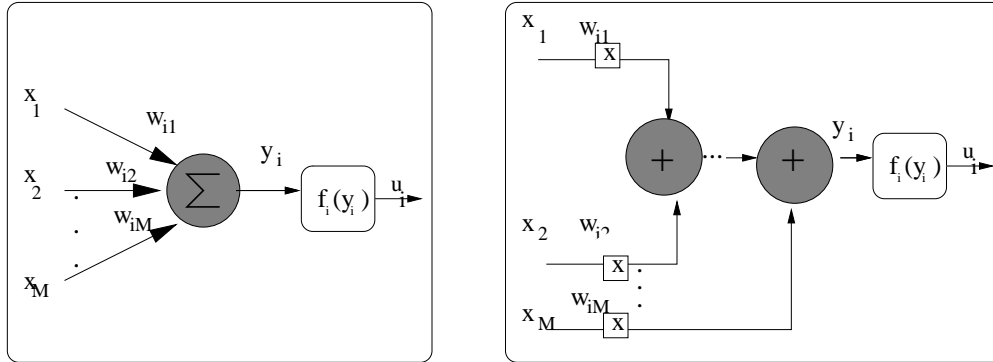


Figura 1.2: Modelo de una neurona artificial

sistema de separación \mathbf{W} para que las salidas sean estadísticamente independientes entre sí es suficiente para garantizar la separación.

El criterio de independencia estadística, sin embargo, es insuficiente para separar las fuentes si estas tienen una distribución gaussiana. Esto es fácil de demostrar si consideramos un vector de N fuentes gaussianas, estadísticamente independientes entre sí, de media cero y varianza unidad. Es bien conocido que las señales gaussianas quedan totalmente caracterizadas por su matriz de correlación que en este caso es inmediato comprobar que adopta la forma $\mathbf{R}_s = E[\mathbf{s}\mathbf{s}^H] = \mathbf{I}$. Tanto las observaciones (1.1) como las salidas (1.3) son combinaciones lineales de señales gaussianas de media nula y, por ello, ambas también tendrán distribución gaussiana y media cero. Sus correspondientes matrices de autocorrelación son

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H] = \mathbf{A}\mathbf{R}_s\mathbf{A}^H = \mathbf{A}\mathbf{A}^H \quad (1.6)$$

$$\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^H] = \mathbf{W}^H\mathbf{R}_x\mathbf{W} = \mathbf{W}^H\mathbf{A}\mathbf{A}^H\mathbf{W} \quad (1.7)$$

que, en general, serán distintas de \mathbf{R}_s . Sin embargo, si $\mathbf{W}^H\mathbf{A}$ es una matriz unitaria, entonces $\mathbf{R}_y = \mathbf{I}$ y las salidas serán estadísticamente independientes entre sí sin necesidad de que cada salida se corresponda con una sola fuente.

1.1.1 Conexión con Redes de Neuronas Artificiales

El problema de separación ciega de fuentes está íntimamente relacionado con las Redes de Neuronas Artificiales porque el sistema de separación se puede interpretar como los pesos sinápticos de una red neuronal de una sola capa de modo que los algoritmos adaptativos para separación ciega de fuentes se pueden interpretar como reglas de aprendizaje no supervisadas para redes neuronales.

Las redes neuronales artificiales han surgido como modelos que intentan reproducir el funcionamiento del sistema nervioso desde un punto de vista tanto estructural como funcional [40]. Una red neuronal artificial está formada por elementos simples,

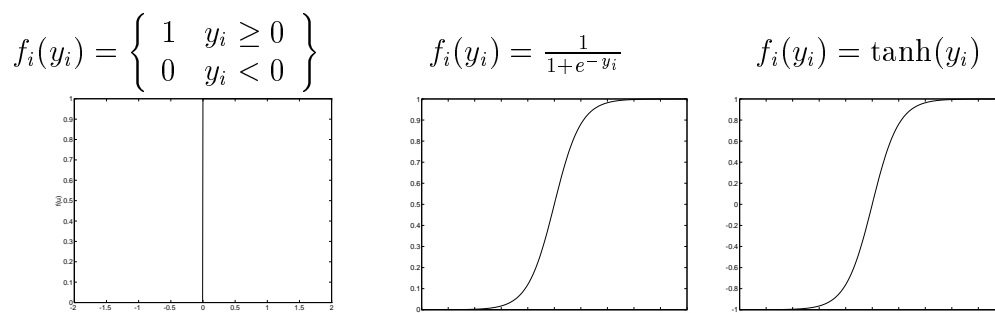


Figura 1.3: Ejemplos de funciones de activación.

habitualmente adaptativos, interconectados entre sí con el objeto de cooperar o competir hacia un fin común. La figura 1.2 muestra el modelo de una neurona artificial con entradas \mathbf{x} , salida y_i y pesos \mathbf{w}_i . La neurona realiza dos operaciones básicas: cálculo de su estado como combinación lineal de entradas ponderadas por unos pesos $y_i = \mathbf{w}_i^H \mathbf{x}$ y cálculo de la salida como función de la combinación lineal de las entradas $u_i = f_i(y_i)$. La función de activación $f_i(\cdot)$ puede tomar diversas formas dependiendo del propósito de la red y de la naturaleza de su salida. Por ejemplo, como se muestra en la figura 1.3, la función escalón proporciona una salida binaria mientras que las otras dos funciones (sigmoide y tangente hiperbólica) proporcionan una salida continua.

Las configuraciones que se pueden obtener mediante la interconexión de las neuronas son muy diversas [40]. Una de las más sencillas es la mostrada en la figura 1.4: N neuronas que procesan el mismo conjunto de observaciones \mathbf{x} . El estado de activación y la salida de esta red puede escribirse en forma matricial como sigue

$$\mathbf{y} = \mathbf{W}^H \mathbf{x} \quad (1.8)$$

$$\mathbf{u} = \mathbf{f}(\mathbf{y}) \quad (1.9)$$

donde \mathbf{W} es una matriz $M \times N$ cuyas columnas son los vectores de pesos \mathbf{w}_i de las neuronas y $\mathbf{f}(\cdot)$ es la función de activación en forma vectorial. Obsérvese que la expresión (1.8) es idéntica a (1.3) lo que nos permite interpretar un sistema de separación de señales como la parte lineal de la red neuronal monocapa y los algoritmos adaptativos para separación de señales como reglas de aprendizaje de redes neuronales [49].

Un aspecto importante respecto al aprendizaje en las redes neuronales es el conocer como se modifican los valores de los pesos, es decir, cuales son los criterios que se siguen para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información. De forma general, se suelen considerar dos tipos de reglas de aprendizaje: supervisado y no supervisado.

En las aproximaciones supervisadas los pesos sinápticos se ajustan según el criterio de minimizar el error cuadrático medio entre la salida y un conjunto de patrones de entrenamiento. Este conjunto de patrones no está disponible en numerosas aplicaciones prácticas por lo que se debe recurrir a aproximaciones no supervisadas que trabajan sin

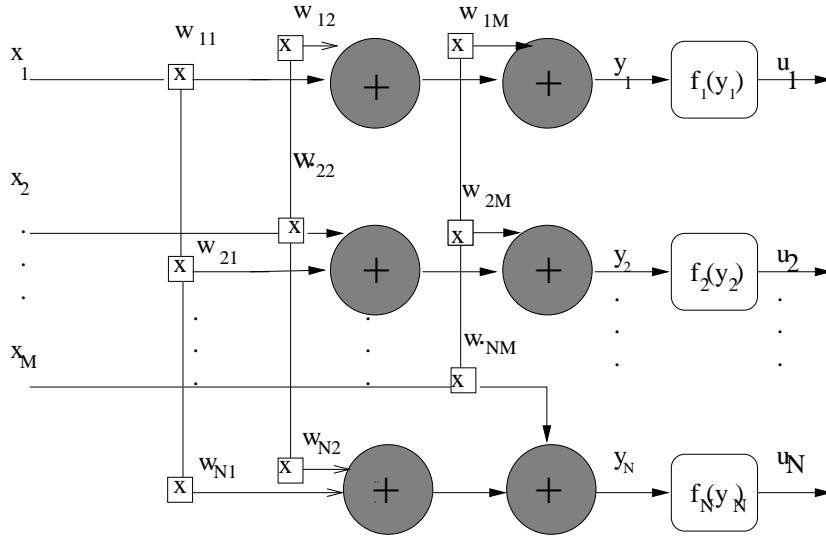


Figura 1.4: Estructura de una red neuronal.

recibir ninguna información que les indique si la salida generada es la deseada. En este tipo de aprendizaje se pretende que las neuronas se auto-organicen aprendiendo a captar las características principales de los datos de entrada sin ninguna ayuda externa.

Uno de los postulados del aprendizaje no-supervisado más antiguo y famoso se debe al neuro-psicólogo Donald Hebb [40] quien estableció lo siguiente: *Cuando un axón de una celda A está suficientemente cerca como para conseguir excitar una celda B y repetida o persistentemente toma parte en su activación, algún proceso de crecimiento o cambio metabólico tiene lugar en una o en ambas celdas, de tal forma que la eficiencia de A, cuando la celda a activar es B, aumenta.* Por celda, Hebb entiende un conjunto de neuronas fuertemente conectadas a través de una estructura compleja. La eficiencia podría identificarse como la intensidad o magnitud de la conexión.

Se puede interpretar, por tanto, que el aprendizaje hebbiano consiste básicamente en ajustar los pesos de las conexiones de acuerdo con la correlación de los valores de las salidas de las neuronas conectadas. Si consideramos las conexiones w_{ij} de la red neuronal de la figura 1.4 con actividad presináptica x_j y postsináptica y_i , entonces los pesos serán modificados utilizando términos del tipo

$$\Delta w_{ij} = x_j y_i \quad (1.10)$$

Esta regla, sin embargo, presenta un crecimiento exponencial lo que puede conducir a la saturación de los pesos \mathbf{W} . Esta situación indeseada es evitada en el algoritmo conocido como GHA (Generalized Hebbian Algorithm) que utiliza el siguiente factor de incremento

$$\Delta w_{ij} = y_i \left(x_j - \sum_{k=1}^i w_{ki} y_k \right) \quad (1.11)$$

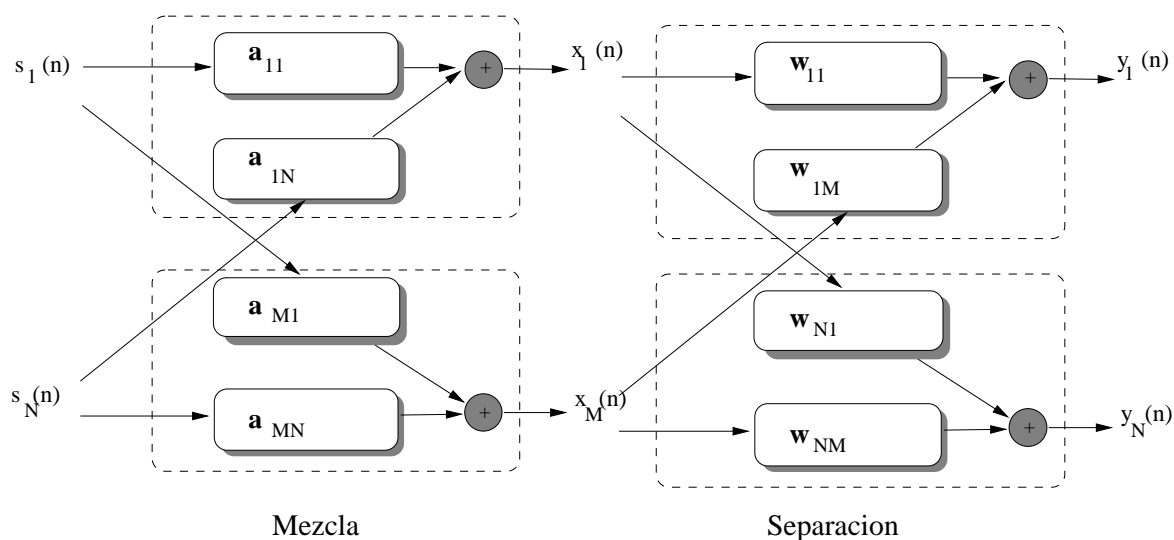


Figura 1.5: Modelo de separación de mezclas convolutivas.

Una característica destacable de este aprendizaje es que se puede demostrar que converge a una situación donde los pesos de una red neuronal monocapa con M entradas y N salidas son los N autovectores más significativos de la matriz de correlación de las entradas \mathbf{R}_x [40]. Por tanto, el GHA permite realizar una descomposición en componentes principales (PCA, del inglés Principal Components Analysis) [33].

Una de las aplicaciones más importantes del GHA es la compresión de datos ya que el ajuste de una red neuronal monocapa de M entradas y N salidas permitirá pasar de bloques \mathbf{x} de tamaño M a otros \mathbf{y} de dimensión N ($N < M$) que contienen, aproximadamente, la misma información. Este algoritmo, sin embargo, resulta ser insuficiente para realizar la separación de fuentes debido a que únicamente emplea estadísticos de orden dos. Esta limitación ha conducido a proponer otras generalizaciones del algoritmo hebbiano incorporando no-linealidades [44, 50]. En particular, más adelante haremos uso de la expresión [44]

$$\Delta w_{ij} = x_j g(y_i) \quad (1.12)$$

donde $g(y_j)$ es una función no-lineal de la salida que puede tomar diversas formas: función tangente hiperbólica, funciones cúbicas, polinomios, etc.

1.2 Modelo de separación de mezclas convolutivas

En la sección anterior se ha descrito el problema de separación ciega de fuentes para mezclas instantáneas. Este modelo sólo es válido para un reducido número de situaciones ya que en la práctica es más habitual encontrarse con que las observaciones son mezclas

convolutivas de las fuentes: el sistema de mezcla suele tener memoria y las observaciones en un instante dependen de las fuentes en ese instante y en instantes anteriores y posteriores.

En la figura 1.5 puede verse el modelo correspondiente a mezclas convolutivas donde s_i son las fuentes que supondremos, al igual que en la sección anterior, procesos aleatorios estacionarios complejos, con distribución no-gaussiana, estadísticamente independientes entre sí, de media cero y potencia unidad. Debido a la existencia de memoria en la mezcla se hace necesario especificar la función de autocorrelación de las fuentes o, equivalentemente, la densidad espectral de potencia. A lo largo del presente trabajo consideraremos únicamente la situación en que las fuentes son blancas. La principal diferencia con el modelo de la sección 1.1 es que ahora las observaciones \mathbf{x} se representan como la convolución entre un sistema desconocido y las fuentes

$$\mathbf{x}(n) = \sum_{k=-\infty}^{\infty} \mathbf{A}(k)\mathbf{s}(n-k) \quad (1.13)$$

donde $\{\mathbf{A}(k)\}$ es una secuencia de matrices de dimensión $M \times N$ que define la respuesta al impulso del sistema de mezcla

$$\mathbf{A}(k) = \begin{bmatrix} a_{11}(k) & \cdots & a_{1N}(k) \\ a_{21}(k) & \cdots & a_{2N}(k) \\ \vdots & \ddots & \vdots \\ a_{M1}(k) & \cdots & a_{MN}(k) \end{bmatrix} \quad (1.14)$$

En ocasiones se considera que (1.13) es la suma de un número finito de términos lo que equivale a tener matrices $\mathbf{A}(k)$ iguales a cero. En el caso extremo en el que solamente una matriz $\mathbf{A}(k)$ es no nula, estaremos ante un problema de mezcla lineal instantánea.

Invertir el proceso de mezcla es una tarea más difícil que en el modelo presentado en la sección 1.1 ya que el sistema separador, por un lado, debe recuperar todas las fuentes y, por otro lado, debe ecualizar las salidas. En la figura 1.5 se ha representado un sistema formado por N filtros IIR (Infinite Impulse Response) cuyas salidas vienen dadas por la siguiente expresión

$$\mathbf{y}(n) = \sum_{k=-\infty}^{\infty} \mathbf{W}^H(k)\mathbf{x}(n-k) \quad (1.15)$$

donde $\{\mathbf{W}(k)\}$ es una secuencia de matrices $M \times N$.

El objetivo en separación de mezclas convolutivas es encontrar los coeficientes \mathbf{W} de forma que la función de transferencia global $\mathbf{G}(z) = \mathbf{W}(z)\mathbf{A}(z)$ satisfaga la siguiente condición

$$\mathbf{G}(z) = \mathbf{D}(z)\mathbf{\Delta}\mathbf{P} \quad (1.16)$$

donde $\mathbf{D}(z) = \text{diag}(z^{-l_1}, \dots, z^{-l_N})$ con l_j $1 \leq j \leq N \in \mathcal{Z}^N$, $\mathbf{\Delta}$ es una matriz diagonal invertible y \mathbf{P} es una matriz de permutación [57]. En otras palabras, las salidas del sistema separador óptimo deben extraer fuentes diferentes y llevar a cabo una igualación.

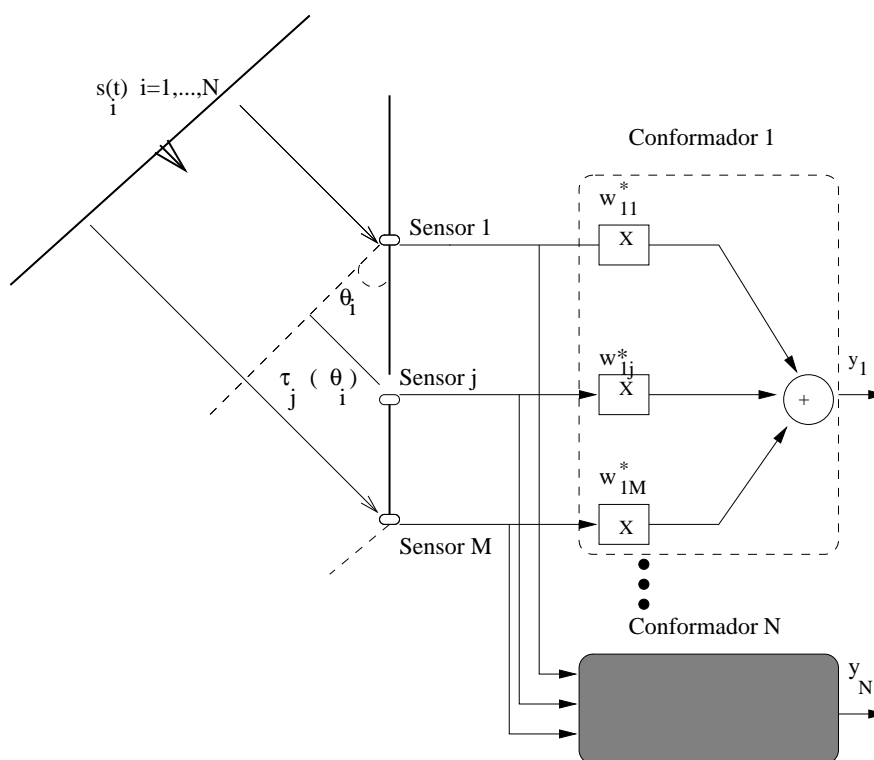


Figura 1.6: Esquema de un conformador de banda estrecha.

1.3 Aplicaciones de separación ciega de fuentes

Muchas son las situaciones prácticas en las que surge la necesidad de separar de forma ciega un conjunto de fuentes a partir de mezclas lineales de ellas. En la presente sección exponemos algunas de las más significativas.

1.3.1 Procesado en array

Una de las aplicaciones más típicas del problema de separación ciega de fuentes es la recuperación de las señales que se propagan a lo largo de un array de sensores (ver figura 1.6). Supongamos un entorno formado por N señales $s_i(t)$ de banda estrecha

$$s_i(t) = u_i(t)e^{j2\pi f_i t} \quad i = 1, \dots, N \quad (1.17)$$

donde $u_i(t)$ y f_i son la envolvente compleja y la frecuencia central de cada una de ellas. Estas señales inciden sobre un array de M sensores equiespaciados. Fijaremos el origen de tiempos de forma que la señal en el primer sensor sea igual a $s_i(t)$. Los restantes sensores recibirán una versión retardada de la señal original. El retardo $\tau_j(\theta_i)$, $i = 1, \dots, N$, $j = 1, \dots, M$ dependerá de la distancia al sensor de referencia, de la velocidad de propagación y del ángulo de incidencia θ_i . En particular, la señal $x_{ji}(t)$ en

el sensor j -ésimo se puede escribir como sigue [66]

$$x_{ji}(t) = s_i(t - \tau_j(\theta_i)) = u_i(t - \tau_j(\theta_i))e^{j2\pi f_i(t - \tau_j(\theta_i))} \quad (1.18)$$

Considerando que el retardo que sufre la envolvente compleja $u_i(t)$ es despreciable se obtiene

$$x_{ji}(t) \approx u_i(t)e^{j2\pi f_i(t - \tau_j(\theta_i))} = s_i(t)e^{-j2\pi f_i\tau_j(\theta_i)} \quad (1.19)$$

De lo anterior se desprende que las salidas de los sensores sólo se diferencian en un cierto desfase. De forma compacta podemos escribir las señales en los sensores de la siguiente forma

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1.20)$$

donde $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_N)]$ es una matriz $M \times N$ que contiene los distintos desfases de la señal: cada vector $\mathbf{a}(\theta_i) = [1, e^{-j2\pi f_i\tau_1(\theta_i)}, \dots, e^{-j2\pi f_i\tau_M(\theta_i)}]^T$ es conocido como vector director o vector de *steering* [66]. Si identificamos las señales que inciden en los sensores como las fuentes de nuestro problema y las señales procedentes de los sensores como las observaciones, obtenemos que la expresión (1.20) es idéntica al modelo de mezcla instantánea (1.1).

En la segunda etapa de la figura 1.6 podemos observar N conformadores que combinan las observaciones \mathbf{x} para producir salidas de la forma

$$y_i = \sum_{j=1}^M w_{ij}^* x_j \quad i = 1, \dots, N \quad (1.21)$$

que puede escribirse en forma matricial como sigue

$$\mathbf{y} = \mathbf{W}^H \mathbf{s} \quad (1.22)$$

La multiplicación de las observaciones por los pesos \mathbf{W} tiene por efecto el cambio de su amplitud y fase. El objetivo en conformación de haz adaptativa es ajustar adecuadamente estos coeficientes de forma que cada salida extraiga una única fuente diferente.

En la discusión anterior se ha supuesto que todas las fuentes que llegan al array de sensores tienen un origen físico distinto. En muchas situaciones reales, sin embargo, esta suposición no se cumple ya que las ondas que se propagan son reflejadas por distintos objetos (mar, edificios, etc.) lo que ocasiona la recepción simultánea de varias versiones retardadas de las fuentes. Estas señales llegan a los sensores a través de distintos caminos e inciden en ángulos diferentes. En particular si suponemos que cada fuente llega a través de P caminos diferentes con un retardo τ_k , se obtiene que las señales proporcionadas por los sensores son

$$\mathbf{x}(t) = \sum_{k=0}^{P-1} \mathbf{A}(k)\mathbf{s}(t - \tau_k) \quad (1.23)$$

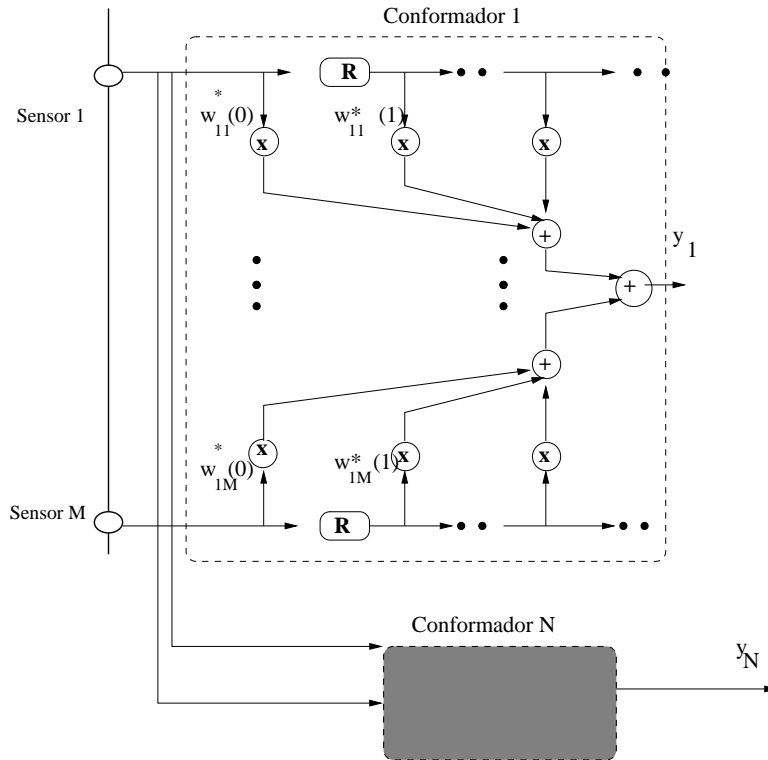


Figura 1.7: Esquema de un conformador de banda ancha.

donde cada columna de $\mathbf{A}(k)$ es el vector de steering asociado a una fuente que llega por el k -ésimo camino, $s_i(t - \tau_k)$. En este caso estamos ante un problema de mezcla convolutiva del tipo (1.13) que requiere emplear un sistema separador como el mostrado en la figura 1.7 para recuperar las fuentes.

1.3.2 Sistemas de comunicación CDMA

Los sistemas de comunicación CDMA (Code Division Multiple Access) utilizan una modulación en espectro ensanchado que consiste en multiplicar los símbolos transmitidos por ciertas secuencias pseudo-aleatorias cíclicas denominadas códigos o firmas de los usuarios. Estos códigos tienen ciertas propiedades matemáticas que permiten la recuperación de los símbolos transmitidos a partir de la suma de todas las señales recibidas [67]. Una técnica muy empleada para imprimir la firma sobre la señal de los usuarios es la conocida como DSSS (Direct-Sequence Spread-Spectrum) y consiste en cambiar la fase de la portadora L veces por cada periodo de bit, T_b , de acuerdo con el código del usuario [74]. De este modo se divide cada bit en L pulsos llamados *chips* de periodo $T_c = T_b/L$. Más concretamente, en la modulación DSSS, la señal del usuario i -ésimo tendrá la forma

$$s_i(t) = \sum_{m=0}^{\infty} A_i b_i(m) c_i(t - mT_b) \quad (1.24)$$

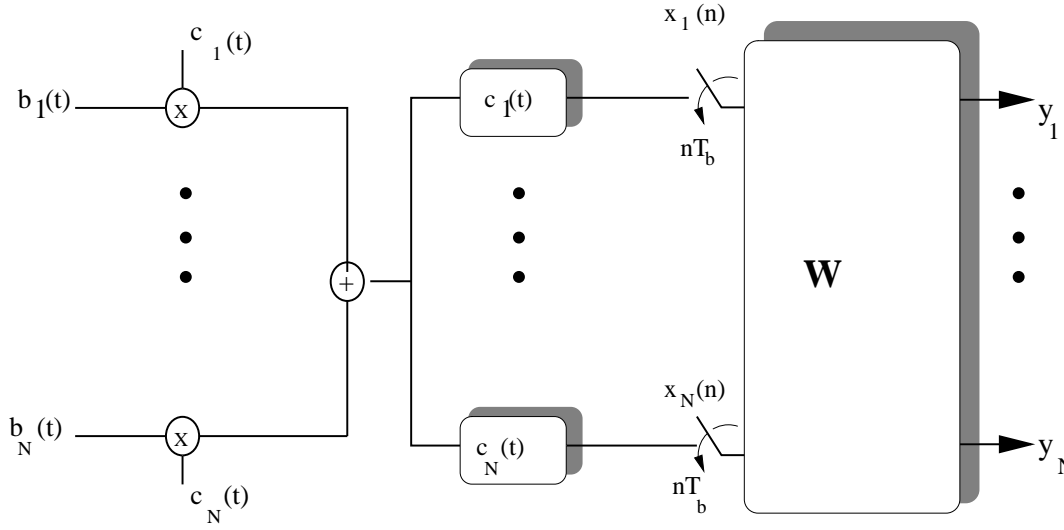


Figura 1.8: Esquema de un receptor CDMA síncrono.

donde A_i es la amplitud con la que se recibe el usuario, $\{b_i(m)\}$ representa la secuencia de símbolos transmitidos y

$$c_i(t) = \sum_{l=0}^{L-1} c_i(l)p(t - lT_b) \quad (1.25)$$

es el resultado de multiplicar la secuencia del código de usuario $\{c_i(n)\}$ por la forma del pulso para cada chip, $p(t)$. En la expresión (1.25) puede observarse que $T_b = LT_c$ y que $c_i(t)$ es cero fuera del intervalo $0 \leq t \leq T_b$.

La señal a la entrada del receptor es la suma de las señales $s_i(t)$ de los N usuarios que transmiten por el mismo canal. Suponiendo un modelo de transmisión síncrona como el mostrado en la figura 1.8 esta señal tendrá la forma

$$s(t) = \sum_{i=1}^N s_i(t) = \sum_{i=1}^N A_i \sum_{m=0}^{\infty} b_i(m)c_i(t - mT_b) \quad (1.26)$$

Si hacemos pasar $s(t)$ por un filtro adaptado al código $c_j(t)$ y, posteriormente, la muestreamos a periodo de bit se obtendrá

$$x_j(n) = \int_{nT_b}^{(n+1)T_b} c_j(t - nT_b)s(t)dt \quad (1.27)$$

que, utilizando (1.26), podemos escribir como sigue

$$\begin{aligned} x_j(n) &= \int_{nT_b}^{(n+1)T_b} \left(\sum_{i=1}^N A_i \sum_{m=0}^{\infty} b_i(m)c_i(t - mT_b) \right) c_j(t - nT_b)dt \\ &= \sum_{i=1}^N A_i \sum_{m=0}^{\infty} b_i(m) \int_{nT_b}^{(n+1)T_b} c_i(t - mT_b)c_j(t - nT_b)dt \end{aligned} \quad (1.28)$$

A su vez podemos simplificar esta expresión aprovechando que $c_i(t)$ es cero fuera del intervalo $(0, T_b)$ y obtener que

$$x_j(n) = \sum_{i=1}^N A_i b_i(n) \rho_{ij} \quad (1.29)$$

donde ρ_{ij} es el cociente de la correlación cruzada entre los códigos

$$\rho_{ij} = \int_0^{T_b} c_i(\tau) c_j(\tau) d\tau \quad (1.30)$$

Agrupando todas las observaciones proporcionadas por los N filtros en el vector $\mathbf{x} = [x_1, \dots, x_N]^T$ obtenemos el siguiente modelo de señal de un sistema CDMA síncrono

$$\mathbf{x} = \mathbf{A} \mathbf{R} \mathbf{b} \quad (1.31)$$

donde $\mathbf{b} = [b_1, \dots, b_N]^T$ son los símbolos transmitidos por los N usuarios, \mathbf{A} es una matriz diagonal que contiene las amplitudes y \mathbf{R} es la matriz de correlación de los códigos

$$\mathbf{A} = \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_N \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1N} \\ \vdots & \ddots & \vdots \\ \rho_{N1} & \cdots & \rho_{NN} \end{bmatrix} \quad (1.32)$$

De este planteamiento se obtiene que la expresión de las observaciones (1.31) es la misma que se asume en el modelo de mezcla lineal instantánea dado por (1.1). Por ello, cabe plantearse el empleo de un sistema MIMO con coeficientes \mathbf{W} y salidas $\mathbf{y} = \mathbf{W}^H \mathbf{s}$ para recuperar los símbolos transmitidos.

Los sistemas de comunicación reales son asíncronos y, por ello, la aplicabilidad del modelo que acabamos de describir resulta muy limitada. Un modelo más realista consiste en suponer que la señal recibida tiene la siguiente forma [74]

$$s(t) = \sum_{i=1}^N s_i(t - \tau_i) = \sum_{i=1}^N A_i \sum_{m=0}^{\infty} b_i(m) c_i(t - mT_b - \tau_i) \quad (1.33)$$

donde τ_i es el retardo que sufre la señal transmitida por el i -ésimo usuario. En este caso las observaciones toman la forma

$$x_j(n) = \int_{nT_b}^{(n+1)T_b} \left(\sum_{i=1}^N A_i \sum_{m=0}^{\infty} b_i(m) c_i(t - mT_b - \tau_i) \right) c_j(t - nT_b) dt \quad (1.34)$$

A diferencia del modelo síncrono, las observaciones $x_j(n)$ contendrán parte de dos símbolos del mismo usuario: la parte final del símbolo anterior y la parte inicial del símbolo actual por lo que podemos reescribir la expresión (1.34) de la siguiente forma

$$x_j(n) = \sum_{i=1}^N \left(A_i b_i(n-1) \rho_{ij}^F + A_i b_i(n) \rho_{ij}^I \right) \quad (1.35)$$

donde ρ_{ij}^F representa la correlación entre $c_j(t)$ y la parte final del código del símbolo anterior y ρ_{ij}^I representa la correlación entre $c_j(t)$ y la parte inicial del código del símbolo actual

$$\begin{aligned}\rho_{ij}^I &= \int_{\tau_i}^{T_b} c_i(t - \tau_i) c_j(t) dt \\ \rho_{ij}^F &= \int_0^{\tau_i} c_i(t + T_b - \tau_i) c_j(t) dt\end{aligned}\quad (1.36)$$

Agrupando todas las observaciones en un vector $\mathbf{x} = [x_1, \dots, x_N]^T$ obtenemos el siguiente modelo de señal de un sistema CDMA asíncrono

$$\mathbf{x}(n) = \mathbf{A}\mathbf{R}(0)\mathbf{b}(n) + \mathbf{A}\mathbf{R}(1)\mathbf{b}(n-1) \quad (1.37)$$

donde $\mathbf{b}(n-l) = [b_1(n-l), \dots, b_N(n-l)]^T$, $l = 0, 1$ es un vector que contiene los símbolos transmitidos por los N usuarios en el instante $n-l$, \mathbf{A} es una matriz diagonal que contiene las amplitudes y $\mathbf{R}(l)$ es una matriz que contiene la correlación entre los códigos

$$\mathbf{R}(0) = \begin{bmatrix} \rho_{11}^I & \cdots & \rho_{1N}^I \\ \vdots & \ddots & \vdots \\ \rho_{N1}^I & \cdots & \rho_{NN}^I \end{bmatrix} \quad \mathbf{R}(1) = \begin{bmatrix} \rho_{11}^F & \cdots & \rho_{1N}^F \\ \vdots & \ddots & \vdots \\ \rho_{N1}^F & \cdots & \rho_{NN}^F \end{bmatrix} \quad (1.38)$$

En este caso estamos ante un problema de mezcla convolutiva del tipo (1.13) con una memoria de una unidad por lo que la recuperación de las señales transmitidas requiere emplear un sistema MIMO como el descrito en la sección 1.2.

1.4 Planteamiento y estructura de la tesis

Esta tesis aborda el problema de separación de fuentes utilizando técnicas ciegas. Nos centraremos en los algoritmos adaptativos ya que, como se comentó anteriormente, tienen numerosas ventajas que los hacen especialmente atractivos para resolver este problema. Por ello, dedicaremos el capítulo 2 a la revisión de algunos conceptos relacionados con sistemas adaptativos, prestando especial interés a aquellos que nos resultan útiles a la hora de analizar la estabilidad de los algoritmos.

En la última década se han propuesto numerosos algoritmos adaptativos para resolver el problema de separación de mezclas instantáneas. Sin embargo, muy pocos trabajos analizan la estabilidad de las implementaciones que proponen y, por ello, quedan abiertas cuestiones sobre las condiciones que aseguran su correcto funcionamiento. Esta tesis pretende contribuir a llenar este vacío analizando en el capítulo 3 la estabilidad de algunos algoritmos adaptativos propuestos por otros autores.

Uno de los objetivos de esta tesis es el de proponer nuevos algoritmos adaptativos para separación de fuentes. A este respecto, en el capítulo 4 se presenta una nueva familia de criterios que han surgido a partir un método de ecualización ciega propuesto a principios

de esta década por Shalvi y Weinstein. Para implementar estos criterios se utilizarán sencillos algoritmos de gradiente. Una de las principales aportaciones de este capítulo es presentar un estudio de la estabilidad de uno de los algoritmos considerando el caso general de entornos con N fuentes y sistemas separadores $N \times N$.

Los criterios propuestos en el capítulo 4 serán extendidos en el capítulo 5 para separación de mezclas convolutivas. Además, se propondrán varios algoritmos de gradiente para encontrar los coeficientes óptimos del sistema separador y se analizará la estabilidad de uno de ellos.

Por último, en el capítulo 6 se exponen las conclusiones más importantes de esta tesis y se proponen líneas futuras de investigación.

Capítulo 2

Fundamentos teóricos de algoritmos adaptativos

2.1 Introducción

El problema de separación ciega de fuentes es un caso particular de problema de filtrado cuyo objetivo es seleccionar los parámetros de un sistema MIMO que permita la recuperación de un conjunto de señales de interés a partir de mezclas lineales de ellas. Según la naturaleza del criterio seguido para encontrar el sistema de separación, las técnicas de filtrado se pueden clasificar en dos grandes grupos: deterministas y estadísticas. Un ejemplo de aproximación determinista a nuestro problema sería el utilizar como sistema de separación el sistema inverso al de mezcla. Esta aproximación, sin embargo, puede no resultar adecuada si existe ruido en las observaciones y, lo que es más importante, requiere un conocimiento exacto del sistema de mezcla, lo cual contradice nuestras hipótesis. Por ello, en el presente trabajo nos centraremos exclusivamente en aproximaciones estadísticas que no necesitan conocer el sistema de mezcla y que, además, tienen la capacidad de hacer seguimiento de sus variaciones temporales.

A su vez, las técnicas de filtrado estadístico pueden subdividirse en otros dos grupos: bloque y adaptativas. Las técnicas bloque calculan el sistema de separación de una forma directa a partir de un bloque de datos de entrada mientras que las adaptativas lo hacen de forma recursiva, lo que les permite aprovechar los resultados obtenidos al procesar bloques anteriores. Ese carácter recursivo hace que las técnicas adaptativas funcionen adecuadamente procesando bloques de datos mucho más pequeños que en las otras técnicas (típicamente una muestra) y, como consecuencia, sus implementaciones necesitan menos memoria, menos coste computacional y proporcionan los resultados con un retardo menor. No todo son ventajas puesto que en la práctica las técnicas adaptativas son más lentas y más complejas de analizar que las bloque. A pesar de ello, en el presente trabajo hemos optado por centrarnos en las técnicas adaptativas.

El empleo de algoritmos adaptativos para actualizar la matriz de separación nos

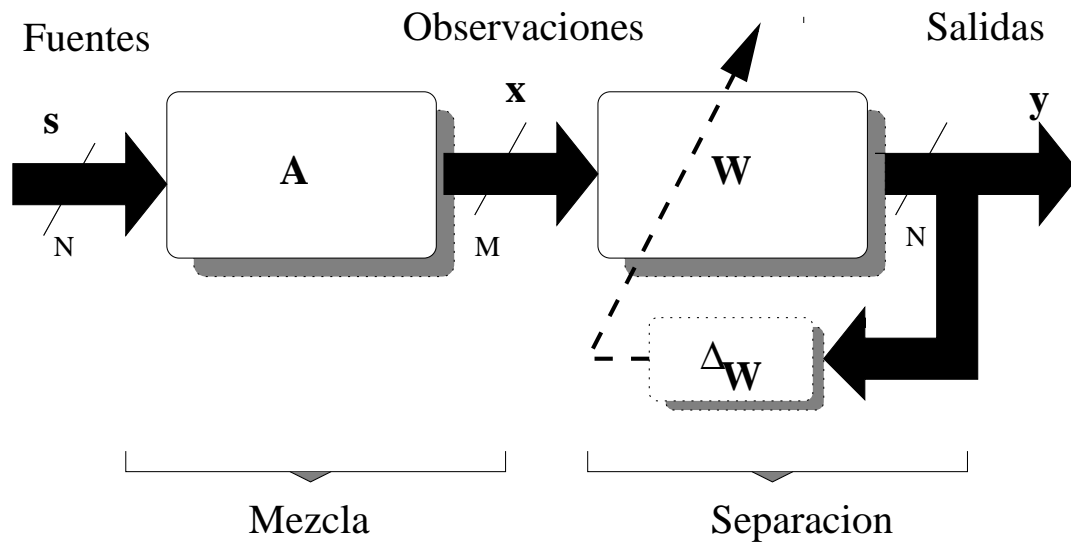


Figura 2.1: Esquema de un sistema adaptativo para separación de señales.

conduce a interpretar el sistema separador como un sistema dinámico regido por una ecuación diferencial ordinaria (ODE, del inglés Ordinary Differential Equation). Como se verá en este capítulo, esta interpretación es de gran utilidad ya que, bajo algunas condiciones poco restrictivas, permite determinar los puntos a los que convergerá el algoritmo (*atractores*) estudiando la ODE y la naturaleza de su jacobiano. El correcto funcionamiento del sistema separador estará asegurado cuando los únicos atractores se correspondan con la perfecta separación de las fuentes. Por el contrario, la existencia de atractores indeseados puede conducir a situaciones donde se deje de extraer alguna fuente o se extraigan combinaciones lineales de ellas.

Este capítulo está estructurado como sigue. La sección 2.2 presenta un enfoque adaptativo al problema de separación. Además recoge conceptos de Teoría de Sistemas Dinámicos que serán utilizados en los capítulos posteriores para estudiar la estabilidad de distintos algoritmos adaptativos. La sección 2.3 está dedicada por completo a los algoritmos de gradiente. La sección 2.4 presenta el empleo del criterio de Mínimo Error Cuadrático Medio como una solución no-ciega al problema de separación. Finalmente, la sección 2.5 recoge las conclusiones más importantes de este capítulo.

2.2 Sistemas adaptativos

La figura 2.1 muestra el esquema básico de un sistema adaptativo para separar mezclas lineales de señales. Inicialmente consideraremos que las mezclas son instantáneas ya que los conceptos que a continuación se van a presentar se pueden generalizar de forma inmediata a mezclas convolutivas. La entrada es el vector de observaciones \mathbf{x} de dimensión $M \times 1$ y la salida es representada por otro vector \mathbf{y} de dimensión $N \times 1$. La matriz \mathbf{W}

de dimensión $M \times N$ contiene los coeficientes libres del sistema de separación. En las aproximaciones adaptativas, este sistema es ajustado utilizando un algoritmo recursivo de la forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \Delta(\mathbf{W}(n), \mathbf{x}(n)) \quad (2.1)$$

donde μ es un valor real que permite controlar la velocidad de convergencia del algoritmo y $\Delta(., .)$ es una matriz formada, en general, por términos no lineales que dependen de \mathbf{W} y del vector de entrada \mathbf{x} en el instante n . La forma exacta de la matriz $\Delta(., .)$ viene determinada por el criterio de separación que se utilice.

Los algoritmos adaptativos de la forma (2.1) son especialmente atractivos para aplicaciones en tiempo real [51] porque suministran de forma inmediata, sin ningún retardo, la salida correspondiente a cada entrada recibida y porque en cada iteración sólo se utiliza la entrada en el instante n , lo cual evita almacenar el valor de las observaciones en instantes anteriores.

Es importante hacer notar la diferencia entre un algoritmo adaptativo de la forma (2.1) y un algoritmo bloque que procese de forma recursiva una colección finita de datos de entrada para calcular el sistema de separación [43, 51]. Por cada colección $\mathbf{x}(1), \dots, \mathbf{x}(N)$ este tipo de algoritmos realizan la siguiente iteración un número $S \neq N$ de veces

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \Delta(\mathbf{W}(n), \mathbf{x}(1), \dots, \mathbf{x}(N)), \quad n = 1, \dots, S \quad (2.2)$$

El valor final de los coeficientes será $\mathbf{W} = \mathbf{W}(S)$. Esta aproximación tiene varias ventajas sobre el algoritmo adaptativo (2.1): utiliza la mayor información disponible del entorno contenida en la entrada del sistema ya que procesa una colección de N muestras y permite realizar operaciones sobre la colección de entradas (por ejemplo, permutar el orden temporal) para adaptarlas al sistema. Sin embargo, la salida se obtiene con un cierto retardo y, además, su implementación requiere tener almacenados N valores de las observaciones \mathbf{x} .

A su vez, tanto los algoritmos adaptativos como los bloques pueden ser deterministas o estocásticos. En una aproximación determinista la actualización de los coeficientes \mathbf{W} se realiza utilizando promedios estadísticos que se suponen conocidos. Sin embargo, en la práctica tales promedios son desconocidos para sistema separador por lo que deben ser estimados a partir de los datos disponibles. Los algoritmos que resultan de emplear tales estimaciones reciben el calificativo de estocásticos. Para evitar confusiones, en lo que sigue utilizaremos $\Delta(., .)$ cuando nos refiramos a los algoritmos estocásticos y $\bar{\Delta}(., .)$ para los algoritmos deterministas.

2.2.1 Conexión con los sistemas dinámicos

Los sistemas adaptativos que acabamos de describir se pueden interpretar como sistemas dinámicos ya que los coeficientes \mathbf{W} evolucionan con el tiempo. En concreto, en separación

de fuentes nos encontramos con $M \times N$ parámetros w_{ij} que, durante la fase de adaptación, evolucionan con el tiempo según un conjunto de $M \times N$ ecuaciones diferenciales ordinarias

$$\frac{dw_{ij}}{dt} = F_{ij}(\mathbf{W}), \quad i = 1, \dots, N, \quad j = 1, \dots, M \quad (2.3)$$

las cuales podemos escribir en forma más compacta como

$$\frac{d\mathbf{W}}{dt} = \mathbf{F}(\mathbf{W}) \quad (2.4)$$

donde $\mathbf{F}(\cdot)$ es ahora una matriz que contiene a las funciones $F_{ij}(\mathbf{W})$, $i = 1, \dots, N$, $j = 1, \dots, M$. Las matrices \mathbf{W}_o que anulan (2.4) se denominan puntos de equilibrio

$$\mathbf{F}(\mathbf{W}_o) = \mathbf{0} \iff F_{ij}(\mathbf{W}_o) = 0, \quad i = 1, \dots, N \quad j = 1, \dots, M \quad (2.5)$$

donde $\mathbf{0}$ es una matriz todo ceros de dimensión $N \times M$. Estos puntos juegan un papel fundamental en el estudio de la estabilidad de los sistemas dinámicos por lo que conviene clasificarlos en tres grupos [42]:

- *Puntos estables:* Un punto de equilibrio \mathbf{W}_o se dice que es estable si para todo entorno U de \mathbf{W}_o existe un entorno U_o contenido en U tal que toda solución $\mathbf{W}(t)$ con $\mathbf{W}(0)$ en U_o permanece en U para todo $t > 0$ como se muestra en la figura 2.2 (A).
- *Puntos asintóticamente estables o atractores:* Un punto de equilibrio estable \mathbf{W}_o es asintóticamente estable cuando los puntos $\mathbf{W}(t)$ en un entorno U_o verifican $\lim_{t \rightarrow \infty} \mathbf{W}(t) = \mathbf{W}_o$ (ver figura 2.2 (B)).
- *Puntos inestables:* Un punto de equilibrio \mathbf{W}_o que no es estable se llama inestable. Esto significa que existe un entorno $U_o \in U$ de \mathbf{W}_o tal que al menos una solución $\mathbf{W}(t)$ que comienza en $\mathbf{W}(0) \in U_o$ no permanece enteramente en U como se muestra en la figura 2.2 (C).

Según esta clasificación, los puntos asintóticamente estables son los que tienen un verdadero interés práctico ya que representan los estados en los que permanecerá el sistema dinámico una vez los haya alcanzado.

Los puntos estacionarios pueden ser asignados a cada uno de los grupos de estabilidad descritos anteriormente atendiendo al signo de los autovalores de la matriz jacobiana, \mathbf{H} , formada por las primeras derivadas de $\mathbf{F}(\mathbf{W})$ [42]:

- *Puntos estables:* los autovalores de \mathbf{H} son imaginarios puros. Este punto tiene la propiedad de que todas las trayectorias son periódicas como se muestra en la figura 2.3 (A) por lo que también se les conoce como centros.

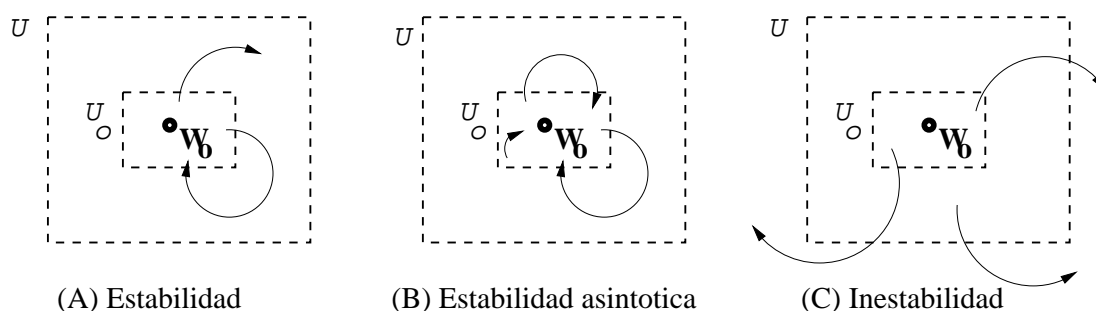


Figura 2.2: Estabilidad de un sistema dinámico.

- *Puntos asintóticamente estables*: todos los autovalores tienen parte real negativa. En este caso el punto de equilibrio es asintóticamente estable ya que las trayectorias tienden hacia él. Dependiendo de los autovalores de la matriz \mathbf{H} se tendrán las distintas trayectorias mostradas en la figura 2.3 (B). Estos puntos también son conocidos como *sumideros*.
- *Puntos inestables*: se pueden dar las dos situaciones siguientes
 - Todos los autovalores de \mathbf{H} tienen parte real positiva. Las trayectorias son opuestas a la de un sumidero como se muestra en la figura 2.3 (C), por lo que el punto estacionario es inestable. A estos puntos se les denomina *fuentes*.
 - La matriz \mathbf{H} tiene autovalores reales de signos opuestos. Como se ilustra en la figura 2.3 (D) existen trayectorias entrantes y salientes. A estos puntos también se les conoce como puntos silla o de ensilladura.

Así pues, atendiendo a esta clasificación podemos establecer que la estabilidad asintótica se consigue sólo en aquellos puntos donde los autovalores del jacobiano tienen parte real negativa.

Método de Liapunov

La clasificación de los puntos estacionarios atendiendo al signo de los autovalores del jacobiano puede resultar un trabajo difícil ya que requiere estudiar la matriz jacobiana. Por ello, en algunos casos se recurre a un método alternativo propuesto por Liapunov en su tesis doctoral de 1892 [42]. Se trata de una generalización de la idea de que para un sumidero \mathbf{W}_o existe una norma, $|\cdot|$, tal que $|\mathbf{W}(t) - \mathbf{W}_o|$ decrece para las soluciones $\mathbf{W}(t)$ cercanas a \mathbf{W}_o . Liapunov propuso construir funciones continuas $V(\mathbf{W})$, definidas en un entorno del punto estacionario, cuyo mínimo sea \mathbf{W}_o .

Una función de Liapunov $V : U \rightarrow \mathcal{W}$ debe cumplir las siguientes condiciones:

- $V(\mathbf{W})$ tiene derivadas parciales continuas con respecto a los elementos de \mathbf{W} .

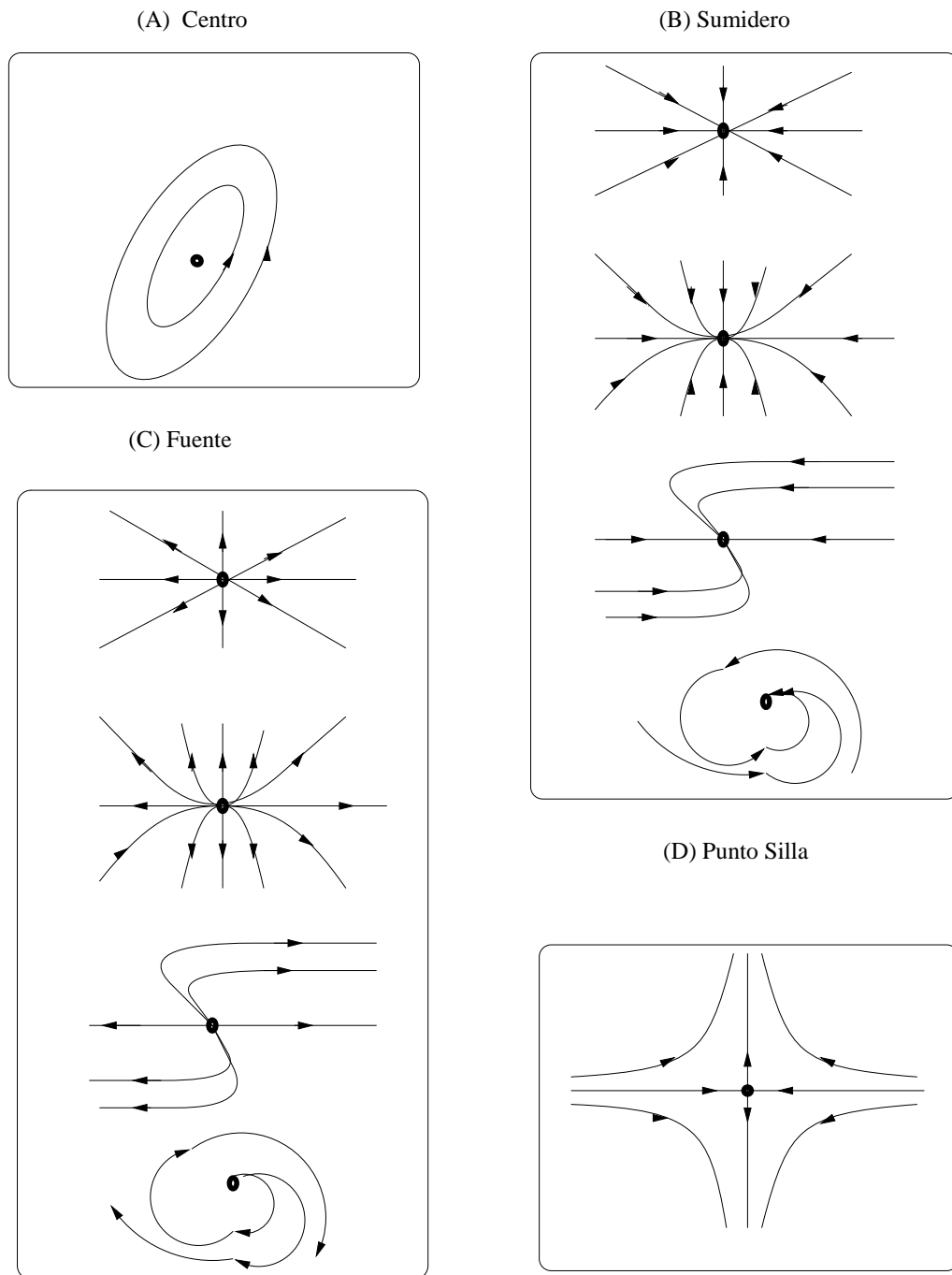


Figura 2.3: Clasificación de los puntos estacionarios.

- $V(\mathbf{W}_o) = 0$.
- $V(\mathbf{W}) > 0$ si $\mathbf{W} \neq \mathbf{W}_o$.

De esta forma tenemos asegurado que \mathbf{W}_o es el mínimo de $V(\cdot)$ pero para garantizar la estabilidad también es necesario que $V(\cdot)$ decrezca a lo largo de la solución (2.4) que pasa por \mathbf{W} de forma que tienda hacia \mathbf{W}_o . Para ello es necesario recurrir a condiciones adicionales establecidas sobre la primera derivada de $V(\cdot)$

$$V'(\mathbf{W}) = \frac{d}{dt}V(\mathbf{W}) = \sum_{i=1}^N \sum_{j=1}^M \frac{dV}{dw_{ij}} \frac{dw_{ij}}{dt} \quad (2.6)$$

Si $V'(\mathbf{W}) < 0$ para todo punto $\mathbf{W} \neq \mathbf{W}_o$ entonces el punto es asintóticamente estable ya que las trayectorias conducen hacia él. Por otro lado, si $V'(\mathbf{W}) \leq 0$ el sistema no presenta una estabilidad asintótica ya que puede llegarse a un punto en el entorno de \mathbf{W}_o donde $V(\mathbf{W})$ es constante y, por tanto, la trayectoria no permite alcanzar \mathbf{W}_o .

El estudio de la estabilidad utilizando el teorema de Liapunov parece más simple que el de analizar los autovalores de la matriz jacobiana \mathbf{H} . Sin embargo, tenemos que destacar varias limitaciones. En primer lugar, no existe un método constructivo para encontrar funciones de Liapunov; es una cuestión de ingenio y de “prueba y error” según cada caso. Adicionalmente, el criterio nos habla del estudio en las cercanías del punto de equilibrio, lo cual no descarta que existan soluciones indeseadas fuera de ese entorno. Por estas razones, el método que preferentemente se utilizará en esta tesis a la hora de analizar la estabilidad de los algoritmos adaptativos será el examen de los autovalores de la matriz jacobiana.

2.2.2 Estabilidad de los algoritmos adaptativos

La estabilidad de los algoritmos adaptativos tanto deterministas como estocásticos puede analizarse utilizando los conceptos de sistemas dinámicos que acabamos de exponer. Comencemos considerando un algoritmo determinista de la forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \bar{\Delta}(\mathbf{W}(n)) \quad (2.7)$$

Esta expresión nos indica que las trayectorias que sigue el sistema de separación están regidas por la siguiente ODE

$$\frac{d\mathbf{W}}{dt} = \bar{\Delta}(\mathbf{W}) \quad (2.8)$$

De la teoría de sistemas dinámicos sabemos que los puntos estacionarios son aquellas matrices \mathbf{W}_o que anulan (2.8), i.e., $\bar{\Delta}(\mathbf{W}_o) = \mathbf{0}$. Estos puntos serán atractores solamente si todos los autovalores del jacobiano, \mathbf{H} , tiene parte real negativa.

Abordemos ahora el estudio de la estabilidad de los algoritmos estocásticos de la forma (2.1) que repetimos aquí por conveniencia

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \Delta(\mathbf{W}(n), \mathbf{x}(n)) \quad (2.9)$$

En estos algoritmos la matriz $\Delta(\cdot, \cdot)$ depende del valor de los coeficientes $\mathbf{W}(n)$ y de la entrada del sistema separador $\mathbf{x}(n)$. Realizar un estudio riguroso de la estabilidad de (2.9) es muy difícil debido a su carácter estadístico y, por ello, se recurre al estudio de un algoritmo determinista obtenido considerando que $\Delta(\cdot, \cdot)$ es una función bien condicionada y que el parámetro μ es suficientemente pequeño [6]. Bajo estos supuestos las variaciones de $\Delta(\cdot, \cdot)$ en (2.9) serán muy suaves y podemos escribir el valor de los coeficientes \mathbf{W} en un instante $n+N$ como sigue

$$\begin{aligned} \mathbf{W}(n+N) &= \mathbf{W}(n) + \mu \sum_{i=0}^{N-1} \Delta(\mathbf{W}(n+i), \mathbf{x}(n+i)) \\ &\approx \mathbf{W}(n) + \mu \sum_{i=0}^{N-1} \Delta(\mathbf{W}(n), \mathbf{x}(n+i)) \\ &= \mathbf{W}(n) + N\mu \frac{1}{N} \sum_{i=0}^{N-1} \Delta(\mathbf{W}(n), \mathbf{x}(n+i)) \end{aligned} \quad (2.10)$$

Si además N es un valor grande y las observaciones $\mathbf{x}(n)$ son procesos aleatorios estacionarios, entonces el segundo término a la derecha de (2.10) puede ser considerado como la esperanza respecto a $\mathbf{x}(n)$ para un determinado valor $\mathbf{W}(n)$ que denotaremos $\bar{\Delta}(\mathbf{W}(n)) = E[\Delta(\mathbf{W}(n), \mathbf{x}(n))]$. De esta forma se obtiene la expresión dada en (2.7).

La discusión anterior nos lleva a concluir que la elección de un parámetro de paso μ suficientemente pequeño asegura que los atractores de un algoritmo de gradiente estocástico se corresponden con los atractores de un algoritmo determinista.

2.3 Algoritmos de gradiente

Muchos de los criterios estadísticos que se plantean para resolver el problema de separación ciega de fuentes consisten en la minimización de una función de coste J de los coeficientes \mathbf{W} del sistema de separación. Los coeficientes óptimos se pueden calcular, entonces, utilizando un algoritmo determinista de gradiente descendente de la forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \nabla_{\mathbf{W}} J \quad (2.11)$$

donde $\nabla_{\mathbf{W}} J$ es una matriz de dimensiones $M \times N$ que representa al gradiente de J respecto a \mathbf{W} y μ es un parámetro que permite controlar la velocidad de convergencia. En efecto, el desarrollo en serie de Taylor de primer orden de J en torno a un punto \mathbf{W} se puede escribir como

$$J(\mathbf{W} + \mathcal{E}) \approx J(\mathbf{W}) + \text{tr} \left(\nabla_{\mathbf{W}}^T J \mathcal{E} \right) \quad (2.12)$$

donde \mathcal{E} es una matriz de dimensión $M \times N$ que contiene pequeñas desviaciones de \mathbf{W} y $\text{tr}(\cdot)$ denota la traza¹ de una matriz.

Ahora bien, si μ es lo suficientemente pequeño el desarrollo (2.12) será válido cuando $\mathcal{E} = -\mu \nabla_{\mathbf{W}} J$ [28] con lo que se obtiene que

$$J(\mathbf{W}(n+1)) \approx J(\mathbf{W}(n)) - \mu \text{tr} \left(\nabla_{\mathbf{W}}^T J \nabla_{\mathbf{W}} J \right) \quad (2.13)$$

Como $\text{tr} \left(\nabla_{\mathbf{W}}^T J \nabla_{\mathbf{W}} J \right)$ es siempre una cantidad positiva entonces $J(\mathbf{W}(n+1)) < J(\mathbf{W}(n))$ lo cual indica que la función de coste disminuye en cada iteración.

En general, la expresión del gradiente involucrará estadísticos que deberán ser estimados a partir de los datos disponibles. Por ello, en la práctica se trabaja con una versión estocástica de (2.11) que resulta de estimar el gradiente de la función de coste, $\widehat{\nabla_{\mathbf{W}} J}(n)$

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \widehat{\nabla_{\mathbf{W}} J}(n) \quad (2.14)$$

2.3.1 Algoritmos de gradiente relativo

Una forma alternativa de minimizar una función de coste es utilizar un algoritmo de gradiente relativo o natural. Este tipo de algoritmos fue simultáneamente propuesto por Cardoso y Amari [1, 13] para separación ciega de fuentes. En este caso, la minimización de la función de coste se realiza mediante la siguiente recursión

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \left(\mathbf{W}(n) \mathbf{W}^T(n) \widehat{\nabla_{\mathbf{W}} J}(n) \right) \quad (2.15)$$

que simplemente consiste en premultiplicar el gradiente por $\mathbf{W}(n) \mathbf{W}^T(n)$. Para comprobar que, en efecto, el algoritmo (2.15) lleva a cabo la minimización de $J(\mathbf{W})$ supondremos que μ es lo suficientemente pequeño. En este caso, el desarrollo de Taylor (2.12) también será válido cuando $\mathcal{E} = -\mu \mathbf{W} \mathbf{W}^T \nabla_{\mathbf{W}} J$ y, como consecuencia, se verificará lo siguiente

$$\text{tr} \left(\nabla_{\mathbf{W}}^T J \mathcal{E} \right) = -\mu \text{tr} \left(\nabla_{\mathbf{W}}^T J \mathbf{W} \mathbf{W}^T \nabla_{\mathbf{W}} J \right) = -\mu \sum_{i=1}^N \sum_{j=1}^N \left(\nabla_{\mathbf{w}_i}^T J \mathbf{w}_j \right)^2 < 0 \quad (2.16)$$

y, por tanto, la función de coste disminuye en cada iteración.

Aunque el algoritmo de máxima pendiente es el que tradicionalmente se ha utilizado para implementar los criterios de separación de fuentes [14, 51, 54], recientemente distintos autores [1, 13, 53] han optado por la versión relativa. La razón principal de este cambio se debe a que el comportamiento de (2.15) es independiente de los parámetros que intervengan en la mezcla y, por ello, presenta un buen comportamiento en situaciones donde la matriz de mezcla está mal condicionada. Esta propiedad es consecuencia de

¹La traza de una matriz se define como la suma de los elementos de su diagonal.

la relación que existe entre las salidas del sistema de separación y las observaciones $\mathbf{y} = \mathbf{W}^T \mathbf{x}$, la cual nos permite (utilizando la regla de la cadena) escribir el gradiente como sigue

$$\nabla_{\mathbf{W}} J = \mathbf{x} \nabla_{\mathbf{y}}^T J \quad (2.17)$$

donde $\nabla_{\mathbf{y}} J$ es ahora un vector que contiene las derivadas de J respecto al vector \mathbf{y} . Utilizando (2.17) el algoritmo de gradiente relativo estocástico (2.15) se transforma en

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) - \mu \mathbf{W}(n) \mathbf{W}^T(n) \mathbf{x}(n) \widehat{\nabla_{\mathbf{y}}^T} J(n) \\ &= \mathbf{W}(n) - \mu \mathbf{W}(n) \mathbf{y}(n) \widehat{\nabla_{\mathbf{y}}^T} J(n) \end{aligned} \quad (2.18)$$

Obsérvese ahora que al multiplicar (2.18) por la matriz de mezcla traspuesta \mathbf{A}^T se tiene que la evolución de la matriz $\mathbf{G} = \mathbf{W}^T \mathbf{A}$ es independiente de los parámetros de la mezcla y viene dada por

$$\mathbf{G}^T(n+1) = \mathbf{G}^T(n) - \mu \mathbf{G}^T(n) \mathbf{y}(n) \widehat{\nabla_{\mathbf{y}}^T} J(n) \quad (2.19)$$

Supongamos ahora que las fuentes \mathbf{s} son mezcladas en dos entornos diferentes modelados por las matrices \mathbf{A}_1 y \mathbf{A}_2 . Las observaciones son

$$\mathbf{x}_1 = \mathbf{A}_1 \mathbf{s}, \quad \mathbf{x}_2 = \mathbf{A}_2 \mathbf{s} \quad (2.20)$$

Para recuperar las fuentes, un sistema con coeficientes \mathbf{W}_1 procesa \mathbf{x}_1 y otro sistema \mathbf{W}_2 procesa \mathbf{x}_2 . La evolución de las ganancias de ambos sistemas será la misma si $\mathbf{G}_1(n)$ es igual que $\mathbf{G}_2(n)$ para todo n , i.e.,

$$\mathbf{W}_1^T(n) \mathbf{A}_1 = \mathbf{W}_2^T(n) \mathbf{A}_2 \quad (2.21)$$

Es inmediato comprobar que esto ocurrirá si inicializamos un sistema de separación en $\mathbf{W}_1(0)$ y el otro en $\mathbf{W}_2(0) = (\mathbf{W}_1^T(0) \mathbf{A}_1 \mathbf{A}_2^{-1})^T$.

Existen otros muchos algoritmos de tipo gradiente con velocidades de convergencia superiores a las de los que hemos presentado aquí pero cuya implementación suele ser computacionalmente más costosa. Una revisión de dichos algoritmos puede ser consultada en [28, 63].

2.3.2 Estabilidad de los algoritmos de gradiente

En el caso particular de utilizar algoritmos de gradiente de máxima pendiente, la ODE asociada a (2.14) es

$$\frac{d \mathbf{W}}{dt} = \bar{\Delta}(\mathbf{W}) = -\nabla_{\mathbf{W}} J \quad (2.22)$$

Los puntos estacionarios pueden ser determinados resolviendo el sistema de ecuaciones $\nabla_{\mathbf{w}}J = \mathbf{0}$ y los autovalores de una matriz \mathbf{H} formada por las segundas derivadas de J nos indicarán si son mínimos (sumideros), máximos (fuentes) o puntos silla. Ahora el jacobiano \mathbf{H} es una matriz formada por las segundas derivadas de J respecto a \mathbf{W} que recibe el nombre de hessiano. Esta matriz es simétrica por su propia definición y, por tanto, todos sus autovalores son siempre reales lo que permite clasificar los puntos estacionarios estudiando el carácter definido positivo de \mathbf{H} . Un método muy útil consiste en calcular los determinantes de todas las matrices superiores izquierdas: el punto será un máximo si alternan su signo (comenzando en negativo), mínimo si son positivos y un punto silla en otro caso. Es importante notar que si se emplea un algoritmo de gradiente ascendente entonces los puntos asintóticamente estables son aquellos puntos estacionarios en los que \mathbf{H} es definida negativa, i.e., son los máximos de la función de coste.

2.4 El criterio de Mínimo Error Cuadrático Medio

El criterio más utilizado en filtrado adaptativo es sin duda el de Mínimo Error Cuadrático Medio (MMSE, del inglés Minimum Mean Square Error) por sus numerosas ventajas tanto teóricas como prácticas. En el caso de separación de señales, la aplicación de este criterio significa ajustar el sistema de separación con el objetivo de minimizar el error cuadrático medio entre la salida \mathbf{y} y las señales que se desean recuperar \mathbf{s} [41, 40].

Matemáticamente, la función de coste que se persigue minimizar tiene la siguiente forma

$$J_{MSE}(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^N E[(y_i - s_i)^2] = \frac{1}{2} E \left[\text{tr} \left((\mathbf{W}^T \mathbf{x} - \mathbf{s})(\mathbf{W}^T \mathbf{x} - \mathbf{s})^T \right) \right] \quad (2.23)$$

La minimización de esta función de coste puede realizarse de forma sencilla utilizando un algoritmo de gradiente descendente que actualiza los pesos en la dirección opuesta al gradiente de (2.23). A partir de las propiedades presentadas en el apéndice A puede obtenerse fácilmente la siguiente expresión de dicho gradiente

$$\nabla_{\mathbf{w}} J_{MSE} = E[\mathbf{x}(\mathbf{W}^T \mathbf{x} - \mathbf{s})^T] \quad (2.24)$$

La esperanza que aparece en $\nabla_{\mathbf{w}} J_{MSE}$ es en general desconocida y debe ser estimada a partir de los datos disponibles. Estimándolo con una sola muestra obtenemos la expresión del algoritmo LMS (Least Mean Squares), también conocido en el contexto de Redes Neuronales Artificiales como regla delta de Widrow-Hopf [40]

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \mathbf{x}(n)(\mathbf{W}^T(n)\mathbf{x}(n) - \mathbf{s}(n))^T \quad (2.25)$$

Como se muestra en la figura 2.4, la actualización de los coeficientes de los combinadores se realiza de forma local a partir de la señal de referencia s_i y del valor de la salida y_i .

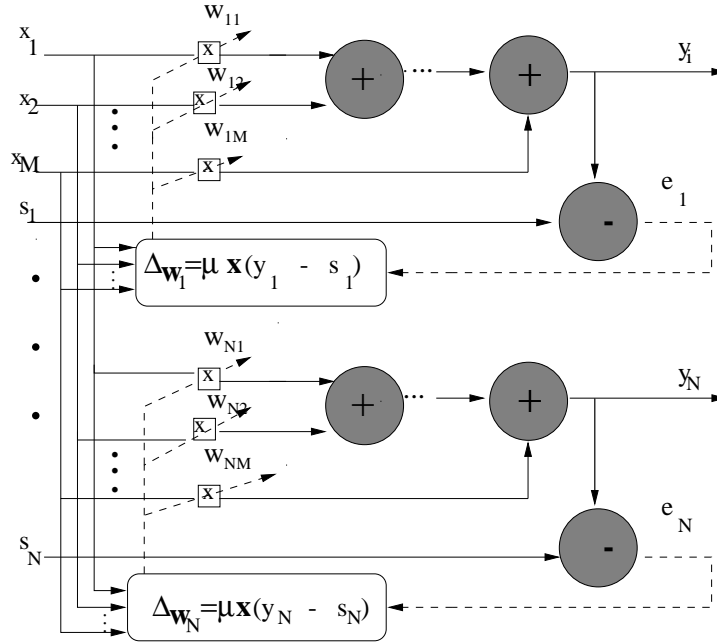


Figura 2.4: Ajuste del sistema de separación utilizando el algoritmo LMS.

Como consecuencia, el orden en que se recuperan las fuentes queda establecida por la señal de referencia empleada para calcular la función de error en cada salida.

De la teoría de sistemas dinámicos sabemos que la recursión (2.25) convergerá a las raíces de la ODE

$$\frac{d\mathbf{W}}{dt} = -E[\mathbf{x}(\mathbf{W}^T \mathbf{x} - \mathbf{s})^T] = -E[\mathbf{x}\mathbf{x}^T] \mathbf{W} + E[\mathbf{x}\mathbf{s}^T] = \mathbf{0} \quad (2.26)$$

Calculando las raíces \mathbf{W}_o de (2.26) obtenemos que el único punto estacionario del algoritmo LMS viene dado por

$$\mathbf{W}_o = \mathbf{R}_x^{-1} \mathbf{P} \quad (2.27)$$

donde $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^T]$ es la matriz de autocorrelación de las observaciones y $\mathbf{P} = E[\mathbf{x}\mathbf{s}^T]$ es la correlación entre las observaciones y las señales de referencia. Esta solución es bien conocida porque corresponde al filtro de Wiener que tanta importancia tiene en filtrado adaptativo [70]. Adicionalmente, estudiaremos el hessiano de J_{MSE} para determinar si el punto (2.27) es estable. A partir de (2.26) puede obtenerse fácilmente que la matriz hessiana formada por las segundas derivadas de J_{MSE} es

$$\mathbf{H} = \begin{bmatrix} \mathbf{R}_x & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{R}_x \end{bmatrix} \quad (2.28)$$

Dado que \mathbf{R}_x es una matriz definida positiva, también lo será \mathbf{H} y, por tanto, el algoritmo converge asintóticamente hacia el punto (2.27).

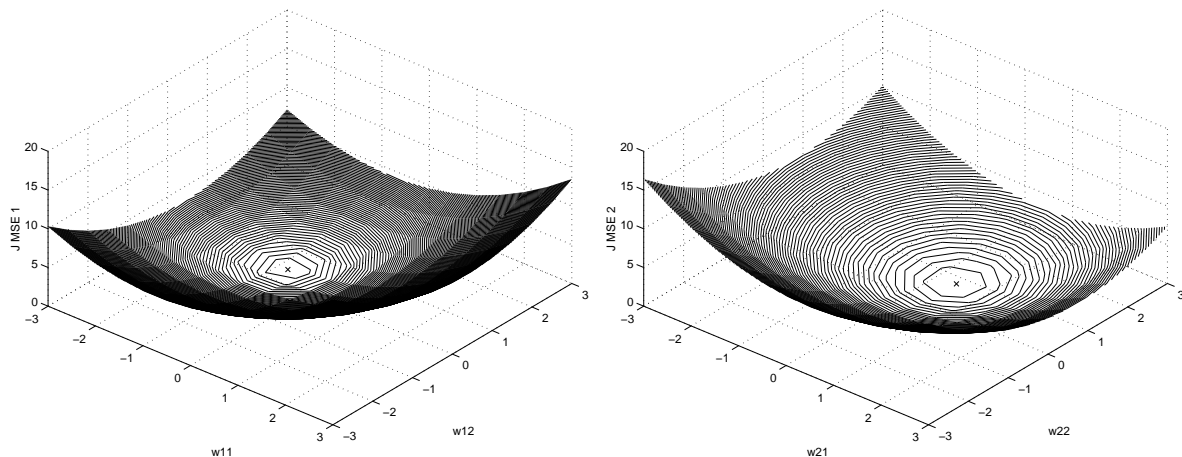


Figura 2.5: Funciones de error cuadrático medio.

Para ilustrar este resultado consideraremos la separación de dos fuentes estadísticamente independientes empleando un sistema con dos entradas y dos salidas. Supondremos que las fuentes son señales binarias y el sistema de mezcla es

$$\mathbf{A} = \begin{bmatrix} -1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (2.29)$$

En la figura 2.5 se ha representado las funciones de error $J_{MSE_i} = E[(\mathbf{w}_i^T \mathbf{x} - s_i)^2]/2$ para las dos salidas. Como puede observarse cada una de ellas tiene un único punto estacionario (mínimo) en $\mathbf{w}_{o1} = [-0.8, 0.4]^T$ y $\mathbf{w}_{o2} = [0.4, 0.8]^T$. Es sencillo de demostrar que este punto se corresponde con el atractor del algoritmo LMS dado por (2.27)

$$\mathbf{W}_o = \mathbf{R}_x^{-1} \mathbf{P} = (\mathbf{A} \mathbf{R}_s \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{R}_s) = (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} = \begin{bmatrix} -0.8 & 0.4 \\ 0.4 & 0.8 \end{bmatrix} \quad (2.30)$$

Además, es interesante notar que el punto (2.30) da lugar a una matriz de ganancias igual a la identidad $\mathbf{G} = \mathbf{W}_o^T \mathbf{A} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} = \mathbf{I}$ lo que se traduce en $y_i = s_i$. Como consecuencia, el algoritmo LMS no presenta las ambigüedades inherentes a la separación de las fuentes ya que el sistema de separación no modifica las amplitudes de las fuentes y, además, el orden en que las extrae viene determinado por las señales de referencia suministradas al sistema separador.

Del planteamiento anterior podemos concluir que el algoritmo LMS es muy sencillo de implementar y de analizar pero, en contrapartida, se deben conocer las fuentes por lo que su empleo es muy limitado en la práctica. Esta limitación puede evitarse utilizando el algoritmo que se deriva a continuación y que surge de forma inmediata a partir ecuación diferencial ordinaria (2.26). Comenzaremos notando que al ser las fuentes estadísticamente independientes con potencia igual a la unidad, $\mathbf{R}_s = \mathbf{I}$, podemos expresar

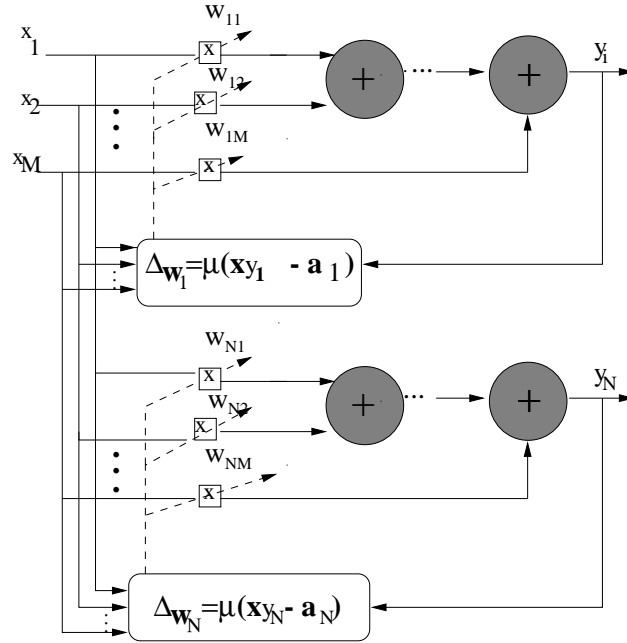


Figura 2.6: Ajuste del sistema de separación utilizando el algoritmo p-vector.

la ecuación (2.26) como sigue

$$\begin{aligned}
 \frac{d\mathbf{W}}{dt} &= -E[\mathbf{x}(\mathbf{W}^T \mathbf{x} - \mathbf{s})^T] = -E[\mathbf{x}(\mathbf{y} - \mathbf{s})^T] = -E[\mathbf{x}\mathbf{y}^T] + E[\mathbf{x}\mathbf{s}^T] \\
 &= -E[\mathbf{x}\mathbf{y}^T] + \mathbf{A}E[\mathbf{s}\mathbf{s}^T] = -E[\mathbf{x}\mathbf{y}^T] + \mathbf{A}
 \end{aligned} \tag{2.31}$$

Estimando la esperanza con una única muestra obtenemos el siguiente algoritmo

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu(\mathbf{x}(n)\mathbf{y}^T(n) - \mathbf{A}) \tag{2.32}$$

que puede interpretarse como una generalización al caso multicanal del algoritmo p-vector descrito en [39]. Como se muestra en la figura 2.6, la actualización de los coeficientes del sistema separador se realiza de forma local ya que el combinador i -ésimo sólo requiere conocer su propia salida y la columna \mathbf{a}_i de la matriz de mezcla \mathbf{A} . Desde un punto de vista práctico el empleo del algoritmo (2.32) es muy limitado ya que, en general, los parámetros de la mezcla serán desconocidos por el sistema separador.

2.5 Conclusiones

En la presente tesis hemos optado por resolver el problema de separación ciega de fuentes utilizando algoritmos adaptativos en los que el cálculo de los coeficientes óptimos del sistema separador se realiza iterativamente de acuerdo a un criterio que, generalmente,

involucra promedios estadísticos. En la mayoría de los casos estos promedios son desconocidos y, por ello, se debe emplear algoritmos estocásticos que trabajan con estimaciones obtenidas a partir los datos disponibles.

Un sistema de separación adaptativo puede ser interpretado como un sistema dinámico cuyos coeficientes w_{ij} evolucionan con el tiempo de acuerdo a una ecuación diferencial ordinaria. Esta interpretación nos permite estudiar la estabilidad del algoritmo en dos pasos. El primero consiste en calcular los puntos estacionarios encontrando las raíces de la ODE. En un segundo paso se determinan los signos de los autovalores de la matriz jacobiana. Un punto será atractor del algoritmo solamente cuando todos los autovalores tengan parte real negativa. Es importante destacar que este estudio es rigurosamente válido únicamente cuando se emplean algoritmo deterministas. Sin embargo, como se ha comentado en este capítulo, la elección de un parámetro de paso suficientemente pequeño lo hace admisible también para el caso de emplear algoritmos estocásticos.

Un tipo de algoritmos adaptativos que tiene gran interés en el presente trabajo es el de gradiente descendente que permite encontrar los mínimos de una función de coste. Cuando el descenso se realiza según la máxima pendiente, el estudio de la estabilidad se reduce a determinar los puntos que anulan las primeras derivadas $\nabla_{\mathbf{w}}J = \mathbf{0}$ y a estudiar el carácter definido positivo de la matriz hessiana formada por las segundas derivadas. Existen otras muchas formas de elegir la dirección de descenso. En particular el empleo de una iteración de la forma $-\mathbf{W}\mathbf{W}^T\nabla_{\mathbf{w}}J$ conduce a un algoritmo conocido con el nombre de gradiente relativo cuya convergencia no depende de la matriz de mezcla. Esta propiedad lo hace especialmente atractivo para abordar situaciones donde las mezclas están mal condicionadas.

Como ejemplos de algoritmos adaptativos para realizar la separación, en este capítulo se ha presentado dos recursiones que surgen de emplear el criterio MMSE para separar las fuentes: el algoritmo LMS y una generalización del algoritmo p-vector. La aplicación de ambos algoritmos es muy reducida en la práctica ya que requieren conocer las fuentes (LMS) o la matriz de mezcla (p-vector). Esta limitación motiva el desarrollo de las técnicas ciegas que se presentan en el siguiente capítulo.

Capítulo 3

Revisión y aportaciones a aproximaciones anteriores

3.1 Introducción

Son numerosos los algoritmos adaptativos que han sido propuestos a lo largo de los últimos años para separar las fuentes de forma ciega. Todos ellos tienen en común el uso de no linealidades para incorporar la información contenida en los estadísticos de orden superior de las salidas. Como veremos a continuación, esta información es la clave para resolver el problema de separación ciega de fuentes.

Una cuestión muy importante de los algoritmos adaptativos es su estabilidad. Este aspecto, sin embargo, apenas ha sido tratado en profundidad en las aproximaciones propuestas debido a las dificultades en el análisis causadas por las no linealidades. En el presente capítulo llevaremos a cabo una revisión de aquellas aproximaciones de mayor relevancia en nuestro trabajo prestando especial interés al aspecto de su estabilidad.

Comenzaremos exponiendo en la sección 3.2 la incapacidad para resolver el problema de separación de los algoritmos que llevan a cabo la decorrelación de las salidas. Tras esta discusión presentaremos algoritmos de separación ciega de fuentes que incorporan estadísticos de orden superior: algoritmo de Héroult y Jutten (sección 3.3), algoritmo de Cichocki y Undebauen (sección 3.4), funciones de contraste (sección 3.5) y algoritmo EASI (sección 3.6). A continuación, en la sección 3.7, se presentarán algoritmos que han surgido a partir de la maximización de la transferencia de información entre las entradas y las salidas del sistema separador. En la sección 3.8 estableceremos una conexión entre estos algoritmos y los estimadores de máxima verosimilitud. Finalmente, la sección 3.9 recoge las conclusiones más importantes de este capítulo.

3.2 Algoritmos de decorrelación

Una gran parte de los criterios de separación que presentaremos en este capítulo persiguen la consecución de la independencia estadística de las salidas ya que, por el teorema de Darmois-Skitovich, sabemos que esta condición garantiza la recuperación de las fuentes. Si sólo se tienen en cuenta los estadísticos de hasta orden dos, se estará buscando la incorrelación de las salidas. Por ello, cabe preguntarse cual es la capacidad de separación de señales de algoritmos que lleven a cabo la decorrelación de las salidas. Para contestar a esta cuestión consideraremos el siguiente algoritmo de gradiente relativo [35]

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \mathbf{W}(n)(\mathbf{y}(n)\mathbf{y}^T(n) - \mathbf{I}) \quad (3.1)$$

Esta recursión puede deducirse (ver apéndice C) a partir de la minimización de la divergencia de Kullback-Leibler entre una distribución gaussiana con matriz de correlación \mathbf{R}_y y otra con matriz de correlación igual a la identidad.

De la teoría de sistemas dinámicos sabemos que la ODE asociada al algoritmo adaptativo es la siguiente

$$\frac{d\mathbf{W}}{dt} = -\mathbf{W} (E[\mathbf{y}\mathbf{y}^T] - \mathbf{I}) \quad (3.2)$$

Es obvio que existe un punto de equilibrio cuando las salidas extraen fuentes diferentes pero también cuando $\mathbf{W}^T \mathbf{A} = \mathbf{U}$ es una matriz unitaria ya que $E[\mathbf{y}\mathbf{y}^T] = \mathbf{W}^T \mathbf{A} (\mathbf{W}^T \mathbf{A})^T = \mathbf{U} \mathbf{U}^T = \mathbf{I}$. Adicionalmente, el estudio del jacobiano revela que el algoritmo puede converger a matrices unitarias que contienen varios elementos distintos de cero en una misma fila (o columna) con lo cual una salida puede extraer una combinación lineal de varias fuentes (o una fuente puede ser extraída por varias salidas). Así pues, la discusión anterior nos indica que el algoritmo de decorrelación no garantiza la separación de las señales bajo nuestras hipótesis de trabajo. Será necesario recurrir a estadísticos de orden superior al segundo y, por tanto, incluir no-linealidades a nuestros algoritmos adaptativos.

No obstante, debemos destacar que los criterios que persiguen la incorrelación de las salidas pueden ser empleados satisfactoriamente en situaciones donde las fuentes no son blancas (i.e., cuando $\mathbf{R}_s(l) = E[\mathbf{s}(n)\mathbf{s}(n-l)^T]$ es diagonal para algún $l \neq 0$) [5, 32]. Nótese que las salidas decorreladas pueden expresarse como sigue

$$\mathbf{y} = \mathbf{U}\mathbf{s} \quad (3.3)$$

y para invertir la mezcla basta con conocer la matriz unitaria \mathbf{U} . En particular, podemos determinarla descomponiendo la matriz $\mathbf{R}_y(l) = E[\mathbf{y}(n)\mathbf{y}(n-l)^T]$ en valores singulares [72]

$$\mathbf{R}_y(l) = \mathbf{U}\mathbf{R}_s(l)\mathbf{U} \quad (3.4)$$

Una vez calculada \mathbf{U} , las fuentes pueden ser recuperadas como $\hat{\mathbf{s}} = \mathbf{U}^T \mathbf{y}$. Es importante destacar que la descomposición será única solamente si $\mathbf{R}_s(l)$ tiene entradas distintas en su diagonal para algún $l \neq 0$.

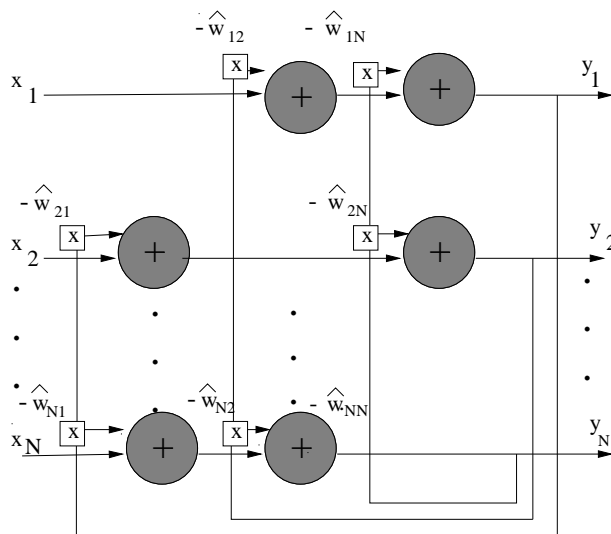


Figura 3.1: Sistema de separación utilizado por Héault y Jutten.

3.3 El algoritmo de Héault y Jutten

El primer algoritmo ciego para separación de señales fue propuesto a principios de los años 90 por los investigadores franceses Héault y Jutten [48]. Estos autores desarrollaron sus trabajos en el contexto de redes de neuronas artificiales y emplearon un sistema de separación como el mostrado en la figura 3.1. Este sistema de separación es ligeramente diferente del planteado en el capítulo 1 porque siempre supone que el número de entradas es igual al de salidas y la existencia de unos caminos de realimentación en los que las salidas se restan, convenientemente ponderados, de las entradas. Precisamente el coeficiente \hat{w}_{ij} representa el factor por el que es multiplicado la salida y_j antes de ser restado de la entrada x_i . También se supone que no existen auto-realimentaciones (i.e., $\hat{w}_{ii} = 0$). Matemáticamente, la relación entre la entrada y la salida viene dada por la siguiente expresión

$$\mathbf{y}(n) = (\mathbf{I} + \hat{\mathbf{W}}^T(n))^{-1} \mathbf{x}(n) \quad (3.5)$$

El criterio que se siguió para seleccionar la matriz $\hat{\mathbf{W}}$ fue el de minimizar la dependencia estadística entre las salidas de la red. Para ello, inicialmente se pensó en ajustar los pesos utilizando momentos de segundo orden pero, como se comentó en la sección 3.2, estos estadísticos resultan insuficientes para conseguir la separación. Conscientes de esta limitación, Héault y Jutten fueron los primeros en utilizar el concepto de Análisis de Componentes Independientes (ICA, del inglés Independent Component Analysis) y propusieron una regla que tuviese en cuenta la dependencia estadística de orden superior entre las salidas y_i y y_j de las neuronas. Para ello, el algoritmo empleado fue el siguiente

$$\hat{\mathbf{W}}(n+1) = \hat{\mathbf{W}}(n) + \mu \mathbf{f}(\mathbf{y}(n)) \mathbf{g}^T(\mathbf{y}(n)) \quad (3.6)$$

donde $\mathbf{f}(\mathbf{y}) = [f(y_1), \dots, f(y_N)]^T$ y $\mathbf{g}(\mathbf{y}) = [g(y_1), \dots, g(y_N)]^T$ son dos vectores formados por funciones no-lineales de las salidas $f(\cdot)$ y $g(\cdot)$. Cuando las señales a separar tienen distribuciones simétricas es necesario que estas no-linealidades sean impares (i.e., $f(y_i) = -f(-y_i)$ y $g(y_i) = -g(-y_i)$) para que sus desarrollos de Taylor sólo tengan potencias impares. Por ejemplo, una elección podría ser $f(y_i) = y_i^m$ y $g(y_i) = y_i^n$ siendo m y n dos números enteros impares.

La teoría de sistemas dinámicos permite explicar fácilmente el funcionamiento del algoritmo de Héroult-Jutten. La ODE asociada al algoritmo (3.6) tiene la siguiente forma

$$\frac{d\hat{\mathbf{W}}}{dt} = E[\mathbf{f}(\mathbf{y}) \mathbf{g}^T(\mathbf{y})] \quad (3.7)$$

y los puntos de equilibrio serán aquellos en que $E[\mathbf{f}(\mathbf{y}) \mathbf{g}^T(\mathbf{y})] = 0$. Esto se consigue cuando el sistema consiga separar las fuentes. En efecto, cuando cada salida se corresponda con una fuente distinta, las salidas serán estadísticamente independientes entre sí y los momentos cruzados serán iguales al producto de los momentos individuales de las componentes del vector \mathbf{y}

$$E[f(y_i) g(y_j)] = E[f(y_i)] E[g(y_j)], \quad i \neq j \quad (3.8)$$

Adicionalmente, si las fuentes tienen distribución simétrica, como ocurre en numerosos casos prácticos, todos sus momentos impares serán cero y la expresión (3.8) se anulará.

El razonamiento anterior muestra cómo los estados en que se consigue la separación de las fuentes son puntos de equilibrio de la ODE asociada al algoritmo Héroult-Jutten. Realizar un análisis general de su estabilidad es una labor difícil debido al carácter recursivo del sistema de separación y a la existencia de no-linealidades. A continuación presentaremos un estudio para el caso $N = 2$ que proporciona ciertas condiciones suficientes para asegurar la estabilidad en los puntos donde el sistema recupera las fuentes. El análisis también muestra que la convergencia del algoritmo depende de las no-linealidades elegidas y de las características estadísticas de las fuentes.

3.3.1 Estabilidad del algoritmo de Héroult y Jutten

A pesar de que el algoritmo de Héroult y Jutten ha sido objeto de numerosos análisis, el estudio de su estabilidad todavía plantea diversas cuestiones que permanecen aún por resolver. En esta sección presentaremos nuestro propio análisis de la estabilidad del algoritmo. A diferencia de otros trabajos como [31, 56, 71] el análisis será realizado para no-linealidades $f(\cdot)$ y $g(\cdot)$ genéricas. Sin embargo, al igual que ellos nos limitaremos al caso una red neuronal con dos entradas y dos salidas.

Comencemos recordando el modelo de señal correspondiente al sistema de separación realimentado utilizado por Héroult y Jutten que para el caso de dos entradas y dos salidas

adopta la forma

$$\mathbf{x} = \mathbf{A}\mathbf{s} \Rightarrow \begin{cases} x_1 = a_{11}s_1 + a_{12}s_2 \\ x_2 = a_{21}s_1 + a_{22}s_2 \end{cases} \quad (3.9)$$

$$\mathbf{y} = (\mathbf{I} + \hat{\mathbf{W}}^T)^{-1}\mathbf{x} \Rightarrow \begin{cases} y_1 = \frac{x_1 - \hat{w}_{12}x_2}{1 - \hat{w}_{12}\hat{w}_{21}} \\ y_2 = \frac{x_2 - \hat{w}_{21}x_1}{1 - \hat{w}_{12}\hat{w}_{21}} \end{cases} \quad (3.10)$$

donde $\hat{w}_{ii} = 0$. Empleando estas expresiones podemos escribir las salidas en función de las fuentes como sigue

$$y_1 = \frac{x_1 - \hat{w}_{12}x_2}{1 - \hat{w}_{12}\hat{w}_{21}} = \frac{a_{11}s_1 + a_{12}s_2 - \hat{w}_{12}(a_{21}s_1 + a_{22}s_2)}{1 - \hat{w}_{12}\hat{w}_{21}} \quad (3.11)$$

$$y_2 = \frac{x_2 - \hat{w}_{21}x_1}{1 - \hat{w}_{12}\hat{w}_{21}} = \frac{a_{21}s_1 + a_{22}s_2 - \hat{w}_{21}(a_{11}s_1 + a_{12}s_2)}{1 - \hat{w}_{12}\hat{w}_{21}} \quad (3.12)$$

De (3.11) y (3.12) se deduce que existen dos puntos de equilibrio en los que el sistema lleva a cabo la separación de las fuentes

$$\text{Sol. 1} \begin{cases} y_1 = \lambda_1 s_1 \Rightarrow (a_{12} - \hat{w}_{12}a_{22})s_2 = 0 \Rightarrow \hat{w}_{12} = \frac{a_{12}}{a_{22}} \\ y_2 = \lambda_2 s_2 \Rightarrow (a_{21} - \hat{w}_{21}a_{11})s_1 = 0 \Rightarrow \hat{w}_{21} = \frac{a_{21}}{a_{11}} \end{cases} \quad (3.13)$$

$$\text{Sol. 2} \begin{cases} y_1 = \lambda_1 s_2 \Rightarrow (a_{11} - \hat{w}_{12}a_{21})s_1 = 0 \Rightarrow \hat{w}_{12} = \frac{a_{11}}{a_{21}} \\ y_2 = \lambda_2 s_1 \Rightarrow (a_{22} - \hat{w}_{21}a_{12})s_2 = 0 \Rightarrow \hat{w}_{21} = \frac{a_{22}}{a_{12}} \end{cases} \quad (3.14)$$

Estudiemos ahora la estabilidad de la recursión propuesta por Héroult y Jutten para ajustar los pesos $\hat{\mathbf{W}}$

$$\hat{\mathbf{W}}(n+1) = \hat{\mathbf{W}}(n) + \mu \mathbf{f}(\mathbf{y}(n)) \mathbf{g}^T(\mathbf{y}(n)) \quad (3.15)$$

En el caso particular de una red neuronal 2×2 , se tiene que sólo son ajustados los términos \hat{w}_{12} y \hat{w}_{21} , lo que conduce a plantear una ODE formada por las siguientes dos ecuaciones

$$\begin{aligned} F_{12} &= \frac{d\hat{w}_{12}}{dt} = E[f(y_2)g(y_1)] \\ F_{21} &= \frac{d\hat{w}_{21}}{dt} = E[f(y_1)g(y_2)] \end{aligned} \quad (3.16)$$

siendo $f(\cdot)$ y $g(\cdot)$ dos funciones impares cualesquiera. Obviamente, las dos soluciones (3.13) y (3.14) producen salidas estadísticamente independientes que anulan (3.16) y, por ello, son puntos de equilibrio del algoritmo (3.15).

Ahora estudiaremos los signos de los autovalores de la matriz jacobiana formada por las primeras derivadas de F_{12} y F_{21} para determinar si los puntos deseados (3.13) y (3.14) son estables. Emplearemos la siguiente notación

$$\begin{aligned} m_{f'}^{\lambda_1} &= E[f'(\lambda_1 s_i)], & m_{f'}^{\lambda_2} &= E[f'(\lambda_2 s_j)], \\ m_{g'}^{\lambda_1} &= E[g'(\lambda_1 s_i)], & m_{g'}^{\lambda_2} &= E[g'(\lambda_2 s_j)], \\ \sigma_f^{\lambda_1} &= E[s_i f(\lambda_1 s_i)], & \sigma_f^{\lambda_2} &= E[s_j f(\lambda_2 s_j)], \\ \sigma_g^{\lambda_1} &= E[s_i g(\lambda_1 s_i)], & \sigma_g^{\lambda_2} &= E[s_j g(\lambda_2 s_j)] \end{aligned} \quad (3.17)$$

A partir de las derivadas de (3.16) calculadas en el apéndice D se obtiene que en las soluciones donde cada salida extrae una única fuente (i.e., $y_1 = \lambda_1 s_i$, $y_2 = \lambda_2 s_j$) la matriz jacobiana es

$$\mathbf{H} = \begin{bmatrix} \frac{\partial F_{12}}{\partial \hat{w}_{12}} & \frac{\partial F_{12}}{\partial \hat{w}_{21}} \\ \frac{\partial F_{21}}{\partial \hat{w}_{12}} & \frac{\partial F_{21}}{\partial \hat{w}_{21}} \end{bmatrix} = \gamma \begin{bmatrix} \lambda_2 m_{g'}^{\lambda_1} \sigma_f^{\lambda_2} & \lambda_1 \sigma_g^{\lambda_1} m_{f'}^{\lambda_2} \\ \lambda_2 m_{f'}^{\lambda_1} \sigma_g^{\lambda_2} & \lambda_1 \sigma_f^{\lambda_1} m_{g'}^{\lambda_2} \end{bmatrix} \quad (3.18)$$

donde $\gamma = -\frac{1}{1 - \hat{w}_{12}\hat{w}_{21}}$. Sin pérdida de generalidad supondremos que γ es negativa ya que el producto de los pesos $\hat{w}_{12}\hat{w}_{21}$ es menor que uno para alguna de las soluciones deseadas (3.13) y (3.14)

Los puntos de equilibrio serán estables si los autovalores del jacobiano \mathbf{H} tienen parte real negativa, lo cual es equivalente a pedir que el determinante de \mathbf{H} sea positivo y su traza sea negativa

$$\det(\mathbf{H}) = \gamma^2 \lambda_1 \lambda_2 (m_{g'}^{\lambda_1} m_{g'}^{\lambda_2} \sigma_f^{\lambda_1} \sigma_f^{\lambda_2} - m_{f'}^{\lambda_1} m_{f'}^{\lambda_2} \sigma_g^{\lambda_1} \sigma_g^{\lambda_2}) > 0 \quad (3.19)$$

$$\text{tr}(\mathbf{H}) = \gamma (\lambda_2 m_{g'}^{\lambda_1} \sigma_f^{\lambda_2} + \lambda_1 \sigma_f^{\lambda_1} m_{g'}^{\lambda_2}) < 0 \quad (3.20)$$

Adicionalmente, como estamos suponiendo que γ es negativa, las condiciones de estabilidad se transforman en las siguientes expresiones

$$\det(\mathbf{H}) > 0 \Rightarrow \lambda_1 \lambda_2 (m_{g'}^{\lambda_1} m_{g'}^{\lambda_2} \sigma_f^{\lambda_1} \sigma_f^{\lambda_2} - m_{f'}^{\lambda_1} m_{f'}^{\lambda_2} \sigma_g^{\lambda_1} \sigma_g^{\lambda_2}) > 0 \quad (3.21)$$

$$\text{tr}(\mathbf{H}) < 0 \Rightarrow \lambda_2 m_{g'}^{\lambda_1} \sigma_f^{\lambda_2} + \lambda_1 m_{g'}^{\lambda_2} \sigma_f^{\lambda_1} > 0 \quad (3.22)$$

Como conclusión, podemos decir que las no-linealidades $f(\cdot)$ y $g(\cdot)$ deben ser elegidas de forma que se verifiquen las condiciones (3.21) y (3.22). En particular, si consideramos las no-linealidades $f(y_i) = y_i$ y $g(y_i) = y_i^3$ y las hipótesis sobre las fuentes (independencia estadística, potencia unidad y momentos impares cero), las condiciones de estabilidad vienen dadas por

$$\lambda_1^4 \lambda_2^4 (9 - E[s_i^4] E[s_j^4]) > 0 \Rightarrow E[s_i^4] E[s_j^4] < 9 \quad (3.23)$$

$$6\lambda_1^2 \lambda_2^2 > 0 \quad (3.24)$$

Es interesante notar que la condición (3.23) siempre se verifica para señales con curtosis negativas, $E[s_1^4] < 3$ y $E[s_2^4] < 3$. Por el contrario, esta condición nunca se cumple cuando las dos señales tienen curtosis positiva. De lo que podemos concluir que las no-linealidades $f(y_i) = y_i$ y $g(y_i) = y_i^3$ son adecuadas para separar fuentes subgaussianas y no debe emplearse cuando todas las fuentes del entorno son supergaussianas. También es interesante notar que existen combinaciones de fuentes con curtosis de distinto signo que verifican (3.23). Un análisis particular para estas no linealidades puede ser consultado en [71].

s_i	$-1/\sqrt{2p_i}$	0	$1/\sqrt{2p_i}$
$p(s_i)$	p_i	$1 - 2p_i$	p_i

Tabla 3.1: Distribución de probabilidad discreta empleada para verificar las condiciones de estabilidad del algoritmo de Hérault y Jutten.

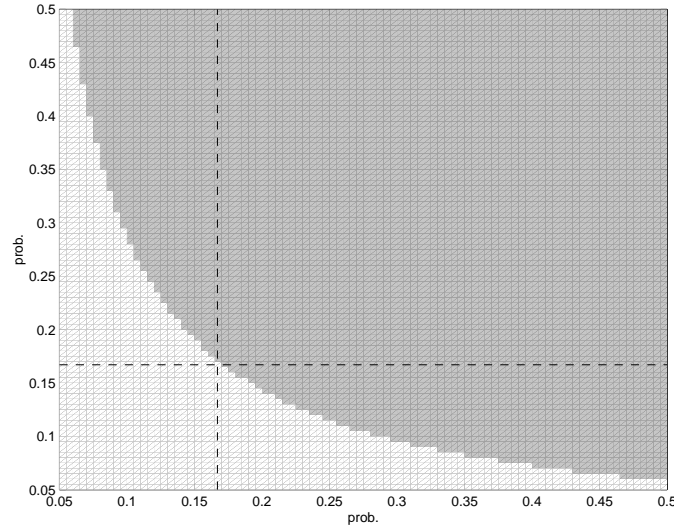


Figura 3.2: Condiciones de estabilidad del algoritmo de Hérault y Jutten.

Para ilustrar estos resultados consideraremos un entorno formado por dos fuentes discretas con la distribución mostrada en la tabla 3.1. Para este caso el momento de cuarto orden viene dado por la siguiente expresión

$$m_{4i} = E[s_i^4] = \frac{2p_i}{(\sqrt{2p_i})^4} = \frac{1}{2p_i} \quad (3.25)$$

La tabla 3.2 recoge el valor del momento de cuarto orden m_{4i} y de la curtosis $k_i = E[s_i^4]/E^2[s_i^2] - 3 = (1/2p_i) - 3$ para distintas probabilidades p_i . Para estos valores, la figura 3.2 muestra la región de estabilidad del algoritmo evaluando la condición (3.23). Puede observarse que el algoritmo no es estable si todas las fuentes tienen curtosis positiva ($p < 0.167$) y sí lo es cuando todas tienen curtosis negativa ($p > 0.167$). Adicionalmente, también es estable en ciertos entornos donde las fuentes tienen curtosis de distinto signos.

p_i	0.10	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
k_i	2	0.333	-0.5	-1	-1.33	-1.57	-1.75	-1.889	-2
m_{4i}	5	3.33	2.50	2	1.66	1.43	1.25	1.11	1

Tabla 3.2: Estadísticos de las fuentes empleadas para verificar las condiciones de estabilidad del algoritmo de Hérault y Jutten.

3.4 El algoritmo de Cichocki y Undebauen

Siguiendo con la idea de incorporar no linealidades en los algoritmos que permitan aprovechar la información de momentos de orden superior al segundo para resolver el problema de separación, Cichocki y Undebauen [17] han utilizado una red neuronal completamente recursiva cuyos coeficientes son actualizados utilizando la siguiente regla de aprendizaje

$$\hat{\mathbf{W}}(n+1) = \hat{\mathbf{W}}(n) - \mu \left(\mathbf{\Lambda} - \mathbf{f}(\mathbf{y}(n)) \mathbf{g}^T(\mathbf{y}(n)) \right) \left(\mathbf{I} + \hat{\mathbf{W}}(n) \right) \quad (3.26)$$

donde $\mathbf{f}(\mathbf{y}) = [f(y_1), \dots, f(y_N)]^T$ y $\mathbf{g}(\mathbf{y}) = [g(y_1), \dots, g(y_N)]^T$ son vectores de funciones no-lineales y $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ es una matriz diagonal con $\lambda_i > 0$ (típicamente $\mathbf{\Lambda} = \mathbf{I}$). A pesar del parecido con el algoritmo de Héault y Jutten, la recursión (3.26) presenta ciertas diferencias con respecto a los puntos de equilibrio.

La ODE asociada al algoritmo (3.26) es

$$\frac{d\hat{\mathbf{W}}}{dt} = - \left(\mathbf{\Lambda} - E[\mathbf{f}(\mathbf{y}) \mathbf{g}^T(\mathbf{y})] \right) \left(\mathbf{I} + \hat{\mathbf{W}} \right) \quad (3.27)$$

Es evidente que (3.27) se anula en aquellos puntos en que se cumple que $E[\mathbf{f}(\mathbf{y}) \mathbf{g}^T(\mathbf{y})] = \mathbf{\Lambda}$, i.e., cuando se lleva a cabo la separación y, además, el momento de la salida $E[f_i(y_i)g_i(y_i)]$ es igual a λ_i . Por tanto, al incluir una matriz diagonal no necesariamente igual a la identidad se puede llevar a cabo una auto-normalización de las salidas del sistema de separación. En particular, si consideramos $f(y_i) = g(y_i) = y_i$, el algoritmo (3.27) convergerá a los puntos donde $\mathbf{R}_y = \mathbf{\Lambda}$. Sin embargo, como hemos visto anteriormente, esta condición es insuficiente para conseguir la separación de las fuentes, por lo que debemos emplear no-linealidades que introduzcan momentos de orden superior.

Debemos destacar que una realización puramente software de los algoritmos anteriores (3.6) y (3.26) requiere invertir la matriz de pesos en cada iteración para calcular las salidas (3.5), lo cual es computacionalmente costoso. Esta operación no es necesaria si se considera una estructura sin realimentaciones o directa para el sistema de separación como la mostrada en la figura 3.3, cuya salida se calcula directamente a partir del vector de entrada

$$\mathbf{y}(n) = \mathbf{W}^T(n)\mathbf{x}(n) \quad (3.28)$$

Comparando (3.28) con (3.5) podemos establecer la siguiente correspondencia entre los pesos de la estructura directa \mathbf{W} y los de la recursiva $\hat{\mathbf{W}}$

$$\mathbf{W} = (\mathbf{I} + \hat{\mathbf{W}})^{-1} \quad (3.29)$$

$$\hat{\mathbf{W}} = (\mathbf{W}^{-1} - \mathbf{I}) \quad (3.30)$$

Por tanto, también existirá una relación entre las ODEs de ambas estructuras

$$\frac{d\hat{\mathbf{W}}}{dt} = \frac{d\mathbf{W}^{-1}}{dt} \quad (3.31)$$

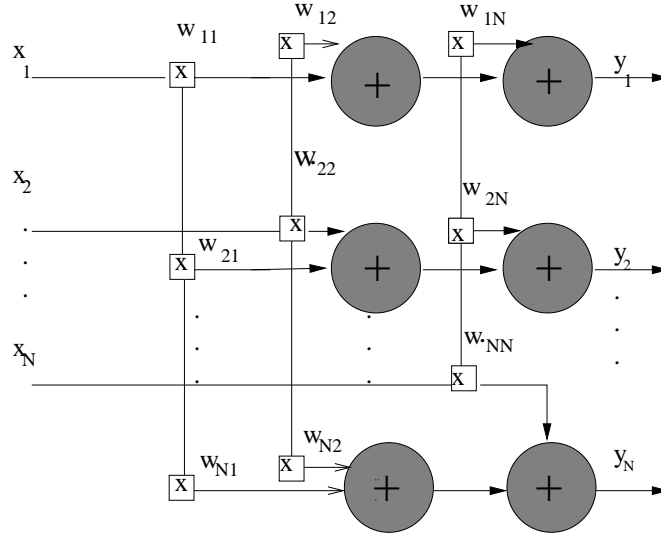


Figura 3.3: Red neuronal directa utilizada en separación ciega.

Por otro lado, considerando la regla de la cadena

$$\frac{d\mathbf{W}^{-1}\mathbf{W}}{dt} = 0 \Rightarrow \mathbf{W}^{-1}\frac{d\mathbf{W}}{dt} + \frac{d\mathbf{W}^{-1}}{dt}\mathbf{W} = 0 \Rightarrow \frac{d\mathbf{W}}{dt} = -\mathbf{W}\frac{d\mathbf{W}^{-1}}{dt}\mathbf{W} \quad (3.32)$$

obtenemos lo siguiente

$$\frac{d\mathbf{W}}{dt} = -\mathbf{W}\frac{d\mathbf{W}^{-1}}{dt}\mathbf{W} = -\mathbf{W}\frac{d\hat{\mathbf{W}}}{dt}\mathbf{W} \quad (3.33)$$

Particularizando esta expresión para (3.27) obtenemos la siguiente expresión de la ODE para actualizar los pesos de la red neuronal directa

$$\begin{aligned} \frac{d\mathbf{W}}{dt} &= -\mathbf{W}\frac{d\hat{\mathbf{W}}}{dt}\mathbf{W} = \mathbf{W}(\mathbf{\Lambda} - E[\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})])\mathbf{W}^{-1}\mathbf{W} \\ &= \mathbf{W}(\mathbf{\Lambda} - E[\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})]) \end{aligned} \quad (3.34)$$

Finalmente, estimando los estadísticos que aparecen en (3.34) con una sola muestra, se obtiene el siguiente algoritmo adaptativo para la red neuronal directa

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu\mathbf{W}(n)(\mathbf{\Lambda} - \mathbf{f}(\mathbf{y}(n))\mathbf{g}^T(\mathbf{y}(n))) \quad (3.35)$$

Es interesante notar que el comportamiento de este algoritmo es invariante en el sentido mencionado en la sección 2.3 del capítulo anterior ya que la recursión que se obtiene al multiplicar (3.35) por la matriz de mezcla traspuesta \mathbf{A}^T es independiente del entorno

$$\mathbf{G}^T(n+1) = \mathbf{G}^T(n) + \mu\mathbf{G}^T(n)(\mathbf{\Lambda} - \mathbf{f}(\mathbf{y}(n))\mathbf{g}^T(\mathbf{y}(n))) \quad (3.36)$$

Por tal motivo, presenta un buen comportamiento en situaciones donde la matriz de mezcla está mal condicionada. Además, para ciertas no linealidades, la expresión (3.35) maximiza la transferencia de información en una red neuronal. Este aspecto será abordado en la sección 3.7 donde también presentaremos un estudio de la estabilidad de (3.35) en los puntos correspondientes a la separación de las fuentes.

3.5 Funciones de contraste

El concepto de función de contraste ha sido introducido por Comon [18] para designar a aquellos criterios estadísticos del vector de salida \mathbf{y} cuyos máximos globales son alcanzados cuando las componentes en \mathbf{y} son estadísticamente independientes. Se supone, entonces, que la maximización de una función de contraste conduce a la separación de las fuentes ya que, por el teorema de Darmois-Skitovich, la separación se produce sí y sólo sí las componentes del vector de salida son estadísticamente independientes entre sí.

Más formalmente, un contraste es una función $J(\cdot)$ de \mathfrak{R}^N en \mathfrak{R} que satisfacen las siguientes condiciones [54]:

- *Condición 1:* $J(\mathbf{y})$ sólo depende de la función densidad de probabilidad de \mathbf{y} y es simétrica respecto a sus argumentos

$$\forall \mathbf{P} \text{ matriz de permutación, } J(\mathbf{P}\mathbf{y}) = J(\mathbf{y})$$

- *Condición 2:* $J(\mathbf{y})$ es invariante respecto a cualquier cambio de escala

$$\forall \mathbf{\Lambda} \text{ matriz diagonal invertible, } J(\mathbf{\Lambda}\mathbf{y}) = J(\mathbf{y})$$

- *Condición 3:* Mezclas de componentes independientes sólo decrementan el contraste

$$\forall \mathbf{s} \text{ vector con componentes independientes, } \forall \mathbf{G}, J(\mathbf{G}\mathbf{s}) \leq J(\mathbf{s})$$

- *Condición 4:* Solamente las permutaciones y cambios de escala mantienen sin cambios el contraste de fuentes independientes

$$\forall \mathbf{s} \text{ vector con componentes independientes, } \forall \mathbf{G}, J(\mathbf{G}\mathbf{s}) = J(\mathbf{s}) \iff \mathbf{G} = \mathbf{\Lambda}\mathbf{P}$$

donde $\mathbf{\Lambda}$ es una matriz diagonal y \mathbf{P} es una matriz de permutación

Es interesante notar que las condiciones impuestas a los contrastes limitan las posibles funciones de coste que se pueden plantear. En primer lugar, las funciones deben ser simétricas en sus argumentos (*condición 1*) y, por tanto, cualquier fuente puede ser extraída por cualquier salida. Sin embargo, si conocemos *a priori* alguna propiedad estadística que diferencie a alguna fuente de las demás, sería más adecuado plantear

contrastes no simétricos que exploten esa información [58]. En segundo lugar, la propiedad de invarianza ante cambios de escala (*condición 2*) implica que los contrastes deben ser maximizados bajo ciertas restricciones. En particular, podríamos pensar en imponer la restricción $E[y_i^2] = 1$ ya que cuando cada salida extrae una única fuente se cumplirá lo siguiente $E[y_i^2] = g_{ij}^2 E[s_i^2] = g_{ij}^2 = 1$ y, por tanto, las matrices \mathbf{A} admisibles tienen entradas $\lambda = \pm 1$. Como se verá más adelante, el añadir restricciones al contraste complica el sistema de separación ya que es necesario incorporar una etapa adicional que las implemente.

La mayoría de las funciones de contraste que se han propuesto hasta la fecha se definen como la suma de cumulantes de las salidas [54, 55, 58] ya que estos estadísticos se adaptan a las condiciones impuestas a los contrastes: son simétricos en sus argumentos y, bajo la restricción $E[y_i^2] = 1$, también son invariantes a cambios de escala. Un ejemplo de función de contraste, propuesto por Moreau y Macchi en [54] y analizado por nosotros en [26], es la siguiente

$$J_c(\mathbf{W}) = \sum_{i=1}^N |K_{y_i}| - \alpha \sum_{\substack{i,j=1 \\ i < j}}^N |K_{y_i y_j}|, \quad \text{sujeto a } E[y_i^2] = 1, \quad i = 1, \dots, N \quad (3.37)$$

donde $\alpha \geq 1$, $K_{y_i} = E[y_i^4] - 3E^2[y_i^2]$ denota la curtosis de la salida y_i y $K_{y_i y_j} = E[y_i^2 y_j^2] - 2E^2[y_i y_j] - E[y_i^2]E[y_j^2]$ es el cumulante cruzado de cuarto orden entre y_i y y_j .

En la función de coste (3.37) podemos distinguir dos partes. El primer término consiste en la suma de los cumulantes de las salidas y su maximización sujeta a las restricciones $E[y_i^2] = 1$ puede interpretarse como una extensión al caso multicanal del criterio propuesto por Shalvi y Weinstein [69] para igualación ciega. De los análisis realizados en [69], se puede concluir que este criterio conduce a la extracción de una única fuente en cada salida. Sin embargo, la maximización de J_c sin el término cruzado es insuficiente para realizar la separación ya que puede ocurrir que varias salidas extraigan la misma fuente. En esta situación indeseada aparece una dependencia estadística entre las salidas que puede ser penalizada añadiendo el segundo término en (3.37): $|K_{y_i y_j}|$ será cero únicamente cuando las salidas sean estadísticamente independientes entre sí.

En la figura 3.4 se representa el esquema adaptativo que lleva a cabo la maximización del contraste (3.37). Pueden distinguirse dos etapas: en la primera se calculan unas salidas intermedias $\mathbf{y}^o = \mathbf{W}^T \mathbf{x}$ y en la segunda se fuerzan las restricciones utilizando algoritmos de control adaptativo de ganancias (AGC, del inglés Adaptive Gain Control).

La forma más sencilla de ajustar los coeficientes \mathbf{W} es utilizar un algoritmo de tipo gradiente estocástico que maximice la función de coste (3.37). Considerando las definiciones dadas en el apéndice A el gradiente de J_c puede ser expresado como sigue

$$\nabla_{\mathbf{W}} J_c = 2\epsilon(\mathbf{C}_i - \alpha \mathbf{C}_{ij}) \quad (3.38)$$

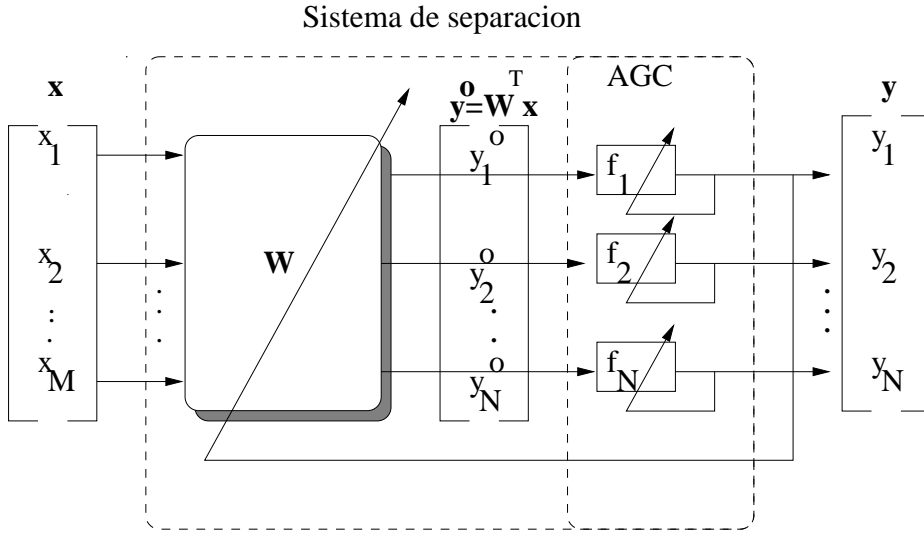


Figura 3.4: Sistema de separación con AGCs.

donde

$$\begin{aligned}
 \mathbf{C}_i &= \begin{bmatrix} 2E[y_1^3 \mathbf{x}] - 6E[y_1^2]E[y_1 \mathbf{x}] \\ \vdots \\ 2E[y_N^3 \mathbf{x}] - 6E[y_N^2]E[y_N \mathbf{x}] \end{bmatrix} \\
 \mathbf{C}_{ij} &= \begin{bmatrix} \sum_{j \neq 1}^N (E[y_1 y_j^2 \mathbf{x}] - E[y_1 \mathbf{x}]E[y_j^2] - 2E[y_j \mathbf{x}]E[y_1 y_j]) \\ \vdots \\ \sum_{j \neq N}^N (E[y_N y_j^2 \mathbf{x}] - E[y_N \mathbf{x}]E[y_j^2] - 2E[y_j \mathbf{x}]E[y_N y_j]) \end{bmatrix} \quad (3.39)
 \end{aligned}$$

Un algoritmo estocástico que resulta de estimar los estadísticos de la expresión anterior es el siguiente

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \epsilon \mu \mathbf{x}(n) (\hat{\mathbf{C}}_i^T(n) - \alpha \hat{\mathbf{C}}_{ij}^T(n)) \quad (3.40)$$

donde un factor 2 ha sido incluido dentro de μ y

$$\begin{aligned}
 \hat{\mathbf{C}}_i(n) &= \begin{bmatrix} 2y_1^3(n) - 6\hat{E}[y_1^2]y_1(n) \\ \vdots \\ 2y_N^3(n) - 6\hat{E}[y_N^2]y_N(n) \end{bmatrix} \\
 \hat{\mathbf{C}}_{ij}(n) &= \begin{bmatrix} \sum_{j \neq 1}^N (y_1(n)y_j^2(n) - y_1(n)\hat{E}[y_j^2] - 2y_j(n)\hat{E}[y_1 y_j]) \\ \vdots \\ \sum_{j \neq N}^N (y_N(n)y_j^2(n) - y_N(n)\hat{E}[y_j^2] - 2y_j(n)\hat{E}[y_N y_j]) \end{bmatrix} \quad (3.41)
 \end{aligned}$$

Los momentos $E[y_j^2]$ y $E[y_i y_j]$ que aparecen en (3.39) no deben estimarse utilizando una única muestra para evitar ambigüedades¹. Una posibilidad es utilizar estimadores con

¹Por ejemplo, $y_i^3(n)\mathbf{x}(n)$ es el estimador de una sola muestra de los estadísticos $E[y_i^3 \mathbf{x}]$ y $E[y_i^2]E[y_i \mathbf{x}]$

una ventana exponencial como los siguientes

$$\begin{aligned}\hat{E}[y_i^2] &= (1 - \mu_l)\hat{E}[y_i^2] + \mu_l y_i^2(n), \quad i = 1, \dots, N \\ \hat{E}[y_i y_j] &= (1 - \mu_l)\hat{E}[y_i y_j] + \mu_l y_i(n) y_j(n), \quad i, j = 1, \dots, N, \quad i \neq j\end{aligned}\quad (3.42)$$

donde μ_l es un valor real positivo.

Finalmente, los elementos escalares f_i de los AGC se pueden actualizar con el algoritmo

$$\begin{aligned}y_i(n) &= y_i^o(n) \sqrt{f_i(n-1)} \\ f_i(n) &= f_i(n-1) + \mu_a(1 - y_i^2(n))\end{aligned}\quad (3.43)$$

donde μ_a es un valor real positivo.

3.5.1 Estabilidad de la función de contraste de Moreau y Macchi

En esta sección estudiaremos la estabilidad del algoritmo (3.40) utilizado para maximizar la función de contraste (3.37). Como hemos comentado en el capítulo 2, los atractores de la ODE de un algoritmo de gradiente ascendente se corresponden con los máximos de la función de coste, i.e., son los puntos donde la primera derivada es cero y la matriz hessiana es definida negativa.

El análisis de la estabilidad para el caso general de N fuentes y N salidas es difícil de realizar debido a la existencia de las restricciones $E[y_i^2] = 1$. Por tal motivo a continuación nos restringiremos al caso $N = 2$. Parte de estos resultados han sido presentados en [26].

La función de contraste (3.37) para el caso de dos salidas toma la siguiente forma

$$J_c(\mathbf{y}) = |K_{y_1}| + |K_{y_2}| - \alpha |K_{y_1 y_2}| \quad \text{sujeto a } E[y_1^2] = 1, \quad E[y_2^2] = 1 \quad (3.44)$$

siendo α un valor real positivo. Comenzaremos expresando $J_c(\mathbf{y})$ en función de los coeficientes de la matriz $\mathbf{G} = \mathbf{W}^T \mathbf{A}$, para lo cual utilizaremos la siguiente relación

$$\begin{aligned}y_1 &= g_{11}s_1 + g_{12}s_2, \\ y_2 &= g_{21}s_1 + g_{22}s_2\end{aligned}\quad (3.45)$$

Empleando la hipótesis de independencia estadística de las fuentes y las propiedades de los cumulantes (ver apéndice B) obtenemos

$$\begin{aligned}K_{y_1} &= g_{11}^4 K_{s_1} + g_{12}^4 K_{s_2} \\ K_{y_2} &= g_{21}^4 K_{s_1} + g_{22}^4 K_{s_2} \\ K_{y_1 y_2} &= g_{11}^2 g_{21}^2 K_{s_1} + g_{12}^2 g_{22}^2 K_{s_2}\end{aligned}\quad (3.46)$$

Sustituyendo estas expresiones en (3.44) resulta

$$J_c(\mathbf{y}) = (g_{11}^4 + g_{21}^4 - \alpha g_{11}^2 g_{21}^2) |K_{s_1}| + (g_{12}^4 + g_{22}^4 - \alpha g_{12}^2 g_{22}^2) |K_{s_2}| \quad (3.47)$$

Adicionalmente, las restricciones sobre las potencia de las salidas se convierten ahora en las siguientes igualdades

$$\begin{aligned} E[y_1^2] = 1 &\Rightarrow g_{11}^2 = 1 - g_{12}^2 \\ E[y_2^2] = 1 &\Rightarrow g_{22}^2 = 1 - g_{21}^2 \end{aligned} \quad (3.48)$$

Al sustituir (3.48) en (3.47), obtenemos una función de coste sin restricciones que depende únicamente de las variables g_{12} y g_{21}

$$\begin{aligned} \phi(g_{12}, g_{21}) \triangleq J_c(\mathbf{y}) &= \left((1 - g_{12}^2)^2 + g_{21}^4 - \alpha(1 - g_{12}^2)g_{21}^2 \right) |K_{s_1}| \\ &+ \left(g_{12}^4 + (1 - g_{21}^2)^2 - \alpha g_{12}^2(1 - g_{21}^2) \right) |K_{s_2}| \end{aligned} \quad (3.49)$$

Los puntos estacionarios de esta función se corresponden con las raíces de las primeras derivadas

$$\begin{aligned} \frac{\partial \phi}{\partial g_{12}} &= g_{12}(4(g_{12}^2 - 1) + 2\alpha g_{21}^2)|K_{s_1}| + g_{12}(4g_{12}^2 - 2\alpha(1 - g_{21}^2))|K_{s_2}| = 0 \\ \frac{\partial \phi}{\partial g_{21}} &= g_{21}(4g_{21}^2 - 2\alpha(1 - g_{12}^2))|K_{s_1}| + g_{21}(4(g_{21}^2 - 1) + 2\alpha g_{12}^2)|K_{s_2}| = 0 \end{aligned} \quad (3.50)$$

Este sistema de ecuaciones tiene múltiples soluciones que pueden clasificarse en cuatro grupos:

- *Solución deseada:* $g_{12} = g_{21} = 0$, i.e., las salidas extraen fuentes diferentes y la matriz \mathbf{G} es diagonal $g_{11}^2 = g_{22}^2 = 1$, $g_{12} = g_{21} = 0$. En el apéndice D se demuestra que la condición $\alpha > 0$ garantiza que el punto es un máximo.
- *Solución indeseada tipo I:* una salida extrae una combinación lineal de las dos fuentes y la otra extrae únicamente una fuente. Hay dos posibles soluciones en este grupo

$$\begin{aligned} \text{Sol. 1: } g_{12} = 0, g_{21}^2 &= \frac{\alpha|K_{s_1}| + 2|K_{s_2}|}{2(|K_{s_1}| + |K_{s_2}|)}, \\ \text{Sol. 2: } g_{12}^2 &= \frac{2|K_{s_1}| + \alpha|K_{s_2}|}{2(|K_{s_1}| + |K_{s_2}|)}, g_{21} = 0 \end{aligned} \quad (3.51)$$

El análisis de ambas soluciones es análogo por lo que nos centraremos en la primera de ellas. Es interesante notar que esta solución no es válida para $\alpha > 2$ ya que, en este caso, g_{21}^2 sería mayor que uno y, por tanto, $g_{22}^2 = 1 - g_{21}^2$ sería negativo. Adicionalmente, en el apéndice D se demuestra que la condición $0 < \alpha < 2$ es suficiente para garantizar que este punto es de ensilladura.

- *Solución indeseada tipo II:* $g_{12} \neq 0$, $g_{21} \neq 0$, i.e, las salidas extraen una combinación lineal de las fuentes:

$$g_{21}^2 = \frac{|K_{s_2}|}{|K_{s_1}| + |K_{s_2}|}, \quad g_{12}^2 = \frac{|K_{s_1}|}{|K_{s_1}| + |K_{s_2}|}$$

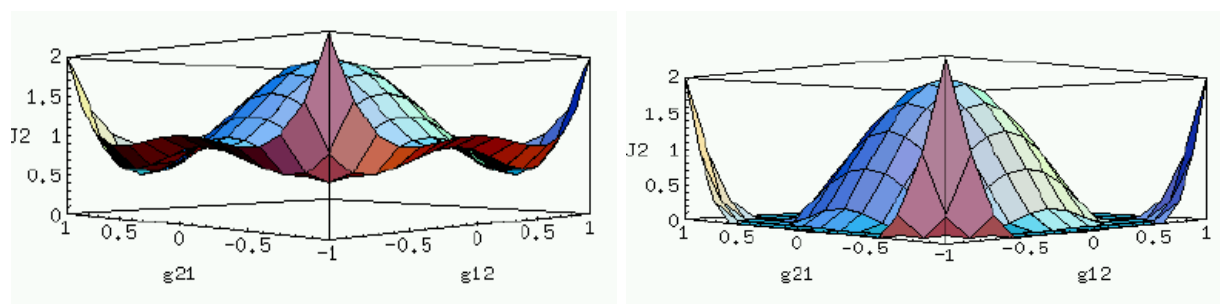


Figura 3.5: Contraste J_c con $\alpha = 1$ (izquierda) y con $\alpha = 2.5$ (derecha).

Esta es otra solución indeseada pero en el apéndice D se demuestra que este punto no se corresponde con un máximo de la función de coste. En particular, será un punto de ensilladura si $\alpha > 2$ y un mínimo si se elige un valor $\alpha < 2$.

- *Puntos en la frontera:* Además de los grupos de soluciones anteriores, la existencia de restricciones en la función de contraste hace necesario considerar la existencia de puntos estacionarios en la frontera, i.e, cuando $g_{12}^2 = 1$ ó $g_{21}^2 = 1$. En el apéndice D se demuestra que la condición $\alpha > 2$ garantiza que la función de coste tiene mínimos cuando ambas salidas extraen la misma fuente ($g_{12}^2 = 1, g_{21} = 0$ ó $g_{12} = 0, g_{21}^2 = 1$) y es creciente hacia las soluciones deseadas $g_{12}^2 = g_{21}^2 = 1$.

Del estudio anterior podemos concluir que la condición $\alpha > 2$ garantiza que los máximos del contraste (3.49) corresponden a la separación de las fuentes y que las otras soluciones son puntos de ensilladura o mínimos. Para ilustrar este resultado, la figura 3.5 muestra la función de coste para dos valores de α considerando $|K_{s_1}| = |K_{s_2}| = 1$. A la izquierda se observa que la función de coste con $\alpha = 1$ contiene tiene cinco máximos deseados ($g_{12} = g_{21} = 0, g_{12}^2 = 1, g_{21} = 0$ y $g_{12} = 0, g_{21}^2 = 1$) pero también presenta atractores indeseados en la frontera. Por otro lado, la gráfica de la derecha muestra que eligiendo $\alpha = 2.5$ se consigue eliminar los máximos indeseados.

Finalizaremos esta sección mostrando el resultado de una simulación realizada por ordenador considerando un entorno formado por dos fuentes binarias mezcladas a través de la siguiente matriz

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad (3.52)$$

En la fase de separación se ha empleado un sistema MIMO con dos entradas y dos salidas cuyos coeficientes, aleatoriamente inicializados, fueron actualizados utilizando el algoritmo (3.40) con $\gamma = 2.5$ y parámetros de paso $\mu = 10^{-4}$, $\mu_l = \mu_a = 5 \times 10^{-3}$. La figura 3.6 muestra la evolución de los coeficientes del sistema después de la etapa de normalización

$$\mathbf{G} = (\mathbf{WF})^T \mathbf{A} \quad (3.53)$$

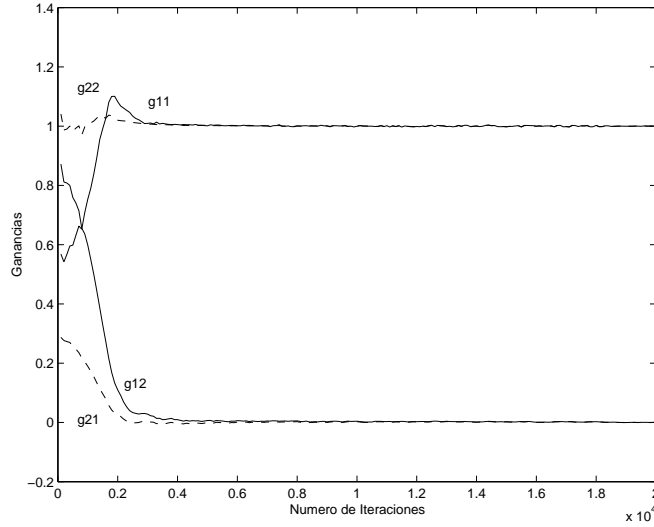


Figura 3.6: Evolución de las ganancias del sistema separador ajustado con el algoritmo de Moreau y Macchi.

donde \mathbf{F} es una matriz diagonal que contiene las ganancias de los AGCs $\mathbf{F} = \text{diag}(\sqrt{f_1}, \sqrt{f_2})$. Como puede observarse el algoritmo ha convergido a los valores teóricos obtenidos del análisis.

3.6 El algoritmo EASI

El algoritmo conocido bajo las siglas EASI (*Equivariant Adaptive Separation via Independence*) ha sido recientemente propuesto por Cardoso y Laheld [13]. Estos autores se plantean ajustar el sistema de separación con el objetivo de optimizar la siguiente función de coste

$$J(\mathbf{W}) = \sum_{i=1}^N E[h(y_i)] \quad \text{sujeto a } \mathbf{R}_y = \mathbf{I} \quad (3.54)$$

donde $h(\cdot)$ es una función no lineal. Esta función de coste puede interpretarse como una generalización de las funciones de contraste propuestas por Comon. De hecho cuando $h(y_i) = y_i^4$ se ha demostrado en [55] que (3.54) es un contraste si el signo de la curtosis de las fuentes es el mismo; cuando la curtosis de las fuentes es positiva el contraste (3.54) debe ser maximizado y cuando es negativa debemos minimizarlo. No obstante, es importante notar que (3.54) no siempre es un contraste para cualquier tipo de no linealidad.

Para poder implementar la restricción $\mathbf{R}_y = \mathbf{I}$, Cardoso y Laheld, al igual que Comon en [18], dividen el sistema de separación en una etapa de blanqueo y una etapa de separación con una matriz unitaria, tal y como se muestra en la figura 3.7. Vamos a representar por la matriz \mathbf{V} de dimensión $N \times N$ a la etapa de blanqueo que transforma

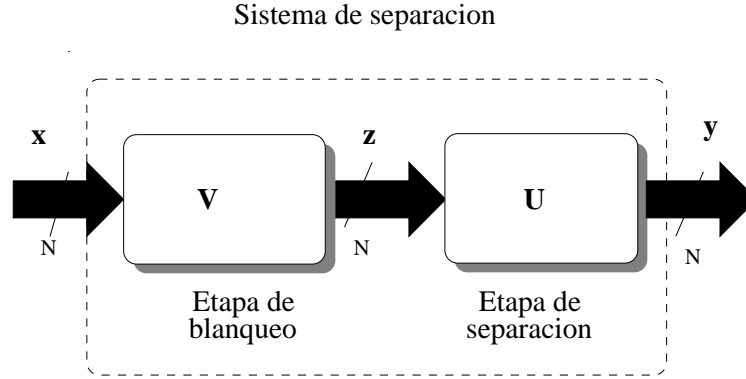


Figura 3.7: Sistema de separación de dos etapas.

el vector de observaciones $\mathbf{x} = \mathbf{A}\mathbf{s}$, con matriz de correlación \mathbf{R}_x , en el vector $\mathbf{z} = \mathbf{V}^T\mathbf{x}$ con matriz de correlación igual a la identidad $\mathbf{R}_z = \mathbf{I}$. Una de las ideas que resultan novedosas en el trabajo de Cardoso y Laheld es que los coeficientes de la matriz \mathbf{V} son actualizados utilizando el algoritmo de decorrelación presentado en la sección 3.2 y que reproducimos aquí de nuevo por conveniencia

$$\mathbf{V}(n+1) = \mathbf{V}(n) - \mu\mathbf{V}(n)(\mathbf{z}(n)\mathbf{z}^T(n) - \mathbf{I}) \quad (3.55)$$

Como ya explicamos, este es un algoritmo de tipo gradiente relativo lo cual va a ser extremadamente útil en los desarrollos posteriores. Adicionalmente, el algoritmo exhibe la propiedad de equivarianza y su convergencia será independiente de la forma de la matriz de mezcla.

La decorrelación por sí sola no garantiza la recuperación de las fuentes. Por esta razón, las observaciones decorreladas, \mathbf{z} , son procesadas por una segunda etapa que vamos a representar por una matriz \mathbf{U} de dimensión $N \times N$. Imponiendo la condición de que \mathbf{U} sea una matriz unitaria se garantiza que la salida $\mathbf{y} = \mathbf{U}^T\mathbf{x}$ siempre cumple la restricción $\mathbf{R}_y = \mathbf{I}$.

El siguiente paso es adaptar la matriz \mathbf{U} con el objetivo de optimizar el contraste (3.54). Para ello, podrían utilizarse métodos de ortogonalización (como rotaciones planas [72]) pero en [13] se opta por utilizar un algoritmo de gradiente relativo

$$\begin{aligned} \mathbf{U}(n+1) &= \mathbf{U}(n) - \mu\mathbf{U}(n)\mathbf{U}^T(n)\hat{\nabla}_{\mathbf{U}}J \\ &= \mathbf{U}(n) - \mu\mathbf{U}(n)\mathbf{U}^T(n)\mathbf{x}(n)\mathbf{f}^T(\mathbf{y}(n)) \\ &= \mathbf{U}(n) - \mu\mathbf{U}(n)\mathbf{y}(n)\mathbf{f}^T(\mathbf{y}(n)) \end{aligned} \quad (3.56)$$

donde $\mathbf{f}(\mathbf{y}) = [h'(y_1), \dots, h'(y_N)]^T$ es un vector de funciones no lineales de \mathbf{y} . En cada iteración de este algoritmo se obtiene una matriz $\mathbf{U} - \mu\mathbf{U}\mathcal{E}$ que cumple lo siguiente

$$(\mathbf{U} - \mu\mathbf{U}\mathcal{E})^T(\mathbf{U} - \mu\mathbf{U}\mathcal{E}) = \mathbf{I} - \mu\mathcal{E} - \mu\mathcal{E}^T + \mu^2\mathcal{E}^T\mathcal{E} \quad (3.57)$$

Si suponemos que el término $\mu^2 \mathcal{E}^T \mathcal{E}$ es despreciable podemos concluir que la matriz \mathbf{U} obtenida en cada iteración será unitaria sólo cuando se verifica la condición $\mathcal{E} = -\mathcal{E}^T$

$$(\mathbf{U} + \mu \mathcal{E} \mathbf{U})^T (\mathbf{U} + \mu \mathcal{E} \mathbf{U}) = \mathbf{I} \iff \mathcal{E} = -\mathcal{E}^T \quad (3.58)$$

Por tanto, la matriz \mathbf{U} será unitaria si es actualizada con una matriz antisimétrica \mathcal{E} . Por lo que el algoritmo finalmente considerado para actualizar la matriz \mathbf{U} es el siguiente

$$\mathbf{U}(n+1) = \mathbf{U}(n) - \mu \mathbf{U}(n) \left(\mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \quad (3.59)$$

Esta regla de adaptación no preserva exactamente la propiedad de ortogonalidad sino sólo hasta el primer orden en μ . Sin embargo, esta dificultad desaparece cuando se consideran la etapa de blanqueo y separación conjuntamente.

El haber utilizado algoritmos de gradiente relativo tiene como principal ventaja el poder agrupar en una única recursión los algoritmo (3.55) y (3.59) propuestos para cada una de las etapas del sistema de separación. En efecto, si tenemos en cuenta que $\mathbf{W} = \mathbf{V} \mathbf{U}$ son los coeficientes conjuntos de las etapas de blanqueo y de separación, podemos multiplicar (3.55) y (3.59) y obtener la recursión que describe la evolución de estos coeficientes

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{V}(n+1) \mathbf{U}(n+1) \\ &= \left(\mathbf{V}(n) - \mu \mathbf{V}(n) (\mathbf{z}(n) \mathbf{z}^T(n) - \mathbf{I}) \right) \\ &\quad \left(\mathbf{U}(n) - \mu \mathbf{U}(n) \left(\mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \right) \\ &= \mathbf{V}(n) \left(\mathbf{I} - \mu (\mathbf{z}(n) \mathbf{z}^T(n) - \mathbf{I}) \right) \\ &\quad \mathbf{U}(n) \left(\mathbf{I} - \mu \left(\mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \right) \\ &= \mathbf{V}(n) \mathbf{U}(n) \mathbf{U}^T(n) \left(\mathbf{I} - \mu (\mathbf{z}(n) \mathbf{z}^T(n) - \mathbf{I}) \right) \\ &\quad \mathbf{U}(n) \left(\mathbf{I} - \mu \left(\mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \right) \\ &= \mathbf{V}(n) \mathbf{U}(n) \left(\mathbf{I} - \mu (\mathbf{y}(n) \mathbf{y}^T(n) - \mathbf{I}) \right) \\ &\quad \left(\mathbf{I} - \mu \left(\mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \right) \end{aligned} \quad (3.60)$$

Desarrollando el producto y despreciando los términos multiplicados por μ^2 obtenemos la expresión del algoritmo EASI

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \mathbf{W}(n) \left(\mathbf{y}(n) \mathbf{y}^T(n) - \mathbf{I} + \mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \quad (3.61)$$

Es interesante notar que el algoritmo EASI es equivariante: si multiplicamos (3.61) por \mathbf{A}^T obtenemos que la evolución de los coeficientes de mezcla/separación no depende de los parámetros que hayan intervenido en la mezcla de las fuentes

$$\mathbf{G}^T(n+1) = \mathbf{G}^T(n) - \mu \mathbf{G}^T(n) \left(\mathbf{y}(n) \mathbf{y}^T(n) - \mathbf{I} + \mathbf{y}(n) \mathbf{f}^T(\mathbf{y}(n)) - \mathbf{f}(\mathbf{y}(n)) \mathbf{y}^T(n) \right) \quad (3.62)$$

Esta propiedad hace que el algoritmo EASI sea especialmente atractivo para situaciones en las que la matriz de mezcla está mal condicionada.

3.6.1 Estabilidad del algoritmo EASI

En esta sección estudiaremos la estabilidad del algoritmo EASI (3.61) en los puntos donde se consigue la separación de las fuentes. Realizar este análisis considerando la expresión del algoritmo en función de los pesos del sistema de separación \mathbf{W} no es sencillo. Sin embargo, la complejidad se reduce notablemente si observamos la relación existente entre los coeficientes \mathbf{W} y la matriz de amplitudes $\mathbf{G} = \mathbf{W}^T \mathbf{A}$. Si \mathbf{A} es una matriz invertible de rango completo entonces existe una correspondencia uno a uno entre \mathbf{G} y \mathbf{W} , lo que permite realizar el estudio de los puntos de equilibrio a partir de la evolución de \mathbf{G} . En [13] se ha presentado un análisis de la estabilidad utilizando el criterio de Liapunov comentado en el capítulo 2. Aquí, sin embargo, presentaremos un estudio de la estabilidad en base al signo de los autovalores de la matriz jacobiana.

Como ya se ha comentado anteriormente, la evolución de $\mathbf{G} = \mathbf{W}^T \mathbf{A}$ viene dada por (3.62). Sus puntos de equilibrio se corresponden con las raíces de la ecuación diferencial ordinaria asociada a la recursión anterior

$$\frac{d\mathbf{G}^T}{dt} = -\mathbf{G}^T \left(E[\mathbf{y}\mathbf{y}^T] - \mathbf{I} + E[\mathbf{y}\mathbf{f}^T(\mathbf{y})] - E[\mathbf{f}(\mathbf{y})\mathbf{y}^T] \right) = \mathbf{0} \quad (3.63)$$

donde $\mathbf{0}$ es una matriz toda cero. En el caso $N = 2$ se obtiene un sistema de cuatro ecuaciones

$$\begin{aligned} F_{11} &= \frac{dg_{11}}{dt} = -g_{11}(E[y_1^2] - 1) - g_{21}(E[y_2 y_1] + E[y_2 f(y_1)] - E[f(y_2) y_1]) \\ F_{12} &= \frac{dg_{12}}{dt} = -g_{12}(E[y_1^2] - 1) - g_{22}(E[y_2 y_1] + E[y_2 f(y_1)] - E[f(y_2) y_1]) \\ F_{21} &= \frac{dg_{21}}{dt} = -g_{11}(E[y_2 y_1] + E[y_1 f(y_2)] - E[f(y_1) y_2]) - g_{21}(E[y_2^2] - 1) \\ F_{22} &= \frac{dg_{22}}{dt} = -g_{12}(E[y_2 y_1] + E[y_1 f(y_2)] - E[f(y_1) y_2]) - g_{22}(E[y_2^2] - 1) \end{aligned} \quad (3.64)$$

Este sistema de ecuaciones tiene múltiples soluciones, de las cuales son deseadas únicamente aquellas en las que cada salida extrae una única fuente

$$\text{Sol. 1 } y_1 = \lambda_1 s_1 \quad y_2 = \lambda_2 s_2 \quad (3.65)$$

$$\text{Sol. 2 } y_1 = \lambda_1 s_2 \quad y_2 = \lambda_2 s_1 \quad (3.66)$$

El estudio de la estabilidad del algoritmo en estos puntos es análogo, por lo que centraremos nuestra atención en el primero: $g_{11} = \lambda_1$, $g_{12} = g_{21} = 0$ y $g_{22} = \lambda_2$. Obviamente, las salidas (3.65) serán estadísticamente independientes por lo que la mayoría

de los términos en (3.64) serán cero

$$\begin{aligned} F_{11} &= -\lambda_1(E[\lambda_1^2 s_1^2] - 1) \\ F_{12} &= F_{21} = 0 \\ F_{22} &= -\lambda_2(E[\lambda_2^2 s_2^2] - 1) \end{aligned} \quad (3.67)$$

Recordemos que estamos suponiendo que las fuentes tiene potencia unidad ya que cambios en el valor de la potencia pueden ser incluidos dentro de la matriz de mezcla. Bajo esta hipótesis es inmediato comprobar que las ecuaciones (3.67) se anulan cuando las amplitudes son iguales a uno, i.e., $\lambda_1^2 = \lambda_2^2 = 1$. En particular, existe una solución en el punto donde las salidas recuperan exactamente las fuentes sin cambiar su amplitud

$$\mathbf{G} = \mathbf{I} \Rightarrow y_1 = s_1, \quad y_2 = s_2 \quad (3.68)$$

Este punto será estable si todos los autovalores de la matriz jacobiana recogida en el apéndice D tienen parte real negativa

$$\mathbf{H} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -1 - E[f'(s_1)] + E[f(s_2)s_2] & -1 - E[f(s_1)s_1] + E[f'(s_2)] & 0 \\ 0 & -1 - E[f(s_2)s_2] + E[f'(s_1)] & -1 - E[f'(s_2)] + E[f(s_1)s_1] & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} \quad (3.69)$$

La matriz \mathbf{H} tiene cuatro autovalores: $e_1 = e_2 = e_3 = -2$ y $e_4 = \mathcal{K}_1 + \mathcal{K}_2$ siendo $\mathcal{K}_i = E[s_i f(s_i)] - E[f'(s_i)]$. Es evidente que todos los autovalores tendrán parte real negativa si se verifica la siguiente condición

$$\mathcal{K}_1 + \mathcal{K}_2 < 0 \Rightarrow E[s_1 f(s_1)] - E[f'(s_1)] + E[s_2 f(s_2)] - E[f'(s_2)] < 0 \quad (3.70)$$

Por tanto, para que los puntos donde se lleva a cabo la separación sean atractores del algoritmo EASI, la no-linealidad $f(\cdot)$ debe ser elegida de forma que se cumpla la condición (3.70).

s_i	$-1/\sqrt{2p_i}$	0	$1/\sqrt{2p_i}$
$p(s_i)$	p_i	$1 - 2p_i$	p_i

Tabla 3.3: Distribución de probabilidad discreta empleada para verificar las condiciones de estabilidad del algoritmo EASI.

A continuación ilustraremos estos resultados considerando una no-linealidad de la forma $f(y_i) = \tanh(y_i)$ y un entorno formado por dos fuentes discretas que tienen la distribución mostrada en la tabla 3.3, que ya hemos empleado anteriormente. Para este caso \mathcal{K}_i viene dada por la siguiente expresión

$$\mathcal{K}_i = \sqrt{2p_i} \tanh\left(\frac{1}{\sqrt{2p_i}}\right) + 2p_i \tanh^2\left(\frac{1}{\sqrt{2p_i}}\right) - 1 \quad (3.71)$$

Evaluando (3.71) para distintos valores de p_i se han obtenido los resultados recogidos en la tabla 3.4 donde k_i denota la curtosis de las fuentes. La región de estabilidad impuesta por la condición (3.70) es la mostrada en la figura 3.8 (izquierda). Como puede observarse, el algoritmo es estable si las fuentes en el entorno tienen curtosis positiva ($p < 0.167$) pero su empleo es muy limitado para aplicaciones donde todas las fuentes del entorno tienen curtosis negativa ($p > 0.167$). Esta limitación puede ser superada fácilmente observando que la elección de la no-linealidad $f(y) = -\tanh(y)$ tiene como región de estabilidad la complementaria a la obtenida para $f(y) = \tanh(y)$ como se refleja en la figura 3.8 (derecha).

p_i	k_i	\mathcal{K}
0.05	7	-0.5856
0.10	2	-0.3718
0.15	0.333	-0.2096
0.20	-0.5	-0.0813
0.25	-1	0.0228
0.30	-1.33	0.1088
0.35	-1.57	0.1810
0.40	-1.75	0.2425
0.45	-1.889	0.2955
0.50	-2	0.3416

Tabla 3.4: Estadísticos de las fuentes empleadas para verificar las condiciones de estabilidad del algoritmo EASI para $f(y) = \tanh(y)$.

Como aplicación de la no-linealidad $f(y) = -\tanh(y)$ hemos considerado la separación de dos señales con $p = 0.5$ (caso binario) mezcladas utilizando la siguiente matriz mal condicionada

$$\mathbf{A} = \begin{bmatrix} 1.01 & 1 \\ 1.01 & 1.01 \end{bmatrix} \quad (3.72)$$

En la fase de separación se ha empleado un sistema MIMO con dos entradas y dos salidas cuyos coeficientes (aleatoriamente inicializados) fueron actualizados utilizando el algoritmo EASI con $\mu = 10^{-2}$. La figura 3.9 muestra la evolución de las amplitudes $\mathbf{G} = \mathbf{W}^T \mathbf{A}$. Como puede verse el algoritmo ha separado las fuentes y ha alcanzado los valores teóricos $\mathbf{G} = \mathbf{W}^T \mathbf{A} = \mathbf{I}$.

3.7 Algoritmos de maximización de la transferencia de información

Uno de los paradigmas del aprendizaje supervisado en redes neuronales es el principio de preservación de la información (*infomax*) propuesto por Linsker [29]. De acuerdo con este

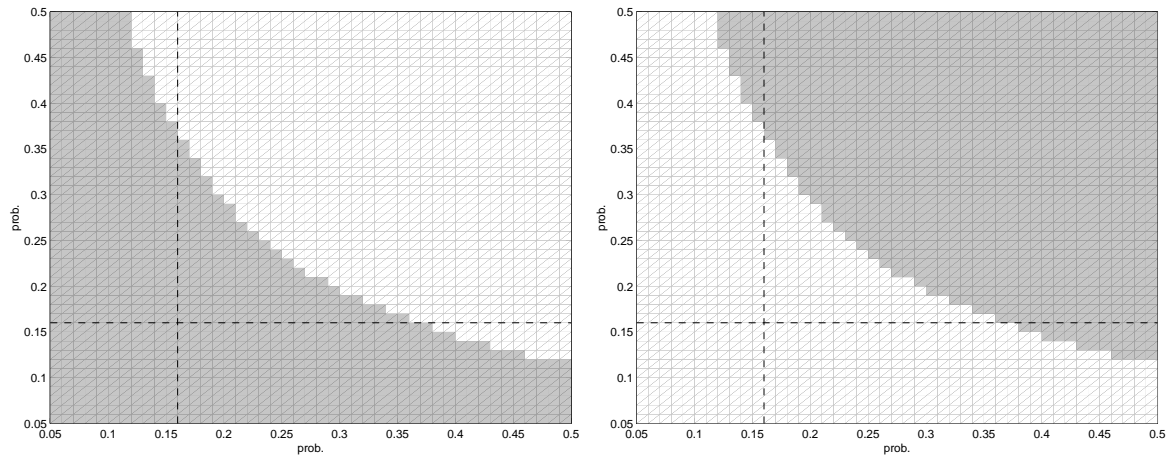


Figura 3.8: Condiciones de estabilidad del algoritmo EASI para $f(y) = \tanh(y)$ (izquierda) y $f(y) = -\tanh(y)$ (derecha).

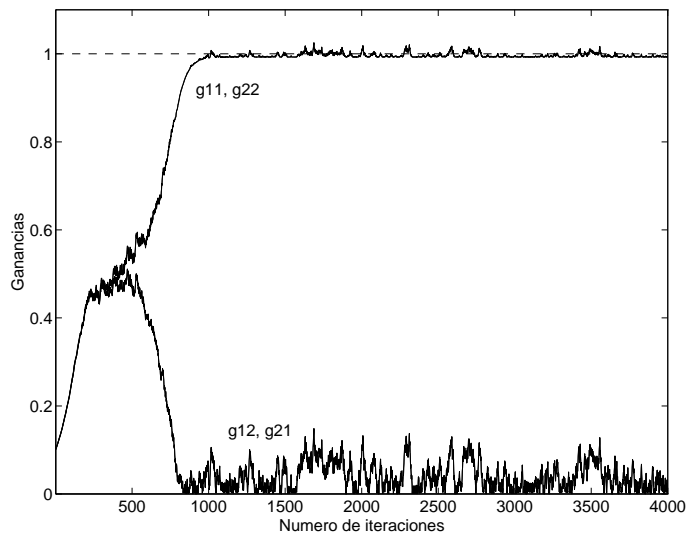


Figura 3.9: Evolución de las ganancias del sistema separador ajustado con el algoritmo EASI empleando $f(y) = -\tanh(y)$.

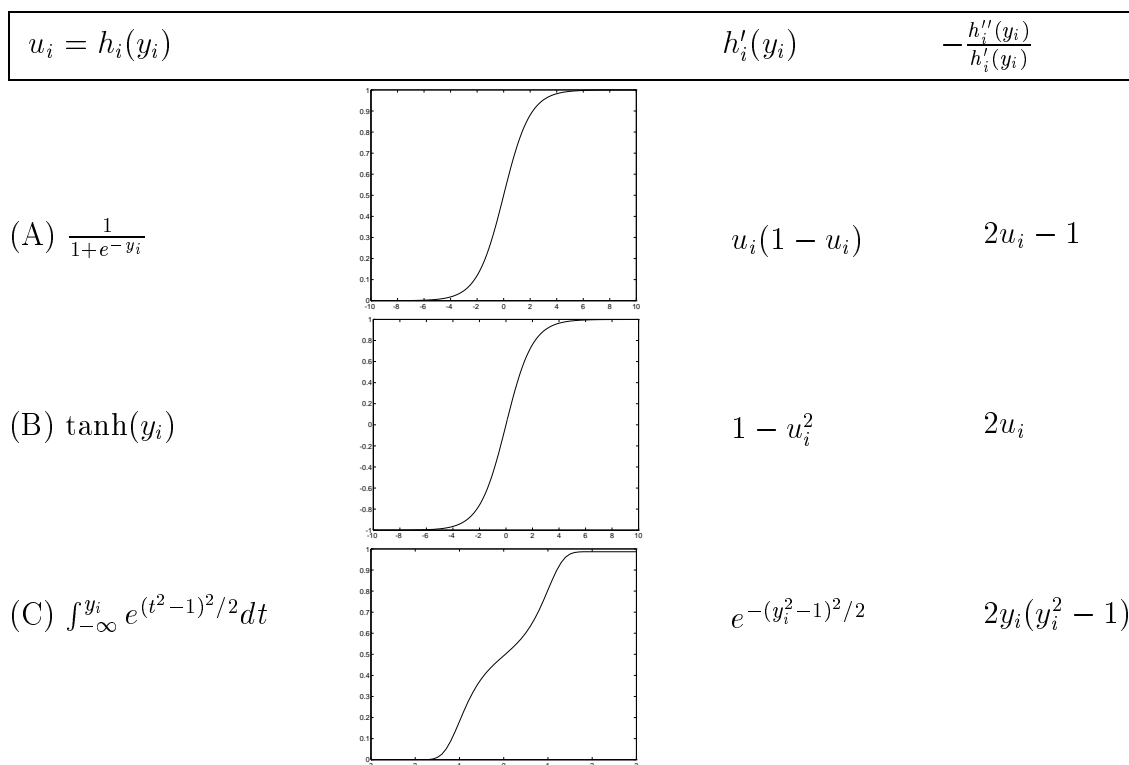


Figura 3.10: Ejemplos de funciones de activación.

principio, los parámetros de una red neuronal se ajustan con el objetivo de maximizar la transferencia de información entre la entrada y la salida. Se conjetura que determinados mecanismos de aprendizaje de los seres vivos funcionan de esta manera.

El algoritmo de Bell y Sejnowski [4] es un algoritmo de gradiente que maximiza la transferencia de información entre la entrada y la salida de una red neuronal no lineal de una sola capa que tenga el mismo número de entradas que de salidas. Se supone que la red es activada por las observaciones \mathbf{x} para producir el vector de estados de activación, $\mathbf{y} = \mathbf{W}^T \mathbf{x}$, donde \mathbf{W} es una matriz cuadrada $N \times N$ que contiene los pesos sinápticos. Posteriormente, la salida final es una función no lineal del estado $\mathbf{u} = \mathbf{h}(\mathbf{y})$. En redes neuronales las funciones de activación que se pueden utilizar son muy diversas. Cuando se manejan muestras de señales continuas se suele pedir que sean monotonamente crecientes, acotadas e invertibles. En la figura 3.10 se muestran algunos ejemplos: las no-linealidades (A) y (B) son dos funciones comúnmente usadas en la construcción de redes neuronales artificiales [40] (sigmoide y tangente hiperbólica) mientras que (C) surge como una aplicación del criterio de filtrado adaptativo conocido como Constant Modulus (CM)[53].

El algoritmo de Bell y Sejnowski persigue la maximización de la transferencia de información entre la entrada \mathbf{x} y la salida después de la no linealidad \mathbf{u} . Si llamamos

$H(\mathbf{u})$ a la entropía de \mathbf{u} y $H(\mathbf{u}|\mathbf{x})$ a la incertidumbre sobre \mathbf{u} que no es resuelta al observar \mathbf{x} entonces podemos escribir (ver apéndice C) la información mutua, $I(\mathbf{u}, \mathbf{x})$, como

$$I(\mathbf{u}, \mathbf{x}) = H(\mathbf{u}) - H(\mathbf{u}|\mathbf{x}) \quad (3.73)$$

Ahora bien, como estamos suponiendo que el comportamiento de la red es determinista (por ejemplo, la red no introduce ruido), el término $H(\mathbf{u}|\mathbf{x})$ es cero y la información mutua coincide con la entropía de la salida

$$I(\mathbf{u}, \mathbf{x}) = H(\mathbf{u}) \quad (3.74)$$

A continuación vamos a expresar la entropía de la salida $H(\mathbf{u})$ en función de la entropía de la entrada $H(\mathbf{x})$. La entropía de \mathbf{u} se define como

$$H(\mathbf{u}) = -E[\ln(p(\mathbf{u}))] \quad (3.75)$$

donde $p(\mathbf{u})$ es la función densidad de probabilidad de \mathbf{u} . Como las funciones de activación $h_i(y_i)$, $i = 1 \cdots N$ son invertibles, podemos escribir la f.d.p. de la salida \mathbf{u} en función de la f.d.p. del estado \mathbf{y} como sigue [65]

$$p(\mathbf{u}) = \frac{p(\mathbf{y})}{\prod_{i=1}^N h'_i(y_i)} \quad (3.76)$$

Por tanto,

$$H(\mathbf{u}) = -E[\ln p(\mathbf{u})] = -E \left[\ln \left(\frac{p(\mathbf{y})}{\prod_{i=1}^N h'_i(y_i)} \right) \right] = H(\mathbf{y}) + E \left[\ln \left(\prod_{i=1}^N h'_i(y_i) \right) \right] \quad (3.77)$$

donde $H(\mathbf{y})$ es la entropía del estado \mathbf{y} . A su vez, de la relación lineal entre las salidas y las observaciones $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ resulta que $p(\mathbf{y}) = p(\mathbf{x}) / \det(\mathbf{W}^T)$ [65] y, por tanto, podemos expresar $H(\mathbf{y})$ como

$$H(\mathbf{y}) = -E \left[\ln \left(\frac{p(\mathbf{x})}{\det(\mathbf{W}^T)} \right) \right] = H(\mathbf{x}) + \ln(\det(\mathbf{W}^T)) \quad (3.78)$$

donde $H(\mathbf{x})$ es la entropía de la entrada \mathbf{x} . Finalmente, sustituyendo (3.78) en (3.77) se obtiene que

$$\begin{aligned} H(\mathbf{u}) &= H(\mathbf{x}) + \ln(\det(\mathbf{W}^T)) + E \left[\ln \left(\prod_{i=1}^N h'_i(y_i) \right) \right] \\ &= H(\mathbf{x}) + \ln(\det(\mathbf{W}^T)) + \sum_{i=1}^N E[\ln(h'_i(y_i))] \end{aligned} \quad (3.79)$$

Esta expresión nos indica que la incertidumbre sobre la salida de la red \mathbf{u} es la misma que existe sobre la entrada \mathbf{x} más un conjunto de términos que dependen de los pesos de la red

neuronal. El término $H(\mathbf{x})$ que aparece en la expresión anterior puede eliminarse ya que no depende de \mathbf{W} con lo que podemos establecer que la maximización de la información mutua $I(\mathbf{u}, \mathbf{x})$ equivale a maximizar la siguiente función de coste

$$J_{MI}(\mathbf{W}) \stackrel{def}{=} \ln(\det(\mathbf{W}^T)) + \sum_{i=1}^N E[\ln(h'_i(y_i))] \quad (3.80)$$

El siguiente paso es obtener un algoritmo adaptativo que maximice $J_{MI}(\mathbf{W})$. Para ello se puede utilizar un algoritmo de gradiente de máxima pendiente [4] o un algoritmo de gradiente relativo [1, 53]. A partir de las derivadas dadas en el apéndice A es sencillo obtener que el gradiente de (3.80) es el siguiente

$$\begin{aligned} \nabla_{\mathbf{W}} J &= \nabla_{\mathbf{W}} (\ln(\det(\mathbf{W}^T))) + \nabla_{\mathbf{W}} \left(\sum_{i=1}^N E[\ln(h'_i(y_i))] \right) \\ &= \frac{\text{adj}(\mathbf{W}^T)}{\det(\mathbf{W}^T)} - E[\mathbf{x}\mathbf{g}^T(\mathbf{y})] = \mathbf{W}^{-T} - E[\mathbf{x}\mathbf{g}^T(\mathbf{y})] \end{aligned} \quad (3.81)$$

donde $\mathbf{g}(\mathbf{y}) = [-h''_1(y_1)/h'_1(y_1), \dots, -h''_N(y_N)/h'_N(y_N)]^T$ es un vector de no linealidades que depende de la función de activación que se esté utilizando (la tercera columna de la figura 3.10 recoge algunas posibilidades). Finalmente, estimando las esperanzas que aparecen en (3.81) con una única muestra se obtienen las siguientes reglas de aprendizaje:

- *Algoritmo de gradiente ascendente de máxima pendiente:*

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu (\mathbf{W}(n)^{-T} - \mathbf{x}(n) \mathbf{g}^T(\mathbf{y}(n))) \quad (3.82)$$

- *Algoritmo de gradiente relativo:* se obtiene a partir del anterior multiplicando el factor de incremento en (3.82) por $\mathbf{W}\mathbf{W}^T$

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu \mathbf{W}(n) \mathbf{W}^T(n) (\mathbf{W}(n)^{-T} - \mathbf{x}(n) \mathbf{g}^T(\mathbf{y}(n))) \\ &= \mathbf{W}(n) + \mu \mathbf{W}(n) (\mathbf{I} - \mathbf{y}(n) \mathbf{g}^T(\mathbf{y}(n))) \end{aligned} \quad (3.83)$$

Bell y Sejnowski [4] conjeturan que el algoritmo (3.82) es útil para separación ciega de fuentes si se interpreta la entrada \mathbf{x} como las observaciones, la parte lineal de la red \mathbf{W} como el sistema de separación y el estado $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ como las salidas que deben recuperar las fuentes. Su conjetura se basa en que la expresión (3.77) de la información mutua entre \mathbf{u} y \mathbf{x} , y que reproducimos a continuación,

$$I(\mathbf{u}, \mathbf{x}) = H(\mathbf{u}) = -E \left[\ln \left(\frac{p(\mathbf{y})}{\prod_{i=1}^N h'_i(y_i)} \right) \right] \quad (3.84)$$

se puede interpretar como la distancia de Kullback-Leibler entre la f.d.p. de la salida $p(\mathbf{y})$ y una f.d.p. de referencia $\hat{p}(\mathbf{y}) = \prod_{i=1}^N h'_i(y_i)$. Si ahora escogemos cada no linealidad de forma que $h'_i(\cdot)$ se corresponda con la f.d.p. de cada fuente, entonces $\hat{p}(\mathbf{s})$ es la f.d.p. de las fuentes y al maximizar $I(\mathbf{u}, \mathbf{x})$ se está forzando a que \mathbf{y} tenga la misma f.d.p. que las fuentes. En [59] se llega a esta misma elección de las no linealidades partiendo del principio de reducción de la redundancia entre las salidas cuyo objetivo es extraer las características estadísticamente independientes que mejor representan a las entradas.

Del razonamiento anterior Bell y Sejnowski deducen que para llevar a cabo la separación de las fuentes las no linealidades deben escogerse de forma que sean las funciones de distribución de las fuentes. No obstante, al no realizar un estudio de la estabilidad de los algoritmos resultantes, numerosas cuestiones permanecen sin resolver. Por ejemplo, la elección de las no linealidades como funciones de distribución de las fuentes en absoluto garantiza que no existan otros máximos no deseados que puedan dañar la convergencia del algoritmo. Más aún, las distribuciones de las fuentes es una información que en muchas situaciones prácticas es desconocida. Por otro lado, si en el algoritmo se utilizan unas no linealidades distintas de las funciones de distribución de las fuentes, no está claro bajo qué condiciones los puntos donde se consigue la separación son atractores del algoritmo. En la siguiente subsección se presenta un análisis de la estabilidad del algoritmo (3.83) en el que se da una respuesta a algunas de estas cuestiones.

Es interesante notar la relación que existe entre los algoritmos (3.82) y (3.83) y otros presentados anteriormente. En primer lugar, el algoritmo de máxima pendiente (3.82) tiene ciertas semejanzas con el algoritmo p-vector dado en la ecuación (2.32): cuando el sistema separa las fuentes sin incorporar ninguna amplitud ($\mathbf{W}^T \mathbf{A} = \mathbf{I}$) el primer término en (3.82) conduce a $\mathbf{W}^{-T} = \mathbf{A}$. Además el término $-\mathbf{x}\mathbf{g}^T(\mathbf{y})$ puede entenderse como una generalización del término $-\mathbf{xy}^T$ que aparece en el algoritmo p-vector. Por otro lado, el algoritmo de gradiente relativo (3.83) se corresponde con la recursión (3.35) para el caso $\mathbf{f}(\mathbf{y}) = \mathbf{y}$ y $\mathbf{A} = \mathbf{I}$.

Un caso particular de (3.80) que admite una interesante interpretación es el que resulta de elegir la no linealidad (C) de la figura 3.10 como función de activación de la red neuronal [16]. Para este caso la maximización de la transferencia de la información (3.80) es equivalente a minimizar la siguiente función de coste

$$J_{CM}(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^N E \left[(y_i^2 - 1)^2 \right] - \ln(\det(\mathbf{W})) \quad (3.85)$$

Obsérvese que el primer término de esta función de coste es una extensión para sistemas MIMO de un criterio ampliamente utilizado en igualación ciega conocido con el nombre de Constant Modulus [38] cuyo algoritmo de gradiente asociado se denomina CMA (Constant Modulus Algorithm). De los estudios de estabilidad realizados acerca de este algoritmo en sistemas MISO (Multiple Input Single Output) se sabe que converge a aquellas situaciones en las que se extrae una única fuente. Sin embargo, la minimización del primer sumando en (3.85) es por sí sola insuficiente ya que varias salidas pueden extraer la misma fuente.

Esta situación indeseada es penalizada por el segundo término ya que la dependencia entre columnas de \mathbf{W} da como resultado un valor negativo muy grande de $\ln(\det(\mathbf{W}))$. Alternativamente, este término ha sido sustituido en [14, 36] por la suma de los valores absolutos de las correlaciones entre las salidas.

3.7.1 Estabilidad del algoritmo de Bell y Sejnowski

El algoritmo basado en estadísticos de orden superior propuesto por Amari y Cichocki (3.35) tiene la misma expresión que la versión relativa del desarrollado por Bell y Sejnowski (3.83) considerando la maximización de la transferencia de información en una red neuronal. Ambos actualizan los pesos mediante la siguiente regla de aprendizaje

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \mathbf{W}(n) (\mathbf{I} - \mathbf{y}(n) \mathbf{g}^T(\mathbf{y}(n))) \quad (3.86)$$

donde μ es el parámetro de paso y $\mathbf{g}(\cdot)$ es un vector formado por una transformación no-lineal de la salida cuya forma concreta depende del algoritmo bajo estudio. En particular, en aquellos que maximizan la transferencia de información, la no-linealidad $g(y_i) = -h''(y_i)/h'(y_i)$ se determina a partir de la primera y segunda derivada de la función de activación de una red neuronal. En este apartado consideraremos el caso en que todas las no linealidades son idénticas $h_i(\cdot) = h(\cdot)$. En [2] se ha realizado un primer análisis de la estabilidad de (3.86) del cual se deducen algunas condiciones suficientes que garantizan la correspondencia entre los atractores del algoritmo y los puntos de separación. Estas condiciones son más restrictivas que las que obtendremos a continuación. Parte de este análisis ha sido presentado en [27].

En esta sección analizaremos la estabilidad del algoritmo (3.86) en los puntos donde se lleva a cabo la separación considerando la evolución de los coeficientes del sistema conjunto de mezcla-separación obtenida multiplicando la recursión (3.86) por la transpuesta de la matriz de mezcla

$$\mathbf{G}^T(n+1) = \mathbf{G}^T(n) + \mu \mathbf{G}^T(n) (\mathbf{I} - \mathbf{y}(n) \mathbf{g}^T(\mathbf{y}(n))) \quad (3.87)$$

La ODE asociada a (3.87) es la siguiente

$$\frac{d\mathbf{G}^T}{dt} = \mathbf{G}^T (\mathbf{I} - E[\mathbf{y} \mathbf{g}^T(\mathbf{y})]) \quad (3.88)$$

Esta ecuación tiene múltiples raíces que se corresponden con los puntos estacionarios del algoritmo. De ellas son deseadas aquellas en las que cada salida extrae una única fuente. En particular, para el caso de dos fuentes y dos salidas existen dos puntos

$$\text{Sol. 1 } y_1 = \lambda_1 s_1 \quad y_2 = \lambda_2 s_2 \quad (3.89)$$

$$\text{Sol. 2 } y_1 = \lambda_1 s_2 \quad y_2 = \lambda_2 s_1 \quad (3.90)$$

El estudio para cualquiera de las dos soluciones es análogo por lo que centraremos nuestra atención en la primera. Haciendo uso de la independencia estadística de las

fuentes obtenemos que (3.88) se anula cuando los parámetros λ_i satisfacen las siguientes igualdades

$$\begin{aligned}\lambda_1 E[s_1 g(\lambda_1 s_1)] &= 1 \\ \lambda_2 E[s_2 g(\lambda_2 s_2)] &= 1\end{aligned}\quad (3.91)$$

El siguiente paso consiste en estudiar los autovalores de la matriz formada por las primeras derivadas de la ecuación diferencial ordinaria. Considerando las expresiones recogidas en el apéndice D se obtiene que esa matriz tiene la siguiente forma

$$\mathbf{H} = \begin{bmatrix} -1 - \lambda_1^2 \sigma_1 & 0 & 0 & 0 \\ 0 & -\lambda_2^2 m_1 & -\frac{\lambda_2}{\lambda_1} & 0 \\ 0 & -\frac{\lambda_1}{\lambda_2} & -\lambda_1^2 m_2 & 0 \\ 0 & 0 & 0 & -1 - \lambda_2^2 \sigma_2 \end{bmatrix}\quad (3.92)$$

donde

$$\begin{aligned}\sigma_1 &= E[s_1^2 g'(\lambda_1 s_1)] & \sigma_2 &= E[s_2^2 g'(\lambda_2 s_2)] \\ m_1 &= [g'(\lambda_1 s_1)] & m_2 &= [g'(\lambda_2 s_2)]\end{aligned}\quad (3.93)$$

La matriz (3.92) tiene cuatro autovalores dos de los cuales serán negativos si se cumple la siguiente condición

$$\lambda_i^2 \sigma_i > -1 \quad i = 1, 2\quad (3.94)$$

Adicionalmente, podemos garantizar que la parte real de los otros dos autovalores será negativa si la matriz central de (3.92) tiene traza real negativa y determinante real positivo lo que nos conduce a plantear las siguientes condiciones de estabilidad

$$m_i > 0, \quad i = 1, 2\quad (3.95)$$

$$\lambda_1^2 \lambda_2^2 m_1 m_2 > 1\quad (3.96)$$

Por tanto, aquellas no-linealidades $g(y_i)$ que verifiquen las condiciones (3.91), (3.94), (3.95) y (3.96) serán adecuadas para efectuar separación ciega de fuentes ya que los puntos de separación serán atractores de (3.87).

A continuación ilustraremos estos resultados considerando una no-linealidad de la forma $g(y_i) = 2 \tanh(y_i)$ que resulta de utilizar la función de activación $h(y_i) = \tanh(y_i)$ en una red que maximiza la transferencia de información. En este caso, la función de coste (3.80) toma la siguiente forma

$$J_{MI}(\mathbf{W}) = \sum_{i=1}^N E[\ln(1 - \tanh^2(y_i))] + \ln(\det(\mathbf{W}^T))\quad (3.97)$$

s_i	$-1/\sqrt{2p_i}$	0	$1/\sqrt{2p_i}$
$p(s_i)$	p_i	$1 - 2p_i$	p_i

Tabla 3.5: Distribución de probabilidad discreta utilizada para verificar las condiciones de estabilidad del algoritmo de Bell y Sejnowski.

Es imposible resolver (3.91) de forma general para encontrar los valores de las amplitudes λ_i en los puntos donde se consigue la separación. Por ello, a continuación supondremos un entorno concreto con fuentes de naturaleza discreta distribuidas como se muestra en la tabla 3.5.

Bajo estas consideraciones, la ecuación (3.91) y los promedios (3.93) toman la siguiente forma

$$2\lambda\sqrt{2p_i} \tanh\left(\frac{\lambda}{\sqrt{2p_i}}\right) = 1 \quad (3.98)$$

$$\sigma_i = 2 \left(1 - \tanh^2\left(\frac{\lambda}{\sqrt{2p_i}}\right)\right) \quad (3.99)$$

$$m_i = -4p_i \tanh^2\left(\frac{\lambda}{\sqrt{2p_i}}\right) + 2 \quad (3.100)$$

Resolviendo numéricamente la ecuación (3.91) para distintos valores de p_i , se han obtenido los valores de λ_i mostrados en la tabla 3.6. Esta tabla también recoge los valores de σ_i y m_i obtenidos a partir de (3.99) y (3.100) y el valor de la curtosis k_i para distintas probabilidades. De los valores en esta tabla se deduce que las condiciones (3.94) y (3.95) se verifican independientemente del valor de p_i ya que σ_i y m_i son siempre positivos.

p_i	k_i	λ_i	σ_i	m_i
0.0500	7	1.5868	0.0004	1.8000
0.1000	2	1.1323	0.0499	1.6100
0.1500	0.333	0.9678	0.2205	1.4661
0.2000	-0.5	0.8910	0.4256	1.3702
0.2500	-1	0.8483	0.6104	1.3052
0.3000	-1.33	0.8214	0.7649	1.2590
0.3500	-1.57	0.8030	0.8924	1.2246
0.4000	-1.75	0.7897	0.9978	1.1983
0.4500	-1.889	0.7796	1.0859	1.1774
0.5000	-2	0.7717	1.1604	1.1604

Tabla 3.6: Parámetros de estabilidad del algoritmo de Bell y Sejnowski para $g(y) = 2 \tanh(y)$.

La gráfica 3.11 muestra la región de estabilidad impuesta por la condición (3.96). Puede observarse que la estabilidad del algoritmo está garantizada si todas las fuentes

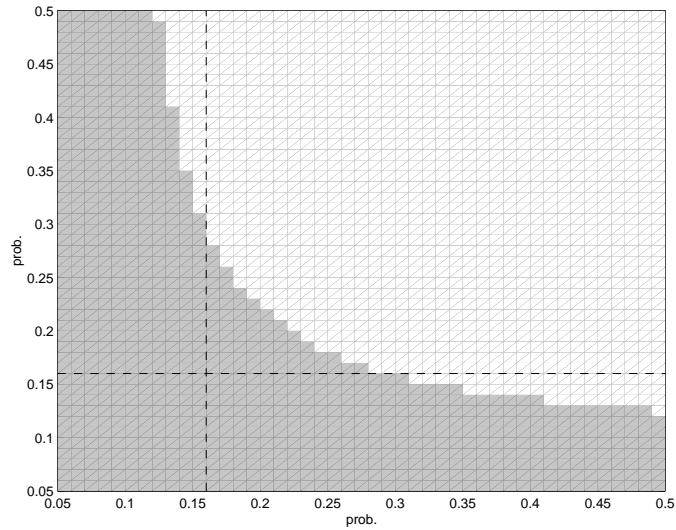


Figura 3.11: Condiciones de estabilidad del algoritmo de Bell y Sejnowski para $g(y) = 2 \tanh(y)$.

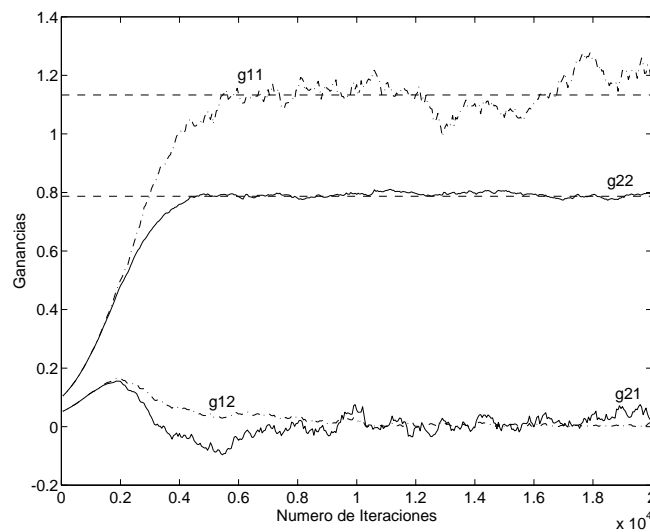


Figura 3.12: Evolución de las ganancias del sistema separador ajustado con el algoritmo de Bell y Sejnowski empleando $g(y) = 2 \tanh(y)$.

en el entorno tienen curtosis positiva ($p < 0.167$). Por otro lado, el algoritmo no es asintóticamente estable en la mayoría de los casos en los que alguna de las fuentes tiene curtosis negativa ($p > 0.167$), por lo que el empleo de $g(y) = 2 \tanh(y)$ en aplicaciones de comunicaciones es muy limitada. También es interesante notar que existen ciertas combinaciones de fuentes super-gaussianas y sub-gaussianas para las cuales se cumple la condición (3.96). Por ejemplo, el algoritmo es estable si las observaciones son mezclas de dos fuentes, una de ellas sub-gaussiana con curtosis $k_1 = -1.75$ ($p_1 = 0.4$) y otra super-gaussiana con curtosis $k_2 = 2$ ($p = 0.1$). Para esta situación, la figura 3.12 muestra la evolución de las amplitudes del sistema $\mathbf{G} = \mathbf{W}^T \mathbf{A}$ obtenidas con la siguiente matriz mal condicionada

$$\mathbf{A} = \begin{bmatrix} 1.01 & 1 \\ 1.01 & 1.01 \end{bmatrix} \quad (3.101)$$

En esta simulación se ha utilizado un sistema MIMO de dos entradas y dos salidas cuyos coeficientes (aleatoriamente inicializados) han sido actualizados empleando el algoritmo (3.83) con $\mu = 5 \times 10^{-4}$. Puede observarse como el algoritmo ha convergido a los coeficientes óptimos λ_i calculados numéricamente: $\lambda_1 = 0.7897$ y $\lambda_2 = 1.1323$ lo que refuerza nuestros resultados analíticos.

3.8 Maximización de la verosimilitud

Los criterios de separación ciega presentados anteriormente intentan encontrar un vector \mathbf{y} que tenga una f.d.p. lo más parecida posible al vector de fuentes \mathbf{s} . En esta sección nos planteamos un enfoque alternativo que consiste en determinar el estimador de máxima verosimilitud de los parámetros que han intervenido en la mezcla $\hat{\mathbf{A}}$, de forma que $\hat{\mathbf{A}}^{-1} \mathbf{A} \mathbf{s} = \mathbf{s}$ [10, 11].

Recordemos que el modelo de mezcla que estamos utilizando supone que las observaciones son mezclas lineales de las fuentes $\mathbf{x} = \mathbf{A} \mathbf{s}$ donde tanto \mathbf{A} como \mathbf{s} son desconocidas, las f.d.p. $p_{\mathbf{x}}(\mathbf{x}|\mathbf{A})$ y $p_{\mathbf{s}}(\mathbf{s})$ son también desconocidas. Establezcamos ahora el siguiente modelo de las observaciones

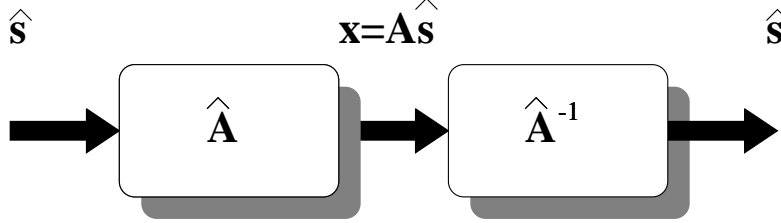
$$\mathbf{x} = \hat{\mathbf{A}} \hat{\mathbf{s}} \quad (3.102)$$

donde $\hat{\mathbf{s}}$ es un vector de fuentes hipotéticas con una f.d.p. conocida $p_{\hat{\mathbf{s}}}(\hat{\mathbf{s}})$ y $\hat{\mathbf{A}}$ es una estimación del sistema de mezcla que escogeremos de acuerdo al principio de máxima verosimilitud tal y como se explica a continuación.

Consideremos que disponemos de L realizaciones independientes de las observaciones $\mathbf{x}(1), \dots, \mathbf{x}(L)$. Al ser las observaciones estadísticamente independientes entre sí, su f.d.p. puede ser escrita de la siguiente forma

$$p_{\mathbf{x}}(\mathbf{x}(1), \dots, \mathbf{x}(L)|\hat{\mathbf{A}}) = \prod_{i=1}^L p_{\mathbf{x}}(\mathbf{x}(i)|\hat{\mathbf{A}}) \quad (3.103)$$

Modelo de mezcla hipotético:



Modelo de mezcla real:

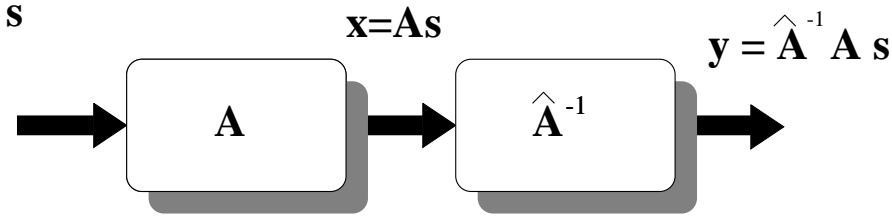


Figura 3.13: Modelo del sistema de mezcla/separación para el criterio de máxima verosimilitud.

Tomando logaritmos y dividiendo por el número de muestras obtenemos la verosimilitud logarítmica (log-likelihood) normalizada

$$\mathcal{L}(\hat{\mathbf{A}}) = \frac{1}{L} \log \left(\prod_{i=1}^L p_{\mathbf{x}}(\mathbf{x}(i)|\hat{\mathbf{A}}) \right) = \frac{1}{L} \sum_{i=1}^L \log(p_{\mathbf{x}}(\mathbf{x}(i)|\hat{\mathbf{A}})) \quad (3.104)$$

Si el número de observaciones es grande, $L \rightarrow \infty$, la función $\mathcal{L}(\hat{\mathbf{A}})$ tiende a su esperanza

$$\begin{aligned} \mathcal{L}(\hat{\mathbf{A}}) \rightarrow E[\mathcal{L}(\hat{\mathbf{A}})] &= \int p_{\mathbf{x}}(\mathbf{x}|\mathbf{A}) \log(p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}})) d\mathbf{x} = \int p_{\mathbf{x}}(\mathbf{x}|\mathbf{A}) \log \left(\frac{p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}}) p_{\mathbf{x}}(\mathbf{x}|\mathbf{A})}{p_{\mathbf{x}}(\mathbf{x}|\mathbf{A})} \right) d\mathbf{x} \\ &= -K(p_{\mathbf{x}}(\mathbf{x}|\mathbf{A}), p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}})) - H(\mathbf{x}; \mathbf{A}) \end{aligned} \quad (3.105)$$

donde $K(\cdot, \cdot)$ denota la divergencia de Kullback-Leibler y $H(\mathbf{x}; \mathbf{A})$ es la incertidumbre sobre las observaciones que no depende de $\hat{\mathbf{A}}$. Por ello, maximizar $\mathcal{L}(\hat{\mathbf{A}})$ es equivalente a minimizar la distancia de Kullback-Leibler entre $p_{\mathbf{x}}(\mathbf{x}|\mathbf{A})$ y $p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}})$. En otras palabras, el estimador de máxima verosimilitud de \mathbf{A} será una matriz $\hat{\mathbf{A}}$ que minimice la siguiente función de coste

$$J_{MV}(\hat{\mathbf{A}}) \stackrel{def}{=} K(p_{\mathbf{x}}(\mathbf{x}|\mathbf{A}), p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}})) \quad (3.106)$$

El razonamiento anterior nos conduce a plantear el sistema inversor (del modelo real e hipotético) mostrado en la figura 3.13. Como puede verse, al procesar el vector de observaciones hipotético $\mathbf{x} = \hat{\mathbf{A}}\hat{\mathbf{s}}$ con un sistema inversor representado por la matriz $\hat{\mathbf{A}}^{-1}$

se obtiene exactamente el vector de fuentes $\hat{\mathbf{s}}$ cuya f.d.p. viene dada por la siguiente expresión

$$p(\hat{\mathbf{s}}) = \frac{p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}})}{|\det(\hat{\mathbf{A}}^{-1})|} \quad (3.107)$$

Por otro lado, al aplicar el vector de observaciones real al mismo sistema inversor $\hat{\mathbf{A}}^{-1}$ se obtiene una salida $\mathbf{y} = \hat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{s}$ cuya f.d.p. es, en general, distinta a la de las fuentes

$$p(\hat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{s}) = \frac{p_{\mathbf{x}}(\mathbf{x}|\mathbf{A})}{|\det(\hat{\mathbf{A}}^{-1})|} \quad (3.108)$$

Sustituyendo (3.107) y (3.108) en la función de coste (3.106) resulta lo siguiente

$$J_{MV}(\hat{\mathbf{A}}) = K(p_{\mathbf{x}}(\mathbf{x}|\mathbf{A}), p_{\mathbf{x}}(\mathbf{x}|\hat{\mathbf{A}})) = K(p(\mathbf{y}), p(\hat{\mathbf{s}})) \quad (3.109)$$

Como consecuencia, al maximizar $J_{MV}(\hat{\mathbf{A}})$ estamos forzando que un vector $\mathbf{y} = \hat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{s}$ tenga componentes independientes.

Es importante hacer notar que el criterio de máxima verosimilitud es equivalente al de máxima transferencia de información de la sección 3.7. En efecto, anteriormente hemos determinado que la información mutua entre la salida $\mathbf{u} = \mathbf{f}(\mathbf{y})$ y las observaciones \mathbf{x} viene dada por la expresión (3.77) y que repetimos a continuación

$$I(\mathbf{u}, \mathbf{x}) = -E \left[\ln \left(\frac{p(\mathbf{y})}{\prod_{i=1}^N h_i'(y_i)} \right) \right] \quad (3.110)$$

Si elegimos como no-linealidades $h_i(y_i)$ las f.d.p. de las fuentes hipotéticas \hat{s}_i que, recordemos, son independientes entre sí

$$p_{\hat{\mathbf{s}}}(\mathbf{y}) = \prod_{i=1}^N h_i'(y_i) \quad (3.111)$$

y sustituimos (3.111) en (3.110) obtenemos la siguiente expresión de la información mutua

$$I(\mathbf{u}, \mathbf{x}) = -E \left[\ln \left(\frac{p_{\mathbf{y}}(\mathbf{y})}{\prod_{i=1}^N p_{\hat{\mathbf{s}}}(\mathbf{y})} \right) \right] \quad (3.112)$$

Como consecuencia, también la maximización de la información mutua conduce a la minimización de la distancia de Kullback-Leibler entre la f.d.p. de las salidas $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ y la f.d.p. conjunta de las fuentes hipotéticas. De la discusión anterior podemos deducir que tanto el criterio de máxima verosimilitud como el de maximización de la transferencia de información tienen como objetivo encontrar una salida estadísticamente independientes. Como se muestra en la figura 3.14, el papel de la matriz \mathbf{W} es el mismo que el de $\hat{\mathbf{A}}$. Adicionalmente, este es el mismo objetivo que persiguen los criterios basados en estadísticos de orden superior.

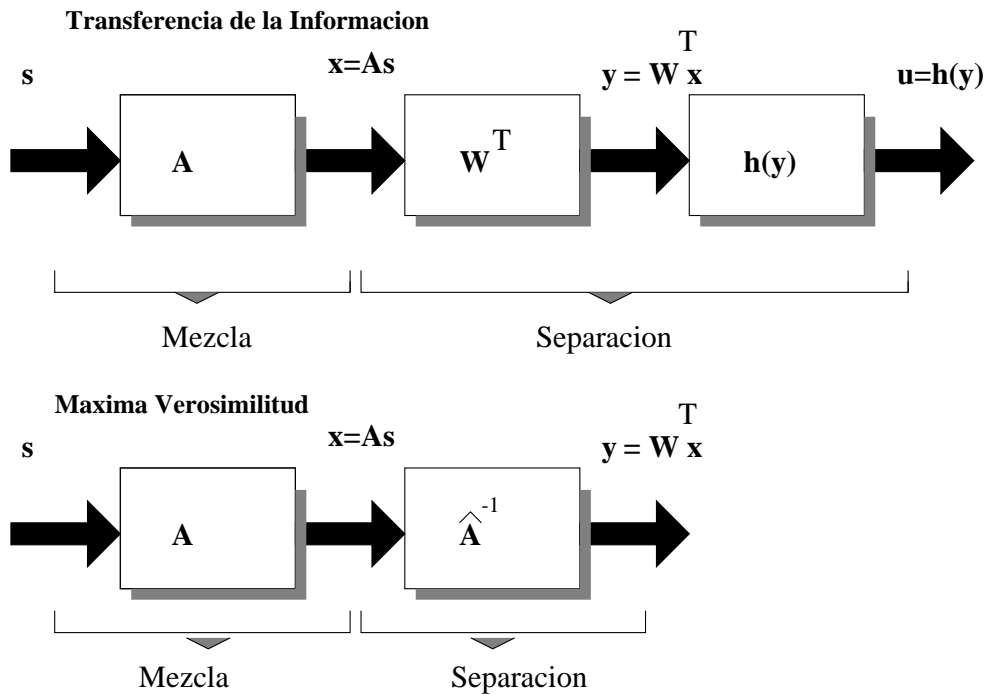


Figura 3.14: Comparación entre el criterio de máxima transferencia de información y el de máxima verosimilitud.

3.9 Conclusiones

En este capítulo se han revisado varios métodos de separación ciega de fuentes. Hemos comenzado exponiendo las limitaciones de los algoritmos de decorrelación para resolver el problema de separación bajo nuestras hipótesis de trabajo. Conscientes de esta limitación, Héroult y Jutten propusieron aprovechar la información contenida en los estadísticos de orden superior de las señales. El sistema de separación que utilizan estos autores es ligeramente diferente al que hemos presentado en el capítulo anterior ya que se consideran realimentaciones entre las salidas de los combinadores lineales. Los pesos son ajustados con una recursión que incorpora funciones no-lineales y que persigue la independencia estadística de las salidas. La principal ventaja de este método es su sencillez pero, en contrapartida, el análisis presentado en este capítulo revela que su convergencia depende de diversos factores como las características de las señales y las no-linealidades empleadas.

A partir del algoritmo de Héroult-Jutten han sido propuestos diversos métodos de separación que también utilizan estadísticos de orden superior. En particular, Cichocki y Undebauen han propuesto de forma heurística una modificación del algoritmo de Héroult y Jutten para actualizar los coeficientes del sistema MIMO presentado en el capítulo 1. Este algoritmo admite una interesante interpretación utilizando conceptos de Teoría de la Información. En concreto, puede entenderse como un algoritmo que maximiza la transferencia de información en una red neuronal monocapa. En este capítulo

hemos analizado la estabilidad de este algoritmo y hemos obtenido algunas condiciones suficientes que garantizan que los puntos donde se consiguen la separación son atractores del algoritmo.

Otro de los criterios que han centrado nuestro interés ha sido el propuesto por Comon bajo el nombre contraste. Básicamente, un contraste es una función de coste que cumple ciertas restricciones que garantizan la correspondencia entre sus máximos y los estados de separación. La mayoría de los contrastes propuestos hasta la fecha utilizan cumulantes de orden superior. En particular, en este capítulo hemos presentado y analizado una función de coste definida a partir de cumulantes de orden cuatro de las salidas. La maximización de esta función se realiza bajo la restricción de que las salidas tengan potencia igual a la unidad lo que trae consigo la necesidad de un sistema separador de dos etapas: la primera ajusta \mathbf{W} y la segunda blanquea las salidas. El análisis de la estabilidad que hemos presentado revela la inexistencia de puntos indeseados en esta función de coste.

A partir de las funciones de contraste, Cardoso y Laheld han propuesto un algoritmo conocido como EASI. Este ingenioso algoritmo se caracteriza por no necesitar ninguna etapa de blanqueo y, por tanto, su implementación es más sencilla que los basados en contrastes. Además, es un algoritmo de tipo relativo y, por ello, es adecuado para situaciones donde la matriz de mezcla está mal condicionada. En este capítulo hemos analizado su estabilidad y hemos deducido algunas condiciones que garantizan la convergencia a las soluciones deseadas.

Por último, hemos abordado la separación de las fuentes empleando estimadores de máxima verosimilitud de la matriz de mezcla. Aunque el planteamiento parece inicialmente distinto al de los otros métodos estudiados existe una clara relación entre ellos ya que todos buscan salidas estadísticamente independientes entre sí.

Capítulo 4

Nuevos criterios para separación ciega de fuentes

4.1 Introducción

Este capítulo tiene por objeto presentar una nueva familia de criterios estadísticos para la separación de fuentes no-gaussianas a partir únicamente de mezclas lineales e instantáneas de ellas. La idea de nuestra aproximación es adaptar los coeficientes del sistema separador de acuerdo a la maximización de una nueva función de coste que surge como extensión de un criterio para igualación ciega propuesto en [69] por Shalvi y Weinstein. El criterio propuesto por estos autores ha demostrado ser adecuado para extraer una señal de interés cancelando las interferencias. Sin embargo, resulta insuficiente para conseguir separación ya que la adaptación de los combinadores lineales se realiza de forma individual y, por ello, puede ocurrir que varias salidas extraigan la misma fuente.

Para superar la limitación del criterio de Shalvi y Weinstein en el presente capítulo proponemos añadir un conjunto de cumulantes cruzados que penalicen la dependencia estadística entre las salidas que aparece cuando la misma fuente es extraída por varios combinadores. Una de las principales ventajas de nuestra aproximación es que los coeficientes del sistemas de separación pueden ajustarse de forma adaptativa utilizando un sencillo algoritmo de gradiente cuya estabilidad ha sido analizada para el caso general de sistemas de mezcla y separación con N entradas y N salidas. El estudio permite establecer las condiciones bajo las cuales los puntos donde se consigue la separación se corresponden con máximos de la función de coste y, por tanto, son atractores del algoritmo. La función de coste contiene otros puntos estacionarios pero también se demuestra que son mínimos o puntos silla lo que se traduce en puntos inestables a los que nunca convergerá el algoritmo.

Como alternativa al empleo del cumulante de cuarto orden como término cruzado proponemos añadir la correlación entre las salidas o una función logarítmica del determinante de la matriz de separación. Al igual que en la aproximación anterior, los coeficientes del sistema separador son ajustados utilizando algoritmos de gradiente.

Desafortunadamente, estos algoritmos son muy difíciles de analizar y, hasta la fecha, no hemos sido capaces de estudiar de forma completa su estabilidad.

Este capítulo está estructurado como sigue. En la sección 4.2 proponemos las nuevas familias de funciones de coste. En la sección 4.3 presentamos algoritmos de gradiente para ajustar los coeficientes del sistema separador de acuerdo a la maximización de las funciones de coste. En la sección 4.4 analizamos la estabilidad de uno de estos algoritmos. La sección 4.5 recoge los resultados de diversas simulaciones realizadas por ordenador, los cuales muestran el comportamiento de los algoritmos y ratifican la validez de nuestro estudio. Finalmente, la sección 4.6 recoge las conclusiones más importantes.

4.2 Criterios para separación de mezclas instantáneas

A principios de esta década, Shalvi y Weinstein [68, 69] propusieron un criterio de igualación ciega que utiliza momentos de segundo orden y cumulantes de cuarto orden. En [69] los coeficientes de un sistema Multiple Input Single Output (MISO) son adaptados utilizando un algoritmo de gradiente que maximiza una función de coste que tiene la siguiente forma

$$\begin{aligned} J_i &= |K_{y_i}| - f(E[|y_i|^2]) \\ &= \left| E[|y_i|^4] - 2E^2[|y_i|^2] - |E[y_i^2]|^2 \right| - f(E[|y_i|^2]) \end{aligned} \quad (4.1)$$

donde K_{y_i} denota la curtosis de la salida y_i y $f(\cdot)$ es una función continua, real y derivable $f: \Re \rightarrow \Re$ elegida de forma que el siguiente conjunto de funciones tengan un único máximo en $x > 0$

$$p_i(x) = |\rho_i|x^2 - f(x), \quad i = 1, \dots, N \quad (4.2)$$

donde $\rho_i = K_{s_i}/E^2[|s_i|^2]$ denota la curtosis normalizada de la fuente s_i . En [69] se demuestra que los únicos atractores de un algoritmo de gradiente que maximice (4.1) se corresponden con la extracción de una única señal contenida en las observaciones $\mathbf{x} = \mathbf{A}\mathbf{s}$. Además se demuestra que los puntos estacionarios de la función de coste donde se cancelan todas las fuentes o se extrae una combinación lineal de ellas son puntos de ensilladura o mínimos y, por tanto, un algoritmo de gradiente nunca convergerá a ellos.

El criterio (4.1) puede tomar formas muy diversas dependiendo de la elección de la función $f(\cdot)$. En particular, consideraremos el siguiente polinomio de segundo orden

$$f(x) = \alpha x^2 - 2\beta x + 1 \quad (4.3)$$

Sustituyendo (4.3) en (4.2) es inmediato comprobar que cada función $p_i(\cdot)$ tiene un único máximo en

$$x_i = \frac{\beta}{\alpha - |\rho_i|}, \quad i = 1, \dots, N \quad (4.4)$$

Adicionalmente, para que $p_i(\cdot)$ tenga un único máximo en $x_i > 0$ debemos elegir los parámetros α y β de forma que se cumplan las siguientes condiciones

$$\alpha > |\rho_i|, \quad i = 1, \dots, N, \quad \beta > 0 \quad (4.5)$$

Como puede apreciarse, la elección del parámetro α depende del valor de la curtosis de las fuentes. Esta condición no limita en absoluto la aplicabilidad del criterio ya que no es necesario conocer el valor exacto de los ρ_i . Solamente debemos tener una idea del rango en el que están comprendidos y, en cualquier caso, siempre podemos escoger un parámetro α suficientemente grande.

Una particularización muy interesante de (4.3) resulta de considerar que todas las fuentes en el entorno son complejas, circulares y de curtosis negativa ($\rho_i \in (-2, 0)$). Para esta aplicación los parámetros $\alpha = 2$ y $\beta = 1$ satisfacen las condiciones (4.5) y dan lugar al siguiente criterio

$$\max_{\mathbf{w}_i} J_i = - \left(E[|y_i|^4] - 2E[|y_i|^2] + 1 \right) = -E[|y_i|^2 - 1]^2 \quad (4.6)$$

donde el término $E[y_i^2]$ ha sido eliminado de la expresión como consecuencia de la simetría circular asumida en las fuentes. La maximización de (4.6) es el bien conocido criterio Constant Modulus propuesto inicialmente por Godard [38] para encontrar los coeficientes del igualador en un receptor de comunicaciones y, recientemente, empleado en separación ciega de fuentes [36, 53].

La propiedad del criterio de Shalvi y Weinstein de extraer una única señal cancelando las señales interferentes nos puede conducir a plantear un criterio de separación consistente en ajustar los coeficientes de un sistema MIMO mediante la maximización de la suma de funciones J_i dadas en (4.1)

$$\max_{\mathbf{W}} J = \sum_{i=1}^N J_i = \sum_{i=1}^N \left(|K_{y_i}| - f(E[|y_i|^2]) \right) \quad (4.7)$$

La figura 4.1 muestra el esquema de un sistema separador cuyos parámetros se ajustan utilizando un algoritmo de gradiente que maximiza (4.7). Como puede verse, cada combinador lineal es adaptado de forma individual sin considerar lo que sucede en los demás. Esta falta de cooperación puede conducir a una situación donde la misma fuente sea extraída por varios combinadores. Este problema puede ser evitado añadiendo a (4.7) la suma de cumulantes cruzados de cuarto orden entre las salidas del sistema separador

$$\begin{aligned} T_1 &= -\gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N |K_{y_i y_j}| \\ &= -\gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \left| E[|y_i|^2 |y_j|^2] - E[|y_i|^2] E[|y_j|^2] - |E[y_i y_j^*]|^2 - |E[y_i y_j]|^2 \right| \quad (4.8) \end{aligned}$$

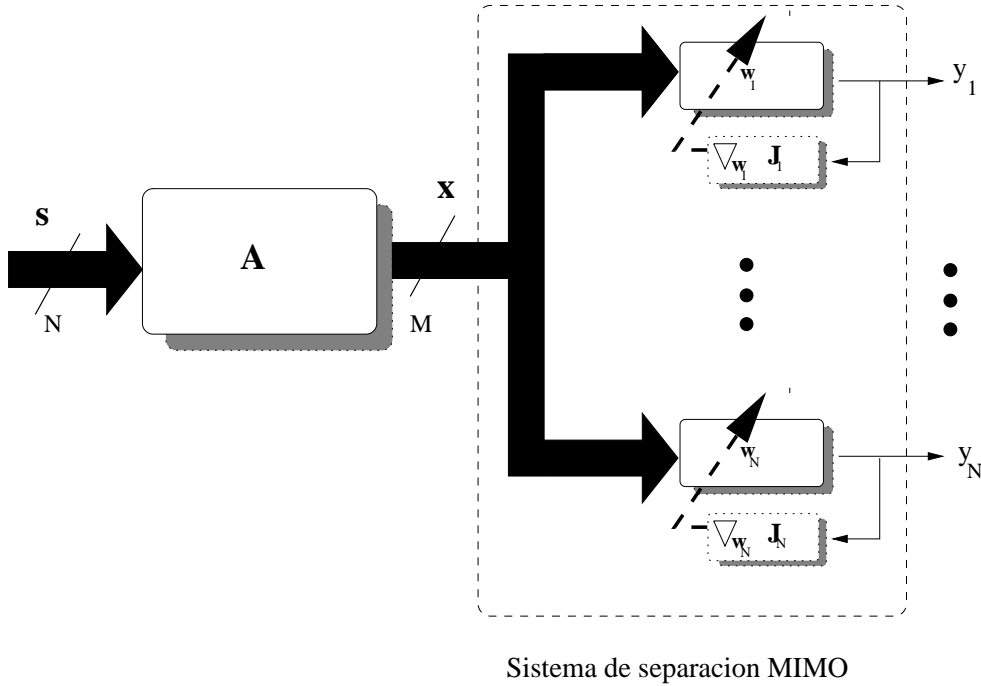


Figura 4.1: Sistema MIMO ajustado con el criterio de Shalvi y Weinstein.

donde γ es un parámetro real positivo. La función de coste (4.8) será negativa cuando algún par de salidas y_i y y_j extraigan la misma fuente y se anulará cuando todas las salidas sean estadísticamente independientes por lo que parece una buena candidata para conseguir la separación de las fuentes. Sin embargo, debemos notar que también se anulará cuando las salidas y_i , $i = 1, \dots, N$ cancelen todas las fuentes, por lo que la maximización de (4.8) es por sí sola insuficiente para conseguir la separación a no ser que se introduzcan restricciones sobre el sistema de separación como las propuestas en [52, 60].

En el presente trabajo proponemos incorporar en una única función de coste las características de las funciones (4.7) y (4.8) comentadas anteriormente. Nuestro criterio de optimización es el siguiente

$$\max_{\mathbf{W}} J_1(\mathbf{W}) = J + T_1 = \sum_{i=1}^N (|K_{y_i}| - f(E[|y_i|^2])) - \gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N |K_{y_i y_j}| \quad (4.9)$$

La función de coste que acabamos de definir nos recuerda al contraste (3.40) presentado en el capítulo anterior. Sin embargo, la inclusión de la función $f(\cdot)$ hace innecesarias las restricciones sobre las salidas, lo que se traduce en una menor complejidad del sistema separador al eliminarse los AGCs que normalizan la potencia de las salidas. Además el análisis de los puntos estacionarios de la función de coste (4.9) es notablemente más simple que el de (3.37).

Siguiendo con la idea de utilizar el criterio de Shalvi y Weinstein en separación ciega de

fuentes, se ha pensado penalizar la dependencia entre las salidas utilizando otros términos cruzados:

- La primera alternativa consiste en añadir la correlación cruzada entre las salidas [23]

$$T_2 = -\gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N |E[y_i y_j^*]|^2 \quad (4.10)$$

Este término es negativo cuando existe una dependencia estadística entre las salidas y se anula cuando las salidas extraen fuentes diferentes. El criterio de separación que resulta de emplear este término cruzado es el siguiente

$$\max_{\mathbf{W}} J_2(\mathbf{W}) = J + T_2 = \sum_{i=1}^N (|K_{y_i}| - f(E[|y_i|^2])) - \gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N |E[y_i y_j^*]|^2 \quad (4.11)$$

- Al observar los criterios de maximización de la información presentados en el capítulo anterior, parece lógico pensar en emplear como término cruzado una función logarítmica que penalice la dependencia entre las filas de \mathbf{W} que se produce cuando varias salidas extraen la misma fuente

$$T_3 = \ln(\det(|\mathbf{W}^H|)) \quad (4.12)$$

El criterio que resulta de añadir este término a (4.7) es la siguiente

$$\max_{\mathbf{W}} J_3(\mathbf{W}) = J + T_3 = \sum_{i=1}^N (|K_{y_i}| - f(E[|y_i|^2])) + \ln(\det(|\mathbf{W}^H|)) \quad (4.13)$$

4.3 Algoritmos adaptativos

Una de las principales ventajas de nuestra aproximación es que la matriz de separación óptima puede ser calculada utilizando un sencillo algoritmo de gradiente estocástico de la forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \hat{\nabla}_{\mathbf{W}} J_k(n) \quad (4.14)$$

donde μ es un valor real que determina la velocidad de convergencia del algoritmo y $\hat{\nabla}_{\mathbf{W}} J_k(n)$ es una estimación del gradiente de la función de coste (J_1 , J_2 o J_3) respecto a la matriz \mathbf{W} que denotaremos $\nabla_{\mathbf{W}} J_k$. Como las funciones de coste con las que estamos trabajando dependen de la matriz \mathbf{W} , el gradiente $\nabla_{\mathbf{W}} J_k$ es una matriz formada por los vectores gradiente respecto a cada columna de \mathbf{W}

$$\nabla_{\mathbf{W}} J_k = [\nabla_{\mathbf{w}_1} J_k \dots \nabla_{\mathbf{w}_N} J_k]$$

Cada uno de estos vectores gradiente complejo se define como

$$\nabla_{\mathbf{w}_i} J_k = \left. \frac{\partial J_k}{\partial w_{ij}^*} \right|_{j=1}^N \quad i = 1, \dots, N$$

Un estudio riguroso de este operador, así como algunas de sus principales propiedades puede ser consultado en [7]. Como las funciones de coste que hemos propuesto son la suma de dos términos, podemos descomponer el gradiente en dos partes

$$\nabla_{\mathbf{w}} J_k = \nabla_{\mathbf{w}} J + \nabla_{\mathbf{w}} T_k \quad (4.15)$$

Para determinar las expresiones de cada vector gradiente supondremos que la curtosis de todas las fuentes tienen el mismo signo ϵ y que éste es conocido por el sistema de separación. En este caso se obtiene

$$\nabla_{\mathbf{w}} J_k = \begin{bmatrix} 2\epsilon(E[\mathbf{x}y_1^*|y_1|^2] - 2E[\mathbf{x}y_1^*]E[|y_1|^2] - E[\mathbf{x}y_1]E[y_1^{*2}]) - \nabla_{\mathbf{w}_1} f(E[|y_1|^2]) \\ \vdots \\ \epsilon(E[\mathbf{x}y_N^*|y_N|^2] - 2E[\mathbf{x}y_N^*]E[|y_N|^2] - E[\mathbf{x}y_N]E[y_N^{*2}]) - \nabla_{\mathbf{w}_N} f(E[|y_N|^2]) \end{bmatrix} \quad (4.16)$$

donde $\nabla_{\mathbf{w}_i} f(E[|y_i|^2])$ depende de la elección de $f(\cdot)$. En particular, si consideramos que $f(\cdot)$ es el polinomio de segundo orden dado en (4.3) y que repetimos a continuación

$$f(x) = \alpha x^2 - 2\beta x + 1 \quad (4.17)$$

obtenemos

$$\nabla_{\mathbf{w}_i} f(E[|y_i|^2]) = 2E[y_i^* \mathbf{x}](\alpha E[|y_i|^2] - \beta) \quad (4.18)$$

Sustituyendo (4.18) en (4.16) y reemplazando los promedios por sus estimaciones, podemos escribir el gradiente como

$$\hat{\nabla}_{\mathbf{w}} J(n) = 2\mathbf{x}(n)\hat{\mathbf{c}}^T(n) \quad (4.19)$$

donde la matriz $\hat{\mathbf{c}}(n)$ tiene la siguiente forma

$$\hat{\mathbf{c}}(n) = \begin{bmatrix} \epsilon(y_1^*(n)|y_1(n)|^2 - 2y_1^*(n)\hat{E}[|y_1|^2] - y_1(n)\hat{E}[y_1^{*2}]) - y_1^*(n)(\alpha\hat{E}[|y_1|^2] - \beta) \\ \vdots \\ \epsilon(y_N^*(n)|y_N(n)|^2 - 2y_N^*(n)\hat{E}[|y_N|^2] - y_N(n)\hat{E}[y_N^{*2}]) - y_N^*(n)(\alpha\hat{E}[|y_N|^2] - \beta) \end{bmatrix} \quad (4.20)$$

Los promedios en (4.20) pueden ser calculados iterativamente con algoritmos de ventana exponencial como los siguientes

$$\begin{aligned} \hat{E}[|y_i|^2] &= (1 - \mu_l) \hat{E}[|y_i|^2] + \mu_l |y_i(n)|^2 \\ \hat{E}[y_i^{*2}] &= (1 - \mu_l) \hat{E}[y_i^{*2}] + \mu_l y_i^*(n)^2 \end{aligned} \quad (4.21)$$

Por otro lado, los diferentes términos cruzados utilizados en (4.9), (4.11) y (4.13) dan lugar a los siguientes algoritmos de gradiente.

4.3.1 Algoritmo I

El algoritmo de gradiente para la función de coste (4.9) puede expresarse como sigue

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \mathbf{x}(n)(\hat{\mathbf{c}}^T(n) - \gamma \epsilon \hat{\mathbf{c}}_1^T(n)) \quad (4.22)$$

donde un factor 2 ha sido incluido dentro del parámetro μ , $\hat{\mathbf{c}}(n)$ viene dado por (4.20) y $\hat{\mathbf{c}}_1(n)$ es una estimación del gradiente del término cruzado (4.8)

$$\nabla_{\mathbf{w}} T_1 = -2\epsilon \gamma \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N (E[\mathbf{x}y_1^* | y_j|^2] - E[\mathbf{x}y_1^*]E[|y_j|^2] - E[\mathbf{x}y_j^*]E[y_1^* y_j] - E[\mathbf{x}y_j]E[y_1^* y_j^*]) \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N (E[\mathbf{x}y_N^* | y_j|^2] - E[\mathbf{x}y_N^*]E[|y_j|^2] - E[\mathbf{x}y_j^*]E[y_N^* y_j] - E[\mathbf{x}y_j]E[y_N^* y_j^*]) \end{bmatrix}$$

Estimando los promedios se obtiene la expresión de la matriz $\mathbf{c}_1(n)$ en (4.22)

$$\hat{\mathbf{c}}_1(n) = \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N (y_1^*(n)|y_j(n)|^2 - y_1^*(n)\hat{E}[|y_j|^2] - y_j^*(n)\hat{E}[y_1^* y_j] - y_j(n)\hat{E}[y_1^* y_j^*]) \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N (y_N^*(n)|y_j(n)|^2 - y_N^*(n)\hat{E}[|y_j|^2] - y_j^*(n)\hat{E}[y_N^* y_j] - y_j(n)\hat{E}[y_N^* y_j^*]) \end{bmatrix} \quad (4.23)$$

Los promedios en (4.23) pueden ser calculados iterativamente utilizando (4.21) y los siguientes algoritmos

$$\begin{aligned} \hat{E}[y_i^* y_j] &= (1 - \mu_l) \hat{E}[y_i^* y_j] + \mu_l y_i^*(n) y_j(n) \\ \hat{E}[y_i^* y_j^*] &= (1 - \mu_l) \hat{E}[y_i^* y_j^*] + \mu_l y_i^*(n) y_j^*(n) \end{aligned}$$

donde μ_l es un número real positivo menor que 1.

Al emplear este algoritmo, los vectores \mathbf{w}_i del sistema separador son actualizados utilizando información proveniente de los otros conformadores como puede verse en la figura 4.2. Esta cooperación entre los elementos de proceso es la clave para conseguir la separación de las fuentes.

4.3.2 Algoritmo II

El algoritmo de gradiente para maximizar (4.11) puede expresarse como sigue

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \mathbf{x}(n)(\hat{\mathbf{c}}^T(n) - \gamma \hat{\mathbf{c}}_2^T(n)) \quad (4.24)$$

donde un factor 2 ha sido incluido dentro de μ , $\hat{\mathbf{c}}(n)$ viene dado por (4.20) y $\hat{\mathbf{c}}_2(n)$ es una estimación del término cruzado (4.10) dado por la siguiente expresión

$$\nabla_{\mathbf{w}} T_2 = -2\gamma \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N E[\mathbf{x}y_j^*]E[y_1^* y_j] \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N E[\mathbf{x}y_j^*]E[y_N^* y_j] \end{bmatrix} \quad (4.25)$$

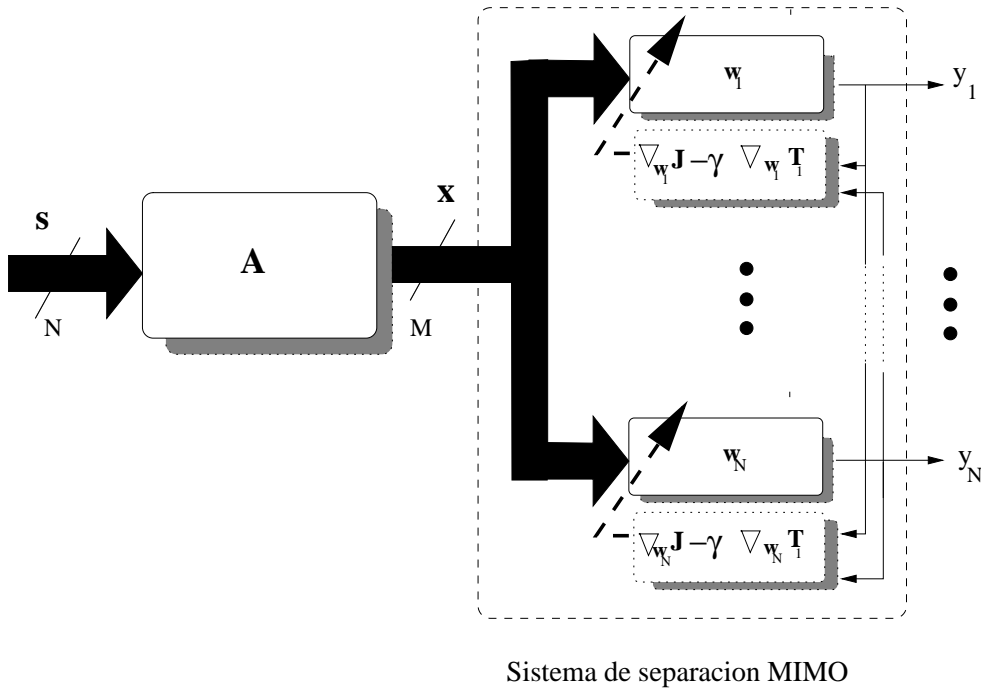


Figura 4.2: Sistema MIMO adaptado considerando la dependencia entre sus salidas.

Estimando los promedios se obtiene lo siguiente

$$\hat{\mathbf{c}}_2(n) = \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N y_j^*(n) \hat{E}[y_1^* y_j] \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N y_j^*(n) \hat{E}[y_N^* y_j] \end{bmatrix} \quad (4.26)$$

donde

$$\hat{E}[y_i^* y_j] = (1 - \mu_i) \hat{E}[y_i^* y_j] + \mu_i y_i^*(n) y_j(n) \quad (4.27)$$

Si comparamos (4.23) y (4.26) podemos concluir que la estimación de los momentos cruzados requiere menor número de operaciones. Esta importante ventaja del empleo de (4.26) se ve reducida por la complejidad de realizar un análisis de la estabilidad del algoritmo adaptativo. En la sección E.2 del apéndice E se presenta un análisis para sistemas separadores 2×2 en el que se demuestra que la condición $\gamma > \max_i(|\rho_i|)$ garantiza que los únicos atractores del algoritmo sean los puntos donde se consigue la perfecta separación de las fuentes. Esta demostración, sin embargo, es extremadamente difícil de generalizar al caso de sistemas $N \times N$.

4.3.3 Algoritmo III

La maximización de la función de coste (4.13) conduce al siguiente algoritmo de gradiente

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(\mathbf{x}(n)\hat{\mathbf{c}}^T(n) + \hat{\mathbf{c}}_3(n)) \quad (4.28)$$

donde $\hat{\mathbf{c}}(n)$ viene dado por (4.20) y $\hat{\mathbf{c}}_3(n)$ se corresponde con el gradiente de la función logarítmica (4.12) respecto a la matriz de separación \mathbf{W}

$$\hat{\mathbf{c}}_3(n) = \nabla_{\mathbf{W}} \ln(|\det(\mathbf{W}^H(n))|) = \mathbf{W}^{-H}(n) \quad (4.29)$$

Este algoritmo requiere la inversión de la matriz pesos, lo cual puede resultar costoso y, además, limita la aplicación de (4.28) a sistemas con el mismo número de entradas que de salidas, i.e., $M = N$.

4.3.4 Algoritmo IV

Para ajustar los coeficientes del sistema separador mediante la maximización de las funciones de coste propuestas anteriormente también pueden emplearse algoritmos de gradiente relativo. Esta aproximación tiene especial interés como alternativa al algoritmo (4.28) ya que evita invertir la matriz de coeficientes. Multiplicando la recursión (4.28) por $\mathbf{W}\mathbf{W}^H$ se obtiene la siguiente expresión

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu\mathbf{W}(n)\mathbf{W}^H(n)(\mathbf{x}(n)\hat{\mathbf{c}}^T(n) + \mathbf{W}^{-H}(n)) \\ &= \mathbf{W}(n) + \mu\mathbf{W}(n)(\mathbf{y}(n)\hat{\mathbf{c}}^T(n) + \mathbf{I}) \end{aligned} \quad (4.30)$$

Este algoritmo, al igual que (4.28), pueden emplearse solamente cuando la matriz de separación es cuadrada, i.e., la dimensión del vector de observaciones coincide con el de salidas.

Es interesante notar que, a diferencia de los algoritmos propuestos anteriormente, la recursión (4.30) es equivariante ya que si multiplicamos (4.30) por la matriz de mezcla traspuesta obtenemos que la evolución de los coeficientes del sistema de mezcla/separación no depende de los parámetros \mathbf{A}

$$\begin{aligned} \mathbf{A}^T\mathbf{W}(n+1) &= \mathbf{A}^T\mathbf{W}(n) + \mu\mathbf{A}^T\mathbf{W}(n)(\mathbf{y}(n)\hat{\mathbf{c}}^T(n) + \mathbf{I}) \\ \mathbf{G}^T(n+1) &= \mathbf{G}^T(n) + \mu\mathbf{G}^T(n)(\mathbf{y}(n)\hat{\mathbf{c}}^T(n) + \mathbf{I}) \end{aligned} \quad (4.31)$$

Esta propiedad hace que el algoritmo (4.30) sea especialmente atractivo para situaciones donde la matriz de mezcla está mal condicionada.

4.4 Análisis de estabilidad del algoritmo I

Los problemas de optimización presentados anteriormente consisten en la maximización de una función de coste no cuadrática de la matriz de coeficientes. Esto significa que

los coeficientes óptimos del sistema separador no se corresponden con la solución de un sistema de ecuaciones lineales; más aún, la superficie de error asociada a cada función de coste puede contener máximos indeseados que dañen la convergencia de los algoritmos de gradiente.

A continuación determinaremos los atractores del algoritmo I dado por (4.22) propuesto para maximizar la función de coste J_1 . Primero encontraremos los puntos estacionarios de la función de coste resolviendo el sistema de ecuaciones que resulta de anular las primeras derivadas y, posteriormente, analizaremos el carácter definido negativo de la matriz hessiana para determinar cuáles de ellos son máximos. Supondremos el caso general de N fuentes no gaussianas, estadísticamente independientes, estacionarias, de media cero y varianza unidad. Las observaciones son mezclas lineales de las fuentes $\mathbf{x} = \mathbf{A}\mathbf{s}$ siendo \mathbf{A} una matriz $M \times N$ de rango completo. Como sistema separador consideraremos una matriz de dimensión $M \times N$ con coeficientes complejos y salidas $\mathbf{y} = \mathbf{W}^H \mathbf{x}$.

Realizar el estudio de la estabilidad respecto a la matriz de coeficientes \mathbf{W} es un trabajo arduo ya que deberíamos considerar $M \times N$ ($M \geq N$) variables complejas. El análisis se simplifica notablemente expresando la función en términos de la matriz $\mathbf{G} = \mathbf{W}^H \mathbf{A}$ en la que el elemento g_{ij} representa la amplitud con la que la salida i extrae a la fuente j . La dimensión de esta matriz se reduce a N^2 variables. Debemos hacer notar que este análisis es estrictamente válido solamente cuando existe una correspondencia uno a uno entre \mathbf{G} y \mathbf{W} , i.e., cuando \mathbf{A} es una matriz cuadrada invertible.

Para expresar la función de coste J_1 en términos de la matriz de amplitudes \mathbf{G} utilizaremos la relación lineal que existe entre las salidas del sistema de separación y las fuentes

$$y_i = \sum_{l=1}^N g_{il} s_l, \quad i = 1, \dots, N \quad (4.32)$$

Teniendo en cuenta las hipótesis sobre las fuentes y utilizando las propiedades de los cumulantes recogidas en el apéndice B, podemos expresar los estadísticos que aparecen en la función de coste J_1 de la siguiente forma

$$\begin{aligned} E[|y_i|^2] &= \sum_{l=1}^N |g_{il}|^2 E[|s_l|^2] = \sum_{l=1}^N |g_{il}|^2 \\ K_{y_i y_i} &= \sum_{l=1}^N K_{s_l} |g_{il}|^4 = \sum_{l=1}^N \rho_l |g_{il}|^4 \\ K_{y_i y_j} &= \sum_{l=1}^N K_{s_l} |g_{il}|^2 |g_{jl}|^2 = \sum_{l=1}^N \rho_l |g_{il}|^2 |g_{jl}|^2 \end{aligned} \quad (4.33)$$

donde $\rho_l = K_{s_l} / E^2[|s_l|^2]$ es el cumulante de cuarto orden normalizado de la fuente s_l .

Sustituyendo (4.33) en (4.9) obtenemos lo siguiente

$$\psi(\mathbf{G}) \triangleq J_1(\mathbf{W}) = \sum_{i=1}^N \left(\sum_{l=1}^N |\rho_l| |g_{il}|^4 - f(\|\mathbf{g}_i\|^2) \right) - \gamma \sum_{\substack{i,j=1 \\ i \neq j}}^N \sum_{l=1}^N |\rho_l| |g_{il}|^2 |g_{jl}|^2 \quad (4.34)$$

donde $\|\mathbf{g}_i\|^2 = \sum_{l=1}^N |g_{il}|^2$.

De la teoría de sistemas dinámicos sabemos que los puntos de equilibrio del algoritmo de gradiente (4.22) se corresponden con los puntos estacionarios de la función de coste ψ . Para encontrarlos, resolveremos el siguiente sistema de N^2 ecuaciones

$$\frac{\partial \psi}{\partial g_{ik}} = 2|\rho_k| g_{ik}^* |g_{ik}|^2 - g_{ik}^* f'(\|\mathbf{g}_i\|^2) - 2\gamma |\rho_k| g_{ik}^* \sum_{\substack{j=1 \\ i \neq j}}^N |g_{jk}|^2 = 0 \quad i, k = 1, \dots, N \quad (4.35)$$

donde $f'(\cdot)$ es la primera derivada de $f(\cdot)$. Por otro lado, al ser \mathbf{H} una matriz hermítica el estudio del signo sus autovalores puede realizarse analizando el carácter definido negativo de la siguiente matriz de dimensión $2N^2 \times 2N^2$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1111} & \cdots & \mathbf{H}_{111N} & \mathbf{H}_{1121} & \cdots & \mathbf{H}_{112N} & \cdots & \mathbf{H}_{11N1} & \cdots & \mathbf{H}_{11NN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{1N11} & \cdots & \mathbf{H}_{1N1N} & \mathbf{H}_{1N21} & \cdots & \mathbf{H}_{1N2N} & \cdots & \mathbf{H}_{1NN1} & \cdots & \mathbf{H}_{1NNN} \\ \mathbf{H}_{2111} & \cdots & \mathbf{H}_{211N} & \mathbf{H}_{2121} & \cdots & \mathbf{H}_{212N} & \cdots & \mathbf{H}_{21N1} & \cdots & \mathbf{H}_{21NN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{2N11} & \cdots & \mathbf{H}_{2N1N} & \mathbf{H}_{2N21} & \cdots & \mathbf{H}_{2N2N} & \cdots & \mathbf{H}_{2NN1} & \cdots & \mathbf{H}_{2NNN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N111} & \cdots & \mathbf{H}_{N11N} & \mathbf{H}_{N121} & \cdots & \mathbf{H}_{N12N} & \cdots & \mathbf{H}_{N1N1} & \cdots & \mathbf{H}_{N1NN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{NN11} & \cdots & \mathbf{H}_{NN1N} & \mathbf{H}_{NN21} & \cdots & \mathbf{H}_{NN2N} & \cdots & \mathbf{H}_{NNN1} & \cdots & \mathbf{H}_{NNNN} \end{bmatrix} \quad (4.36)$$

donde \mathbf{H}_{iknp} es la siguiente matriz compleja de dimensión 2×2

$$\mathbf{H}_{iknp} = \begin{bmatrix} \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ik}^*} & \frac{\partial^2 \psi}{\partial g_{ik}^* \partial g_{np}^*} \\ \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{np}} & \frac{\partial^2 \psi}{\partial g_{ik}^* \partial g_{np}} \end{bmatrix} \quad (4.37)$$

Las segundas derivadas en (4.36) vienen dadas por las siguientes expresiones

$$\begin{aligned} \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ik}^*} &= 4|\rho_k| |g_{ik}|^2 - f'(\|\mathbf{g}_i\|^2) - |g_{ik}|^2 f''(\|\mathbf{g}_i\|^2) - 2\gamma |\rho_k| \sum_{\substack{j=1 \\ i \neq j}}^N |g_{jk}|^2 \\ \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ip}^*} &= -g_{ik}^* g_{ip} f''(\|\mathbf{g}_i\|^2) \quad k \neq p \\ \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{nk}^*} &= -2\gamma |\rho_k| g_{ik}^* g_{nk} \quad i \neq n \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{np}^*} &= 0 \quad i \neq n, k \neq p \\
 \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ik}} &= 2|\rho_k|g_{ik}^{*2} - g_{ik}^{*2} f''(\|\mathbf{g}_i\|^2) \\
 \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ip}} &= -g_{ik}^* g_{ip}^* f''(\|\mathbf{g}_i\|^2) \quad k \neq p \\
 \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{nk}} &= -2\gamma|\rho_k|g_{ik}^* g_{nk}^* \quad i \neq n \\
 \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{np}} &= 0 \quad i \neq n, k \neq p
 \end{aligned} \tag{4.38}$$

donde $f''(\cdot)$ es la segunda derivada de $f(\cdot)$.

El sistema (4.35) es no lineal y contiene múltiples soluciones, de las cuales $N!$ se corresponden con soluciones deseadas donde la matriz de amplitudes \mathbf{G} tiene un único elemento distinto de cero por fila y columna¹. El análisis de la estabilidad del algoritmo en todas las soluciones deseadas es análogo por lo que a continuación centraremos nuestra atención en el punto donde la matriz \mathbf{G} es diagonal

$$g_{ii} \neq 0 \quad \text{y} \quad g_{ik} = 0, \quad i, k = 1, \dots, N, \quad i \neq k$$

En este caso, el sistema de ecuaciones (4.35) se reduce a las siguientes expresiones

$$\frac{\partial \psi}{\partial g_{ii}} = 2|\rho_i|g_{ii}^*|g_{ii}|^2 - g_{ii}^* f'(|g_{ii}|^2) = g_{ii}^* p'_i(|g_{ii}|^2) \quad i = 1, \dots, N \tag{4.39}$$

$$\frac{\partial \psi}{\partial g_{ik}} = 0 \quad i, k = 1, \dots, N, \quad i \neq k \tag{4.40}$$

donde $p'_i(\cdot)$, $i = 1, \dots, N$ es la primera derivada de $p_i(\cdot)$. Recordemos que estamos suponiendo que cada función $p_i(\cdot)$, $i = 1, \dots, N$ tiene un único máximo en $x > 0$. Esto implica que la expresión (4.39) se anulará solamente cuando cada valor $|g_{ii}|^2$, $i = 1, \dots, N$ se corresponde con el máximo de $p_i(\cdot)$, $i = 1, \dots, N$.

Para determinar si el punto donde se consigue la separación es un atractor analizaremos la matriz hessiana (4.36). De (4.38) se obtiene que la mayoría de las segundas derivadas son cero. Las únicas que no se anulan son las siguientes

$$\begin{aligned}
 \frac{\partial^2 \psi}{\partial g_{ii} \partial g_{ii}^*} &= 4|\rho_i||g_{ii}|^2 - f'(|g_{ii}|^2) - |g_{ii}|^2 f''(|g_{ii}|^2) \\
 \frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ik}^*} &= -f'(|g_{ii}|^2) - 2\gamma|\rho_k||g_{kk}|^2 \quad i \neq k \\
 \frac{\partial^2 \psi}{\partial g_{ii} \partial g_{ii}} &= 2|\rho_i|g_{ii}^{*2} - g_{ii}^{*2} f''(|g_{ii}|^2)
 \end{aligned} \tag{4.41}$$

¹El número de soluciones deseadas $N!$ no considera el signo (o fase) de las amplitudes no nulas.

Teniendo en cuenta la forma que (4.39) se anula en el punto de separación, podemos escribir las ecuaciones (4.41) como sigue

$$\begin{aligned}
\frac{\partial^2 \psi}{\partial g_{ii} \partial g_{ii}^*} &= \frac{\partial \psi}{\partial g_{ii}} \frac{1}{g_{ii}^*} + |g_{ii}|^2 (2|\rho_i| - f''(|g_{ii}|^2)) = |g_{ii}|^2 p_i''(|g_{ii}|^2) \\
\frac{\partial^2 \psi}{\partial g_{ik} \partial g_{ik}^*} &= \frac{\partial \psi}{\partial g_{ii}} \frac{1}{g_{ii}^*} - 2|\rho_i| |g_{ii}|^2 - 2\gamma |\rho_k| |g_{kk}|^2 = -2|\rho_i| |g_{ii}|^2 - 2\gamma |\rho_k| |g_{kk}|^2 \quad k \neq i \\
\frac{\partial^2 \psi}{\partial g_{ii} \partial g_{ii}} &= g_{ii}^{*2} (2|\rho_i| - f''(|g_{ii}|^2)) = g_{ii}^{*2} p_i''(|g_{ii}|^2)
\end{aligned} \tag{4.42}$$

Como consecuencia, todas las matrices \mathbf{H}_{ijkl} son cero excepto

$$\begin{aligned}
\mathbf{H}_{iiii} &= \begin{bmatrix} |g_{ii}|^2 p_i''(|g_{ii}|^2) & g_{ii}^2 p_i''(|g_{ii}|^2) \\ g_{ii}^{*2} p_i''(|g_{ii}|^2) & |g_{ii}|^2 p_i''(|g_{ii}|^2) \end{bmatrix} \quad i = 1, \dots, N \\
\mathbf{H}_{ikik} &= \begin{bmatrix} -2|\rho_i| |g_{ii}|^2 - 2\gamma |\rho_k| |g_{kk}|^2 & 0 \\ 0 & -2|\rho_i| |g_{ii}|^2 - 2\gamma |\rho_k| |g_{kk}|^2 \end{bmatrix} \\
&\quad i, k = 1, \dots, N, i \neq k
\end{aligned} \tag{4.43}$$

$$(4.44)$$

La forma que toma esta matriz hessiana nos permite determinar la estabilidad del punto donde se lleva a cabo la separación analizando el carácter definido negativo de cada submatriz \mathbf{H}_{iiii} , $i = 1, \dots, N$ y \mathbf{H}_{ikik} , $i, k = 1, \dots, N$, $i \neq k$. Para estudiar las submatrices \mathbf{H}_{iiii} , $i = 1, \dots, N$, consideraremos un resultado obtenido anteriormente que indica que el gradiente (4.39) se anula únicamente cuando cada punto $|g_{ii}|^2$ se corresponde con un máximo de $p_i(\cdot)$, $i = 1, \dots, N$, lo cual nos permite afirmar que los elementos en la diagonal de \mathbf{H}_{iiii} son negativos. Además el determinante de \mathbf{H}_{iiii} es cero y, por tanto, las matrices (4.43) son definidas negativas. Por otro lado, la condición $\gamma > 0$ asegura que todas las matrices \mathbf{H}_{ikik} son definidas negativas. Como consecuencia el hessiano es definido negativo y el punto que estamos analizando es un máximo de la función de coste. Estos resultados pueden ser extendidos a todas las soluciones de (4.35) correspondientes a la separación de las fuentes.

El estudio que acabamos de presentar nos permite afirmar que la condición $\gamma > 0$ asegura que los puntos donde se consigue la separación de las fuentes son atractores del algoritmo de gradiente (4.22). Sin embargo, la convergencia del algoritmo hacia estos puntos puede verse dañada por la existencia de máximos indeseados. En el apéndice E se analizan las otras soluciones del sistema de ecuaciones (4.35) y se demuestra que la condición $\gamma > 1$ garantiza que no son máximos de la función de coste y, por tanto, el algoritmo (4.22) nunca convergerá a ellos. La figura 4.3 recoge los grupos analizados en el apéndice E y las condiciones suficientes que aseguran la inestabilidad del algoritmo en ellos.

Grupos de puntos estacionarios		
<p>Alguna salida cancela todas las fuentes y una fuente es cancelada por todas las salidas</p>	$\mathbf{G} = \begin{bmatrix} 0 & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{bmatrix}$	<p>Punto inestable si $p_i(x)$ $i=1,\dots,N$ tiene un unico maximo en $x>0$</p>
<p>Todas las salidas extraen una unica fuente y varias salidas extraen la misma fuente</p>	$\mathbf{G} = \begin{bmatrix} x & 0 & \dots & 0 & 0 \\ 0 & x & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & x & \dots & 0 & 0 \end{bmatrix}$	<p>Punto inestable si $\gamma > 1$</p>
<p>Alguna salida extrae una combinacion lineal de varias fuentes</p>	$\mathbf{G} = \begin{bmatrix} \dots & \dots & \dots & \dots \\ \dots & x & \dots & x \\ \dots & \dots & \dots & \dots \end{bmatrix}$	<p>Punto inestable.</p>

Figura 4.3: Clasificación de los puntos estacionarios indeseados del algoritmo I.

4.5 Resultados de simulaciones

Con el objeto de ilustrar el funcionamiento de los algoritmos propuestos en la sección 4.3 se han llevado a cabo diversas simulaciones por ordenador cuyos resultados más importantes presentamos a continuación. Estas simulaciones han sido divididas en dos grupos: el primero muestra el comportamiento de los algoritmos en situaciones donde no hay ruido y el segundo grupo abarca simulaciones realizadas para analizar el efecto del ruido blanco gaussiano sobre el rendimiento del algoritmo. En un último apartado se presentan varias simulaciones realizadas utilizando una aproximación tipo bloque para maximizar la función de coste J_1 .

4.5.1 Entornos sin ruido

En el primer grupo de simulaciones se evaluaron las prestaciones en entornos sin ruido de los algoritmos de máxima pendiente (4.22), (4.24) y (4.28) propuestos en este capítulo y del algoritmo (3.40) presentado en el capítulo anterior. El algoritmo de gradiente relativo será abordado en una subsección posterior.

Como medida de rendimiento se ha utilizado el índice $P(\mathbf{G})$ definido en [55]

$$P(\mathbf{G}) = \sum_{i=1}^N \left(\sum_{j=1}^N \frac{|g_{ij}|^2}{\max_l(|g_{il}|^2)} - 1 \right) + \sum_{j=1}^N \left(\sum_{i=1}^N \frac{|g_{ij}|^2}{\max_l(|g_{lj}|^2)} - 1 \right)$$

donde $\max_l(|g_{il}|^2)$ y $\max_l(|g_{lj}|^2)$ representa los máximos por fila y columna, respectivamente. Este índice mide la distancia entre la matriz $\mathbf{G} = \mathbf{W}^H \mathbf{A}$ obtenida en cada iteración del algoritmo y una matriz $\mathbf{G} = \mathbf{\Lambda} \mathbf{P}$ que tiene una única componente distinta de cero por fila y columna. Para presentar la evolución del índice $P(\mathbf{G})$ hemos promediado 10 realizaciones independientes del experimento.

Las observaciones empleadas en todas las simulaciones que constituyen este grupo son combinaciones lineales e instantáneas de las fuentes obtenidas con la siguiente matriz

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \quad (4.45)$$

En la etapa de separación se ha utilizado un sistema MIMO de cuatro entradas y cuatro salidas cuyos coeficientes han sido aleatoriamente inicializados. Para reducir el ruido de desajuste hemos multiplicado por 0.6 cada 10.000 iteraciones todos los parámetros de paso de los algoritmos (μ , μ_l y μ_a en el algoritmo que maximiza el contraste y μ y μ_l en los algoritmo propuestos en el presente capítulo).

En el primer experimento hemos considerado la separación de las cuatro imágenes de dimensión 256×256 mostradas en la figura 4.4, las cuales han sido escaneadas horizontalmente para dar lugar a secuencias de longitud 65.536. Las cuatro fuentes tienen curtosis negativa con valores estimados: $\rho_1 = \rho_2 = \rho_4 = -2$ y $\rho_3 = -0.5133$. En la misma figura pueden observarse las observaciones obtenidas utilizando la matriz de mezcla (4.45).

En la figura 4.5 puede observarse como el algoritmo I ha conseguido recuperar las fuentes. Comparemos los resultados obtenidos en esta simulación con las conclusiones del análisis de la estabilidad del algoritmo (4.24) presentado en la sección 4.4. El análisis predice que los atractores de este algoritmo son los máximos de las funciones $p_i(\cdot)$ dados por la expresión (4.4), es decir, serán matrices de amplitudes de la forma $\mathbf{G}_o = \mathbf{\Lambda}_o \mathbf{P}$ siendo \mathbf{P} una matriz de permutación y $\mathbf{\Lambda}_o$ la siguiente matriz diagonal

$$\mathbf{\Lambda}_o = \begin{bmatrix} \pm 0.5270 & 0 & 0 & 0 \\ 0 & \pm 0.5270 & 0 & 0 \\ 0 & 0 & \pm 0.5065 & 0 \\ 0 & 0 & 0 & \pm 0.5270 \end{bmatrix} \quad (4.46)$$

Por otro lado, después de procesar todas las observaciones se ha alcanzado la siguiente matriz de amplitudes

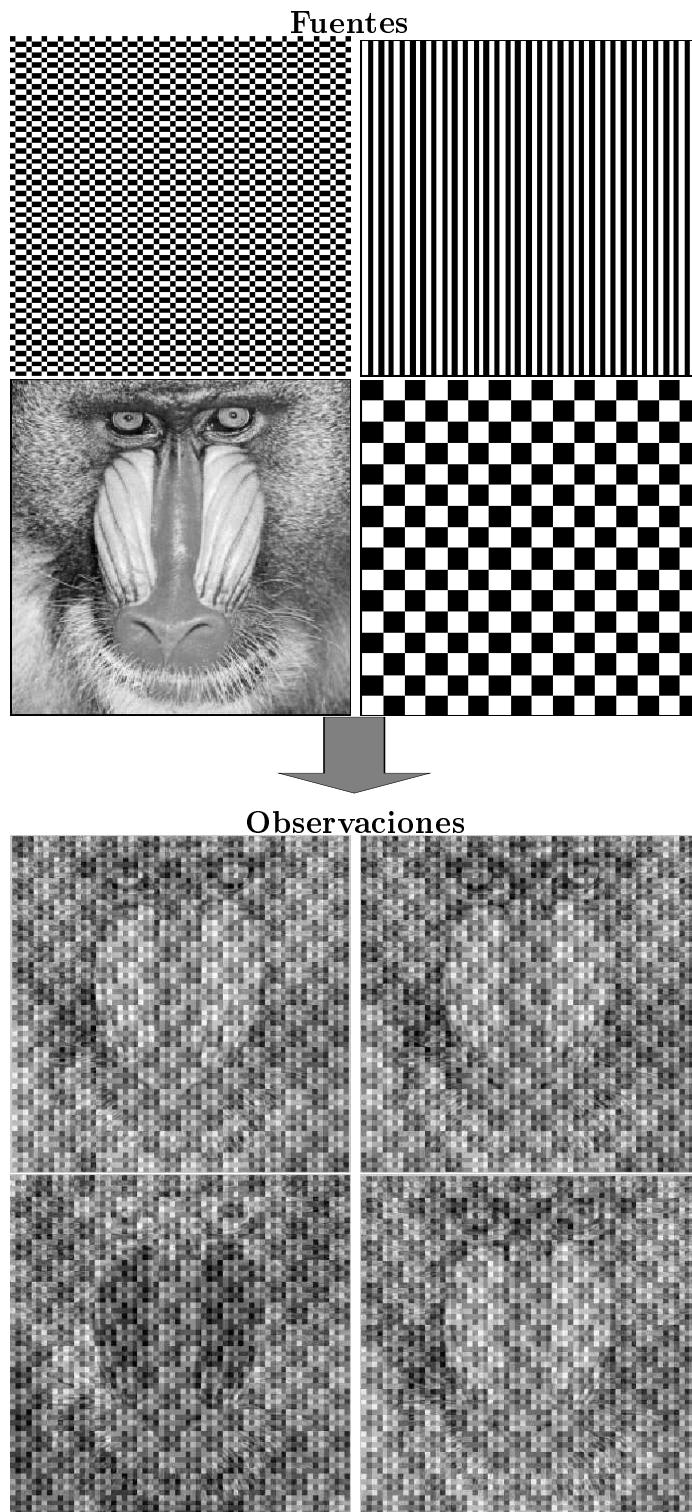


Figura 4.4: Separación de imágenes: fuentes y observaciones.

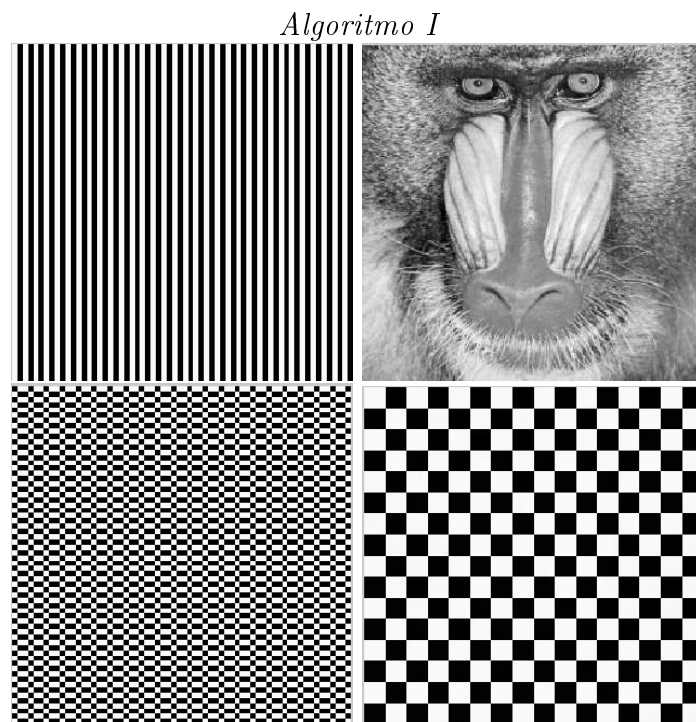


Figura 4.5: Separación de imágenes: señales recuperadas con el algoritmo I.

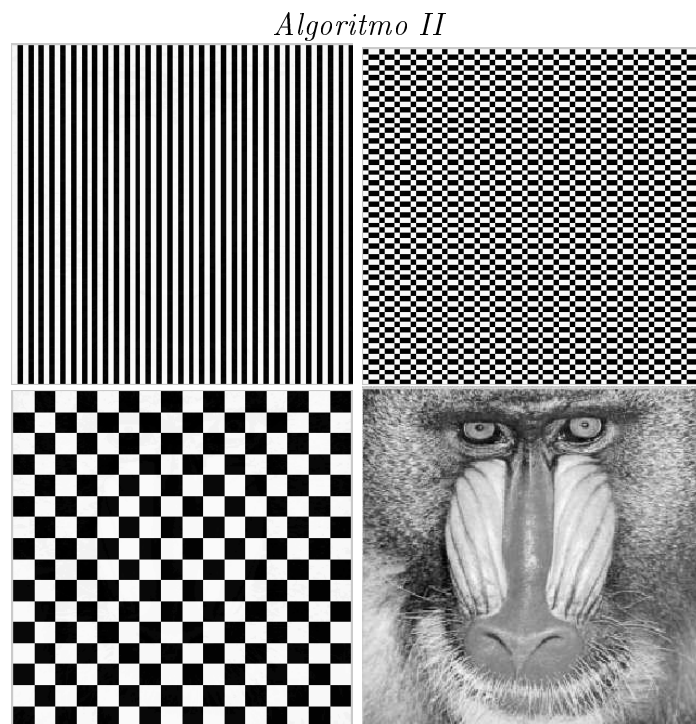


Figura 4.6: Separación de imágenes: señales recuperadas con el algoritmo II.

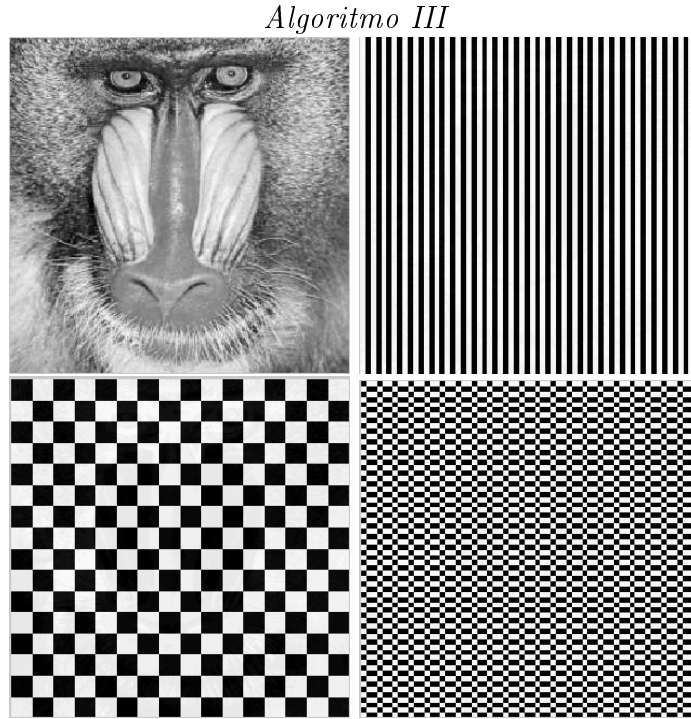


Figura 4.7: Separación de imágenes: señales recuperadas con el algoritmo III.

$$\mathbf{G} = \begin{bmatrix} 0.0005 & \boxed{0.5285} & -0.0036 & -0.0009 \\ -0.0247 & 0.0213 & \boxed{0.5552} & -0.0045 \\ \boxed{0.5283} & 0.0003 & 0.0071 & -0.0008 \\ -0.0009 & -0.0011 & 0.0037 & \boxed{0.5285} \end{bmatrix} \quad (4.47)$$

Como puede observarse, esta matriz es muy cercana a la óptima, lo cual refuerza la validez de nuestro análisis.

Las figuras 4.6 y 4.7 muestran las señales recuperadas utilizando, respectivamente, los algoritmos II y III. Estos algoritmos también han separado adecuadamente las fuentes. Para finalizar este primer experimento mostramos en la figura 4.8 la evolución del índice de rendimiento $P(\mathbf{G})$ para los cuatro algoritmos considerados. Como puede observarse, el algoritmo I y el que maximiza la función de contraste son los que presentan un mejor comportamiento en el entorno que estamos considerando. Los parámetros utilizados en este grupo de simulaciones son los presentados en la tabla 4.1.

En el segundo experimento hemos considerado la separación de tres señales de voz (s_1 , s_2 y s_4) y una de música (s_3) mostradas en la figura 4.9. En este caso, todas las señales tienen curtosis positivas siendo sus valores estimados: $\rho_1 = 7.25$, $\rho_2 = 4.36$, $\rho_3 = 2.76$ y $\rho_4 = 2.70$. En la misma figura se muestran las observaciones obtenidas utilizando la matriz de mezcla \mathbf{A} dada en (4.45). Desafortunadamente, la aplicación directa de estas

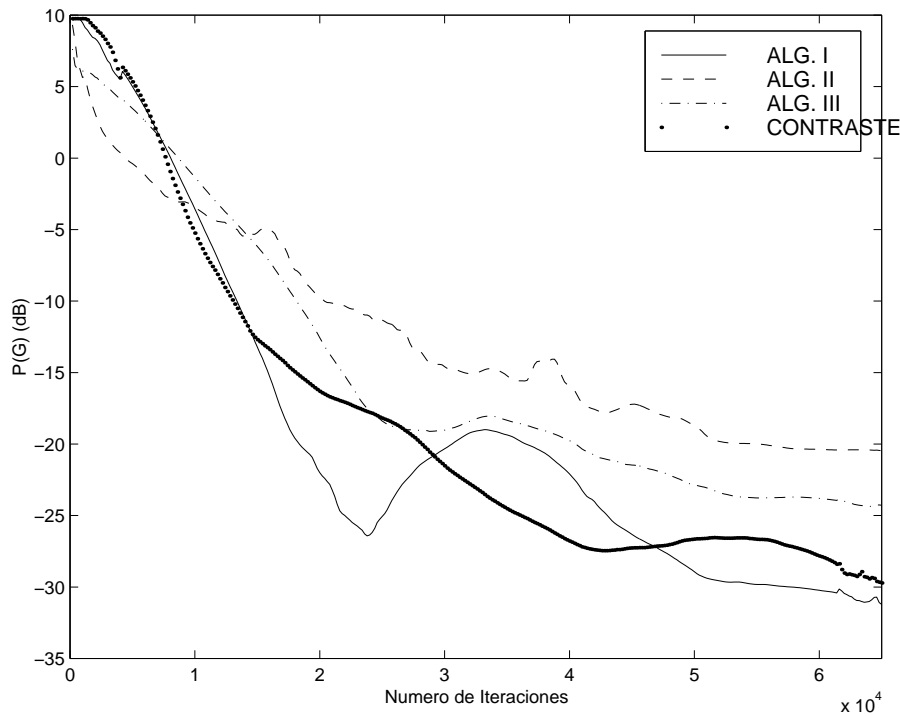


Figura 4.8: Separación de imágenes: evolución del índice $P(\mathbf{G})$.

Algoritmo	Parámetros	
	Paso inicial	Función de coste
<i>Alg.I</i>	$\mu = 10^{-4}, \mu_l = 5 \times 10^{-3}$	$\epsilon = -1, \alpha = 20, \beta = 5, \gamma = 2$
<i>Alg.II</i>	$\mu = 10^{-4}, \mu_l = 5 \times 10^{-3}$	$\epsilon = -1, \alpha = 20, \beta = 5, \gamma = 10$
<i>Alg.III</i>	$\mu = 10^{-4}, \mu_l = 1 \times 10^{-2}$	$\epsilon = -1, \alpha = 20, \beta = 5$
<i>Contraste</i>	$\mu = 10^{-5}, \mu_l = 10^{-2}, \mu_a = 10^{-2}$	$\epsilon = -1, \alpha = 2.5$

Tabla 4.1: Separación de imágenes: parámetros de los algoritmos.

observaciones a cualquiera de los cuatro algoritmos produce una convergencia muy lenta: después de 40.000 iteraciones no se ha conseguido realizar la separación de las fuentes. Este comportamiento indeseable es debido al carácter no estacionario de las observaciones, el cual afecta a los estadísticos involucrados en los algoritmos estocásticos con los que estamos trabajando. Para mejorar el rendimiento hemos permutado aleatoriamente el índice de las observaciones, lo cual equivale a tomar observaciones no consecutivas en el tiempo. Al aplicar estas observaciones a los algoritmos propuestos se han conseguido recuperar las fuentes como se muestra en la figura 4.10.

Comparemos ahora la matriz de amplitudes obtenida con el algoritmo I con la que

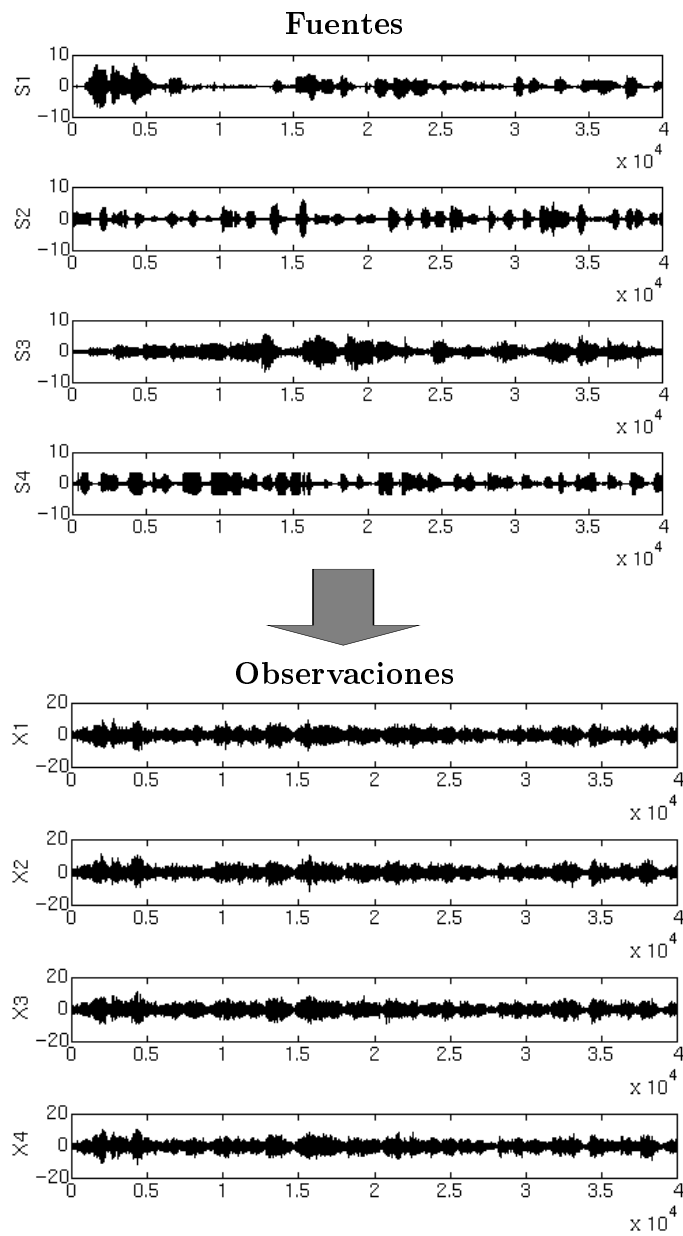


Figura 4.9: Separación de voz y música: fuentes y observaciones.

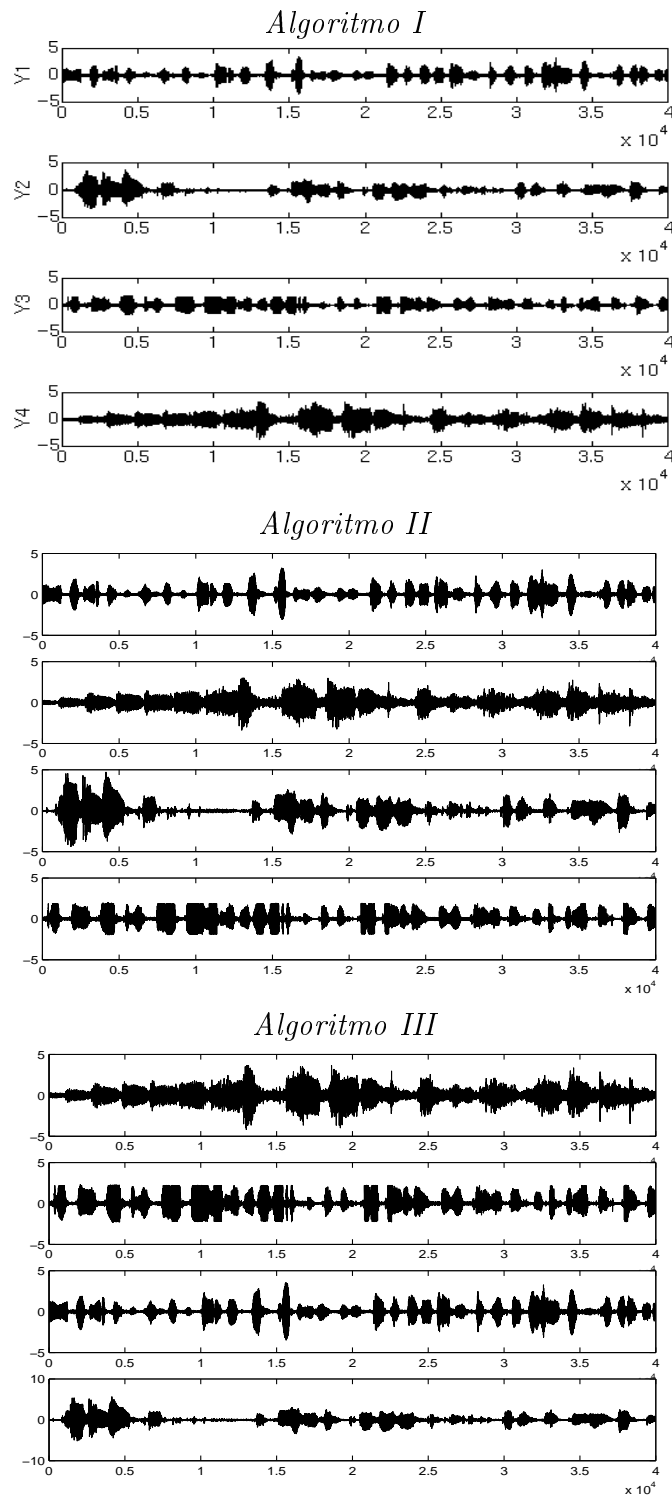


Figura 4.10: Separación de voz y música: señales recuperadas con los algoritmo propuestos.

predice el estudio de su estabilidad presentado en la sección 4.4. Esta última es

$$\mathbf{\Lambda}_o = \begin{bmatrix} \pm 0.6262 & 0 & 0 & 0 \\ 0 & \pm 0.5655 & 0 & 0 \\ 0 & 0 & \pm 0.5385 & 0 \\ 0 & 0 & 0 & \pm 0.5376 \end{bmatrix} \quad (4.48)$$

mientras que después de procesar todo el registro de observaciones (40.000 muestras) el algoritmo ha convergido al siguiente punto

$$\mathbf{G} = \begin{bmatrix} -0.0095 & \boxed{0.5581} & -0.0086 & 0.0233 \\ \boxed{0.5846} & 0.0076 & 0.0020 & -0.0183 \\ 0.0234 & 0.0175 & -0.0048 & \boxed{0.5080} \\ -0.0111 & 0.0120 & \boxed{0.5628} & 0.0119 \end{bmatrix} \quad (4.49)$$

Como puede observarse, en este entorno la matriz que se alcanza con el algoritmo es muy cercana a la teórica.

Finalmente, en la figura 4.11 se muestra la evolución del índice de rendimiento $P(\mathbf{G})$ obtenida con los cuatro algoritmos considerados. Es de destacar la gran similitud de la convergencia de los tres algoritmos presentados en este trabajo y la ligera superioridad sobre el que maximiza la función de contraste. En este grupo de simulaciones se han considerado los parámetros mostrados en la tabla 4.2.

Algoritmo	Parámetros	
	Paso inicial	Función de coste
<i>Alg.I</i>	$\mu = 10^{-4}, \mu_l = 10^{-2}$	$\epsilon = 1, \alpha = 20, \beta = 5, \gamma = 2$
<i>Alg.II</i>	$\mu = 5 \times 10^{-5}, \mu_l = 10^{-3}$	$\epsilon = 1, \alpha = 20, \beta = 5, \gamma = 10$
<i>Alg.III</i>	$\mu = 10^{-4}, \mu_l = 10^{-2}$	$\epsilon = 1, \alpha = 20, \beta = 5$
Contraste	$\mu = 10^{-5}, \mu_l = 10^{-2}, \mu_a = 10^{-2}$	$\epsilon = 1, \gamma = 2.5$

Tabla 4.2: Separación de voz y música: parámetros de los algoritmos.

4.5.2 Mezclas mal condicionadas

Una de las propiedades más importantes del algoritmo IV que hemos propuesto anteriormente es su carácter invariante ante cambios en la matriz de mezcla por lo que es especialmente atractivo para realizar la separación en situaciones donde la matriz \mathbf{A} está mal condicionada.

Para ilustrar esta propiedad, en la figura 4.12 mostramos la evolución del índice de separación $P(\mathbf{G})$ obtenida al utilizar el algoritmo IV para separar las cuatro imágenes de la figura 4.4 a partir de las observaciones proporcionadas por tres sistemas de mezcla

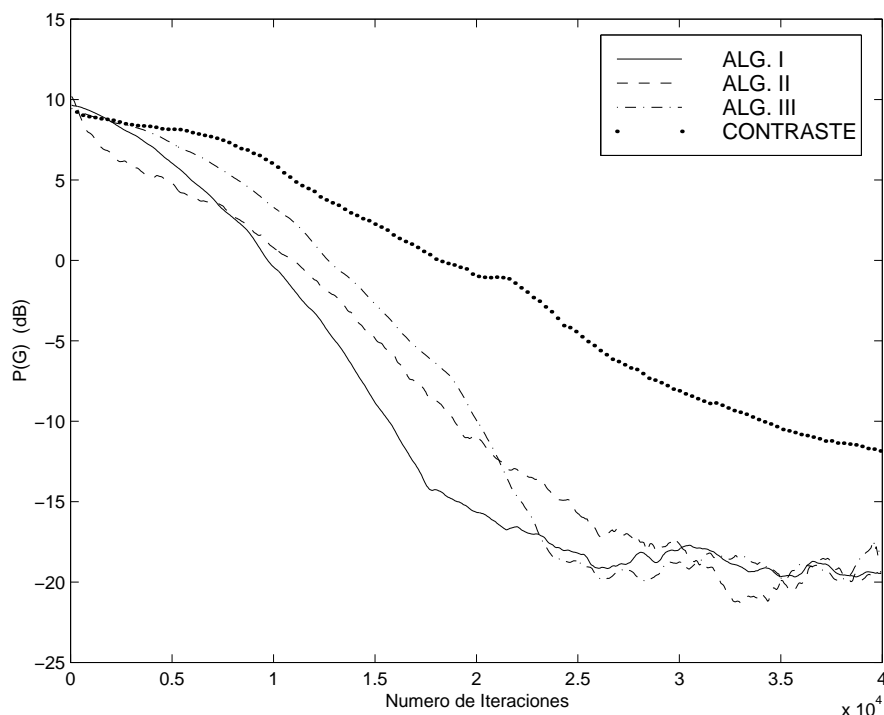


Figura 4.11: Separación de voz y música: evolución del índice $P(\mathbf{G})$.

diferentes: la matriz \mathbf{A} dada por (4.45) y las siguiente matrices mal condicionadas

$$\mathbf{A}_1 = \begin{bmatrix} 1.1 & 1 & 1 & 1 \\ 1 & 1.1 & 1 & 1 \\ 1 & 1 & 1.1 & 1 \\ 1 & 1 & 1 & 1.1 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 1.01 & 1 & 1 & 1 \\ 1 & 1.01 & 1 & 1 \\ 1 & 1 & 1.01 & 1 \\ 1 & 1 & 1 & 1.01 \end{bmatrix} \quad (4.50)$$

Los parámetros del algoritmo utilizados han sido: $\epsilon = -1$, $\alpha = 20$, $\beta = 1$, $\mu = 2 \times 10^{-3}$ y $\mu_l = 2 \times 10^{-2}$. Como puede observarse en la figura 4.12, la evolución del algoritmo en las tres situaciones ha sido similar.

4.5.3 Entornos con ruido blanco gaussiano

En las secciones anteriores hemos ignorado por completo la presencia del ruido que inevitablemente aparece en las situaciones reales. A continuación presentaremos los resultados de varias simulaciones que muestran el comportamiento de los algoritmos propuestos en situaciones donde las observaciones son perturbadas por ruido blanco gaussiano. El modelo que se considerará es el siguiente

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v}$$

donde $\mathbf{v} = [v_1, \dots, v_M]^T$ es un vector de dimensión $M \times 1$ que contiene componentes estadísticamente independientes e idénticamente distribuidas; es decir, su matriz de

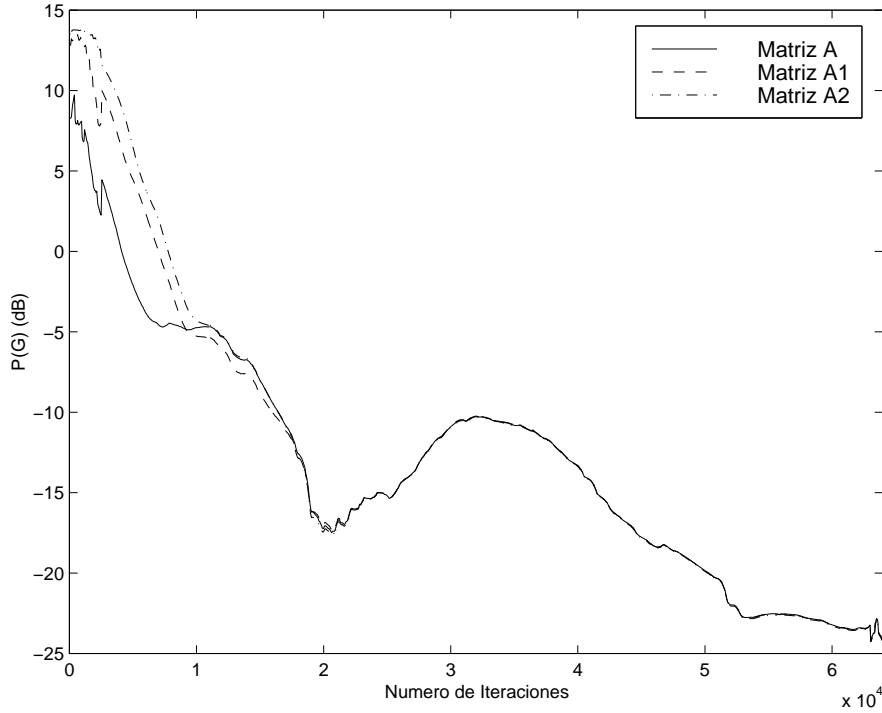


Figura 4.12: Separación de imágenes: evolución del índice $P(\mathbf{G})$ para el algoritmo IV y distintas matrices de mezcla.

autocorrelación es $\mathbf{R}_v = E[\mathbf{v}\mathbf{v}^H] = \sigma_v^2 \mathbf{I}$ donde σ_v^2 es la potencia del ruido y \mathbf{I} es la matriz identidad. En entornos con ruido resulta más adecuado medir el rendimiento del algoritmo en término de la relación señal a interferencia y ruido (*SINR*, del inglés *Signal to Interference and Noise Ratio*) de las fuentes s_j en cada salida y_i

$$SINR_{ij} = \frac{|g_{ij}|^2}{\sum_{\substack{l=1 \\ l \neq j}}^N |g_{il}|^2 + \mathbf{w}_i^H \mathbf{R}_v \mathbf{w}_i} \quad (4.51)$$

Como se demuestra en [66], el valor máximo de este índice es

$$(SINR_{ij})_{max} = \mathbf{a}_j^H (\mathbf{R}_x - \mathbf{a}_j \mathbf{a}_j^H)^{-1} \mathbf{a}_j \quad (4.52)$$

donde $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H]$ es la matriz de auto-correlación de las observaciones.

En este grupo de simulaciones hemos considerado una aplicación de procesado en array donde las observaciones se corresponden con las muestras proporcionadas por un array de M sensores que recibe N fuentes que se propagan como ondas planas. Los sensores son asumidos linealmente equiespaciados separados media longitud de onda lo cual se traduce en una matriz de mezcla compleja cuyos elementos tienen la siguiente forma

$$a_{ik} = \sigma_k e^{-j\pi(i-1)\sin(\theta_k)} \quad i = 1, \dots, M, \quad k = 1, \dots, N \quad (4.53)$$

donde θ_k y $\sigma_k^2 = E[|s_k|^2]$ son los ángulos de llegada y la potencia de la fuente s_k . Para recuperar las fuentes se ha utilizado un sistema que procesa el vector de observaciones de dimensión $M \times 1$ para producir una salida y de dimensión $N \times 1$ lo cual se traduce en una matriz \mathbf{W} no cuadrada. Por tal motivo, solamente pueden emplearse los algoritmos I y II.

Señal	Modulación	Curtosis normalizada	Dirección	σ_k^2/σ_v^2
Fuente #1	4-QAM	-1	-40°	0 dB
Fuente #2	4-QAM	-1	40°	5 dB
Fuente #3	16-QAM	-0.68	0°	10 dB

Tabla 4.3: Parámetros de las fuentes utilizadas en las simulaciones del entorno con ruido.

En un primer experimento hemos supuesto un entorno en el que las tres fuentes descritas en la tabla 4.3 inciden sobre un array de 10 sensores. La figura 4.13 muestra la evolución de las *SINRs* obtenida utilizando el algoritmo I y el algoritmo II. Como puede verse la *SINR* se aproxima al valor óptimo donde se separan las fuentes y se minimiza el efecto del ruido. Los parámetros utilizados en esta simulación son los recogidos en la tabla 4.4.

Algoritmo	Parámetros	
	Paso inicial	Función de coste
<i>Alg.I</i>	$\mu = 10^{-4}, \mu_l = 10^{-1}$	$\epsilon = -1, \alpha = 3, \beta = 1, \gamma = 1.5$
<i>Alg.II</i>	$\mu = 10^{-4}, \mu_l = 10^{-1}$	$\epsilon = -1, \alpha = 3, \beta = 1, \gamma = 1.5$

Tabla 4.4: Parámetros de los algoritmos empleados en las simulaciones del entorno con ruido.

En un segundo experimento hemos considerado un array de 10 sensores sobre el que inciden dos fuentes por ángulos de $\theta_1 = 20^\circ$ y $\theta_2 = -20^\circ$. Las tablas 4.5 y 4.6 muestran el valor máximo *SINR* alcanzable y los *SINR* (promediados en 10 realizaciones) obtenidos en diferentes iteraciones (2000, 4000, 6000, 8000 y 10000) de los algoritmos I y II. Los valores obtenidos revelan que los algoritmos han conseguido la separación de las fuentes para diferentes tipos de modulaciones y *SINRs* de las fuentes (σ_k^2/σ_v^2). En todas las simulaciones se han utilizado los parámetros recogidos en la tabla 4.4.

4.5.4 Aproximación tipo bloque

Como alternativa a los algoritmos presentados en este capítulo se ha pensado en implementar los criterios utilizando una técnica tipo bloque como la presentada en [43]. La idea es estimar los estadísticos que aparecen en los algoritmos empleando L observaciones y, posteriormente, realizar la actualización de la matriz de coeficientes. Este proceso se repite un número S de veces para cada bloque. Las pruebas realizadas en este sentido

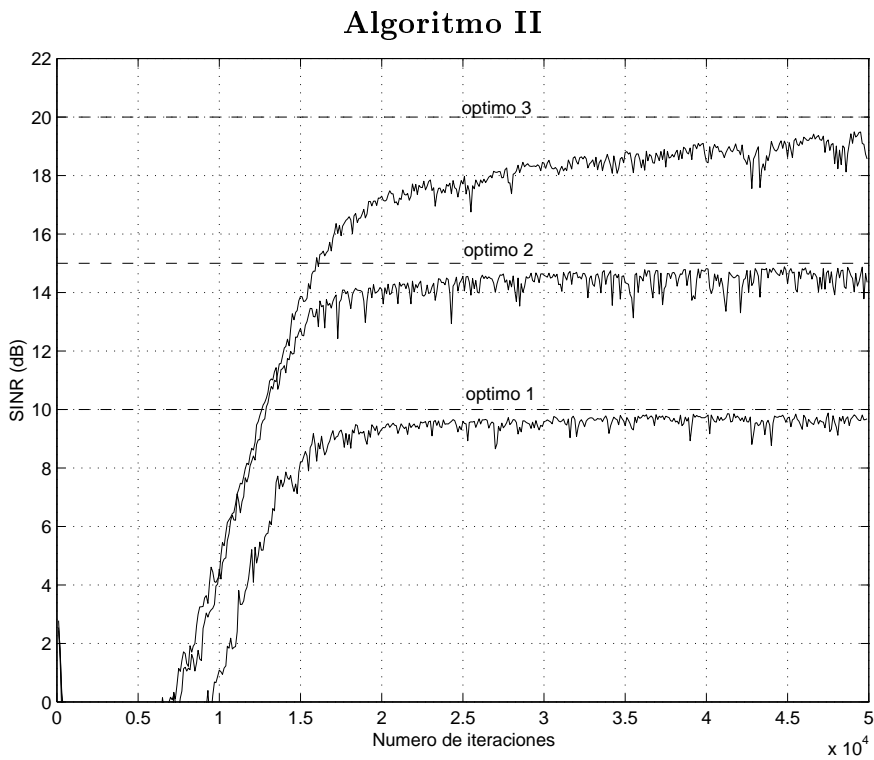
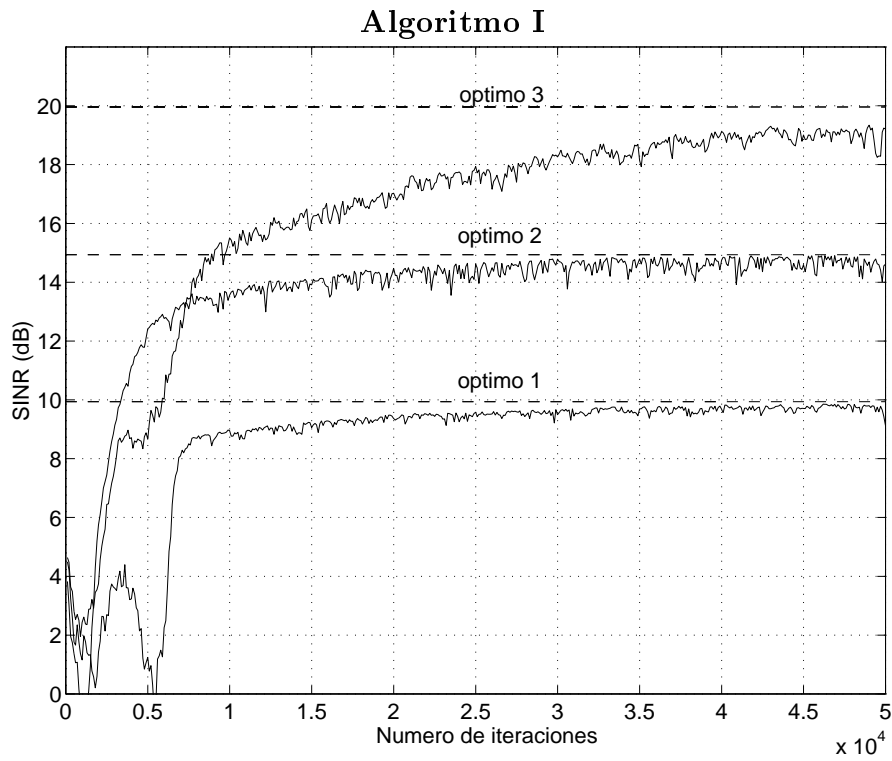


Figura 4.13: Evolución de la SINR para los algoritmos I y II.

Algoritmo I

Fuentes		$SINR_{max}$ (dB)	$SINR$ promedio para diferentes iteraciones				
Modulación	σ_k^2/σ_v^2 (dB)		2000	4000	6000	8000	10000
4-QAM	0	10	6.7952	8.6498	9.2631	9.4401	9.6052
16-QAM	0	10	6.5863	8.6353	9.2085	9.4220	9.5812
4-QAM	0	10	7.0675	8.6512	9.2668	9.3741	9.5281
16-QAM	5	15	10.5891	12.6064	13.6280	13.6752	14.3298
4-QAM	5	15	11.5489	13.3322	14.0410	13.9827	14.4591
16-QAM	5	15	11.5360	12.9556	13.9999	14.0179	14.4562
4-QAM	0	10	8.2282	9.4584	9.4158	9.5896	9.6610
4-QAM	0	10	8.1184	9.3625	9.5742	9.6405	9.6064
4-QAM	0	10	7.4873	8.7939	9.2520	9.4480	9.4261
4-QAM	5	15	13.3196	14.3763	14.0854	14.1882	14.4582
4-QAM	5	15	12.6589	13.6800	14.2519	14.3399	14.6131
4-QAM	5	15	12.6855	13.4508	14.4131	14.2174	14.6274

Tabla 4.5: Entornos con ruido: $SINRs$ obtenidas con el algoritmo I.

muestran que los algoritmos convergen en muy pocas iteraciones con bloques de datos pequeños. A continuación presentamos algunos de estos resultados.

En un primer experimento se ha considerado la separación de las cuatro imágenes de la figura 4.4 a partir de las observaciones mostradas en la misma figura. El sistema separador ha sido ajustado utilizando el algoritmo I con los parámetros: $\epsilon = -1$, $\alpha = 20$, $\beta = 5$, $\gamma = 2$, $\mu = 5 \times 10^{-2}$. La figura 4.14 muestra la evolución del índice $P(\mathbf{G})$ obtenido para distintos tamaños de bloque $L = 128$, $L = 256$, $L = 384$ y $L = 512$. Como puede observarse, se ha conseguido la separación en menos de $S = 25$ iteraciones salvo en el caso de bloques de tamaño $L = 128$. Además, las matrices de amplitudes \mathbf{G} obtenidas son muy cercanas a la que predice el análisis de estabilidad presentado en la sección 4.4. Por ejemplo, para un tamaño de bloque $L = 256$ se ha alcanzado la siguiente matriz después de $S = 40$ iteraciones.

$$\mathbf{G} = \begin{bmatrix} 0.0050 & -0.0208 & -0.0449 & \boxed{0.5349} \\ 0.0026 & \boxed{0.5439} & -0.0016 & 0.0002 \\ \boxed{0.5372} & 0.0036 & -0.0194 & 0.0078 \\ 0.0140 & 0.1252 & \boxed{0.5491} & -0.0354 \end{bmatrix} \quad (4.54)$$

En un segundo experimento se ha considerado una aplicación de procesado en array donde se persigue separar tres señales complejas (dos 4-QAM y una 16-QAM) que inciden sobre un array de 10-sensores por ángulos: $\theta_1 = -40^\circ$, $\theta_2 = 0^\circ$ y $\theta_3 = 40^\circ$. Los coeficientes fueron actualizados utilizando el algoritmo I con $\epsilon = -1$, $\alpha = 2$, $\beta = 1$, $\gamma = 2$ y $\mu = 5 \times 10^{-2}$. En la figura 4.15 puede observarse el índice $P(\mathbf{G})$ obtenido en la iteración $S = 40$ para tamaños de bloque desde $L = 25$ hasta $L = 55$. Al igual que en el experimento

Algoritmo II

Fuentes		$SINR_{max}$ (dB)	$SINR$ promedio para diferentes iteraciones				
Modulación	σ_k^2/σ_v^2 (dB)		2000	4000	6000	8000	10000
4-QAM	0	10	7.0588	9.0948	9.3953	9.6728	9.7592
16-QAM	0	10	6.9248	8.8335	9.1750	9.4887	9.5673
4-QAM	0	10	7.0415	8.9414	9.5880	9.4031	9.6498
16-QAM	5	15	9.7420	12.3151	13.5749	14.0709	14.2266
4-QAM	5	15	10.9969	12.7272	13.6943	13.9481	14.4591
16-QAM	5	15	12.9802	13.2533	14.3521	14.194	14.7814
4-QAM	0	10	8.0470	9.4322	9.6835	9.8011	9.8660
4-QAM	0	10	7.0610	9.3833	9.6838	9.8231	9.8319
4-QAM	0	10	8.4426	8.9010	9.2547	9.4523	9.5277
4-QAM	5	15	13.6454	14.2412	14.6411	14.8307	14.7082
4-QAM	5	15	11.9049	13.3055	14.4597	14.2407	14.4207
4-QAM	5	15	12.6855	13.7953	14.0068	13.9342	14.3687

Tabla 4.6: Entornos con ruido: $SINRs$ obtenidas con el algoritmo II.

anterior, se ha observado que la matriz de amplitudes alcanzada es muy cercana a la que predice el análisis de la estabilidad de este algoritmo.

Por último, se ha considerado separar las señales de voz y música de la figura 4.9 utilizando esta aproximación bloque. Los resultados, sin embargo, no han sido nada satisfactorios debido a que estas señales son altamente no estacionarias y el signo de su curtosis depende del bloque de datos que se procese. Esto hace imposible la aplicabilidad de estas técnicas.

4.6 Conclusiones

En el presente capítulo se ha desarrollado una familia de criterios de separación que hacen uso de momentos de segundo orden y cumulantes de orden superior. Los nuevos criterios surgen a partir de un método de filtrado ciego propuesto por Shalvi y Weinstein a principios de esta década. La aplicación directa de dicho criterio al problema de separación resulta infructuosa ya que puede ocurrir que varias salidas extraigan la misma fuente. En esta situación indeseada aparecen dos tipos de dependencias que, como se ha visto en este capítulo, son la clave para conseguir la separación de las fuentes. En primer lugar existe una dependencia estadística entre las salidas que se traduce en que sus estadísticos cruzados sean distintos de cero. Y, en segundo lugar, aparece una dependencia lineal entre las filas de la matriz de coeficientes que da lugar a un determinante igual a cero. Para aprovechar estas dependencias, hemos propuesto añadir ciertos términos cruzados al criterio de Shalvi y Weinstein.

En el criterio que ha ocupado la mayor parte de este capítulo hemos añadido un

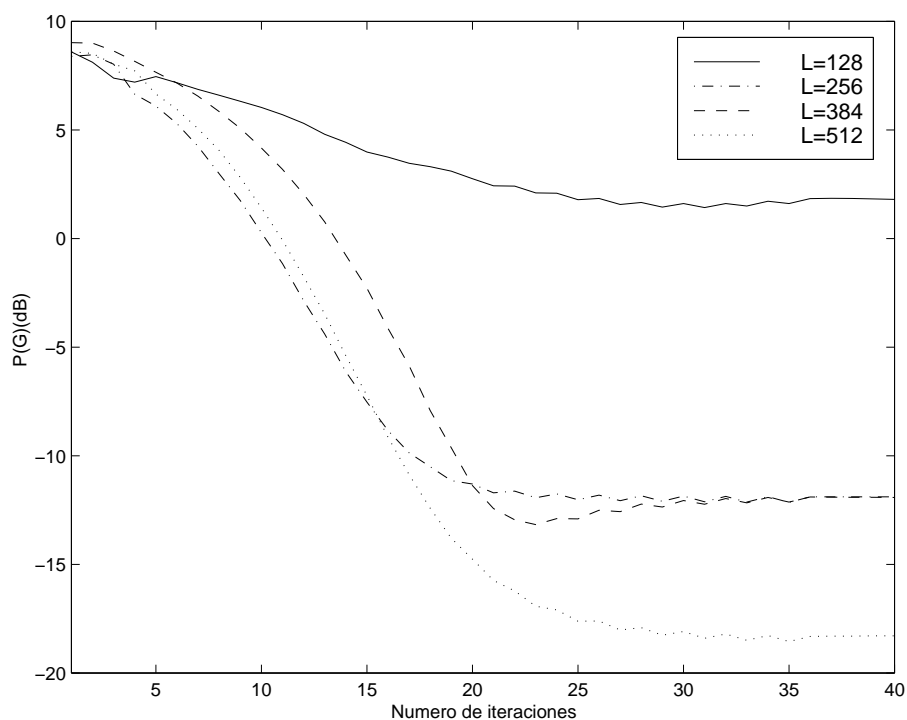


Figura 4.14: Evolución del índice $P(\mathbf{G})$ para separación de imágenes utilizando una aproximación tipo bloque.

conjunto de cumulantes de cuarto orden entre las salidas. Estos cumulantes son distintos de cero cuando varias salidas extraen la misma fuente y se anularán cuando se consigue la separación. Una de las ventajas de la nueva aproximación es que los coeficientes del sistema separador pueden ajustarse utilizando un algoritmo de gradiente estocástico cuya convergencia ha sido estudiada en este mismo capítulo. El análisis demuestra que los estados donde se consigue la perfecta separación de las fuentes son atractores del algoritmo. Existen otros puntos estacionarios pero también se ha demostrado que son inestables y, por tanto, el algoritmo de gradiente nunca convergerá a ellos.

Otra forma de aprovechar la dependencia estadística entre las salidas para conseguir la separación consiste en añadir la correlación entre las salidas, la cual únicamente se anula cuando se consigue la separación. El algoritmo de gradiente que resulta de considerar este criterio es ligeramente más simple que el anterior pero, desafortunadamente, sólo hemos probado su estabilidad para el caso de sistemas separadores con dos entradas y dos salidas.

Por otro lado, la dependencia entre las filas de la matriz de pesos nos ha conducido a emplear una función logarítmica del determinante de la matriz de separación como término cruzado. La implementación de este criterio con un algoritmo de máxima pendiente requiere invertir la matriz de coeficientes y, por ello, resulta más conveniente emplear un algoritmo de gradiente relativo. Independientemente del algoritmo que se emplee, esta

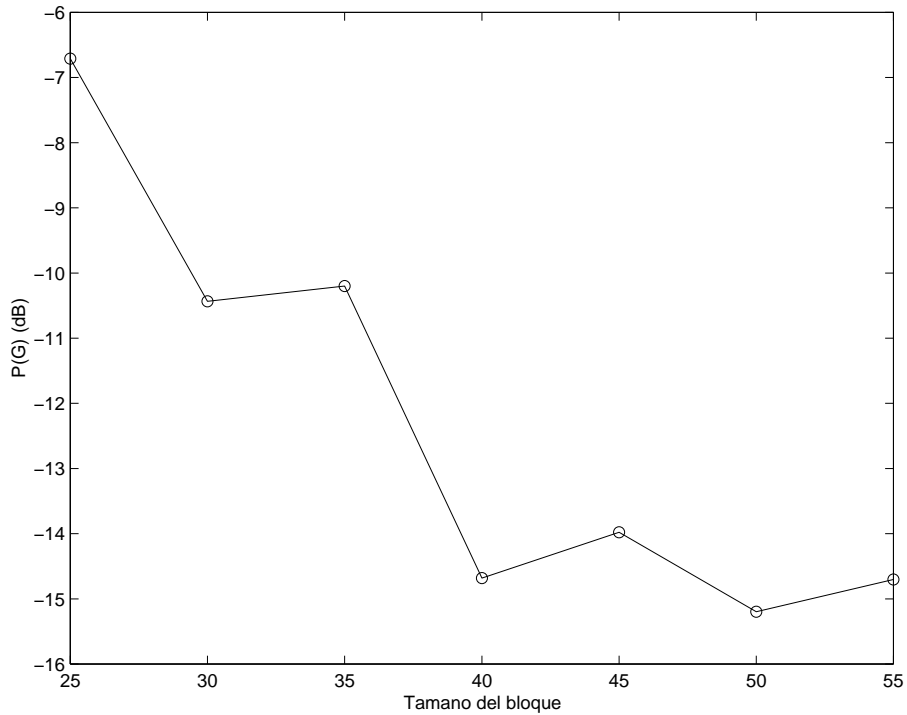


Figura 4.15: Índice $P(\mathbf{G})$ en la iteración $S = 40$ para separación de señales de comunicación utilizando una aproximación tipo bloque.

aproximación solamente puede ser utilizada cuando la matriz de separación es cuadrada. Adicionalmente, la inexistencia de atractores indeseados es una cuestión que todavía no ha sido demostrada.

Finalmente, hemos presentado diversas simulaciones realizadas por ordenador que muestran un buen comportamiento de los algoritmos propuestos tanto en entornos con fuentes de curtosis positiva como negativa. Además revelan que las prestaciones de nuestros criterios no se ven degradadas cuando las observaciones son perturbadas por ruido blanco gaussiano.

Capítulo 5

Criterios de separación para mezclas convolutivas

5.1 Introducción

Los criterios de separación que hemos presentado en los capítulos anteriores asumen que las observaciones son mezclas lineales e instantáneas de las fuentes. Sin embargo, en la mayoría de las situaciones prácticas es más acertado asumir un modelo de mezcla con memoria en el que las observaciones se representan como la convolución entre los parámetros del sistema de mezcla y las fuentes. Como se comentó en el capítulo 1, el objetivo del sistema separador es recuperar todas las fuentes eliminando la distorsión introducida por el sistema de mezcla.

Una gran parte de los criterios para separación de mezclas convolutivas han surgido como extensiones de los criterios propuestos inicialmente para mezclas instantáneas. En particular, en [61] se ha generalizado el algoritmo de Héroult y Jutten y en [20, 57] se ha extendido la idea de funciones de contrastes. Siguiendo en esta línea en la sección 5.2 extenderemos los criterios I y II propuestos en la sección 4.2 del capítulo anterior. A continuación, en la sección 5.3, desarrollaremos dos algoritmos de gradiente ascendente que permiten calcular los coeficientes del sistema separador y en la sección 5.4 analizaremos la estabilidad de uno de ellos. La sección 5.5 recoge algunas simulaciones que ilustran el comportamiento de los algoritmos. Finalmente, la sección 5.6 está dedicada a las conclusiones más importantes de este capítulo.

5.2 Criterios para separación de mezclas convolutivas

Como ya hemos comentado en el capítulo 1, el modelo de mezclas convolutivas asume que las observaciones $\mathbf{x}(n)$ tienen la siguiente forma

$$\mathbf{x}(n) = \sum_{k=-\infty}^{\infty} \mathbf{A}(k)\mathbf{s}(n-k) \quad (5.1)$$

donde $\{\mathbf{A}(k)\}$ es una secuencia de matrices de dimensión $M \times N$ que define la respuesta al impulso y $\mathbf{s}(k)$ es un vector $N \times 1$ que contiene las fuentes que, recordemos, son procesos estocásticos estacionarios, con distribución no-gaussiana, estadísticamente independientes, de media cero y varianza unidad. Adicionalmente, en este capítulo también asumiremos que las fuentes son blancas, i.e., $E[s_i(n)s_i(n+\tau)] = \delta(\tau)$. Esta hipótesis es válida, por ejemplo, en aplicaciones de comunicaciones donde cada símbolo transmitido es independiente de los anteriores y posteriores.

Resolver el problema de separación de fuentes a partir de observaciones de la forma (5.1) es más difícil que cuando se trabaja con mezclas instantáneas. Obsérvese que ahora las observaciones (5.1) pueden interpretarse como la suma de infinitos problemas de mezclas instantáneas de la forma $\mathbf{A}(k)\mathbf{s}(n-k)$

$$\mathbf{x}(n) = \mathbf{A}(0)\mathbf{s}(n) + \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \mathbf{A}(k)\mathbf{s}(n-k) \quad (5.2)$$

Si consideramos que $\mathbf{s}(n)$ contienen las fuentes que se desean recuperar en el instante n y suprimimos el segundo sumando llegaremos a un problema de separación de mezcla instantáneas $\mathbf{A}(0)\mathbf{s}(n)$. Por tanto, el sistema separador de mezclas convolutivas debe, por un lado, eliminar el efecto debido a la memoria del sistema mezclador y, por otro lado, separar las fuentes. Para conseguir este objetivo, las observaciones son procesadas por un sistema MIMO con memoria cuya salida podemos escribir como sigue

$$\mathbf{y}(n) = \sum_{k=-\infty}^{\infty} \mathbf{W}^H(k)\mathbf{x}(n-k) \quad (5.3)$$

donde $\{\mathbf{W}(k)\}$ es una secuencia de matrices $M \times N$ que representa la respuesta al impulso del sistema.

En esta sección proponemos seleccionar el sistema MIMO utilizando extensiones de los criterios de separación I y II que hemos presentado en el capítulo anterior. Recordemos que nuestra aproximación para separación ciega de mezclas instantáneas ha surgido como una generalización del siguiente criterio de igualación ciega

$$J = \sum_{i=1}^N |K_{y_i}(0)| - f(E[|y_i(n)|^2])$$

$$= \sum_{i=1}^N \left(\left| E[|y_i(n)|^4] - 2E^2[|y_i(n)|^2] - |E[y_i(n)^2]|^2 - f(E[|y_i(n)|^2]) \right| \right) \quad (5.4)$$

Esta expresión es la misma que (4.7) con la salvedad de que hemos incorporado la dependencia con el tiempo. Así $K_{y_i}(0) = cum(y_i(n), y_i^*(n), y_i(n), y_i^*(n))$ denota la curtosis de la salida $y_i(n)$. Recordemos que $f(\cdot)$ es una función continua, de dominio y rango real, derivable que debe ser elegida de forma que el siguiente conjunto de funciones tengan un único máximo en $x > 0$

$$p_i(x) = |\rho_i|x^2 - f(x), \quad i = 1, \dots, N \quad (5.5)$$

donde $\rho_i = K_{s_i}(0)/E^2[|s_i(n)|^2]$ denota la curtosis normalizada de la fuente $s_i(n)$.

Una importante propiedad de la función de coste (5.4) es que por sí sola es capaz de eliminar el efecto causado por la memoria del sistema mezclador. Un análisis similar al presentado en [68] demuestra que los únicos máximos de (5.4) se corresponden con la recuperación de las fuentes deseadas. Esta función tiene otros puntos estacionarios pero también se demuestra que son mínimos o puntos sillas y, por tanto, no se corresponden con atractores de un algoritmo de gradiente. Sin embargo, como ya se ha expuesto en el capítulo anterior la maximización de la función (5.4) es insuficiente para llevar a cabo la separación ya que varias salidas pueden extraer la misma fuente. Para evitar esta situación indeseada, en el capítulo 4 hemos añadido ciertos términos cruzados que penalizan la dependencia estadística entre salidas diferentes. En particular, el criterio I emplea cumulantes de cuarto orden entre las salidas y el criterio II utiliza la correlación cruzada.

Estos criterios han demostrado ser adecuados para el caso de mezclas instantáneas pero resultan insuficientes cuando se trabaja con mezclas convolutivas ya que puede ocurrir que las salidas extraigan la misma fuente con un cierto retardo τ . En esta situación indeseada aparece una dependencia estadística entre el valor de la salida $y_i(n)$ y el de la salida $y_j(n + \tau)$ que puede ser penalizada utilizando los siguientes términos cruzados

- *Criterio I generalizado:* Como primer término cruzado proponemos emplear el cumulante cruzado de cuarto orden entre $y_i(n)$ y $y_j(n + \tau)$ definido como sigue

$$T_{1G} = -\gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{\tau=-\infty}^{\infty} |K_{y_i y_j}(0, \tau)| \quad (5.6)$$

donde γ es un valor real positivo y $K_{y_i y_j}(0, \tau)$ es la curtosis entre $y_i(n)$ y $y_j(n + \tau)$

$$\begin{aligned} K_{y_i y_j}(0, \tau) &= cum(y_i(n), y_i^*(n), y_j(n + \tau), y_j^*(n + \tau)) \\ &= E[|y_i(n)|^2 |y_j(n + \tau)|^2] - E[|y_i(n)|^2] E[|y_j(n + \tau)|^2] \\ &\quad - |E[y_i(n) y_j^*(n + \tau)]|^2 - |E[y_i(n) y_j(n + \tau)]|^2 \end{aligned} \quad (5.7)$$

Este término se anulará únicamente cuando las salidas $y_i(n)$ y $y_j(n + \tau)$ sean estadísticamente independientes para cualquier retardo. El criterio de separación que resulta de añadir este término cruzado a la expresión (5.4) es el siguiente

$$\begin{aligned} \max_{\mathbf{W}} J_{1G}(\mathbf{W}) &= J + T_{1G} \\ &= \sum_{i=1}^N \left(|K_{y_i}(0)| - f(E[|y_i(n)|^2]) \right) - \gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{\tau=-\infty}^{\infty} |K_{y_i y_j}(0, \tau)| \end{aligned} \quad (5.8)$$

- *Criterio II generalizado:* Otra forma de penalizar la dependencia estadística entre las salidas consiste en utilizar la correlación cruzada entre ellas

$$T_{2G} = -\gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{\tau=-\infty}^{\infty} |E[y_i(n)y_j^*(n + \tau)]|^2 \quad (5.9)$$

El criterio de separación que resulta de emplear (5.9) es el siguiente

$$\begin{aligned} \max_{\mathbf{W}} J_{2G}(\mathbf{W}) &= J + T_{2G} \\ &= \sum_{i=1}^N \left(|K_{y_i}(0)| - f(E[|y_i(n)|^2]) \right) - \gamma \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{\tau=-\infty}^{\infty} |E[y_i(n)y_j^*(n + \tau)]|^2 \end{aligned} \quad (5.10)$$

Otros autores [47, 64] han propuesto criterios estadísticos similares a (5.10). En concreto, el criterio propuesto en [64] puede entenderse como un caso particular del (5.10) en el caso de que la función $f(\cdot)$ se elija de forma que dé lugar al criterio Constant Modulus.

5.3 Algoritmos adaptativos

Los problemas de optimización que acabamos de describir permiten calcular los coeficientes del sistema separador utilizando algoritmos de gradiente estocástico de la forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \hat{\nabla}_{\mathbf{W}} J_{kG}(n) = \mathbf{W}(n) + \mu \left(\hat{\nabla}_{\mathbf{W}} J(n) + \hat{\nabla}_{\mathbf{W}} T_{kG}(n) \right) \quad (5.11)$$

donde $\hat{\nabla}_{\mathbf{W}} J(n)$ y $\hat{\nabla}_{\mathbf{W}} T_{kG}(n)$ son estimaciones de los gradientes de J dada en (5.4) y de los términos cruzados empleados por los criterios de separación, T_{1G} ó T_{2G} .

Utilizando las propiedades dadas en el apéndice A se obtiene que el gradiente $\nabla_{\mathbf{w}} J$ viene dado por la siguiente expresión

$$\nabla_{\mathbf{w}} J = \begin{bmatrix} 2\epsilon(E[\mathbf{x}(n)y_1^*(n)|y_1(n)|^2] - 2E[\mathbf{x}(n)y_1^*(n)]E[|y_1(n)|^2] - E[\mathbf{x}(n)y_1(n)]E[y_1^*(n)^2]) - \nabla_{\mathbf{w}_1} f(E[|y_1(n)|^2]) \\ \vdots \\ 2\epsilon(E[\mathbf{x}(n)y_N^*(n)|y_N(n)|^2] - 2E[\mathbf{x}(n)y_N^*(n)]E[|y_N(n)|^2] - E[\mathbf{x}(n)y_N(n)]E[y_N^*(n)^2]) - \nabla_{\mathbf{w}_N} f(E[|y_N(n)|^2]) \end{bmatrix} \quad (5.12)$$

donde ϵ es el signo de la curtosis de las fuentes que supondremos conocido por el sistema de separación. El término $\nabla_{\mathbf{w}_i} f(E[|y_i(n)|^2])$ depende de la función $f(\cdot)$. En particular, para el polinomio de segundo orden

$$f(x) = \alpha x^2 - 2\beta x + 1 \quad (5.13)$$

obtenemos la siguiente expresión

$$\nabla_{\mathbf{w}_i} f(E[|y_i(n)|^2]) = 2E[y_i^*(n)\mathbf{x}(n)](\alpha E[|y_i(n)|^2] - \beta) \quad (5.14)$$

Sustituyendo (5.14) en (5.12) y reemplazando los promedios por sus estimaciones, podemos escribir el gradiente como sigue

$$\hat{\nabla}_{\mathbf{w}} J(n) = 2\mathbf{x}(n)\hat{\mathbf{c}}^T(n) \quad (5.15)$$

donde

$$\hat{\mathbf{c}}(n) = \begin{bmatrix} \epsilon(y_1^*(n)|y_1(n)|^2 - 2y_1^*(n)\hat{E}[|y_1(n)|^2] - y_1(n)\hat{E}[y_1^*(n)^2]) - y_1^*(n)(\alpha\hat{E}[|y_1(n)|^2] - \beta) \\ \vdots \\ \epsilon(y_N^*(n)|y_N(n)|^2 - 2y_N^*(n)\hat{E}[|y_N(n)|^2] - y_N(n)\hat{E}[y_N^*(n)^2]) - y_N^*(n)(\alpha\hat{E}[|y_N(n)|^2] - \beta) \end{bmatrix} \quad (5.16)$$

Los estimadores de los promedios en (5.16) pueden ser calculados iterativamente con los siguientes algoritmos

$$\begin{aligned} \hat{E}[|y_i(n)|^2] &= (1 - \mu_i) \hat{E}[|y_i(n)|^2] + \mu_i |y_i(n)|^2 \\ \hat{E}[y_i^*(n)^2] &= (1 - \mu_i) \hat{E}[y_i^*(n)^2] + \mu_i y_i^*(n)^2 \end{aligned} \quad (5.17)$$

Determinemos ahora las expresiones particulares de los gradientes de las funciones de coste (5.8) y (5.10).

5.3.1 Algoritmo I generalizado

Al considerar la función de coste (5.8), el algoritmo de gradiente (5.11) adquiere la siguiente forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(\mathbf{x}(n)\hat{\mathbf{c}}^T(n) - \epsilon\gamma\hat{\mathbf{c}}_{1G}^T(n)) \quad (5.18)$$

donde un factor 2 ha sido incluido dentro del parámetro de paso y $\hat{\mathbf{c}}(n)$ viene dado por la expresión (5.16). Utilizando las propiedades dadas en el apéndice A se obtiene la siguiente expresión del gradiente del término cruzado (5.6)

$$\begin{aligned} \nabla_{\mathbf{W}} T_{1G} = & \\ - 2\gamma \sum_{\tau \geq 0} & \left[\begin{array}{l} \sum_{\substack{j=1 \\ j \neq 1}}^N (E[\mathbf{x}(n)y_1^*(n)|y_j(n-\tau)|^2] - E[\mathbf{x}(n)y_1^*(n)]E[|y_j(n-\tau)|^2] - \\ E[\mathbf{x}(n)y_j^*(n-\tau)]E[y_1^*(n)y_j(n-\tau)] - E[\mathbf{x}(n)y_j(n-\tau)]E[y_1^*(n)y_j^*(n-\tau)]) \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N (E[\mathbf{x}(n)y_N^*(n)|y_j(n-\tau)|^2] - E[\mathbf{x}(n)y_N^*(n)]E[|y_j(n-\tau)|^2] - \\ E[\mathbf{x}(n)y_j^*(n-\tau)]E[y_N^*(n)y_j(n-\tau)] - E[\mathbf{x}(n)y_j(n-\tau)]E[y_N^*(n)y_j^*(n-\tau)]) \end{array} \right] \\ - 2\gamma \sum_{\tau \geq 0} & \left[\begin{array}{l} \sum_{\substack{j=1 \\ j \neq 1}}^N (E[\mathbf{x}(n-\tau)y_1^*(n-\tau)|y_j(n)|^2] - E[\mathbf{x}(n-\tau)y_1^*(n-\tau)]E[|y_j(n)|^2] - \\ E[\mathbf{x}(n-\tau)y_j^*(n)]E[y_1^*(n-\tau)y_j(n)] - E[\mathbf{x}(n-\tau)y_j(n)]E[y_1^*(n-\tau)y_j^*(n)]) \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N (E[\mathbf{x}(n-\tau)y_N^*(n-\tau)|y_j(n)|^2] - E[\mathbf{x}(n-\tau)y_N^*(n-\tau)]E[|y_j(n)|^2] - \\ E[\mathbf{x}(n-\tau)y_j^*(n)]E[y_N^*(n-\tau)y_j(n)] - E[\mathbf{x}(n-\tau)y_j(n)]E[y_N^*(n-\tau)y_j^*(n)]) \end{array} \right] \end{aligned} \quad (5.19)$$

Sustituyendo los promedios que aparecen en la expresión anterior por sus estimaciones se obtiene la matriz $\hat{\mathbf{c}}_{1G}(n)$ que aparece en (5.18)

$$\begin{aligned} \hat{\mathbf{c}}_{1G}(n) = \mathbf{x}(n) \sum_{\tau \geq 0} & \left[\begin{array}{l} \sum_{\substack{j=1 \\ j \neq 1}}^N (y_1^*(n)|y_j(n-\tau)|^2 - y_1^*(n)\hat{E}[|y_j(n-\tau)|^2] - \\ y_j^*(n-\tau)\hat{E}[y_1^*(n)y_j(n-\tau)] - y_j(n-\tau)\hat{E}[y_1^*(n)y_j^*(n-\tau)]) \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N (y_N^*(n)|y_j(n-\tau)|^2 - y_N^*(n)\hat{E}[|y_j(n-\tau)|^2] - \\ y_j^*(n-\tau)\hat{E}[y_N^*(n)y_j(n-\tau)] - y_j(n-\tau)\hat{E}[y_N^*(n)y_j^*(n-\tau)]) \end{array} \right] \\ + \sum_{\tau \geq 0} \mathbf{x}(n-\tau) & \left[\begin{array}{l} \sum_{\substack{j=1 \\ j \neq 1}}^N (y_1^*(n-\tau)|y_j(n)|^2 - y_1^*(n-\tau)\hat{E}[|y_j(n)|^2] - \\ y_j^*(n)\hat{E}[y_1^*(n-\tau)y_j(n)] - y_j(n)\hat{E}[y_1^*(n-\tau)y_j^*(n)]) \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N (y_N^*(n-\tau)|y_j(n)|^2 - y_N^*(n-\tau)\hat{E}[|y_j(n)|^2] - \\ y_j^*(n)\hat{E}[y_N^*(n-\tau)y_j(n)] - y_j(n)\hat{E}[y_N^*(n-\tau)y_j^*(n)]) \end{array} \right] \end{aligned} \quad (5.20)$$

Los estimadores de los promedios que aparecen en esta expresión pueden ser calculados iterativamente con algoritmos de ventana exponencial similares a los presentados en el capítulo anterior.

$$\begin{aligned}\hat{E}[y_i^*(n)y_j^*(n-\tau)] &= (1-\mu_l)\hat{E}[y_i^*(n)y_j^*(n-\tau)] + \mu_l y_i^*(n)y_j^*(n-\tau) \\ \hat{E}[y_i(n-\tau)y_j^*(n)] &= (1-\mu_l)\hat{E}[y_i(n-\tau)y_j^*(n)] + \mu_l y_i(n-\tau)y_j^*(n)\end{aligned}\quad (5.21)$$

donde μ_l es un número real positivo menor que 1.

5.3.2 Algoritmo II generalizado

El algoritmo de gradiente correspondiente a la función de coste (5.10) tiene la siguiente forma

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(\mathbf{x}(n)\hat{\mathbf{c}}^T(n) - \gamma\hat{\mathbf{c}}_{2G}^T(n)) \quad (5.22)$$

donde un factor 2 ha sido incluido dentro de μ , $\hat{\mathbf{c}}(n)$ viene dado por (5.16) y $\hat{\mathbf{c}}_{2G}(n)$ es una estimación del gradiente del término cruzado

$$\begin{aligned}\nabla_{\mathbf{W}}T_{2G} &= -2\gamma \sum_{\tau \geq 0} \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N E[\mathbf{x}(n)y_j^*(n-\tau)]E[y_1^*(n)y_j(n-\tau)] \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N E[\mathbf{x}(n)y_j^*(n-\tau)]E[y_N^*(n)y_j(n-\tau)] \end{bmatrix} \\ &\quad -2\gamma \sum_{\tau \geq 0} \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N E[\mathbf{x}(n-\tau)y_j^*(n)]E[y_1^*(n-\tau)y_j(n)] \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N E[\mathbf{x}(n-\tau)y_j^*(n)]E[y_N^*(n-\tau)y_j(n)] \end{bmatrix}\end{aligned}\quad (5.23)$$

Estimando los promedios que aparecen en esta expresión obtenemos que el vector $\hat{\mathbf{c}}_{2G}(n)$ en (5.22) tiene la siguiente forma

$$\begin{aligned}\hat{\mathbf{c}}_{2G}(n) &= \mathbf{x}(n) \sum_{\tau \geq 0} \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N y_j^*(n-\tau)\hat{E}[y_1^*(n)y_j(n-\tau)] \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N y_j^*(n-\tau)\hat{E}[y_N^*(n)y_j(n-\tau)] \end{bmatrix} \\ &\quad + \sum_{\tau \geq 0} \mathbf{x}(n-\tau) \begin{bmatrix} \sum_{\substack{j=1 \\ j \neq 1}}^N y_j^*(n)\hat{E}[y_1^*(n-\tau)y_j(n)] \\ \vdots \\ \sum_{\substack{j=1 \\ j \neq N}}^N y_j^*(n)\hat{E}[y_N^*(n-\tau)y_j(n)] \end{bmatrix}\end{aligned}\quad (5.24)$$

donde

$$\begin{aligned}\hat{E}[y_i(n)y_j^*(n-\tau)] &= (1-\mu_l)\hat{E}[y_i(n)y_j^*(n-\tau)] + \mu_l y_i(n)y_j^*(n-\tau) \\ \hat{E}[y_i(n-\tau)y_j^*(n)] &= (1-\mu_l)\hat{E}[y_i(n-\tau)y_j^*(n)] + \mu_l y_i(n-\tau)y_j^*(n)\end{aligned}\quad (5.25)$$

El número de operaciones necesarias para evaluar los promedios en (5.24) es notablemente menor que las involucradas en el algoritmo I generalizado, lo que se traduce en un menor coste computacional.

5.4 Análisis de estabilidad del algoritmo I generalizado

Al igual que sucedía con los algoritmos para separación de mezclas instantánea propuestos en el capítulo anterior, los problemas de optimización que acabamos de presentar consisten en la maximización de funciones de coste que dependen de los coeficientes del sistema MIMO de separación. Estas funciones son no cuadráticas y, por ello, los coeficientes óptimos no se corresponden con la solución de un sistema de ecuaciones lineales. Además, la superficie de error asociada a cada algoritmo de gradiente puede contener máximos indeseados que dañen su convergencia. Por tal motivo, se hace necesario realizar un estudio de la estabilidad de los algoritmos propuestos. En esta sección analizaremos la estabilidad del algoritmo I generalizado. Parte de este análisis ha sido presentado en [24].

Partiremos de la relación lineal que existe entre las salidas del sistema de separación y las fuentes

$$y_i(n) = \sum_{l=1}^N \sum_{k=-\infty}^{\infty} g_{il}(k) s_l(n-k), \quad i = 1, \dots, N \quad (5.26)$$

el coeficiente $g_{ij}(n)$ representa la amplitud con la que la salidas $y_i(n)$ extrae el símbolo $s_l(n-k)$

$$g_{ij}(n) = \sum_{k=-\infty}^{\infty} w_{ij}^*(k) a_{ji}(n-k) \quad (5.27)$$

Teniendo en cuenta la expresión (5.26) y las hipótesis realizadas sobre las fuentes podemos expresar los estadísticos que aparecen en (5.8) de la siguiente forma

$$\begin{aligned} E[|y_i(n)|^2] &= E \left[\left| \sum_{l=1}^N \sum_{k=-\infty}^{\infty} g_{il}(k) s_l(n-k) \right|^2 \right] = \sum_{l=1}^N \sum_{k=-\infty}^{\infty} |g_{il}(k)|^2 = \|\mathbf{g}_i\|^2 \\ K_{y_i}(0) &= cum(y_i(n), y_i^*(n), y_i(n), y_i^*(n)) = \sum_{l=1}^N \sum_{k=-\infty}^{\infty} |g_{il}(k)|^4 \rho_{s_l} \\ K_{y_i y_j}(0, \tau) &= cum(y_i(n), y_i^*(n), y_j(n-\tau), y_j^*(n-\tau)) = \sum_{l=1}^N \sum_{k=-\infty}^{\infty} |g_{il}(k)|^2 |g_{jl}(k-\tau)|^2 \rho_{s_l} \end{aligned} \quad (5.28)$$

donde ρ_l es el cumulante de cuarto orden normalizado de la fuente s_l . Sustituyendo (5.28) en (5.8) podemos expresar la función de coste en términos de las amplitudes introducidas

por el sistema separador

$$\begin{aligned} \psi_G(\mathbf{G}) \triangleq J_{1G}(\mathbf{W}) &= \sum_{i=1}^N \left(\sum_{l=1}^N \sum_{k=-\infty}^{\infty} |\rho_l| |g_{il}(k)|^4 - f(\|\mathbf{g}_i\|^2) \right) \\ &- \gamma \sum_{\substack{i,j=1 \\ i \neq j}}^N \sum_{l=1}^N \sum_{k=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} |\rho_l| |g_{il}(k)|^2 |g_{jl}(k - \tau)|^2 \end{aligned} \quad (5.29)$$

El gradiente de la función $\psi_G(\mathbf{G})$ respecto a las ganancias viene dado por la siguiente expresión

$$\begin{aligned} \frac{\partial \psi_G}{\partial g_{ik}(n)} &= 2|\rho_k| g_{ik}^*(n) |g_{ik}(n)|^2 - g_{ik}^*(n) f'(\|\mathbf{g}_i\|^2) - 2\gamma |\rho_k| g_{ik}^*(n) \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{\tau=-\infty}^{\infty} |g_{jk}(n - \tau)|^2 \\ & \quad i, k = 1, \dots, N, \quad n \in (-\infty, \infty) \end{aligned} \quad (5.30)$$

donde $f'(\cdot)$ es la primera derivada de $f(\cdot)$.

El sistema anterior tiene muchas soluciones de las cuales son deseadas únicamente aquellas donde cada salida $y_i(n)$ extrae una única fuente diferente. Para determinar cuales de las soluciones se corresponden con máximos de la función de coste estudiaremos el carácter definido negativo de la siguiente matriz

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1111}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} & \cdots & \mathbf{H}_{111N}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} & \cdots & \mathbf{H}_{11NN}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{1N11}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} & \cdots & \mathbf{H}_{1N1N}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} & \cdots & \mathbf{H}_{1NNN}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{NN11}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} & \cdots & \mathbf{H}_{NN1N}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} & \cdots & \mathbf{H}_{NNNN}(l_1, l_2)|_{l_1, l_2=-\infty}^{\infty} \end{bmatrix} \quad (5.31)$$

donde $\mathbf{H}_{iknp}(l_1, l_2)$ es la siguiente matriz compleja de dimensión 2×2

$$\mathbf{H}_{iknp}(l_1, l_2) = \begin{bmatrix} \frac{\partial^2 \psi_G}{\partial g_{ik}(l_1) \partial g_{np}^*(l_1)} & \frac{\partial^2 \psi_G}{\partial g_{ik}^*(l_1) \partial g_{np}^*(l_2)} \\ \frac{\partial^2 \psi_G}{\partial g_{ik}(l_1) \partial g_{np}(l_2)} & \frac{\partial^2 \psi_G}{\partial g_{ik}^*(l_2) \partial g_{np}(l_2)} \end{bmatrix} \quad (5.32)$$

Las segundas derivadas que aparecen en el hessiano vienen dadas por las siguientes expresiones

$$\begin{aligned} \frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{ik}^*(n)} &= 4|\rho_k| |g_{ik}(n)|^2 - f'(\|\mathbf{g}_i\|^2) - |g_{ik}(n)|^2 f''(\|\mathbf{g}_i\|^2) \\ &- 2\gamma |\rho_k| \sum_{\substack{j=1 \\ i \neq j}}^N \sum_{\tau=-\infty}^{\infty} |g_{jk}(n - \tau)|^2 \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{ik}^*(m)} &= -g_{ik}^*(n) g_{ik}(m) f''(\|\mathbf{g}_i\|^2) \quad n \neq m \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{ip}^*(m)} &= -g_{ik}^*(n) g_{ip}(m) f''(\|\mathbf{g}_i\|^2) \quad k \neq p \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{jk}^*(m)} &= -2\gamma |\rho_k| g_{ik}^*(n) g_{jk}(m) \quad i \neq j \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{jp}^*(m)} &= 0 \quad i \neq j, k \neq p \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{ik}(n)} &= 2|\rho_k| g_{ik}^*(n)^2 - g_{ik}^*(n)^2 f''(\|\mathbf{g}_i\|^2) \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{ik}(m)} &= -g_{ik}^*(n) g_{ik}^*(m) f''(\|\mathbf{g}_i\|^2) \quad n \neq m \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{ip}(m)} &= -g_{ik}^*(n) g_{ip}^*(m) f''(\|\mathbf{g}_i\|^2) \quad k \neq p \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{jk}(m)} &= -2\gamma |\rho_k| g_{ik}^*(n) g_{jk}^*(m) \quad i \neq j \\
\frac{\partial^2 \psi_G}{\partial g_{ik}(n) \partial g_{jp}(m)} &= 0 \quad i \neq j, k \neq p
\end{aligned} \tag{5.33}$$

donde $i, j, k, p = 1, \dots, N$ y $n, p \in (-\infty, \infty)$.

Consideremos ahora la solución deseada donde $y_i(n)$ extrae la fuentes $s_i(n)$ sin ningún retardo, i.e.,

$$g_{ii}(0) \neq 0, \quad g_{ii}(n) = 0 \quad n \in (-\infty, \infty), n \neq 0$$

y además cancela las otras fuentes

$$g_{ik}(n) = 0, \quad i, k = 1, \dots, N, \quad i \neq k, \quad n \in (-\infty, \infty)$$

Empezaremos notando que en este punto el sistema de ecuaciones (5.30) se reduce a la siguiente expresión

$$\begin{aligned}
\frac{\partial \psi_G}{\partial g_{ii}(0)} &= 2|\rho_i| g_{ii}^*(0) |g_{ii}(0)|^2 - g_{ii}^*(0) f'(|g_{ii}(0)|^2) = g_{ii}^*(0) p_i'(|g_{ii}(0)|^2) = 0 \quad i = 1, \dots, N \\
\frac{\partial \psi_G}{\partial g_{ii}(n)} &= 0 \quad i, k = 1, \dots, N, \quad i \neq k, \quad n \in (-\infty, \infty), n \neq 0 \\
\frac{\partial \psi_G}{\partial g_{ik}(n)} &= 0 \quad i, k = 1, \dots, N, \quad i \neq k, n \in (-\infty, \infty)
\end{aligned} \tag{5.34}$$

donde $p_i'(\cdot)$, $i = 1, \dots, N$, es la primera derivada de $p_i(\cdot)$. Recordemos que estamos suponiendo que cada función $p_i(\cdot)$, $i = 1, \dots, N$ tiene un único máximo en $x > 0$. Por tanto (5.34) se anulará solamente cuando cada valor $|g_{ii}(0)|^2$, $i = 1, \dots, N$, se corresponda con el máximo de $p_i(\cdot)$, $i = 1, \dots, N$.

Estudiemos ahora la expresión de la matriz hessiana en este punto. Las únicas segundas derivadas (5.33) que no se anulan son las siguientes

$$\begin{aligned}
 \frac{\partial^2 \psi_G}{\partial g_{ii}(0) \partial g_{ii}^*(0)} &= 4|\rho_i| |g_{ii}(0)|^2 - f'(|g_{ii}(0)|^2) - |g_{ii}(0)|^2 f''(|g_{ii}(0)|^2) \\
 &= |g_{ii}(0)|^2 p_i''(|g_{ii}(0)|^2) \\
 \frac{\partial^2 \psi_G}{\partial g_{ii}(m) \partial g_{ii}^*(m)} &= -f'(|g_{ii}(0)|^2) = -2\gamma |\rho_i| |g_{ii}(0)|^2 \quad m \neq 0 \\
 \frac{\partial^2 \psi_G}{\partial g_{ik}(m) \partial g_{ik}^*(m)} &= -f'(|g_{ii}(0)|^2) - 2\gamma |\rho_k| |g_{kk}(0)|^2 \\
 &= -2|\rho_i| |g_{ii}(0)|^2 - 2\gamma |\rho_k| |g_{kk}(0)|^2 \quad i \neq k \\
 \frac{\partial^2 \psi_G}{\partial g_{ii}(0) \partial g_{ii}(0)} &= 2|\rho_i| g_{ii}^*(0)^2 - g_{ii}^*(0)^2 f''(|g_{ii}(0)|^2) = g_{ii}^*(0)^2 p_i''(|g_{ii}(0)|^2)
 \end{aligned} \tag{5.35}$$

Como consecuencia, todas las submatrices (5.32) son cero excepto

$$\begin{aligned}
 \mathbf{H}_{iii}(0, 0) &= \begin{bmatrix} |g_{ii}(0)|^2 p_i''(|g_{ii}(0)|^2) & g_{ii}(0)^2 p_i''(|g_{ii}(0)|^2) \\ g_{ii}^*(0)^2 p_i''(|g_{ii}(0)|^2) & |g_{ii}(0)|^2 p_i''(|g_{ii}(0)|^2) \end{bmatrix} \quad i = 1, \dots, N \\
 \mathbf{H}_{iii}(m, m) &= \begin{bmatrix} -2|\rho_i| |g_{ii}(0)|^2 & 0 \\ 0 & -2|\rho_i| |g_{ii}(0)|^2 \end{bmatrix} \\
 &\quad i = 1, \dots, N, \quad l \in (-\infty, \infty) \neq 0 \\
 \mathbf{H}_{ikik}(0, 0) &= \begin{bmatrix} -2|\rho_i| |g_{ii}(0)|^2 - & & & \\ 2\gamma |\rho_k| |g_{kk}(0)|^2 & & 0 & \\ 0 & & -2|\rho_i| |g_{ii}(0)|^2 - & \\ 2\gamma |\rho_k| |g_{kk}(0)|^2 & & & \end{bmatrix} \quad i, k = 1, \dots, N, \quad i \neq k
 \end{aligned}$$

Para determinar si el punto deseado es un máximo de la función de coste, estudiaremos el carácter definido negativo de cada una de las submatrices anteriores. Es inmediato comprobar que las submatrices $\mathbf{H}_{iii}(0, 0)$ $i = 1, \dots, N$ son definidas negativas ya que el gradiente (5.34) se anula únicamente cuando cada punto $|g_{ii}(0)|^2$ se corresponde con un máximo de $p_i(\cdot)$, $i = 1, \dots, N$. Las submatrices $\mathbf{H}_{iii}(m, m)$ también son definidas negativas. Por otro lado, la condición $\gamma > 0$ asegura que todas las matrices $\mathbf{H}_{ikik}(0, 0)$ son definidas negativas. Como consecuencia, el punto que estamos analizando es un máximo de la función de coste.

El estudio que acabamos de presentar indica que, al igual que sucedía con mezclas instantáneas, la condición $\gamma > 0$ asegura que los puntos donde se consigue la separación de las fuentes son máximos de la función de coste. Este resultado puede ser fácilmente extendido a todas las soluciones de (5.30) correspondientes a la separación de las fuentes realizando transformaciones congruentes sobre la matriz hessiana. En el apéndice F se analizan las otras soluciones del sistema de ecuaciones (5.30) y se demuestra que la

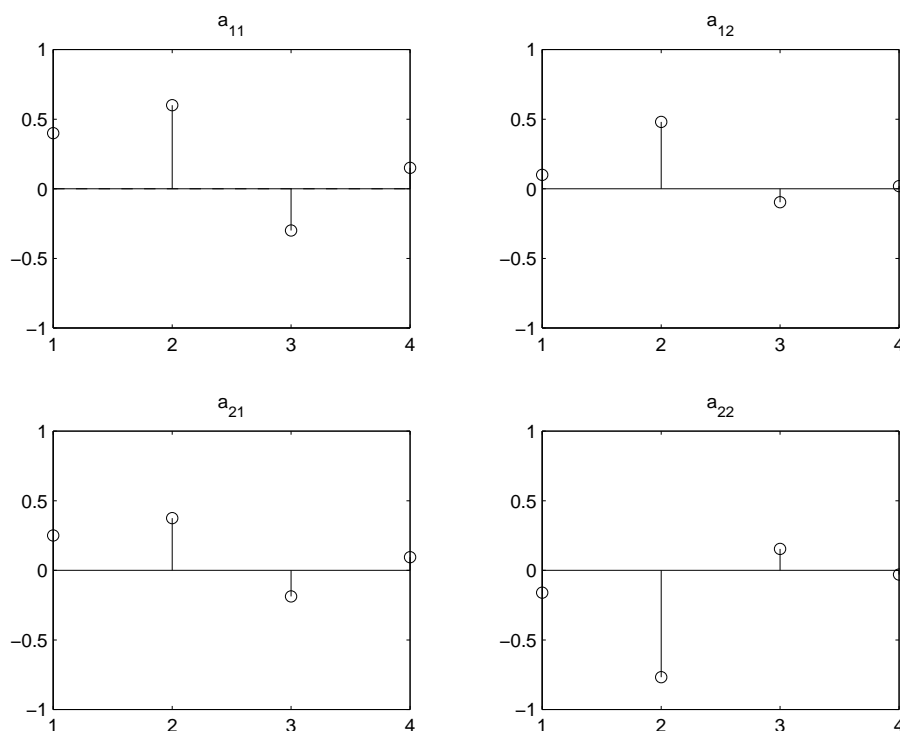


Figura 5.2: Respuesta al impulso del sistema de mezcla.

Es bien conocido que la utilización de sistemas finitos pueden dañar la perfecta recuperación de las fuentes [34]. Sin embargo, los resultados de distintas simulaciones realizadas por ordenador muestran que los algoritmos propuestos anteriormente trabajan adecuadamente en estas situaciones. Más aún, los puntos alcanzados por los algoritmos son muy próximos a los que predice el análisis de la sección 5.4.

Como medida de rendimiento se ha utilizado el índice $P(\mathbf{G})$ definido en [46]

$$P(\mathbf{G}) = \sum_{i=1}^N \left| \sum_{j=1}^N \sum_{n=-\infty}^{\infty} \frac{|g_{ij}(n)|^2}{\max_{l,n} (|g_{il}(n)|^2)} - 1 \right| + \sum_{j=1}^N \left| \sum_{i=1}^N \sum_{n=-\infty}^{\infty} \frac{|g_{ij}(n)|^2}{\max_{l,n} (|g_{lj}(n)|^2)} - 1 \right|$$

donde $\max_{l,n} (|g_{il}(n)|^2) = \max_{l=1,\dots,N} \max_{-\infty < n < \infty} |g_{il}(n)|^2$ y $\max_{l,n} (|g_{lj}(n)|^2) = \max_{l=1,\dots,N} \max_{-\infty < n < \infty} |g_{lj}(n)|^2$. Para presentar la evolución del índice $P(\mathbf{G})$ hemos promediado 10 realizaciones independientes del experimento.

En todas las simulaciones se ha utilizado un sistema mezclador descrito por

$$\mathbf{A}(z) = \begin{bmatrix} 0.8 & 0.5 \\ 0.5 & -0.8 \end{bmatrix} \begin{bmatrix} \frac{0.5+z^{-1}}{1+0.5z^{-1}} & 0 \\ 0 & \frac{0.2+z^{-1}}{1+0.2z^{-1}} \end{bmatrix}$$

que corresponde a la respuesta al impulso presentada en la figura 5.2. En la etapa de separación se ha utilizado un sistema MIMO de dos entradas, dos salidas y $L = 8$

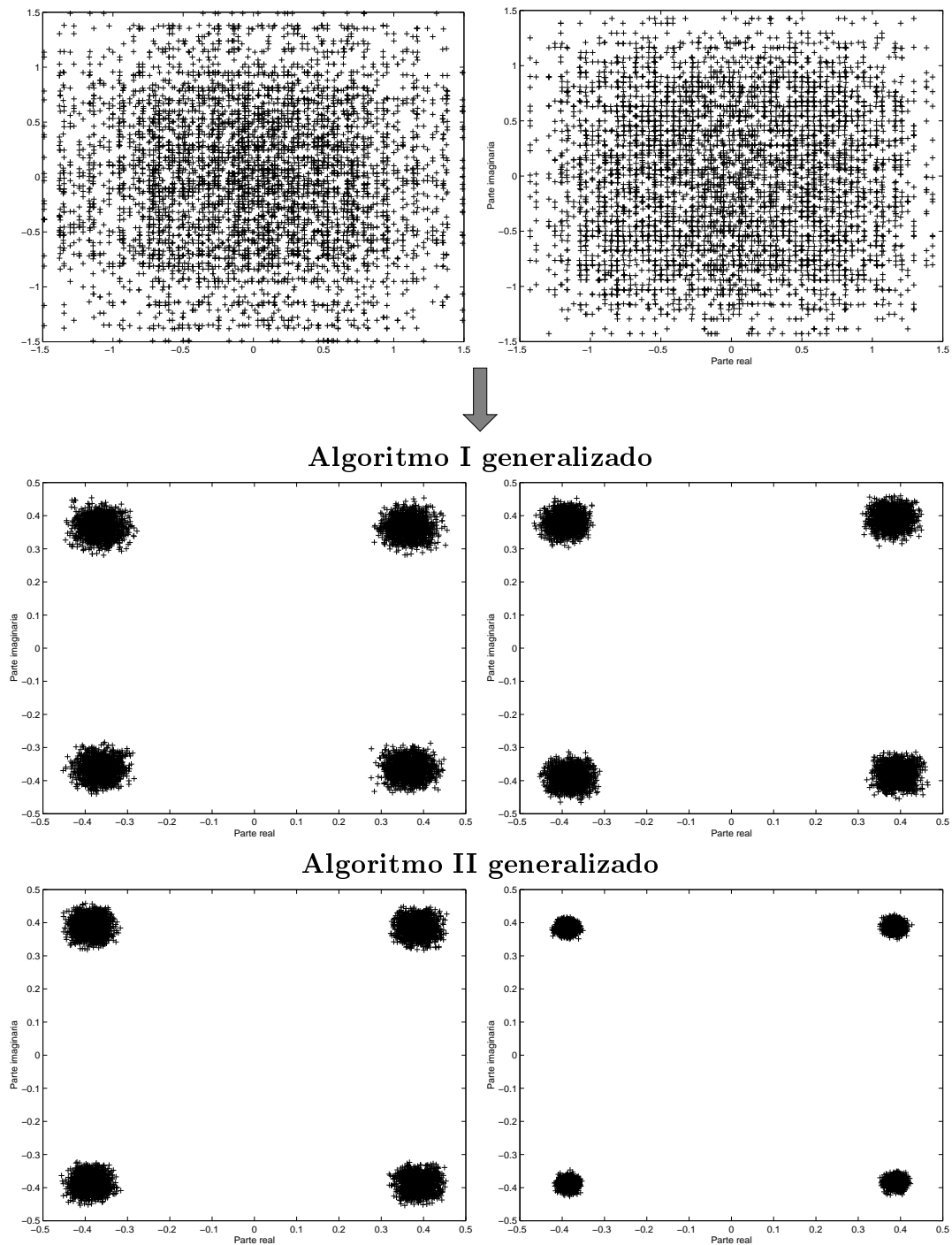


Figura 5.3: Observaciones y salidas del primer experimento.

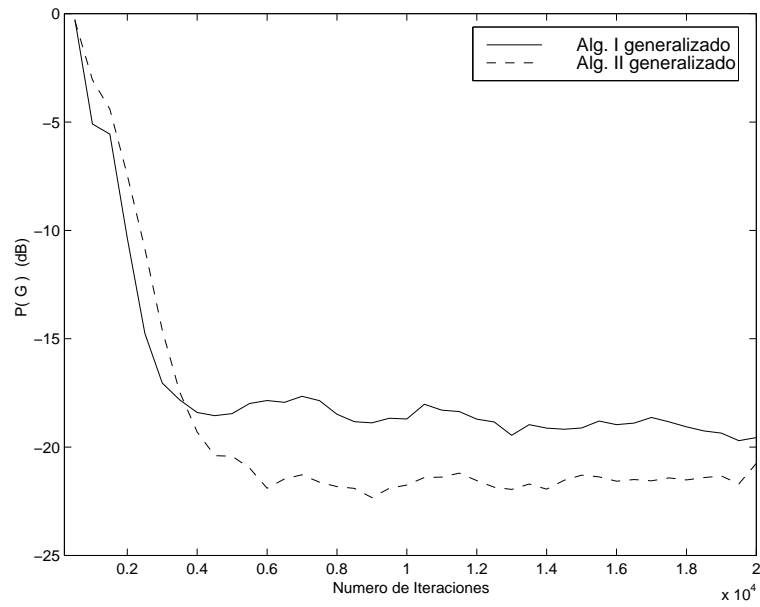


Figura 5.4: Evolución del índice $P(\mathbf{G})$ para el primer experimento.

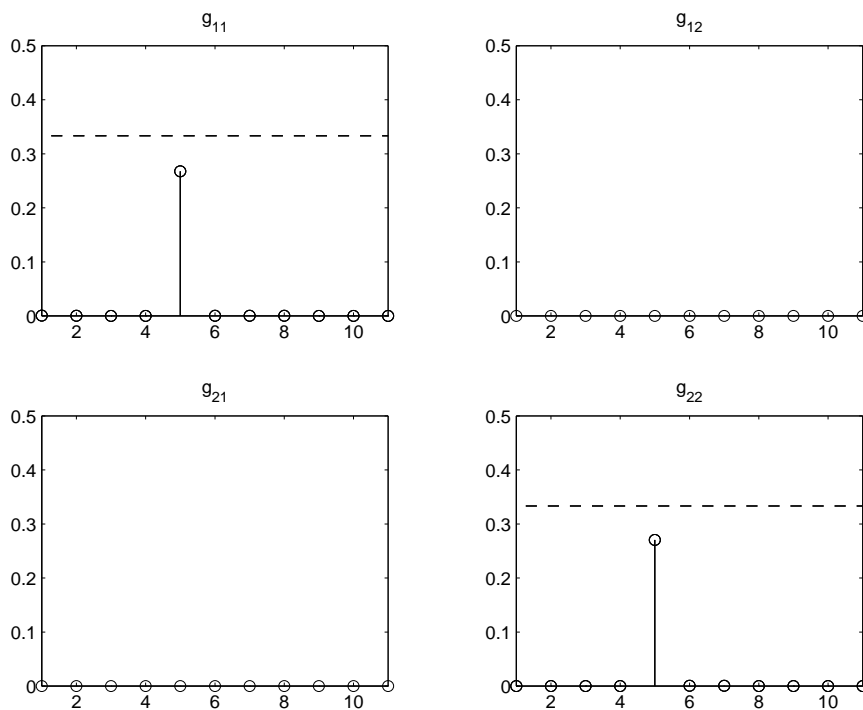


Figura 5.5: Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo I generalizado en el primer experimento.

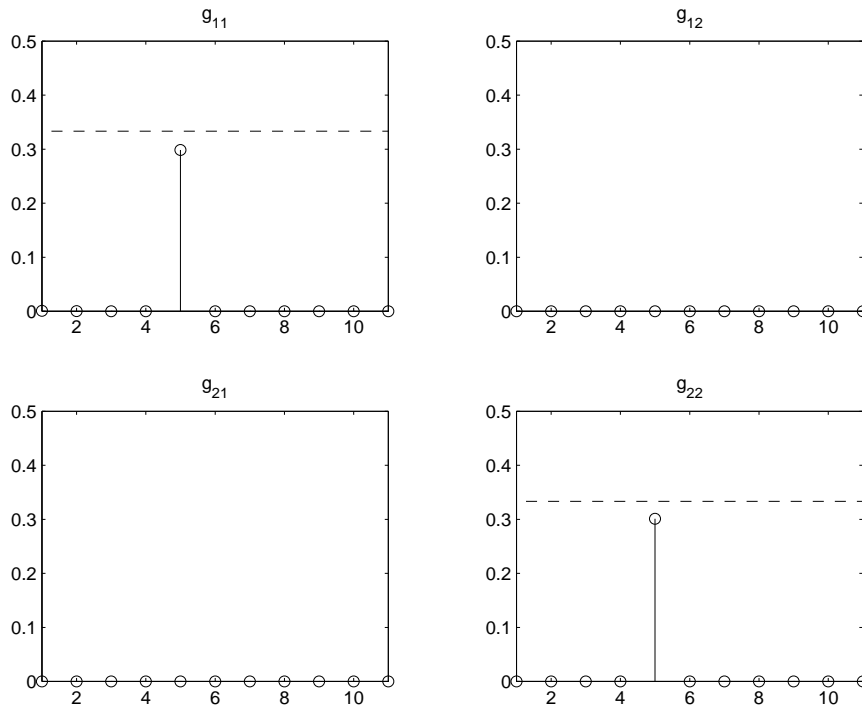


Figura 5.6: Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo II generalizado en el primer experimento.

Algoritmo	Parámetros	
	Paso inicial	Función de coste
<i>Alg.IG</i>	$\mu = 1 \times 10^{-3}, \mu_l = 5 \times 10^{-2}$	$\epsilon = -1, \alpha = 4, \beta = 1, \gamma = 1.5$
<i>Alg.IIG</i>	$\mu = 1 \times 10^{-4}, \mu_l = 5 \times 10^{-2}$	$\epsilon = -1, \alpha = 4, \beta = 1, \gamma = 1.5$

Tabla 5.1: Parámetros de los algoritmos empleados en las simulaciones.

etapas por observación. Los coeficientes han sido inicializados a cero excepto $w_{11}(4) = w_{22}(4) = 1$.

En el primer experimento se ha considerado un entorno formado por dos señales 4-QAM que al pasarlas por el canal dan lugar a las observaciones mostradas en la figura 5.3. En esta figura también se pueden observar las señales recuperadas empleando los algoritmos I y II generalizados para ajustar los coeficientes del sistema separador. Los parámetros utilizados están recogidos en la tabla 5.1. La figura 5.4 muestra la evolución del índice $P(\mathbf{G})$ donde puede observarse que ambos algoritmos han recuperado las fuentes.

El análisis de estabilidad presentado en la sección 5.4 predice que las fuentes serán extraídas con ganancias $|g_{i1}(l_1)|^2 = |g_{j2}(l_2)|^2 = 0.333$. Por otro lado, después de 20.000 iteraciones el algoritmo ha alcanzado los valores que se muestran en la figura 5.5. Como puede verse existe un único valor significativo que, además, es muy cercano al teórico. En la figura 5.6 puede observarse como el algoritmo II también ha separado las fuentes.

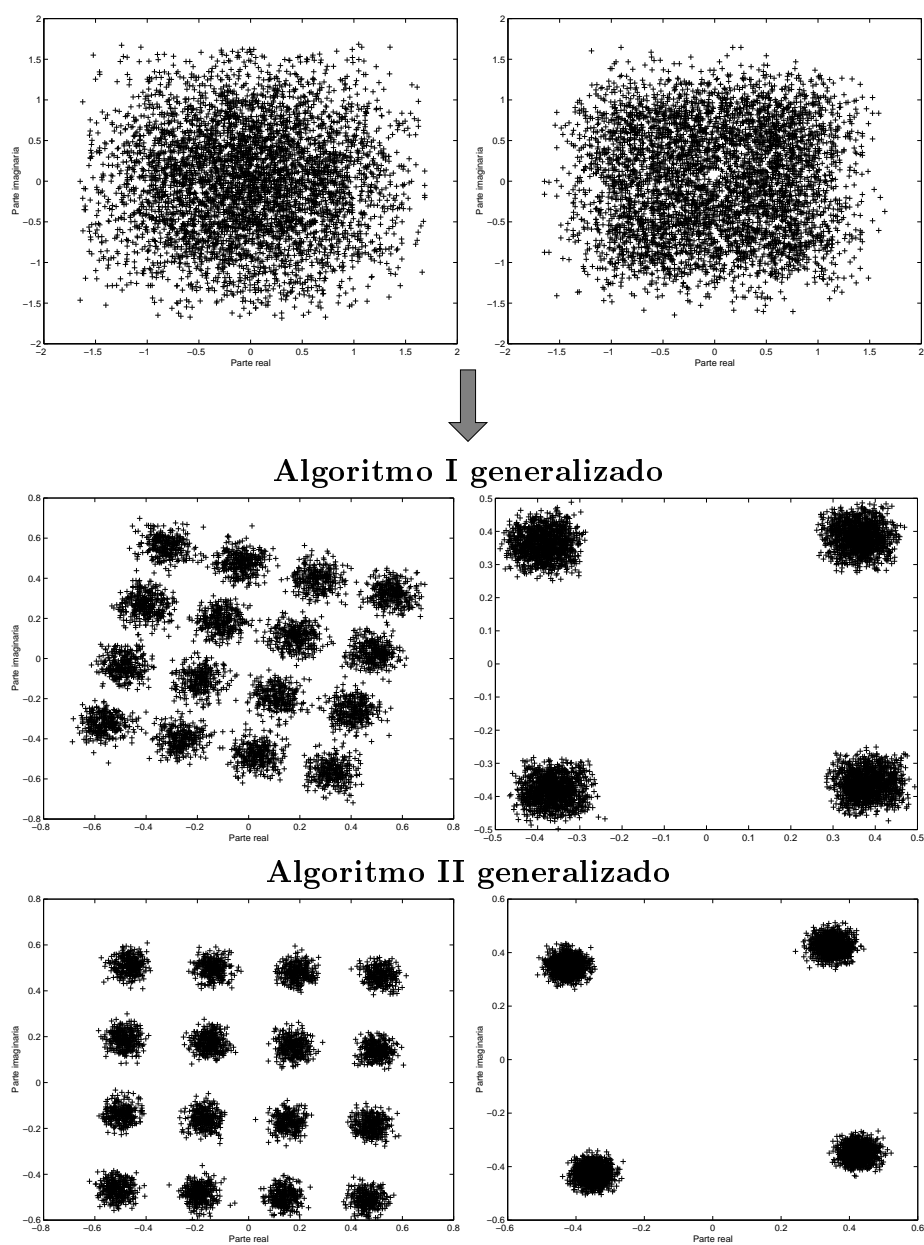


Figura 5.7: Observaciones y salidas del segundo experimento.

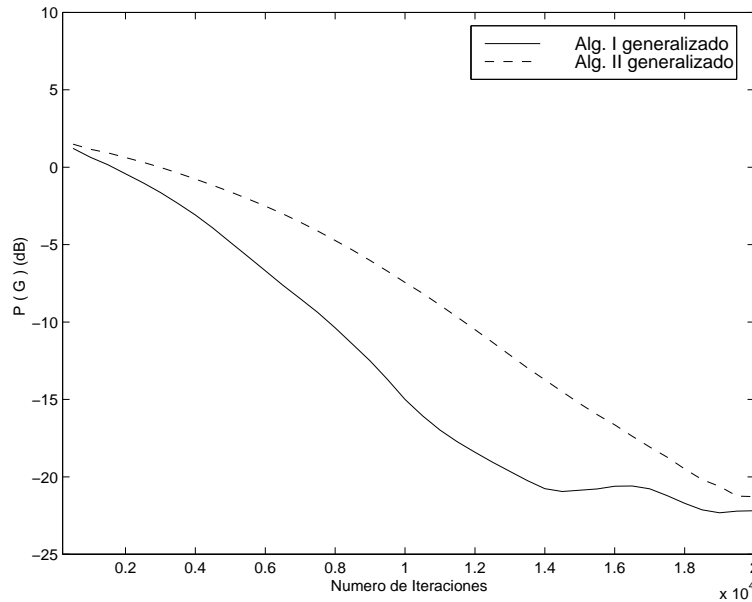


Figura 5.8: Evolución del índice $P(\mathbf{G})$ para el segundo experimento.

En un segundo experimento se han considerado dos fuentes: s_1 con modulación 16-QAM y una señal s_2 con modulación 4-QAM. Al hacer pasar estas señales por el canal (5.37) se han obtenido las observaciones mostradas en la figura 5.7. Esta figura también muestra las señales recuperadas al procesar estas observaciones con sistemas separadores adaptados con los algoritmos I y II generalizados (los parámetros empleados son los mostrados en la tabla 5.1). Para este entorno también se ha obtenido un buen comportamiento de los algoritmos como refleja en el índice $P(\mathbf{G})$ en la figura 5.8.

Comparemos ahora las ganancias $|g_{ij}(n)|^2$ obtenidas utilizando el algoritmo I después de 20.000 iteraciones con los teóricos. El análisis de estabilidad predice la siguientes amplitudes: $|g_{i1}(l_1)|^2 = 0.3012$ y $|g_{i2}(l_2)|^2 = 0.333$. Por otro lado, en la simulación se han obtenido los valores mostrados en la figura 5.9. Como puede verse, al igual que en el experimento anterior, son muy cercanos a los teóricos. Finalmente, la figura 5.10 muestra como el algoritmo II generalizado también ha recuperado las fuentes.

5.6 Conclusiones

En este capítulo se ha abordado el problema de separación de fuentes a partir de mezclas convolutivas. Este problema aparece, por ejemplo, cuando varias señales que se propagan por un medio abierto sufren distorsiones que se traducen en la recepción simultánea de varias versiones retardadas de las mismas.

La recuperación de las fuentes es un problema más complejo que en el caso de mezclas instantáneas debido a que ahora el sistema además de separar las fuentes debe eliminar el

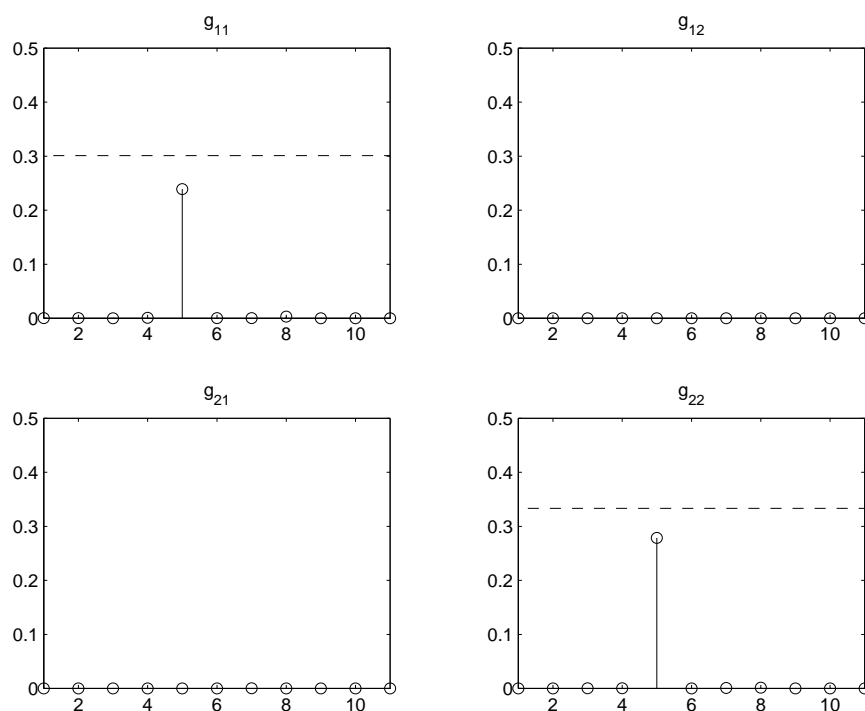


Figura 5.9: Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo I generalizado en el segundo experimento.

efecto de otras versiones retardadas de las fuentes. Sin embargo, en general, los criterios de separación de mezclas instantáneas pueden extenderse al caso convolutivo. En concreto, en este capítulo hemos generalizado los criterios I y II presentados en el capítulo anterior.

La generalización ha surgido a partir de la dependencia estadística que aparece entre las salidas cuando varias salidas extraen la misma fuente. Al igual que en el caso de mezclas instantáneas, los criterios de separación se definen como el criterio de Shalvi y Weinstein para igualación más un término cruzado. La principal diferencia con el caso de mezclas instantáneas radica en que varias salidas pueden extraer la misma fuente con diferentes retardos y, por ello, el término cruzado debe ser extendido para penalizar estas situaciones indeseadas.

Los dos problemas de optimización presentados en este capítulo pueden ser implementados utilizando algoritmos de gradiente estocástico cuya complejidad es notablemente superior al caso de mezcla instantánea. La estabilidad de uno de estos algoritmos ha sido estudiada en este mismo capítulo obteniéndose ciertas condiciones que aseguran la correspondencia entre la perfecta separación de las fuentes y los atractores del algoritmo. Existen otros puntos estacionarios pero también se ha demostrado que son inestables y, por tanto, el algoritmo de gradiente nunca convergerá a ellos.

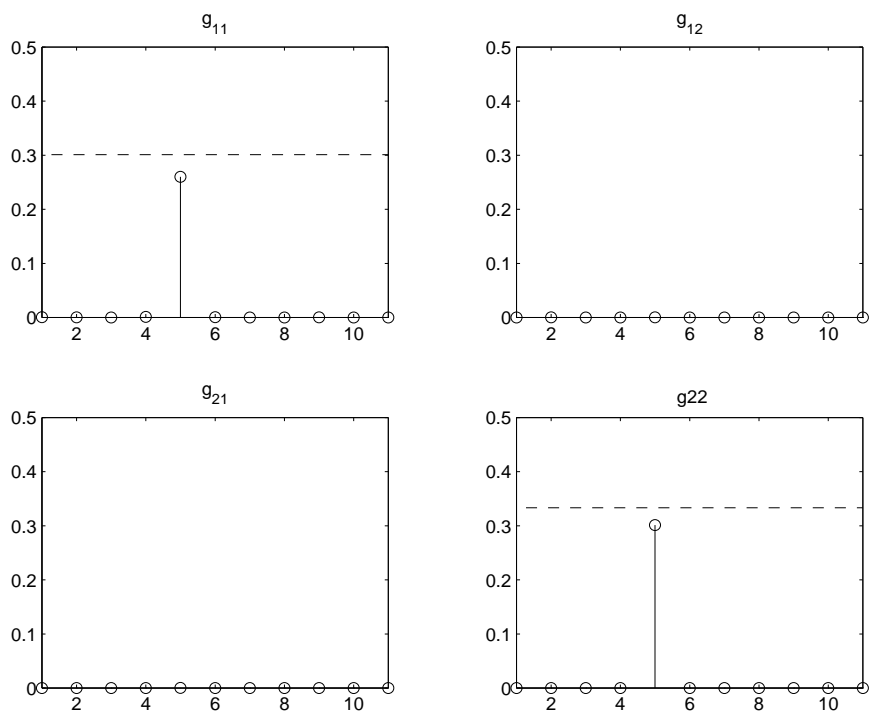


Figura 5.10: Respuesta conjunta del sistema de mezcla/separación obtenida con el algoritmo II generalizado en el segundo experimento.

Capítulo 6

Conclusiones y líneas futuras de trabajo

6.1 Conclusiones

En la presente tesis doctoral se aborda el problema de recuperar un conjunto de señales originales (fuentes) a partir únicamente de observaciones de mezclas lineales de ellas. Este problema se conoce con el nombre de separación ciega de fuentes y resulta de enorme interés en procesamiento de señal por dos razones. La primera es que en muchas aplicaciones prácticas debe llevarse a cabo la separación de mezclas lineales de señales y la segunda es que el problema puede resolverse utilizando una cantidad de información *a priori* mínima: la invertibilidad del sistema de mezcla y la independencia mutua y la distribución no-gaussiana de las fuentes.

Se ha contemplado que la mezcla pueda ser de dos tipos: instantánea o convolutiva. En el primer caso, la mezcla no tiene memoria y las señales observadas en un determinado instante son una suma ponderada de las fuentes en ese mismo instante. Por el contrario, la mezcla convolutiva tiene memoria y las observaciones son mezclas de las fuentes en instantes anteriores y posteriores al instante presente. Independientemente del tipo de mezcla, las fuentes pueden ser recuperadas ajustando los parámetros de un sistema MIMO con el objetivo de que cada salida extraiga una única fuente diferente. Una manera práctica y eficiente de calcular el sistema de separación óptimo es mediante un algoritmo adaptativo y esos algoritmos constituyen el objeto de estudio de esta tesis.

A lo largo de los últimos años han sido propuestos numerosos algoritmos adaptativos para separación ciega de fuentes pero pocos son los trabajos que garantizan la estabilidad de las recursiones propuestas. Parte de esta tesis se centra en el estudio de la estabilidad de algunos de esos algoritmos para el caso de mezclas instantáneas, lo cual nos ha permitido determinar condiciones suficientes que garantizan que los puntos donde se consigue la perfecta recuperación de las fuentes sean atractores de las recursiones.

En esta tesis también se ha desarrollado una nueva familia de algoritmos para

separación ciega de fuentes. Entre las aportaciones más importantes debemos destacar el haber analizado la estabilidad de uno de los algoritmos para el caso general de N fuentes complejas y sistemas separadores de N entradas y N salidas. El estudio garantiza que los únicos atractores del algoritmo son los puntos donde se consigue la perfecta recuperación de las fuentes. Finalmente, estos resultados han sido extendidos al caso de mezclas convolutivas.

A continuación se comentan más detalladamente las conclusiones y aportaciones más importantes de este trabajo.

6.1.1 Revisión y aportaciones a anteriores aproximaciones

Una de las contribuciones de esta tesis es el estudio de la estabilidad de varios algoritmos adaptativos para separación de mezclas instantáneas propuestos previamente por diversos autores. Estos algoritmos han surgido a partir de consideraciones de muy diversa índole (heurísticas, maximización de la transferencia de información, estimación de máxima verosimilitud, etc.) pero todos ellos tienen en común el emplear estadísticos de orden superior ya que, como se ha descrito en el capítulo 3, los algoritmos que buscan la decorrelación de las salidas resultan insuficientes para conseguir la separación. El empleo de estos estadísticos se traduce en la aparición de no-linealidades en las recursiones que dificulta notablemente el análisis de su estabilidad. Por ello, en el capítulo 3 hemos considerado únicamente el caso de sistemas separadores con dos entradas y dos salidas.

El primer algoritmo estudiado se debe a Héroult y Jutten quienes propusieron utilizar un sistema separador cuya salida es realimentada a la entrada. La existencia de realimentaciones complica la implementación software del algoritmo ya que se hace necesario realizar la inversión de la matriz de coeficientes. Para evitar esta operación, Cichocki y Undebauen realizaron una sencilla modificación del algoritmo que permite llevar a cabo la separación utilizando una estructura sin realimentaciones. En esta tesis se ha analizado la estabilidad de ambos algoritmos y se han obtenido ciertas condiciones suficientes que deben cumplir las no-linealidades para garantizar que los puntos donde se consigue la separación sean atractores.

Otro de los criterios que han centrado nuestro interés se debe a Comon quien introdujo el término de *contraste* para designar a ciertas funciones de coste que cumplen una serie de condiciones que garantizan la correspondencia entre sus máximos y los estados de separación. Esta aproximación permite conseguir la separación ajustando los coeficientes del sistema MIMO mediante algoritmos de gradiente ascendente. En esta tesis se ha presentado el algoritmo correspondiente a una función de contraste y se ha analizado su estabilidad para sistemas separadores con dos entradas y dos salidas. El estudio ha consistido en determinar todos los puntos estacionarios de la función de coste y, posteriormente, se ha estudiado el carácter definido negativo de la matriz hessiana. Como conclusión, se ha obtenido una condición poco restrictiva que garantiza que únicamente los puntos donde se lleva a cabo la separación son máximos y, por ello, el algoritmo

de gradiente siempre convergerá a una solución deseada. Extender estos resultados al caso general de sistemas separadores $N \times N$ es un trabajo arduo que todavía no ha sido realizado.

A partir de las funciones de contraste, Cardoso y Laheld han propuesto un algoritmo conocido como EASI cuyo comportamiento resulta ser independiente de la matriz de mezcla y, por ello, es especialmente atractivo para situaciones donde la mezcla está mal condicionada. Al igual que las aproximaciones anteriores, el análisis del algoritmo EASI revela que la elección de las no-linealidades juega un papel crucial en su estabilidad.

6.1.2 Nuevos criterios de separación

Otra de las contribuciones de esta tesis radica en presentar una nueva familia de algoritmos para separación de fuentes que se basa en la maximización de funciones de coste no-lineales que se componen de la suma de dos términos. El primero de ellos es una extensión de un criterio de igualación ciega propuesto por Shalvi y Weinstein a principios de esta década. Estos autores han demostrado que al adaptar los pesos del sistema separador utilizando dicho criterio se consigue que cada salida extraiga una única fuente. Sin embargo, su maximización resulta insuficiente para llevar a cabo la separación ya que puede ocurrir que varias salidas extraigan simultáneamente la misma fuente. Esta situación indeseada se traduce en la aparición de ciertas dependencias en el sistema separador (dependencia estadística entre las salidas del sistema y dependencia lineal entre las filas de la matriz de separación) las cuales pueden aprovecharse para conseguir la separación de las fuentes.

La dependencia estadística puede ser penalizada utilizando un segundo término que contiene algunos estadísticos cruzados entre las salidas. En el criterio que ha centrado la mayor parte del capítulo 4 se ha empleado como término cruzado la suma de cumulantes cruzados de cuarto orden entre las salidas. Para esta elección, se ha demostrado que los puntos donde se consigue la perfecta separación de las fuentes son máximos de la función de coste. Existen otros puntos estacionarios pero también se ha obtenido que son mínimos o puntos de ensilladura y, por tanto, un algoritmo de gradiente nunca convergerá a ellos. Este estudio ha sido realizado para el caso más general de señales complejas y sistemas separadores con N entradas y N salidas.

El segundo criterio propuesto emplea como término cruzado la correlación entre las salidas. La función de coste que resulta es más sencilla que la anterior ya que involucra un menor número de momentos estadísticos pero, en contrapartida, solamente se ha demostrado la inexistencia de máximos indeseados para el caso de sistemas separadores con dos entradas y dos salidas.

En el último criterio presentado se ha aprovechado la dependencia lineal entre las filas del sistema separador que aparece cuando varias salidas extraen la misma fuente. Se ha propuesto utilizar como término cruzado una función logarítmica de la matriz de separación. La implementación de esta aproximación empleando un algoritmo de máxima pendiente requiere realizar la inversión de la matriz de coeficientes. Tal y como se

demuestra en el capítulo 4 esta limitación se supera empleando un algoritmo de gradiente relativo. Adicionalmente, esta opción resulta ser invariante a cambios en la matriz de mezcla y puede ser utilizada en entornos donde la mezcla está mal condicionada.

Con el objeto de mostrar el comportamiento de los algoritmos propuestos se han llevado a cabo simulaciones por ordenador. Se ha contemplado la separación de señales de voz y música (caracterizadas por una curtosis positiva) y de imágenes (fuentes con curtosis negativa). Los resultados obtenidos muestran que todos los criterios propuestos llevan a cabo la separación y, además, corroboran las conclusiones obtenidas en el análisis de su estabilidad. También se han realizado algunas simulaciones considerando que las observaciones están perturbadas por ruido blanco gaussiano. Al igual que antes, se ha conseguido la separación de las fuentes.

Finalmente, se ha discutido la forma de extender al caso de mezcla convolutiva los criterios que emplean la dependencia estadística entre las salidas. La diferencia principal con el caso de mezcla instantánea radica en que ahora varias salidas pueden extraer la misma fuente en distintos instantes de tiempo y, por ello, se hace necesario extender los términos cruzados. Se han derivado dos algoritmos de gradiente para conseguir los coeficientes óptimos y se ha analizado la estabilidad de uno de ellos. El análisis demuestra que, para sistemas separadores no-causales e infinitos, el algoritmo es estable en los puntos de separación y no contiene atractores indeseados. Las simulaciones realizadas por ordenador muestran que los resultados teóricos son muy cercanos a los que se obtienen cuando se emplean sistemas causales de longitud finita.

6.2 Líneas futuras de investigación

Posibles direcciones hacia las que encaminar posteriores investigaciones son las siguientes

Separación de mezclas convolutivas en el dominio de la frecuencia

Otra manera de abordar la separación de mezclas convolutivas es trabajar en el dominio de la frecuencia utilizando un esquema separador como el mostrado en la figura 6.1. En una primera etapa se calcula la DFT (Discrete Fourier Transform) de las observaciones. De esta forma se consigue transformar el problema de mezcla convolutiva en L problemas de mezclas instantáneas. Posteriormente, se pueden utilizar algoritmos como los presentados en el capítulo 3 y 4 de esta tesis para ajustar los coeficientes del sistema separador para cada una de las L frecuencias. Finalmente, se realiza la transformada de Fourier inversa para recuperar las señales en el dominio temporal.

La aplicación directa de los algoritmos para separación de mezclas instantáneas, sin embargo, resulta insuficiente para realizar la separación de mezclas convolutivas ya que la reconstrucción de la señal requiere resolver las indeterminaciones inherentes a ellos: cambio del orden y de la amplitud de las señales recuperadas.

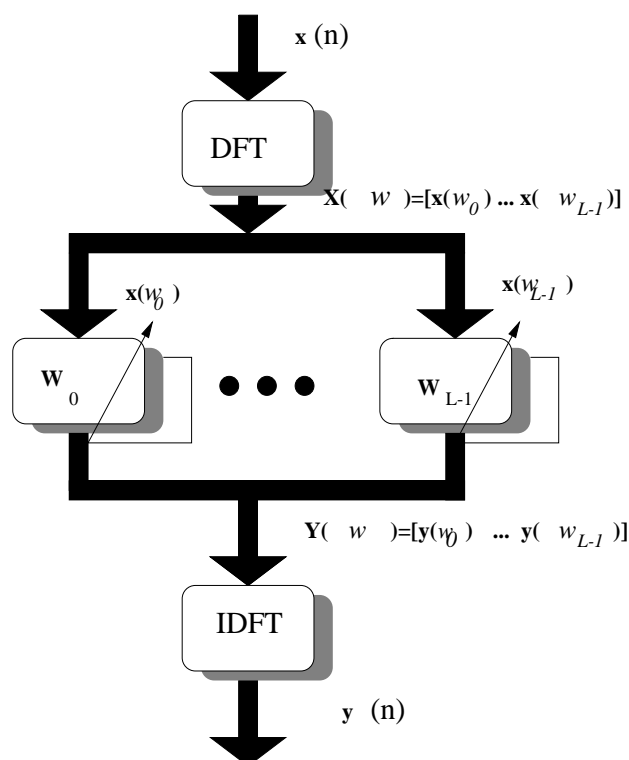


Figura 6.1: Sistema separador de mezclas convolutivas en el dominio de la frecuencia.

Alternativa a PCA

El Análisis de Componentes Principales (PCA, del inglés Principal Component Analysis) es una técnica estadística que permite representar la información recogida en un grupo inicial de M variables mediante un número menor N ($N < M$) de nuevas variables denominadas componentes principales. Esta técnica es de gran utilidad en aplicaciones de extracción de características principales, comprensión de datos, etc.. Sin embargo, los algoritmos basados en PCA emplean estadísticos de segundo orden y, por ello, resultan insuficientes para realizar la separación de fuentes.

Como alternativa a PCA, Comon ha introducido el concepto de ICA (Independent Component Analysis) para designar a criterios que permiten descomponer una señal en componentes estadísticamente independientes [18]. Como se ha visto en esta tesis, las hipótesis de independencia estadística de las fuentes y la distribución no gaussiana de las mismas son suficientes para resolver el problema de separación empleando esta técnica. Sin embargo, el campo de aplicación de los algoritmos ICA no se reduce a la separación de fuentes. En particular, nos planteamos aplicarlos como alternativa a PCA para comprensión de datos ya que es de esperar que si los datos de entrada cumplen las hipótesis que hemos establecido para el problema de separación de fuentes entonces estos algoritmos extraerán componentes independientes que, teóricamente, contienen toda la información

de los datos. Sin embargo, en la práctica nos encontramos con datos de entrada que no cumplen nuestras hipótesis de trabajo como, por ejemplo, las observaciones pueden corresponderse a mezclas no-lineales de las fuentes.

Estudio del efecto del ruido

En la mayor parte de esta tesis se han considerado modelos libres de ruido, lo cual no sucede en la práctica. Los resultados de algunas simulaciones recogidas en el capítulo 4 parecen indicar que los algoritmos basados en el criterio de Shalvi y Weinstein presentan un comportamiento óptimo en el sentido de que maximizan la relación señal a ruido o, equivalentemente, minimizan el error cuadrático medio entre las fuentes y las salidas. Este importante resultado está todavía pendiente de ser demostrado analíticamente. Existen trabajos al respecto [36] realizados eligiendo la no-linealidad que aparece en el criterio II de forma que el primer término de la función de coste se corresponda con el Constant Modulus.

Desarrollo de algoritmos basados en cicloestacionariedad

Se ha pensado en desarrollar criterios de separación especialmente dedicados a aplicaciones de comunicación que incorporen el carácter cicloestacionario que tienen la mayor parte de las señales utilizadas [37]. Como una primera aproximación en [25] se ha extendido al problema de separación de fuentes un criterio propuesto inicialmente para conformación de haz adaptativa en [15, 45]. Este criterio ha mostrado ser adecuado para abordar el problema de mezclas instantáneas pero todavía no ha sido estudiado su comportamiento para el caso de mezclas convolutivas.

Apéndice A

Definiciones de derivadas de funciones

En los estudios de la estabilidad de los algoritmos presentados en esta tesis se han calculado las derivadas de funciones escalares respecto a vectores o matrices. Este apéndice recoge las definiciones de estas derivadas así como algunos resultados que han sido utilizados en capítulos anteriores.

A.1 Definiciones de derivadas reales

La derivada de una función escalar J de variable vectorial o matricial real se define como sigue

- *Definición 1:* Sea $J = J(\mathbf{x})$ una función real de variable vectorial de dimensión $N \times 1$ real, la derivada de J respecto a \mathbf{x} es el vector columna

$$\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial J(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial J(\mathbf{x})}{\partial x_N} \end{bmatrix} \quad (\text{A.1})$$

- *Definición 2:* Sea $J = J(\mathbf{X})$ una función escalar de variable matricial \mathbf{X} de dimensión $P \times N$ real. La derivada de J respecto a la matriz \mathbf{X} es la matriz de dimensión $P \times N$ dada por

$$\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial J(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial J(\mathbf{X})}{\partial x_{1N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial J(\mathbf{X})}{\partial x_{P1}} & \cdots & \frac{\partial J(\mathbf{X})}{\partial x_{PN}} \end{bmatrix} \quad (\text{A.2})$$

Ejemplos:

Sea \mathbf{A} una matriz cuadrada de dimensión $M \times M$ y \mathbf{X} una matriz de dimensión $M \times P$. Entonces

- Si $J(\mathbf{X}) = \mathbf{X}^T \mathbf{A}$, se tiene que

$$\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^T \quad (\text{A.3})$$

- Si $J(\mathbf{A}) = \det(\mathbf{A})$, se tiene que

$$\frac{\partial J(\mathbf{A})}{\partial \mathbf{A}} = \text{adj}(\mathbf{A}^T) \quad (\text{A.4})$$

donde $\text{adj}(\cdot)$ representa la matriz adjunta.

Prueba:

Dada una matriz cuadrada de dimensión $M \times M$, si para cada elemento a_{ij} , $i, j = 1, \dots, N$ se denota \mathbf{A}_{ij} su cofactor correspondiente, se tiene que [3]

$$\det(\mathbf{A}) = \sum_{k=1}^N a_{ik} \mathbf{A}_{ik} \quad (\text{A.5})$$

Entonces

$$\frac{\partial \det(\mathbf{A})}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}} \left(\sum_{k=1}^N a_{ik} \mathbf{A}_{ik} \right) = \mathbf{A}_{ij} \quad (\text{A.6})$$

Por tanto, teniendo en cuenta la definición de derivada de una función escalar respecto a una matriz obtenemos

$$\frac{\partial \det(\mathbf{A})}{\partial \mathbf{A}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{N1} & \mathbf{A}_{N2} & \cdots & \mathbf{A}_{NN} \end{bmatrix} = \text{adj}(\mathbf{A})^T \quad (\text{A.7})$$

- Si \mathbf{A} es simétrica y $J(\mathbf{X}) = \ln(\det(\mathbf{X}^T \mathbf{A} \mathbf{X}))$ con $\mathbf{X}^T \mathbf{A} \mathbf{X}$ no singular, se tiene que

$$\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = \frac{2}{\det(\mathbf{X}^T \mathbf{A} \mathbf{X})} \mathbf{A} \mathbf{X} \text{adj}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = 2 \mathbf{A} \mathbf{X} (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \quad (\text{A.8})$$

Prueba:

Aplicando la regla de la cadena

$$\begin{aligned} \frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} &= \frac{\partial \ln(\det(\mathbf{X}^T \mathbf{A} \mathbf{X}))}{\partial \det(\mathbf{X}^T \mathbf{A} \mathbf{X})} = \frac{\partial \det(\mathbf{X}^T \mathbf{A} \mathbf{X})}{\partial \mathbf{X}} \\ &= \frac{1}{\det(\mathbf{X}^T \mathbf{A} \mathbf{X})} 2 \mathbf{A} \mathbf{X} \text{adj}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = 2 \mathbf{A} \mathbf{X} (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \end{aligned} \quad (\text{A.9})$$

donde se ha utilizado (A.6) y $\mathbf{A}^{-1} = \text{adj}(\mathbf{A}) / \det(\mathbf{A})$.

- Si \mathbf{A} es simétrica y $J(\mathbf{X}) = \text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X})$ entonces

$$\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{A}\mathbf{X} \quad (\text{A.10})$$

Prueba:

La función $J(\mathbf{X})$ puede escribirse de la siguiente forma

$$J(\mathbf{X}) = \text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \sum_{i=1}^N \mathbf{x}_i^T \mathbf{A} \mathbf{x}_i \quad (\text{A.11})$$

donde \mathbf{x}_i es la i -ésima columna de \mathbf{X} . Utilizando la definición de derivada de una función escalar de variable matricial, obtenemos lo siguiente

$$\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = 2[\mathbf{A}\mathbf{x}_1, \dots, \mathbf{A}\mathbf{x}_N] = 2\mathbf{A}\mathbf{X} \quad (\text{A.12})$$

A.2 Definiciones de derivadas complejas

La derivada de una función escalar J de variable vectorial o matricial compleja se definen como sigue

- *Definición 1:* Sea $J = J(\mathbf{z})$ una función real de variable compleja de dimensión $N \times 1$, $\mathbf{z} = [z_1, \dots, z_N]^T = [x_1 + jy_1, \dots, x_N + jy_N]^T$, se define la derivada de J respecto a \mathbf{z} como el vector columna

$$\frac{\partial J(\mathbf{z})}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial J(\mathbf{z})}{\partial z_1} \\ \vdots \\ \frac{\partial J(\mathbf{z})}{\partial z_N} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{\partial J(\mathbf{z})}{\partial x_1} - j \frac{\partial J(\mathbf{z})}{\partial y_1} \\ \vdots \\ \frac{\partial J(\mathbf{z})}{\partial x_N} - j \frac{\partial J(\mathbf{z})}{\partial y_N} \end{bmatrix} \quad (\text{A.13})$$

y de forma similar su derivada conjugada es

$$\frac{\partial J(\mathbf{z})}{\partial \mathbf{z}^*} = \begin{bmatrix} \frac{\partial J(\mathbf{z})}{\partial z_1^*} \\ \vdots \\ \frac{\partial J(\mathbf{z})}{\partial z_N^*} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{\partial J(\mathbf{z})}{\partial x_1} + j \frac{\partial J(\mathbf{z})}{\partial y_1} \\ \vdots \\ \frac{\partial J(\mathbf{z})}{\partial x_N} + j \frac{\partial J(\mathbf{z})}{\partial y_N} \end{bmatrix} \quad (\text{A.14})$$

- *Definición 2:* Sea $J = J(\mathbf{Z})$ una función escalar de variable compleja matricial \mathbf{Z} de dimensión $P \times N$. La derivada de J respecto a la matriz \mathbf{Z} es la matriz de dimensión $P \times N$ dada por

$$\begin{aligned} \frac{\partial J(\mathbf{Z})}{\partial \mathbf{Z}} &= \begin{bmatrix} \frac{\partial J(\mathbf{Z})}{\partial z_{11}} & \dots & \frac{\partial J(\mathbf{Z})}{\partial z_{1N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial J(\mathbf{Z})}{\partial z_{P1}} & \dots & \frac{\partial J(\mathbf{Z})}{\partial z_{PN}} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \frac{\partial J(\mathbf{Z})}{\partial x_{11}} & \dots & \frac{\partial J(\mathbf{Z})}{\partial x_{1N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial J(\mathbf{Z})}{\partial x_{P1}} & \dots & \frac{\partial J(\mathbf{Z})}{\partial x_{PN}} \end{bmatrix} - \frac{j}{2} \begin{bmatrix} \frac{\partial J(\mathbf{Z})}{\partial z_{11}} & \dots & \frac{\partial J(\mathbf{Z})}{\partial z_{1N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial J(\mathbf{Z})}{\partial z_{P1}} & \dots & \frac{\partial J(\mathbf{Z})}{\partial z_{PN}} \end{bmatrix} \end{aligned} \quad (\text{A.15})$$

De acuerdo con estas definiciones el operador gradiente complejo con respecto al vector $\mathbf{z} = [z_1, \dots, z_N]^T$ se define como $\nabla_{\mathbf{z}} = [\partial/\partial z_1^*, \dots, \partial/\partial z_N^*]^T$.

Ejemplos:

Sean \mathbf{c} y \mathbf{w} dos vector $N \times 1$ y \mathbf{R} es una matriz $N \times N$ hermítica. Entonces

$$\frac{\partial}{\partial \mathbf{w}^*}(\mathbf{w}^T \mathbf{c}) = \frac{\partial}{\partial \mathbf{w}^*}(\mathbf{c}^T \mathbf{w}) = \mathbf{0} \quad (\text{A.16})$$

$$\frac{\partial}{\partial \mathbf{w}^*}(\mathbf{w}^H \mathbf{R} \mathbf{w}) = \mathbf{R} \mathbf{w} \quad (\text{A.17})$$

$$\frac{\partial}{\partial \mathbf{w}^*}(\mathbf{w}^H \mathbf{c}) = \mathbf{c} \quad (\text{A.18})$$

Apéndice B

Revisión de estadísticos de orden superior

Como se ha visto a lo largo del presente trabajo, la separación ciega de fuentes se puede resolver recurriendo al uso de estadísticos de orden superior (HOS, del inglés Higher Order Statistics). A continuación, comentaremos brevemente las definiciones y propiedades de los HOS. Una exposición más detallada al respecto puede ser consultada en [62, 65, 73].

B.1 Definiciones de HOS

Sean $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$ y $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ dos colecciones de variables aleatorias reales. El cumulante de orden k de estas variables se define como los coeficientes asociados al polinomio de orden k en serie de Taylor, con respecto a \mathbf{w} , de la función generadora

$$\Theta(\mathbf{w}) = \ln(\Phi(\mathbf{w})) \quad (\text{B.1})$$

donde $\Phi(\cdot)$ es la función característica conjunta de las variables aleatorias \mathbf{x}

$$\Phi(\mathbf{w}) = E[\exp(j\mathbf{w}^T \mathbf{x})] \quad (\text{B.2})$$

y su desarrollo en serie de Taylor proporciona los momentos de \mathbf{x} . En base a esta definición, se puede establecer la siguiente relación entre los cumulantes de segundo, tercer y cuarto orden de una variable \mathbf{x} (de media cero) y sus momentos

$$\begin{aligned} cum(x_1, x_2) &= E[x_1 x_2] \\ cum(x_1, x_2, x_3) &= E[x_1 x_2 x_3] \\ cum(x_1, x_2, x_3, x_4) &= E[x_1 x_2 x_3 x_4] - E[x_1 x_2]E[x_3 x_4] - E[x_1 x_3]E[x_2 x_4] - E[x_1 x_4]E[x_2 x_3] \end{aligned} \quad (\text{B.3})$$

Si las variables no tienen media cero, debe sustituirse x_i por $x_i - E[x_i]$ en las expresiones anteriores. En este trabajo tienen especial interés el cumulante cruzado de cuarto orden

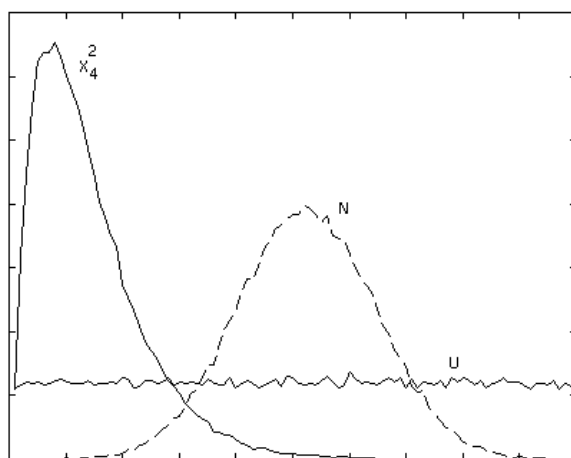


Figura B.1: Comparación de distribuciones con la gaussiana

entre dos variables x_1 y x_2 , que denotaremos $K_{x_1x_2}$, y el cumulante de cuarto orden (curtosis) de x_1 , que denotaremos K_{x_1} . A partir de (B.3) obtenemos las expresiones de ambos cumulantes

$$\begin{aligned} K_{x_1x_2} &= cum(x_1, x_2, x_1, x_2) = E[x_1^2x_2^2] - 2E^2[x_1x_2] - E[x_1^2]E[x_2^2] \\ K_{x_1} &= cum(x_1, x_1, x_1, x_1) = E[x_1^4] - 3E^2[x_1^2] \end{aligned} \quad (B.4)$$

Otras definiciones importantes son las de momentos y cumulantes de un proceso estacionario. Si suponemos que x es una señal estacionaria real que tiene momentos hasta orden k , entonces

$$m_k^x(\tau_1, \tau_2, \dots, \tau_{k-1}) = m_k[x(n)x(n + \tau_1)\dots x(n + \tau_{k-1})] \quad (B.5)$$

representa la función momento de orden k de la señal y depende solamente de las diferencias $\tau_1, \tau_2, \dots, \tau_{k-1}$. Claramente, la función momento de orden dos es la autocorrelación $E[x(n)x(n + \tau)]$. Los cumulantes de orden $k > 2$ de una señal aleatoria estacionaria no gaussiana pueden escribirse de la siguiente forma

$$cum_k^x(\tau_1, \tau_2, \dots, \tau_{k-1}) = m_k^x[x(n)x(n + \tau_1)\dots x(n + \tau_{k-1})] - m_k^G[x(n)x(n + \tau_1)\dots x(n + \tau_{k-1})] \quad (B.6)$$

donde m_k^G es el momento de orden k de una señal gaussiana que tiene la misma media y secuencia de correlación que x . Obviamente, si x tiene distribución gaussiana todos sus cumulantes de orden superior a dos serán nulos, $cum_k^x(\tau_1, \tau_2, \dots, \tau_{k-1}) = 0$.

La definición (B.6) nos permite interpretar los cumulantes de orden superior a dos como una medida de la distancia entre la distribución de una variable aleatoria no gaussiana con respecto a la gaussiana. En particular, la curtosis mide el mayor o menor grado de apuntamiento de la distribución. Las señales con curtosis positivas reciben el nombre

de *supergaussianas* y las de curtosis negativas de *subgaussianas*. En la figura B.1 se ha representado una distribución gaussiana, una supergaussiana y una subgaussiana. La supergaussiana corresponde a una distribución \mathcal{X}^2 de Pearson con cuatro grados de libertad (curtosis teórica positiva $K_x = 3$) y, como puede observarse, su distribución es más apuntada que la gaussiana. Por otro lado, la distribución subgaussiana es una variable uniforme (curtosis teórica $K_x = -1.2$) cuya distribución es más achatada que la gaussiana.

Son numerosos los algoritmos de deconvolución [8] y de separación que emplean momentos de orden superior [2, 13, 14, 17] y cumulantes de orden superior [18, 22, 52, 55, 60]. A la hora de utilizar cualquiera de ellos debe tenerse en cuenta varios factores. En primer lugar, los momentos de orden superior de un proceso gaussiano no son cero y los cumulantes sí. En segundo lugar, sea cual sea el orden del cumulante, el cumulante de la suma de cada uno de los procesos aleatorios estadísticamente independientes es igual a la suma de los cumulantes de cada uno por separado, lo cual no es cierto para los momentos. Como consecuencia, las expresiones que resultan de utilizar cumulantes son más fáciles de manejar que las de los momentos. Por otro lado, los momentos de orden superior son, en general, más fáciles de estimar que los cumulantes.

B.2 Propiedades de los cumulantes

Las propiedades más importante de los cumulantes son las siguientes

- *Propiedad 1:* Los cumulantes de cantidades escaladas son iguales al producto de todos los factores de escala por el cumulante de las cantidades sin escalar

$$cum(\lambda_1 x_1, \dots, \lambda_N x_N) = \left(\prod_{i=1}^N \lambda_i \right) cum(x_1, \dots, x_N) \quad (\text{B.7})$$

donde λ_i , $i = 1, \dots, N$ son constantes y x_i , $i = 1, \dots, N$ son variables aleatorias.

- Los cumulantes son simétricos en sus argumentos

$$cum(x_1, \dots, x_N) = cum(x_{i_1}, \dots, x_{i_N}) \quad (\text{B.8})$$

donde (i_1, \dots, i_N) es una permutación de los índices $1, \dots, N$.

- *Propiedad 2:* Los cumulantes son ciegos a constantes aditivas, es decir, si α es una constante

$$cum(\alpha + x_1, \dots, x_N) = cum(x_1, \dots, x_N) \quad (\text{B.9})$$

- *Propiedad 3:* Los cumulantes de las sumas de variables independientes y_i , $i = 1, \dots, N$ y x_i , $i = 1, \dots, N$ es igual a la suma de los cumulantes de cada variable

$$cum(x_1 + y_1, \dots, x_N + y_N) = cum(x_1, \dots, x_N) + cum(y_1, \dots, y_N) \quad (\text{B.10})$$

- *Propiedad 4:* El cumulante cruzado de variables aleatorias independientes x_i , $i = 1, \dots, N$ y y_i , $i = 1, \dots, N$ es cero

$$\text{cum}(x_1, \dots, x_N, y_1, \dots, y_N) = 0 \quad (\text{B.11})$$

Apéndice C

Revisión de Teoría de la Información

Este apéndice recoge diversos conceptos de teoría de la información los cuales han sido utilizados en el capítulo 3 para derivar algoritmos de separación ciega de fuentes. Una revisión más detallada puede ser consultada en [30].

Sea x una variable aleatoria que toma valores discretos x_k en un alfabeto \mathcal{X} con probabilidad $p(x_k)$, $\sum p(x_k) = 1$. El valor de $p(x_k)$ expresa la probabilidad de que ocurra un determinado valor x_k pero también nos está indicando la cantidad de información que se adquiere cuando ese valor es observado. Por ejemplo, el valor $p(x_k) = 1$ indica que sabemos con absoluta certeza el valor que va a ocurrir y, por tanto, al observarlo no adquirimos ninguna información. Si tenemos dos posibles valores x_1 con probabilidad $p(x_1)$ y x_2 con probabilidad $p(x_2) > p(x_1)$, entonces es mayor la incertidumbre sobre el evento $x = x_1$ que sobre $x = x_2$ y, por ello, también es mayor la información que se adquiere cuando éste ocurre. Como consecuencia, podemos establecer que la cantidad de información adquirida al observar x_k está relacionada con la inversa de su probabilidad. Es usual medir la cantidad de información adquirida después de observar un evento $x = x_k$, con probabilidad $p(x_k)$, como una función logarítmica

$$I(x_k) = \log \left(\frac{1}{p(x_k)} \right) = -\log(p(x_k)) \quad (\text{C.1})$$

donde la base del logaritmo es arbitraria. La cantidad media de información de x recibe el nombre de entropía y se define como sigue

$$H(x) = E[I(x_k)] = - \sum_{x \in \mathcal{X}} p(x) \log(p(x)) = -E[\log p(x)] \quad (\text{C.2})$$

La entropía es igual a cero sólo si la variable aleatoria x describe un proceso determinista ya que se conoce *a priori* el valor que va a ocurrir y toma su valor máximo cuando la variable aleatoria es uniforme ya que todos los valores $x = x_k$ tienen la misma probabilidad.

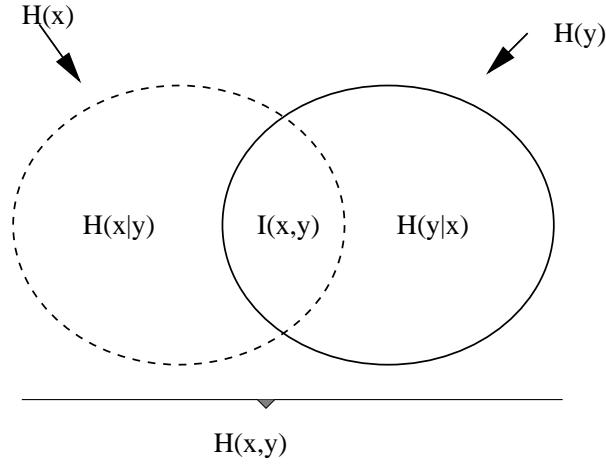


Figura C.1: Relación entre la entropía de dos variables aleatorias y la información mutua.

En algunos casos es interesante medir la cantidad de información que una variable aleatoria y contiene sobre otra variable aleatoria x . Consideremos que tenemos un sistema cuya entrada x toma valores en un alfabeto \mathcal{X} y su salida y es una versión perturbada de x que toma valores en otro alfabeto \mathcal{Y} . La cuestión que nos planteamos ahora es cuanta información adquirimos sobre x cuando observamos la salida del sistema. Para contestar a esta interrogante primero definiremos el concepto de entropía conjunta $H(x, y)$ como sigue

$$H(x, y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log(p(x, y)) = -E [\log(p(x, y))] \quad (\text{C.3})$$

donde $p(x, y)$ es la probabilidad conjunta de ambas variables aleatorias. Ahora, como se muestra en la figura C.1, podemos definir la incertidumbre sobre la entrada que permanece sin resolver después de observar la salida de la siguiente forma

$$H(x|y) = H(x, y) - H(y) \quad (\text{C.4})$$

Análogamente, la cantidad de incertidumbre sobre la salida que no es resuelta al observar la entrada viene dada por

$$H(y|x) = H(x, y) - H(x) \quad (\text{C.5})$$

Es interesante notar que esta última expresión nos da información sobre la perturbación que ha experimentado la entrada del sistema x .

Otro problema que aparece con frecuencia es medir la distancia entre dos distribuciones $p(x)$ y $q(x)$. Para ello, en este trabajo utilizaremos la divergencia de Kullback-Leibler $K(p, q)$, también llamada entropía relativa o entropía cruzada, definida como sigue

$$K(p, q) = \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right) = E \left[\log \left(\frac{p(x)}{q(x)} \right) \right] \quad (\text{C.6})$$

Esta cantidad es no negativa tomando el valor cero cuando las distribuciones coinciden $q(x) = p(x)$. Además, no es simétrica $K(p, q) \neq K(q, p)$ por lo que no puede considerarse una medida de distancia en un sentido estricto.

La divergencia de Kullback-Leibler suele ser utilizada para estudiar la independencia estadística entre varias variables. Si la probabilidad conjunta entre x y y es $p(x, y)$ entonces sabemos que si son estadísticamente independientes se cumple que $p(x, y) = p(x)p(y)$. Por tanto, la dependencia estadística puede ser medida como la “distancia” entre la probabilidad conjunta y el producto de las probabilidades individuales

$$I(x; y) = K(p(x, y), p(x)p(y)) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (\text{C.7})$$

Esta medida recibe el nombre de información mutua y es siempre no negativa anulándose cuando las variables son estadísticamente independientes ya que $\log(p(x, y)/p(x)p(y)) = 0$. En la figura C.1 puede verse que la información mutua y las entropías definidas anteriormente están relacionadas por la siguiente expresión

$$I(x, y) = H(x) - H(x|y) \quad (\text{C.8})$$

Como $H(x)$ nos indica la incertidumbre sobre la entrada del sistema x que existe antes de observar la salida y y $H(x|y)$ representa la incertidumbre sobre x que permanece sin resolver después de observar y , entonces $I(x, y)$ debe representar la cantidad de incertidumbre sobre la entrada que es resuelta al observar la salida. De forma análoga podemos escribir la información mutua en función de la entropía de la salida $H(y)$ y de $H(y|x)$

$$I(x, y) = H(y) - H(y|x) \quad (\text{C.9})$$

Las definiciones anteriores de entropías e información mutua pueden plantearse para variables aleatorias continuas sustituyendo los sumatorios por integrales y la función de probabilidad discreta $p(x)$ por la función de densidad $f(x)$. En la tabla C.1 se resumen las definiciones anteriores y se indica su equivalente continuo.

C.1 Minimización de la divergencia de Kullback-Leibler

En esta sección obtendremos la expresión del algoritmo de decorrelación presentado en el capítulo 3 a partir de la minimización de la divergencia de Kullback-Leibler entre las funciones densidad de probabilidad de dos variables gaussianas. Antes de ello, conviene introducir los siguiente teoremas.

Teorema C.1 Si $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^T]$ es la matriz de autocorrelación de N variables aleatorias \mathbf{x} , entonces

$$E[\mathbf{x}^T \mathbf{R}_x^{-1} \mathbf{x}] = N$$

Variable discreta	Variable continua
<i>Entropía</i>	
$H(x) = -\sum_{x \in \mathcal{X}} p(x) \log(p(x))$	$h(x) = -\int_{\mathcal{X}} f(x) \log(f(x)) dx$
<i>Entropía conjunta</i>	
$H(x, y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log(p(x, y))$	$h(x, y) = -\int_{\mathcal{X}} \int_{\mathcal{Y}} f(x, y) \log(f(x, y)) dy dx$
<i>Entropía condicionada</i>	
$H(x y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log(p(y x))$	$h(x y) = -\int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} p(x, y) \log(p(y x)) dy dx$
<i>Divergencia de Kullback-Leibler</i>	
$K(p, q) = \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right)$	$K(f, g) = \int_{x \in \mathcal{X}} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$
<i>Información mutua</i>	
$I(x, y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right)$	$I(x, y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x, y) \log\left(\frac{f(x, y)}{f(x)f(y)}\right) dy dx$

Tabla C.1: Entropías e información mutua para variables discretas y continuas.

Demostración:

Para demostrar este teorema expresaremos la inversa y el determinante de la matriz \mathbf{R} (con términos $r_{ij} = E[x_i x_j]$) en función de sus cofactores R_{ij} [72]:

$$\mathbf{R}^{-1} = \frac{\text{adj}(\mathbf{R})}{\det(\mathbf{R})} \quad (\text{C.10})$$

$$\det(\mathbf{R}) = \sum_{j=1}^N r_{ij} R_{ij} \quad (\text{C.11})$$

donde $\text{adj}(\mathbf{R})$ es la matriz adjunta

$$\text{adj}(\mathbf{R}) = \begin{bmatrix} R_{11} & R_{21} & \cdots & R_{N1} \\ R_{12} & R_{22} & \cdots & R_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1N} & R_{2N} & \cdots & R_{NN} \end{bmatrix} \quad (\text{C.12})$$

Utilizando estas definiciones podemos escribir

$$\begin{aligned} E[\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}] &= \frac{E[\mathbf{x}^T \text{adj}(\mathbf{R}) \mathbf{x}]}{\det(\mathbf{R})} = \frac{E\left[\sum_{i=1}^N \sum_{j=1}^N x_i x_j R_{ij}\right]}{\det(\mathbf{R})} \\ &= \frac{\sum_{i=1}^N \sum_{j=1}^N E[x_i x_j] R_{ij}}{\det(\mathbf{R})} = \frac{\sum_{i=1}^N \sum_{j=1}^N r_{ij} R_{ij}}{\det(\mathbf{R})} \\ &= \frac{\sum_{j=1}^N \det(\mathbf{R})}{\det(\mathbf{R})} = N \end{aligned} \quad (\text{C.13})$$

Con lo que queda demostrado el teorema.

Teorema C.2 *Sea \mathbf{y} un vector de N variables aleatorias continuas gaussianas de media cero y autocorrelación $E[\mathbf{y}\mathbf{y}^T] = \mathbf{W}^T \mathbf{R}_x \mathbf{W}$ y sea \mathbf{s} un vector de N variables aleatorias gaussianas de media cero y autocorrelación $E[\mathbf{s}\mathbf{s}^T] = \mathbf{I}$. La divergencia de Kullback-Leibler entre las f.d.p de ambas distribuciones viene dada por*

$$K(f_y, f_s) = \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{R}_x \mathbf{W}) - \frac{1}{2} \ln(\det(\mathbf{W}^T \mathbf{R}_x \mathbf{W})) - \frac{N}{2} \quad (\text{C.14})$$

y es siempre no negativa anulándose cuando $\mathbf{R}_y = \mathbf{I}$.

Demostración:

La función de densidad de probabilidad f_y y f_s vienen dadas por las siguientes expresiones

$$f_s(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^N}} \exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{y}\right) \quad (\text{C.15})$$

$$f_y(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{R}_y)}} \exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y}\right) \quad (\text{C.16})$$

Dividiendo (C.15) entre (C.16) obtenemos lo siguiente

$$\frac{f_y(\mathbf{y})}{f_s(\mathbf{y})} = \frac{1}{\sqrt{\det(\mathbf{R}_y)}} \exp\left(-\frac{1}{2}\mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y} + \frac{1}{2}\mathbf{y}^T \mathbf{y}\right) \quad (\text{C.17})$$

Por tanto, la divergencia de Kullback-Leibler puede escribirse de la siguiente forma

$$\begin{aligned} K(f_y, f_s) &= E\left[\ln\left(\frac{f_y(\mathbf{y})}{f_s(\mathbf{y})}\right)\right] = -\frac{1}{2}\ln(\det(\mathbf{R}_y)) + E\left[-\frac{1}{2}\mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y} + \frac{1}{2}\mathbf{y}^T \mathbf{y}\right] \\ &= -\frac{1}{2}\ln(\det(\mathbf{R}_y)) - \frac{1}{2}E[\mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y}] + \frac{1}{2}E[\mathbf{y}^T \mathbf{y}] \end{aligned} \quad (\text{C.18})$$

Utilizando el teorema C.1 presentado en este mismo apéndice podemos escribir $K(f_y, f_s)$ de la siguiente forma

$$K(f_y, f_s) = -\frac{1}{2}\ln(\det(\mathbf{R}_y)) - \frac{N}{2} + \frac{1}{2}E[\mathbf{y}^T \mathbf{y}] \quad (\text{C.19})$$

Utilizando $E[\mathbf{y}^T \mathbf{y}] = \text{tr}(\mathbf{W}^T \mathbf{R}_x \mathbf{W})$ y $\mathbf{R}_y = \mathbf{W}^T \mathbf{R}_x \mathbf{W}$ podemos expresar la distancia de Kullback-Leibler de la siguiente forma

$$K(f_y, f_s) = \frac{1}{2}\text{tr}(\mathbf{W}^T \mathbf{R}_x \mathbf{W}) - \frac{1}{2}\ln(\det(\mathbf{W}^T \mathbf{R}_x \mathbf{W})) - \frac{N}{2} \quad (\text{C.20})$$

Ahora probaremos que (C.20) es no negativa. Llamaremos λ_i a los N autovalores de la matriz de correlación de las salidas \mathbf{R}_y . Por tanto, la traza y el determinante de \mathbf{R}_y pueden escribirse como sigue

$$\begin{aligned} \det(\mathbf{R}_y) &= \prod_{i=1}^N \lambda_i \\ \text{tr}(\mathbf{R}_y) &= \sum_{i=1}^N \lambda_i \end{aligned} \quad (\text{C.21})$$

Sustituyendo estas expresiones en (C.20) obtenemos lo siguiente

$$\begin{aligned} K(f_y, f_s) &= \frac{1}{2} \sum_{i=1}^N \lambda_i - \frac{N}{2} - \frac{1}{2} \ln\left(\prod_{i=1}^N \lambda_i\right) \\ &= \frac{1}{2} \left(\sum_{i=1}^N \lambda_i - N - \sum_{i=1}^N \ln(\lambda_i) \right) = \frac{1}{2} \sum_{i=1}^N \psi(\lambda_i) \end{aligned} \quad (\text{C.22})$$

donde $\psi(\lambda_i) = \lambda_i - 1 - \ln(\lambda_i)$. Cada función $\psi(\lambda_i)$ es no negativa ya que $\lambda_i - 1 \geq \ln(\lambda_i)$. Además toma el valor cero cuando $\lambda_i = 1$, i.e., cuando $\mathbf{R}_y = \mathbf{I}$.

C.1.1 Algoritmo de gradiente relativo

Consideremos un vector \mathbf{s} de N fuentes gaussianas de media cero y varianza unidad. Supondremos que las fuentes son estadísticamente independientes entre sí lo que se traduce en que \mathbf{s} tenga una matriz de autocorrelación diagonal $\mathbf{R}_s = \mathbf{I}$. Estas fuentes son mezcladas por el sistema \mathbf{A} para dar lugar al vector de observaciones $\mathbf{x} = \mathbf{A}\mathbf{s}$ que, a su vez, será procesado por el sistema \mathbf{W} para dar lugar al vector de salidas $\mathbf{y} = \mathbf{W}^T\mathbf{x} = \mathbf{W}^T\mathbf{A}\mathbf{s}$. Obviamente, al ser las observaciones y las salidas combinaciones lineales de señales gaussianas de media nula, entonces ambas también tendrán distribución gaussiana y media cero. Las correspondientes matrices de autocorrelación son

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^T] = \mathbf{A}\mathbf{R}_s\mathbf{A}^T = \mathbf{A}\mathbf{A}^T \quad (\text{C.23})$$

$$\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^T] = \mathbf{W}^T\mathbf{R}_x\mathbf{W} = \mathbf{W}^T\mathbf{A}\mathbf{A}^T\mathbf{W} \quad (\text{C.24})$$

que, en general, serán distintas de \mathbf{R}_s y, como consecuencia, las f.d.p. f_x y f_y serán distintas de f_s . En ausencia de una mayor cantidad de información, es evidente que, desde un punto de vista estadístico, una condición necesaria para recuperar las fuentes es seleccionar \mathbf{W} para que f_y sea igual a f_s . Esto se puede conseguir utilizando el concepto de divergencia de Kullback-Leibler que mide la distancia entre dos funciones densidad de probabilidad. La expresión de la divergencia de Kullback-Leibler entre f_y y f_s viene dada por el teorema C.2

$$K(f_y, f_s) = \frac{1}{2}\text{tr}(\mathbf{W}^T\mathbf{R}_x\mathbf{W}) - \frac{1}{2}\ln(\det(\mathbf{W}^T\mathbf{R}_x\mathbf{W})) - \frac{N}{2} \quad (\text{C.25})$$

El teorema C.2 demuestra que la función (C.25) es siempre no negativa y se anula cuando las salidas tienen la misma distribución que las fuentes (son gaussianas de media cero y matriz de autocorrelación igual a la identidad).

La minimización de (C.25) puede llevarse a cabo con un algoritmo de gradiente relativo. La expresión del gradiente que se obtiene utilizando las derivadas matriciales dadas en el apéndice A es la siguiente

$$\begin{aligned} \nabla_{\mathbf{W}}K(f_y, f_s) &= \frac{1}{2}\nabla_{\mathbf{W}}\text{tr}(\mathbf{W}^T\mathbf{R}_x\mathbf{W}) - \frac{1}{2}\nabla_{\mathbf{W}}\ln(\det(\mathbf{W}^T\mathbf{R}_x\mathbf{W})) \\ &= \mathbf{R}_x\mathbf{W} - \mathbf{R}_x\mathbf{W}(\mathbf{W}^T\mathbf{R}_x\mathbf{W})^{-1} \end{aligned} \quad (\text{C.26})$$

la cual, al multiplicar por $\mathbf{W}\mathbf{W}^T$ y considerando (C.24), permite obtener la versión relativa

$$\begin{aligned} \mathbf{W}\mathbf{W}^T\nabla_{\mathbf{W}}K(f_y, f_s) &= \mathbf{W}\mathbf{W}^T(\mathbf{R}_x\mathbf{W}) - \mathbf{W}\mathbf{W}^T(\mathbf{R}_x\mathbf{W}(\mathbf{W}^T\mathbf{R}_x\mathbf{W})^{-1}) \\ &= \mathbf{W}(\mathbf{W}^T\mathbf{R}_x\mathbf{W} - \mathbf{I}) = \mathbf{W}(E[\mathbf{y}\mathbf{y}^T] - \mathbf{I}) \end{aligned} \quad (\text{C.27})$$

Utilizando una sola muestra para estimar los estadísticos que aparecen en (C.27) se obtiene el algoritmo de gradiente relativo resultante

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) - \mu\mathbf{W}(n)\mathbf{W}^T(n)\widehat{\nabla}_{\mathbf{W}}K(n) \\ &= \mathbf{W}(n) - \mu\mathbf{W}(n)(\mathbf{y}(n)\mathbf{y}^T(n) - \mathbf{I}) \end{aligned} \quad (\text{C.28})$$

Esta expresión se corresponde con el algoritmo de decorrelación (3.1) presentado en el capítulo 3.

Apéndice D

Jacobiano de los algoritmos del capítulo 3

Este apéndice recoge las expresiones de los jacobianos utilizados para analizar la estabilidad de los algoritmos presentados en el capítulo 3.

D.1 Jacobiano del algoritmo de Héroult y Jutten

Comenzaremos obteniendo las expresiones de las primeras derivadas de la ecuación diferencial ordinaria (ODE) asociada al algoritmo (3.6) propuesto por Héroult y Jutten, las cuales han sido utilizadas en el análisis de estabilidad presentado en la sección 3.3.1.

El estudio de la estabilidad del algoritmo (3.6) ha sido realizado considerando una red neuronal recursiva de dos entradas y dos salidas. Para este caso, las salidas $\mathbf{y} = (\mathbf{I} + \hat{\mathbf{W}}^T)^{-1}\mathbf{x}$ (con $\hat{w}_{ii} = 0$) de la red vienen dadas por las siguientes expresiones

$$y_1 = \frac{x_1 - \hat{w}_{12}x_2}{1 - \hat{w}_{12}\hat{w}_{21}} \quad y_2 = \frac{x_2 - \hat{w}_{21}x_1}{1 - \hat{w}_{12}\hat{w}_{21}} \quad (\text{D.1})$$

En el caso $N = 2$, en la sección 3.3.1 se ha determinado que la ODE del algoritmo es

$$\begin{aligned} F_{12} &= \frac{d\hat{w}_{12}}{dt} = E[f(y_2)g(y_1)] \\ F_{21} &= \frac{d\hat{w}_{21}}{dt} = E[f(y_1)g(y_2)] \end{aligned} \quad (\text{D.2})$$

Con el objeto de encontrar la expresión de la matriz jacobiana formada por las primeras derivadas de (D.2) primero determinaremos las derivadas de las salidas y_1 y y_2 dadas en (D.1) con respecto a \hat{w}_{12} y \hat{w}_{21}

$$\begin{aligned} \frac{\partial y_1}{\partial \hat{w}_{12}} &= -\frac{y_2}{1 - \hat{w}_{12}\hat{w}_{21}}, & \frac{\partial y_1}{\partial \hat{w}_{21}} &= \frac{y_1\hat{w}_{12}}{1 - \hat{w}_{12}\hat{w}_{21}} \\ \frac{\partial y_2}{\partial \hat{w}_{12}} &= \frac{y_2\hat{w}_{21}}{1 - \hat{w}_{12}\hat{w}_{21}}, & \frac{\partial y_2}{\partial \hat{w}_{21}} &= -\frac{y_1}{1 - \hat{w}_{12}\hat{w}_{21}} \end{aligned} \quad (\text{D.3})$$

Utilizando la regla de la cadena y las derivadas (D.3) se obtienen las siguientes derivadas de las ecuaciones diferenciales (D.2)

$$\begin{aligned}
\frac{\partial F_{12}}{\partial \hat{w}_{12}} &= \frac{E[f'(y_2)g(y_1)y_2]\hat{w}_{21}}{1 - \hat{w}_{12}\hat{w}_{21}} - \frac{E[f(y_2)g'(y_1)y_2]}{1 - \hat{w}_{12}\hat{w}_{21}} \\
\frac{\partial F_{12}}{\partial \hat{w}_{21}} &= -\frac{E[f'(y_2)g(y_1)y_1]}{1 - \hat{w}_{12}\hat{w}_{21}} + \frac{E[f(y_2)g'(y_1)y_1]\hat{w}_{12}}{1 - \hat{w}_{12}\hat{w}_{21}} \\
\frac{\partial F_{21}}{\partial \hat{w}_{12}} &= -\frac{E[f'(y_1)g(y_2)y_2]}{1 - \hat{w}_{12}\hat{w}_{21}} + \frac{E[f(y_1)g'(y_2)y_2]\hat{w}_{21}}{1 - \hat{w}_{12}\hat{w}_{21}} \\
\frac{\partial F_{21}}{\partial \hat{w}_{21}} &= \frac{E[f'(y_1)g(y_2)y_1]\hat{w}_{12}}{1 - \hat{w}_{12}\hat{w}_{21}} - \frac{E[f(y_1)g'(y_2)y_1]}{1 - \hat{w}_{12}\hat{w}_{21}}
\end{aligned} \tag{D.4}$$

En el punto donde se consigue la separación de las fuentes, las salidas $y_1 = \lambda_1 s_i$, $y_2 = \lambda_2 s_j$, $i \neq j$ son estadísticamente independientes. Adicionalmente, si las fuentes tienen distribución simétrica con momentos impares cero y las no-linealidades son impares, entonces muchos de los momentos estadísticos que aparecen en las expresiones anteriores se anulan, siendo los únicos elementos no cero los siguientes

$$\begin{aligned}
\frac{\partial F_{12}}{\partial \hat{w}_{12}} &= \gamma \lambda_2 E[g'(\lambda_1 s_i)]E[f(\lambda_2 s_j)s_j] \\
\frac{\partial F_{21}}{\partial \hat{w}_{21}} &= \gamma \lambda_1 E[g(\lambda_1 s_i)s_i]E[f'(\lambda_2 s_j)] \\
\frac{\partial F_{21}}{\partial \hat{w}_{12}} &= \gamma \lambda_2 E[f'(\lambda_1 s_i)]E[g(\lambda_2 s_j)s_j] \\
\frac{\partial F_{21}}{\partial \hat{w}_{21}} &= \gamma \lambda_1 E[f(\lambda_1 s_i)s_i]E[g'(\lambda_2 s_j)]
\end{aligned} \tag{D.5}$$

donde $\gamma = -\frac{1}{1 - \hat{w}_{12}\hat{w}_{21}}$. Finalmente, agrupando estas derivadas en una matriz, obtenemos que el jacobiano asociado a (D.2) es el siguiente

$$\mathbf{H} = \begin{bmatrix} \frac{\partial F_{12}}{\partial \hat{w}_{12}} & \frac{\partial F_{12}}{\partial \hat{w}_{21}} \\ \frac{\partial F_{21}}{\partial \hat{w}_{12}} & \frac{\partial F_{21}}{\partial \hat{w}_{21}} \end{bmatrix} = \gamma \begin{bmatrix} \lambda_2 m_{g'}^{\lambda_1} \sigma_f^{\lambda_2} & \lambda_1 \sigma_g^{\lambda_1} m_{f'}^{\lambda_2} \\ \lambda_2 m_{f'}^{\lambda_1} \sigma_g^{\lambda_2} & \lambda_1 \sigma_f^{\lambda_1} m_{g'}^{\lambda_2} \end{bmatrix} \tag{D.6}$$

donde

$$\begin{aligned}
m_{f'}^{\lambda_1} &= E[f'(\lambda_1 s_i)], & m_{f'}^{\lambda_2} &= E[f'(\lambda_2 s_j)], \\
m_{g'}^{\lambda_1} &= E[g'(\lambda_1 s_i)], & m_{g'}^{\lambda_2} &= E[g'(\lambda_2 s_j)], \\
\sigma_f^{\lambda_1} &= E[s_i f(\lambda_1 s_i)], & \sigma_f^{\lambda_2} &= E[s_j f(\lambda_2 s_j)], \\
\sigma_g^{\lambda_1} &= E[s_i g(\lambda_1 s_i)], & \sigma_g^{\lambda_2} &= E[s_j g(\lambda_2 s_j)]
\end{aligned} \tag{D.7}$$

D.2 Puntos estacionarios del contraste de Moreau y Macchi

A continuación determinaremos la naturaleza de los puntos estacionarios del contraste J_c presentado en la sección 3.5. El análisis consistirá en estudiar el carácter definido negativo de la matriz hessiana formada por las segundas derivadas

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \phi}{\partial g_{12}^2} & \frac{\partial^2 \phi}{\partial g_{12} \partial g_{21}} \\ \frac{\partial^2 \phi}{\partial g_{21} \partial g_{12}} & \frac{\partial^2 \phi}{\partial g_{21}^2} \end{bmatrix} \quad (\text{D.8})$$

donde

$$\begin{aligned} \frac{\partial^2 \phi}{\partial^2 g_{12}} &= 12g_{12}^2(|K_{s_1}| + |K_{s_2}|) + 2\alpha g_{21}^2(|K_{s_1}| + |K_{s_2}|) - 4|K_{s_1}| - 2\alpha|K_{s_2}| \\ \frac{\partial^2 \phi}{\partial^2 g_{21}} &= 12g_{21}^2(|K_{s_1}| + |K_{s_2}|) + 2\alpha g_{12}^2(|K_{s_1}| + |K_{s_2}|) - 4|K_{s_2}| - 2\alpha|K_{s_1}| \\ \frac{\partial^2 \phi}{\partial g_{12} g_{21}} &= 4\alpha g_{12} g_{21} (|K_{s_1}| + |K_{s_2}|) \end{aligned} \quad (\text{D.9})$$

En la sección 3.5 se han determinado los siguientes grupos de puntos estacionarios:

- *Solución deseada*: $g_{11} = g_{21} = 0$. La matriz hessiana evaluada en esta solución tiene la siguiente forma

$$\mathbf{H} = \begin{bmatrix} -4|K_{s_1}| - 2\alpha|K_{s_2}| & 0 \\ 0 & -4|K_{s_2}| - 2\alpha|K_{s_1}| \end{bmatrix} \quad (\text{D.10})$$

Para $\alpha > 0$ los elementos de la diagonal serán negativos y el punto será un máximo.

- *Solución indeseada tipo I*: una salida extrae una única fuente y la otra extrae una combinación lineal. El estudio para las dos soluciones dadas en la sección 3.5 es simétrico por lo que nos centraremos en el siguiente punto

$$g_{12} = 0, \quad g_{21}^2 = \frac{\alpha|K_{s_1}| + 2|K_{s_2}|}{2(|K_{s_1}| + |K_{s_2}|)} \quad (\text{D.11})$$

Esta solución será válida únicamente cuando $\alpha < 2$ ya que en otro caso $g_{21}^2 > 1$ y, por tanto, $g_{22}^2 = 1 - g_{21}^2$ será negativo. Consideremos que $\alpha < 2$ y analicemos la matriz hessiana \mathbf{H} obtenida sustituyendo (D.11) en las expresiones de las segundas derivadas (D.9)

$$\mathbf{H} = \begin{bmatrix} (\alpha^2 - 4)|K_{s_1}| & 0 \\ 0 & 4\alpha|K_{s_1}| + 8|K_{s_2}| \end{bmatrix} \quad (\text{D.12})$$

La condición $\alpha > 0$ es suficiente para asegurar que el punto no es un máximo de la función de coste ya que la matriz \mathbf{H} tiene términos positivos en su diagonal. Adicionalmente, si $\alpha < 2$ entonces el primer elemento en la diagonal será negativo y el punto será de ensilladura.

- *Solución indeseada tipo II:* Ambas salidas extraen una combinación lineal de las fuentes

$$g_{12}^2 = \frac{|K_{s_1}|}{|K_{s_1}| + |K_{s_2}|}, \quad g_{21}^2 = \frac{|K_{s_2}|}{|K_{s_1}| + |K_{s_2}|} \quad (\text{D.13})$$

Evaluando las segundas derivadas en esta solución obtenemos la siguiente matriz

$$\mathbf{H} = \begin{bmatrix} 8|K_{s_1}| & \alpha g_{12} g_{21} (|K_{s_1}| + |K_{s_2}|) \\ \alpha g_{12} g_{21} (|K_{s_1}| + |K_{s_2}|) & 8|K_{s_2}| \end{bmatrix} \quad (\text{D.14})$$

La matriz \mathbf{H} tiene términos positivos en la diagonal y, por tanto, el punto no será un máximo de la función de coste. Estudiemos ahora el determinante de la matriz

$$\det(\mathbf{H}) = 16|K_{s_1}||K_{s_2}|(4 - \alpha^2) \quad (\text{D.15})$$

El punto será de ensilladura cuando $\alpha > 2$ y un mínimo cuando $\alpha < 2$.

- *Puntos en la frontera:* Los grupos de soluciones anteriores no abarcan la existencia de puntos estacionarios en la frontera, i.e, cuando $g_{12}^2 = 1$ ó $g_{21}^2 = 1$. El estudio de ambos es análogo por lo que nos limitaremos al caso $g_{21}^2 = 1$ y supondremos $g_{12}^2 \neq 1$ ya que este caso ha sido analizado en el grupo de *soluciones deseadas*. En este punto la función de coste depende de un único coeficiente

$$\phi(g_{12}) = \left((1 - g_{12}^2)^2 + 1 - \alpha(1 - g_{12}^2) \right) |K_{s_1}| + g_{12}^4 |K_{s_2}| \quad (\text{D.16})$$

La primera y segunda derivada de esta función es

$$\begin{aligned} \frac{\partial \phi}{\partial g_{12}} &= \left(4g_{12}(g_{12}^2 - 1) + 2\alpha g_{12} \right) |K_{s_1}| + 4g_{12}^3 |K_{s_2}| \\ \frac{\partial^2 \phi}{\partial g_{12}^2} &= \left(12g_{12}^2 - 4 + 2\alpha \right) |K_{s_1}| + 12g_{12}^2 |K_{s_2}| \end{aligned} \quad (\text{D.17})$$

Existe un punto estacionario cuando $g_{12} = 0$, i.e., cuando ambas salidas extraen la fuente s_1 . Para determinar su naturaleza estudiaremos el signo de la segunda derivada

$$\frac{\partial^2 \psi}{\partial^2 g_{12}} = 2|K_{s_1}|(\alpha - 2) \quad (\text{D.18})$$

Obviamente para que la función de coste no tenga un máximo indeseado en este punto es suficiente con elegir $\alpha > 2$. Finalmente, debemos hacer notar que la primera derivada también se anula en

$$g_{12}^2 = \frac{|K_{s_1}|(2 - \alpha)}{2(|K_{s_1}| + |K_{s_2}|)} \quad (\text{D.19})$$

Sin embargo, la condición $\alpha > 2$ asegura que no es una solución válida ya que $g_{12}^2 < 0$.

D.3 Jacobiano del algoritmo EASI

Esta sección recoge las primeras derivadas de la ecuación diferencial ordinaria asociada al algoritmo EASI (3.61), las cuales han sido utilizadas en el análisis de estabilidad presentado en la sección 3.6.1.

Recordemos que el estudio de la estabilidad ha sido realizado para el caso de un sistema de separación con dos entradas y dos salida. Para este caso, la ODE del algoritmo viene dada por las siguiente cuatro ecuaciones

$$\begin{aligned}
F_{11} &= \frac{dg_{11}}{dt} = -g_{11}(E[y_1^2] - 1) - g_{21}(E[y_2y_1] + E[y_2f(y_1)] - E[f(y_2)y_1]) \\
F_{12} &= \frac{dg_{12}}{dt} = -g_{12}(E[y_1^2] - 1) - g_{22}(E[y_2y_1] + E[y_2f(y_1)] - E[f(y_2)y_1]) \\
F_{21} &= \frac{dg_{21}}{dt} = -g_{11}(E[y_2y_1] + E[y_1f(y_2)] - E[f(y_1)y_2]) - g_{21}(E[y_2^2] - 1) \\
F_{22} &= \frac{dg_{22}}{dt} = -g_{12}(E[y_2y_1] + E[y_1f(y_2)] - E[f(y_1)y_2]) - g_{22}(E[y_2^2] - 1) \quad (D.20)
\end{aligned}$$

Derivando estas expresiones respecto a los coeficientes de la matriz de pesos obtenemos lo siguiente

$$\begin{aligned}
\frac{\partial F_{11}}{\partial g_{11}} &= -E[y_1^2] + 1 - 2g_{11}E[y_1s_1] - g_{21}(E[y_2s_1] + E[y_2f'(y_1)s_1] - E[f(y_2)s_1]) \\
\frac{\partial F_{11}}{\partial g_{12}} &= -2g_{11}E[y_1s_2] - g_{21}(E[y_2s_2] + E[y_2f'(y_1)s_2] - E[f(y_2)s_2]) \\
\frac{\partial F_{11}}{\partial g_{21}} &= -E[y_2y_1] - E[y_2f(y_1)] + E[f(y_2)y_1] - g_{21}(E[s_1y_1] + E[s_1f(y_1)] - E[f'(y_2)s_1y_1]) \\
\frac{\partial F_{11}}{\partial g_{22}} &= -g_{21}(E[s_2y_1] + E[s_2f(y_1)] - E[f'(y_2)s_2y_1]) \\
\frac{\partial F_{12}}{\partial g_{11}} &= -2g_{12}E[y_1s_1] - g_{22}(E[y_2s_1] + E[y_2f'(y_1)s_2] - E[f(y_2)s_1]) \\
\frac{\partial F_{12}}{\partial g_{12}} &= -E[y_1^2] + 1 - 2g_{12}E[y_1s_2] - g_{22}(E[y_2s_2] + E[y_2f'(y_1)s_2] - E[f(y_2)s_2]) \\
\frac{\partial F_{12}}{\partial g_{21}} &= -g_{22}(E[s_1y_1] + E[s_1f(y_1)] - E[f'(y_2)s_1y_1]) \\
\frac{\partial F_{12}}{\partial g_{22}} &= -E[y_2y_1] - E[y_2f(y_1)] + E[f(y_2)y_1] - g_{22}(E[s_2y_1] + E[s_2f(y_1)] - E[f'(y_2)s_2y_1]) \\
\frac{\partial F_{21}}{\partial g_{11}} &= -E[y_1y_2] - E[y_1f(y_2)] + E[f(y_1)y_2] - g_{11}(E[s_1y_2] + E[s_1f(y_2)] - E[f'(y_1)s_1y_2]) \\
\frac{\partial F_{21}}{\partial g_{12}} &= -g_{11}(E[s_2y_2] + E[s_2f(y_2)] - E[f'(y_1)s_2y_2]) \\
\frac{\partial F_{21}}{\partial g_{21}} &= -g_{11}(E[y_1s_1] + E[y_1f'(y_2)s_1] - E[f(y_1)s_1]) - E[y_2^2] + 1 - 2g_{21}E[y_2s_1]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F_{21}}{\partial g_{22}} &= -g_{11} (E[y_1 s_2] + E[y_1 f'(y_2) s_2] - E[f(y_1) s_2]) - 2g_{21} E[y_2 s_2] \\
\frac{\partial F_{22}}{\partial g_{11}} &= -g_{12} (E[s_1 y_2] + E[s_1 f(y_2)] - E[f'(y_1) s_1 y_2]) \\
\frac{\partial F_{22}}{\partial g_{12}} &= -E[y_1 y_2] - E[y_1 f(y_2)] + E[f(y_1) y_2] - g_{12} (E[s_2 y_2] + E[s_2 f(y_2)] - E[f'(y_1) s_2 y_2]) \\
\frac{\partial F_{22}}{\partial g_{21}} &= -g_{21} (E[y_1 s_1] + E[y_1 f'(y_2) s_1] - E[f(y_1) s_1]) - 2g_{22} E[y_2 s_1] \\
\frac{\partial F_{22}}{\partial g_{22}} &= -g_{12} (E[y_1 s_2] + E[y_1 f'(y_2) s_2] - E[f(y_1) s_2]) - E[y_2] + 1 - 2g_{22} E[y_2 s_2]
\end{aligned}$$

En el punto donde cada salida extrae una única fuente, la mayoría de las derivadas anteriores son cero. En particular, si consideramos el punto $\mathbf{G} = \mathbf{I}$ y las hipótesis sobre las fuentes (independencia estadística, potencia unidad y media cero) obtenemos que las derivadas no-nulas son las siguientes

$$\begin{aligned}
\frac{\partial F_{11}}{\partial g_{11}} &= -2 \\
\frac{\partial F_{12}}{\partial g_{12}} &= -1 - E[f'(s_1)] + E[f(s_2) s_2] \\
\frac{\partial F_{12}}{\partial g_{21}} &= -1 - E[f(s_1) s_1] + E[f'(s_2)] \\
\frac{\partial F_{21}}{\partial g_{12}} &= -1 - E[f(s_2) s_2] + E[f'(s_1)] \\
\frac{\partial F_{21}}{\partial g_{21}} &= -1 - E[f'(s_2)] + E[f(s_1) s_1] \\
\frac{\partial F_{22}}{\partial g_{22}} &= -2
\end{aligned} \tag{D.21}$$

Finalmente, podemos escribir la matriz formada por las primeras derivadas como sigue

$$\begin{aligned}
\mathbf{H} &= \begin{bmatrix} \frac{\partial F_{11}}{\partial g_{11}} & \frac{\partial F_{11}}{\partial g_{12}} & \frac{\partial F_{11}}{\partial g_{21}} & \frac{\partial F_{11}}{\partial g_{22}} \\ \frac{\partial F_{12}}{\partial g_{11}} & \frac{\partial F_{12}}{\partial g_{12}} & \frac{\partial F_{12}}{\partial g_{21}} & \frac{\partial F_{12}}{\partial g_{22}} \\ \frac{\partial F_{21}}{\partial g_{11}} & \frac{\partial F_{21}}{\partial g_{12}} & \frac{\partial F_{21}}{\partial g_{21}} & \frac{\partial F_{21}}{\partial g_{22}} \\ \frac{\partial F_{22}}{\partial g_{11}} & \frac{\partial F_{22}}{\partial g_{12}} & \frac{\partial F_{22}}{\partial g_{21}} & \frac{\partial F_{22}}{\partial g_{22}} \end{bmatrix} \\
&= \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -1 - E[f'(s_1)] + E[f(s_2) s_2] & -1 - E[f(s_1) s_1] + E[f'(s_2)] & 0 \\ 0 & -1 - E[f(s_2) s_2] + E[f'(s_1)] & -1 - E[f'(s_2)] + E[f(s_1) s_1] & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}
\end{aligned}$$

D.4 Jacobiano del algoritmo de Bell y Sejnowski

Esta sección recoge las expresiones de las primeras derivadas de la ecuación diferencial ordinaria (ODE) asociada al algoritmo (3.86). Para el caso de dos fuentes y dos salidas, la ecuación diferencial ordinaria asociada al algoritmo es

$$\frac{d\mathbf{G}^T}{dt} = \mathbf{G}^T (\mathbf{I} - E[\mathbf{y}\mathbf{g}^T(\mathbf{y})]) = \mathbf{0} \quad (\text{D.22})$$

da lugar a cuatro ecuaciones

$$\begin{aligned} F_{11} &= \frac{dg_{11}}{dt} = g_{11}(1 - E[y_1g(y_1)]) - g_{21}E[y_2g(y_1)] \\ F_{12} &= \frac{dg_{12}}{dt} = g_{12}(1 - E[y_1g(y_1)]) - g_{22}E[y_2g(y_1)] \\ F_{21} &= \frac{dg_{21}}{dt} = g_{21}(1 - E[y_2g(y_2)]) - g_{11}E[y_1g(y_2)] \\ F_{22} &= \frac{dg_{22}}{dt} = g_{22}(1 - E[y_2g(y_2)]) - g_{12}E[y_1g(y_2)] \end{aligned} \quad (\text{D.23})$$

Derivando estas expresiones respecto a los coeficientes de la matriz \mathbf{G} obtenemos lo siguiente

$$\begin{aligned} \frac{\partial F_{11}}{\partial g_{11}} &= 1 - E[y_1g(y_1)] - g_{11}E[s_1g(y_1)] - g_{11}E[y_1g'(y_1)s_1] - g_{21}E[y_2g'(y_1)s_1] \\ \frac{\partial F_{11}}{\partial g_{12}} &= -g_{11}E[s_2g(y_1)] - g_{11}E[y_1g'(y_1)s_2] - g_{21}E[y_2g'(y_1)s_2] \\ \frac{\partial F_{11}}{\partial g_{21}} &= -E[y_2g(y_1)] - g_{21}E[s_1g(y_1)] \\ \frac{\partial F_{11}}{\partial g_{22}} &= -g_{21}E[s_2g(y_1)] \\ \frac{\partial F_{12}}{\partial g_{11}} &= -g_{12}E[s_1g(y_1)] - g_{12}E[y_1g'(y_1)s_1] - g_{22}E[y_2g'(y_1)s_1] \\ \frac{\partial F_{12}}{\partial g_{12}} &= 1 - E[y_1g(y_1)] - g_{12}E[s_2g(y_1)] - g_{12}E[y_1g'(y_1)s_2] - g_{22}E[y_2g'(y_1)s_2] \\ \frac{\partial F_{12}}{\partial g_{21}} &= -g_{22}E[s_1g(y_1)] \\ \frac{\partial F_{12}}{\partial g_{22}} &= -E[y_2g(y_1)] - g_{22}E[s_2g(y_1)] \\ \frac{\partial F_{21}}{\partial g_{11}} &= -E[y_1g(y_2)] - g_{11}E[s_1g(y_2)] \\ \frac{\partial F_{21}}{\partial g_{12}} &= -g_{11}E[s_2g(y_2)] \\ \frac{\partial F_{21}}{\partial g_{21}} &= 1 - E[y_2g(y_2)] - g_{21}E[s_1g(y_2)] - g_{21}E[y_2g'(y_2)s_1] - g_{11}E[y_1g'(y_2)s_1] \end{aligned}$$

$$\begin{aligned}
\frac{\partial F_{21}}{\partial g_{22}} &= -g_{21}E[s_2g(y_2)] - g_{21}E[y_2g'(y_2)s_2] - g_{11}E[y_1g'(y_2)s_2] \\
\frac{\partial F_{22}}{\partial g_{11}} &= -g_{12}E[s_1g(y_2)] \\
\frac{\partial F_{22}}{\partial g_{12}} &= -E[y_1g(y_2)] - g_{12}E[s_2g(y_2)] \\
\frac{\partial F_{22}}{\partial g_{21}} &= -g_{22}E[s_1g(y_2)] - g_{22}E[g'(y_2)s_1] - g_{12}E[y_1g'(y_2)s_1] \\
\frac{\partial F_{22}}{\partial g_{22}} &= 1 - E[y_2g(y_2)] - g_{22}E[s_2g(y_2)] - g_{22}E[y_2g'(y_2)s_2] - g_{12}E[y_1g'(y_2)s_2]
\end{aligned}$$

En la sección 3.7.1 hemos obtenido que la ODE se anula en el punto donde cada salida extrae una única fuente, $y_1 = \lambda_1 s_1$ y $y_2 = \lambda_2 s_2$, cuando se cumplen las siguientes condiciones

$$\begin{aligned}
\lambda_1 E[s_1g(\lambda_1 s_1)] &= 1 \\
\lambda_2 E[s_2g(\lambda_2 s_2)] &= 1
\end{aligned} \tag{D.24}$$

Utilizando las hipótesis sobre las fuentes (independencia estadística, media cero y potencia unidad) y las expresiones (D.24) se obtiene que las únicas derivadas no-nulas son las siguientes

$$\begin{aligned}
\frac{\partial F_{11}}{\partial g_{11}} &= -1 - \lambda_1^2 E[s_1^2 g'(\lambda_1 s_1)] \\
\frac{\partial F_{12}}{\partial g_{12}} &= -\lambda_2^2 E[g'(\lambda_1 s_1)] \\
\frac{\partial F_{12}}{\partial g_{21}} &= -\lambda_2 E[s_1g(y_1)] = -\frac{\lambda_2}{\lambda_1} \\
\frac{\partial F_{21}}{\partial g_{12}} &= -\lambda_1 E[s_2g(y_2)] = -\frac{\lambda_1}{\lambda_2} \\
\frac{\partial F_{21}}{\partial g_{21}} &= -\lambda_1^2 E[g'(\lambda_2 s_2)] \\
\frac{\partial F_{22}}{\partial g_{22}} &= -1 - \lambda_2^2 E[s_2^2 g'(\lambda_2 s_2)]
\end{aligned} \tag{D.25}$$

Finalmente, podemos escribir la matriz formada por las primeras derivadas como sigue

$$\mathbf{H} = \begin{bmatrix} \frac{\partial F_{11}}{\partial g_{11}} & \frac{\partial F_{11}}{\partial g_{12}} & \frac{\partial F_{11}}{\partial g_{21}} & \frac{\partial F_{11}}{\partial g_{22}} \\ \frac{\partial F_{12}}{\partial g_{11}} & \frac{\partial F_{12}}{\partial g_{12}} & \frac{\partial F_{12}}{\partial g_{21}} & \frac{\partial F_{12}}{\partial g_{22}} \\ \frac{\partial F_{21}}{\partial g_{11}} & \frac{\partial F_{21}}{\partial g_{12}} & \frac{\partial F_{21}}{\partial g_{21}} & \frac{\partial F_{21}}{\partial g_{22}} \\ \frac{\partial F_{22}}{\partial g_{11}} & \frac{\partial F_{22}}{\partial g_{12}} & \frac{\partial F_{22}}{\partial g_{21}} & \frac{\partial F_{22}}{\partial g_{22}} \end{bmatrix}$$

$$= \begin{bmatrix} -1 - \lambda_1^2 \sigma_1 & 0 & 0 & 0 \\ 0 & -\lambda_2^2 m_1 & -\frac{\lambda_2}{\lambda_1} & 0 \\ 0 & -\frac{\lambda_1}{\lambda_2} & -\lambda_1^2 m_2 & 0 \\ 0 & 0 & 0 & -1 - \lambda_2^2 \sigma_2 \end{bmatrix} \quad (\text{D.26})$$

donde

$$\sigma_1 = E[s_1^2 g'(\lambda_1 s_1)] \quad \sigma_2 = E[s_2^2 g'(\lambda_2 s_2)] \quad (\text{D.27})$$

$$m_1 = E[g'(\lambda_1 s_1)] \quad m_2 = E[g'(\lambda_2 s_2)] \quad (\text{D.28})$$

Apéndice E

Puntos estacionarios de los criterios I y II

E.1 Puntos estacionarios del criterio I

En esta sección completaremos el estudio de la naturaleza de los puntos estacionarios de la función de coste (4.9) propuesta en el capítulo 4 para separación ciega de fuentes. Recordemos que en la sección 4.4 hemos expresado la función de coste $J_1(\mathbf{W})$ en términos de la matriz de ganancias \mathbf{G} y hemos determinado las ecuaciones que deben satisfacer sus puntos estacionarios. Además, hemos probado que los puntos donde las fuentes son perfectamente separadas se corresponden con máximos de la función de coste.

A continuación encontraremos las otras soluciones del sistema de ecuaciones (4.35) y determinaremos algunas condiciones suficientes que garantizan que estos puntos no se corresponden con máximos de la función de coste. Un criterio que será de gran utilidad para nuestro objetivo se basa en determinar los signos de los términos en la diagonal de \mathbf{H} y el de los determinantes de las matrices centrales de dimensión 2×2 : el punto no será un máximo si alguno de los términos en la diagonal es positivo o el determinante de alguna matriz central es negativo [72].

Las soluciones indeseadas del sistema de ecuaciones (4.35) pueden ser clasificadas en

(A) Grupo 1:

$$y_n \begin{bmatrix} \overline{0} \\ 0 \ 0 \ \dots \ 0 \\ \vdots \\ 0 \end{bmatrix}$$

s_k

(B) Grupo 2:

$$y_p \begin{bmatrix} x \ \overline{0} \ \dots \ 0 \ 0 \\ 0 \ x_i \ \dots \ 0 \ 0 \\ \vdots \\ 0 \ x_i \ \dots \ 0 \ 0 \end{bmatrix}$$

$s_1 \quad s_q$

(C) Grupo 3:

$$y_p \begin{bmatrix} x \ \dots \ x_i \end{bmatrix}$$

$s_k \quad s_n$

Figura E.1: Clasificación de los puntos estacionarios.

los siguientes tres grupos

- *Grupo 1:* Alguna salida cancela todas las fuentes y alguna fuente es cancelada por todas las salidas, lo que equivale a una matriz \mathbf{G} con al menos una fila y una columna toda ceros. En particular, consideramos la situación mostrada en la figura E.1 (A) donde la salida y_n cancela todas las fuentes y la fuente s_k es cancelada por todas las salidas

$$g_{ni} = 0 \quad i = 1, \dots, N, \quad g_{ik} = 0 \quad i = 1, \dots, N$$

Evaluando las segundas derivadas de la función de coste (4.38) se obtiene el siguiente término de la diagonal de la matriz hessiana

$$\frac{\partial^2 \psi}{\partial g_{nk} \partial g_{nk}^*} = -f'(0) \quad (\text{E.1})$$

Recordemos que una de las premisas de nuestro criterio de separación es la existencia de un único máximo $x > 0$ en las funciones $p_i(x) = |\rho_i|x^2 - f(x)$ y, por tanto, en $x = 0$ se cumple $p'_i(0) = -f'(0) > 0$. Como consecuencia, la matriz hessiana tiene elementos positivos en su diagonal y los puntos en este grupo no son máximos de la función de coste.

- *Grupo 2:* Todas las salidas extraen una única fuente y varias salidas extraen la misma fuente. En este caso cada fila de la matriz \mathbf{G} tiene una única entrada distinta de cero y alguna columna tiene varias entradas no nulas. Consideraremos la matriz \mathbf{G} mostrada en la figura E.1 (B) donde la fuente s_l es extraída por las salidas y_p y y_n

$$g_{pl} \neq 0, \quad g_{nl} \neq 0, \quad g_{pj} = g_{nj} = 0 \quad j \neq l$$

Adicionalmente, como estamos suponiendo que el número de fuentes es igual al número de salidas, debe cumplirse que alguna fuente s_q sea cancelada por todas las salidas, i.e., $g_{iq} = 0$, $i = 1, \dots, N$. Evaluando el sistema (4.35) en este punto obtenemos lo siguiente

$$\begin{aligned} \frac{\partial \psi}{\partial g_{pl}} &= 2|\rho_l|g_{pl}^*|g_{pl}|^2 - g_{pl}^*f'(|g_{pl}|^2) - 2\gamma|\rho_l|g_{pl}^* \sum_{\substack{j=1 \\ j \neq p}}^N |g_{jl}|^2 = 0 \\ \frac{\partial \psi}{\partial g_{nl}} &= 2|\rho_l|g_{nl}^*|g_{nl}|^2 - g_{nl}^*f'(|g_{nl}|^2) - 2\gamma|\rho_l|g_{nl}^* \sum_{\substack{j=1 \\ j \neq n}}^N |g_{jl}|^2 = 0 \end{aligned} \quad (\text{E.2})$$

Obviamente para que las soluciones en este grupo sean puntos estacionarios de la función de coste debe cumplirse lo siguiente

$$\begin{aligned} f'(|g_{pl}|^2) &= 2|\rho_l|(|g_{pl}|^2 - \gamma \sum_{\substack{j=1 \\ j \neq p}}^N |g_{jl}|^2) \\ f'(|g_{nl}|^2) &= 2|\rho_l|(|g_{nl}|^2 - \gamma \sum_{\substack{j=1 \\ j \neq n}}^N |g_{jl}|^2) \end{aligned} \quad (\text{E.3})$$

A continuación demostraremos que la matriz hessiana evaluada en este punto no es definida negativa. Para ello estudiaremos el signo de dos términos en la diagonal de \mathbf{H} obtenidos sustituyendo (E.3) en (4.38)

$$\begin{aligned} \frac{\partial^2 \psi}{\partial g_{pq} \partial g_{pq}^*} &= -f'(|g_{pl}|^2) = -2|\rho_l|(|g_{pl}|^2 - \gamma \sum_{\substack{j=1 \\ j \neq p}}^N |g_{jl}|^2) \\ &= -2|\rho_l|(|g_{pl}|^2 - \gamma |g_{nl}|^2) + 2\gamma |\rho_l| \sum_{\substack{j=1 \\ j \neq p, n}}^N |g_{jl}|^2 \end{aligned} \quad (\text{E.4})$$

$$\begin{aligned} \frac{\partial^2 \psi}{\partial g_{nq} \partial g_{nq}^*} &= -f'(|g_{nl}|^2) = -2|\rho_l|(|g_{nl}|^2 - \gamma \sum_{\substack{j=1 \\ j \neq n}}^N |g_{jl}|^2) \\ &= -2|\rho_l|(|g_{nl}|^2 - \gamma |g_{pl}|^2) + 2\gamma |\rho_l| \sum_{\substack{j=1 \\ j \neq p, n}}^N |g_{jl}|^2 \end{aligned} \quad (\text{E.5})$$

Inspeccionando estas expresiones, podemos concluir que la condición $\gamma > 1$ asegura que alguna de las derivadas (E.4) o (E.5) es positiva. Por tanto, la matriz hessiana contiene entradas positivas en su diagonal y el punto no es un máximo de la función de coste.

- *Grupo 3:* Alguna salida extrae una combinación lineal de varias fuentes, lo cual implica que alguna de las filas de la matriz \mathbf{G} tiene varias entradas no nulas. En particular, asumiremos la situación mostrada en la figura E.1 (C) donde la salida y_p extrae una combinación lineal de las fuentes s_k y s_n ,

$$g_{pk} \neq 0, \quad g_{pn} \neq 0$$

El sistema de ecuaciones (4.35) nos indica que los puntos estacionarios deben cumplir las siguientes igualdades

$$\begin{aligned} \frac{\partial \psi}{\partial g_{pk}} &= 2|\rho_k| g_{pk}^* |g_{pk}|^2 - g_{pk}^* f'(|\mathbf{g}_p|^2) - 2\gamma |\rho_k| g_{pk}^* \sum_{\substack{j=1 \\ j \neq p}}^N |g_{jk}|^2 = 0 \\ \frac{\partial \psi}{\partial g_{pn}} &= 2|\rho_n| g_{pn}^* |g_{pn}|^2 - g_{pn}^* f'(|\mathbf{g}_p|^2) - 2\gamma |\rho_n| g_{pn}^* \sum_{\substack{j=1 \\ j \neq p}}^N |g_{jn}|^2 = 0 \end{aligned} \quad (\text{E.6})$$

Para demostrar que este punto no es un máximo de la función de coste es suficiente con estudiar la siguiente matriz 2×2

$$\mathbf{H}_{pkpn} = \begin{bmatrix} \frac{\partial^2 \psi}{\partial g_{pk} \partial g_{pk}^*} & \frac{\partial^2 \psi}{\partial g_{pk} \partial g_{pn}^*} \\ \frac{\partial^2 \psi}{\partial g_{pk}^* \partial g_{pn}} & \frac{\partial^2 \psi}{\partial g_{pn} \partial g_{pn}^*} \end{bmatrix} \quad (\text{E.7})$$

cuyos elementos se obtienen evaluando (4.38) en el punto que estamos analizando

$$\begin{aligned}
\frac{\partial^2 \psi}{\partial g_{pk} \partial g_{pk}^*} &= \frac{\partial \psi}{\partial g_{pk}} \frac{1}{g_{pk}^*} + |g_{pk}|^2 (2|\rho_k| - f''(\|\mathbf{g}_p\|^2)) \\
&= |g_{pk}|^2 (2|\rho_k| - f''(\|\mathbf{g}_p\|^2)) \\
\frac{\partial^2 \psi}{\partial g_{pk} \partial g_{pn}^*} &= -g_{pk}^* g_{pn} f''(\|\mathbf{g}_p\|^2) \\
\frac{\partial^2 \psi}{\partial g_{pn} \partial g_{pn}^*} &= \frac{\partial \psi}{\partial g_{pn}^*} + |g_{pn}|^2 (2|\rho_n| - f''(\|\mathbf{g}_p\|^2)) \\
&= |g_{pn}|^2 (2|\rho_n| - f''(\|\mathbf{g}_p\|^2))
\end{aligned} \tag{E.8}$$

Sustituyendo en (E.7) resulta una matriz que tiene la siguiente forma

$$\mathbf{H}_{pkpn} = \begin{bmatrix} |g_{pk}|^2 (2|\rho_k| - f''(\|\mathbf{g}_p\|^2)) & -g_{pk}^* g_{pn} f''(\|\mathbf{g}_p\|^2) \\ -g_{pk} g_{pn}^* f''(\|\mathbf{g}_p\|^2) & |g_{pn}|^2 (2|\rho_n| - f''(\|\mathbf{g}_p\|^2)) \end{bmatrix}$$

Puede probarse fácilmente que aquellos puntos donde se cumpla

$$f''(\|\mathbf{g}_p\|^2) < 2|\rho_k| \quad \text{o} \quad f''(\|\mathbf{g}_p\|^2) < 2|\rho_n| \tag{E.9}$$

dan lugar a una matriz \mathbf{H}_{pkpn} con términos positivos en su diagonal y, por tanto, no es definida negativa. Para determinar la naturaleza de \mathbf{H}_{pkpn} en los otros casos, estudiaremos su determinante

$$\begin{aligned}
\det(\mathbf{H}_{pkpn}) &= 2|g_{pk}|^2 |g_{pn}|^2 (2|\rho_k| |\rho_n| - |\rho_k| f''(\|\mathbf{g}_p\|^2) - |\rho_n| f''(\|\mathbf{g}_p\|^2)) \\
&= 2|\rho_k| |\rho_n| |g_{pk}|^2 |g_{pn}|^2 \left(2 - \frac{f''(\|\mathbf{g}_p\|^2)}{|\rho_n|} - \frac{f''(\|\mathbf{g}_p\|^2)}{|\rho_k|} \right)
\end{aligned} \tag{E.10}$$

Es evidente que (E.10) es negativo cuando

$$f''(\|\mathbf{g}_p\|^2) \geq 2|\rho_k| \quad \text{y} \quad f''(\|\mathbf{g}_p\|^2) \geq 2|\rho_n| \tag{E.11}$$

Como consecuencia, \mathbf{H} no es definida negativa y los puntos en este grupo no son máximos de la función de coste.

E.2 Puntos estacionarios del criterio II

El objetivo de esta sección es determinar la naturaleza de los puntos estacionarios de la función de coste (4.11) propuesta en el capítulo 4. Por simplicidad, el estudio será realizado para dos fuentes s_1 y s_2 suponiendo que no existe ruido y que el sistema de separación está formado por dos combinadores lineales cuyos vectores de coeficientes representaremos por \mathbf{w}_1 y \mathbf{w}_2 [21].

Comenzaremos expresando la función de coste J_2 en términos de 4 variables complejas $g_{ij} = \mathbf{w}_i^H \mathbf{a}_j$, que expresan la respuesta del combinador i -ésimo a la fuente s_j . Las salidas de los combinadores pueden escribirse como la suma de las fuentes ponderadas por una cierta ganancia del combinador

$$y_1 = g_{11}s_1 + g_{12}s_2 \quad (\text{E.12})$$

$$y_2 = g_{21}s_1 + g_{22}s_2 \quad (\text{E.13})$$

Al ser las fuentes estadísticamente independientes con potencia unidad, el momento de segundo orden y la curtosis de las salidas son

$$E[|y_i|^2] = \|\mathbf{g}_i\|^2 = |g_{i1}|^2 + |g_{i2}|^2 \quad (\text{E.14})$$

$$K_{y_i} = |g_{i1}|^4 \rho_1 + |g_{i2}|^4 \rho_2 \quad (\text{E.15})$$

donde $\rho_i = K_{s_i}/E^2[|s_i|^2]$ es la curtosis normalizada de s_i . Además, el término cruzado entre dos salidas diferentes es de la forma

$$E[y_1 y_2^*] = g_{11} g_{21}^* + g_{12} g_{22}^* \quad (\text{E.16})$$

Sustituyendo (E.14), (E.15) y (E.16) en (4.11) la función de coste puede expresarse respecto a g_{ij}

$$\begin{aligned} \psi_2(\mathbf{G}) \triangleq J_2(\mathbf{W}) &= |g_{11}|^4 |\rho_1| + |g_{12}|^4 |\rho_2| - f(\|\mathbf{g}_1\|^2) + |g_{21}|^4 |\rho_1| + |g_{22}|^4 |\rho_2| - f(\|\mathbf{g}_2\|^2) \\ &- 2\gamma \left(|g_{11}|^2 |g_{21}|^2 + |g_{12}|^2 |g_{22}|^2 + g_{11} g_{12}^* g_{21}^* g_{22} + g_{11}^* g_{12} g_{21} g_{22}^* \right) \end{aligned}$$

De esta forma, se reduce la dimensión del sistema de ecuaciones que debe resolverse para obtener los puntos estacionarios de la función. Ahora sólo es necesario estudiar un sistema de cuatro ecuaciones

$$\begin{aligned} \frac{\partial \psi_2}{\partial g_{11}} &= 2|g_{11}|^2 g_{11}^* |\rho_1| - g_{11}^* f'(\|\mathbf{g}_1\|^2) - 2\gamma \left(g_{11}^* |g_{21}|^2 + g_{12}^* g_{21}^* g_{22} \right) = 0 \\ \frac{\partial \psi_2}{\partial g_{12}} &= 2|g_{12}|^2 g_{12}^* |\rho_2| - g_{12}^* f'(\|\mathbf{g}_1\|^2) - 2\gamma \left(g_{12}^* |g_{22}|^2 + g_{11}^* g_{21} g_{22}^* \right) = 0 \\ \frac{\partial \psi_2}{\partial g_{21}} &= 2|g_{21}|^2 g_{21}^* |\rho_1| - g_{21}^* f'(\|\mathbf{g}_2\|^2) + 2\gamma \left(g_{21}^* |g_{11}|^2 + g_{11}^* g_{12} g_{22}^* \right) = 0 \\ \frac{\partial \psi_2}{\partial g_{22}} &= 2|g_{22}|^2 g_{22}^* |\rho_2| - g_{22}^* f'(\|\mathbf{g}_2\|^2) - 2\gamma \left(g_{22}^* |g_{12}|^2 + g_{11} g_{12}^* g_{21}^* \right) = 0 \quad (\text{E.17}) \end{aligned}$$

donde $f'(x) = df(x)/dx$ es la primera derivada de $f(x)$.

Por otro lado, el análisis de la naturaleza de los puntos estacionarios puede llevarse a cabo estudiando la matriz hessiana \mathbf{H} formada por las segundas derivadas de ψ_2 respecto a g_{ij} . Estas derivadas son

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} = 4|g_{11}|^2 |\rho_1| - f'(\|\mathbf{g}_1\|^2) - |g_{11}|^2 f''(\|\mathbf{g}_1\|^2) - 2\gamma |g_{21}|^2 \quad (\text{E.18})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{12}^*} = -g_{11}^* g_{12} f''(\|\mathbf{g}_1\|^2) - 2\gamma g_{21}^* g_{22} \quad (\text{E.19})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{21}^*} = -2\gamma g_{11}^* g_{21} - 2\gamma g_{12}^* g_{22} \quad (\text{E.20})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{22}^*} = 0 \quad (\text{E.21})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}} = 2g_{11}^{*2} |\rho_1| - g_{11}^{*2} f''(\|\mathbf{g}_1\|^2) \quad (\text{E.22})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{12}} = -g_{11}^* g_{12}^* f''(\|\mathbf{g}_1\|^2) \quad (\text{E.23})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{21}} = -2\gamma g_{11}^* g_{21}^* \quad (\text{E.24})$$

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{22}} = -2\gamma g_{12}^* g_{21}^* \quad (\text{E.25})$$

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} = 4|g_{12}|^2 |\rho_2| - f'(|g_1|^2) - |g_{12}|^2 f''(\|\mathbf{g}_1\|^2) - 2\gamma |g_{22}|^2 \quad (\text{E.26})$$

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{21}^*} = 0 \quad (\text{E.27})$$

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{22}^*} = -2\gamma g_{12}^* g_{22} - 2\gamma g_{11}^* g_{21} \quad (\text{E.28})$$

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}} = 2g_{12}^{*2} |\rho_2| - g_{12}^{*2} f''(|g_1|^2) \quad (\text{E.29})$$

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{21}} = -2\gamma g_{11}^* g_{22}^* \quad (\text{E.30})$$

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{22}} = -2\gamma g_{12}^* g_{22}^* \quad (\text{E.31})$$

$$\frac{\partial^2 \psi_2}{\partial g_{21} \partial g_{21}^*} = 4|g_{21}|^2 |\rho_1| - f'(\|\mathbf{g}_2\|^2) - |g_{21}|^2 f''(\|\mathbf{g}_2\|^2) - 2\gamma |g_{11}|^2 \quad (\text{E.32})$$

$$\frac{\partial^2 \psi_2}{\partial g_{21} \partial g_{22}^*} = -g_{21}^* g_{22} f''(|g_{21}|^2 + |g_{22}|^2) - 2\gamma g_{11}^* g_{12} \quad (\text{E.33})$$

$$\frac{\partial^2 \psi_2}{\partial g_{21} \partial g_{21}} = 2g_{21}^{*2} |\rho_1| - g_{21}^{*2} f''(\|\mathbf{g}_2\|^2) \quad (\text{E.34})$$

$$\frac{\partial^2 \psi_2}{\partial g_{21} \partial g_{22}} = -g_{21}^* g_{22}^* f''(\|\mathbf{g}_2\|^2) \quad (\text{E.35})$$

$$\frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} = 4|g_{22}|^2 |\rho_2| - f'(\|\mathbf{g}_2\|^2) - |g_{22}|^2 f''(\|\mathbf{g}_2\|^2) - 2\gamma |g_{12}|^2 \quad (\text{E.36})$$

$$\frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}} = 2g_{22}^{*2} |\rho_2| - g_{22}^{*2} f''(\|\mathbf{g}_2\|^2) \quad (\text{E.37})$$

donde $f''(x)$ es la segunda derivada de la función $f(x)$.

El sistema de ecuaciones (E.17) es no lineal y tiene muchas soluciones que clasificaremos en seis grupos

- *Solución deseada:* Cada combinador extrae una fuente diferente. Dentro de este grupo existen dos soluciones

$$g_{11} \neq 0, \quad g_{12} = 0, \quad g_{21} = 0, \quad g_{22} \neq 0 \quad (\text{E.38})$$

$$g_{11} = 0, \quad g_{12} \neq 0, \quad g_{21} \neq 0, \quad g_{22} = 0 \quad (\text{E.39})$$

El estudio para ambas soluciones es similar por lo que sólo presentaremos el del primero. En este punto las primeras derivadas nos indican lo siguiente

$$\frac{1}{g_{11}^*} \frac{\partial \psi_2}{\partial g_{11}} = 2|g_{11}|^2 |\rho_1| - f'(|g_{11}|^2) = 0 \Leftrightarrow p_1'(|g_{11}|^2) = 0 \quad (\text{E.40})$$

$$\frac{1}{g_{12}^*} \frac{\partial \psi_2}{\partial g_{22}} = 2|g_{22}|^2 |\rho_2| + f'(|g_{22}|^2) = 0 \Leftrightarrow p_2'(|g_{22}|^2) = 0 \quad (\text{E.41})$$

donde se ha utilizado el resultado $p_i'(x) = 2|\rho_i|x - f'(x)$. Nótese que $|g_{11}|^2$ coincide con el máximo de $p_1(x)$ y $|g_{22}|^2$ coincide con el máximo de $p_2(x)$.

Para determinar si este punto es un máximo de la función de coste analizaremos el carácter definido negativo de la siguiente matriz

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1111} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{1221} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{1221}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{2222} \end{bmatrix} \quad (\text{E.42})$$

donde

$$\mathbf{H}_{1111} = \begin{bmatrix} \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} & \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} \\ \frac{\partial^2 \psi_2}{\partial g_{11}^* \partial g_{11}^*} & \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} \end{bmatrix}, \quad \mathbf{H}_{1221} = \begin{bmatrix} \frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} & \frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{21}^*} \\ \frac{\partial^2 \psi_2}{\partial g_{12}^* \partial g_{21}^*} & \frac{\partial^2 \psi_2}{\partial g_{21} \partial g_{21}^*} \end{bmatrix}, \quad (\text{E.43})$$

$$\mathbf{H}_{2222} = \begin{bmatrix} \frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} & \frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} \\ \frac{\partial^2 \psi_2}{\partial g_{22}^* \partial g_{22}^*} & \frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} \end{bmatrix} \quad (\text{E.44})$$

Para determinar el carácter de la matriz \mathbf{H} habría que estudiar el signo de los determinantes de todas las matrices superiores izquierdas. Sin embargo, por la forma de \mathbf{H} , basta con determinar si las matrices \mathbf{H}_{1111} , \mathbf{H}_{1221} y \mathbf{H}_{2222} son definidas negativas.

Utilizando (E.40) y $p_1''(x) = 2|\rho_1| - f''(x)$, los términos (E.18) y (E.22) pueden escribirse de la siguiente manera

$$\begin{aligned} \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} &= \frac{1}{g_{11}^*} \frac{\partial \psi_2}{\partial g_{11}} + 2|g_{11}|^2 |\rho_1| - |g_{11}|^2 f''(|g_{11}|^2) \\ &= |g_{11}|^2 p_1''(|g_{11}|^2) \end{aligned} \quad (\text{E.45})$$

$$\begin{aligned} \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}} &= 2g_{11}^{*2} |\rho_1| - g_{11}^{*2} f''(|g_{11}|^2) \\ &= g_{11}^{*2} p_1''(|g_{11}|^2) \end{aligned} \quad (\text{E.46})$$

De esta manera, la matriz \mathbf{H}_{1111} tiene la forma

$$\mathbf{H}_{1111} = \begin{bmatrix} |g_{11}|^2 p_1''(|g_{11}|^2) & g_{11}^{*2} p_1''(|g_{11}|^2) \\ g_{11}^2 p_1''(|g_{11}|^2) & |g_{11}|^2 p_1''(|g_{11}|^2) \end{bmatrix}$$

Siguiendo un procedimiento análogo, podemos escribir la matriz \mathbf{H}_{2222} como

$$\mathbf{H}_{2222} = \begin{bmatrix} |g_{22}|^2 p_2''(|g_{22}|^2) & g_{22}^{*2} p_2''(|g_{22}|^2) \\ g_{22}^2 p_2''(|g_{22}|^2) & |g_{22}|^2 p_2''(|g_{22}|^2) \end{bmatrix}$$

Anteriormente obtuvimos que $|g_{11}|^2$ coincide con el máximo de $p_1(x)$ y que $|g_{22}|^2$ coincide con el máximo de $p_2(x)$. Esto implica que los términos de la diagonal de ambas matrices son negativos. Por otro lado, es inmediato comprobar que el determinante de las matrices es cero. Todo ello nos permite afirmar que las matrices son definidas negativas.

Ahora estudiaremos el carácter de la matriz \mathbf{H}_{1221} . Despejando $f'(\cdot)$ de (E.40) podemos escribir

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} = -f'(|g_{11}|^2) - 2\gamma |g_{22}|^2 = 2|g_{11}|^2 |\rho_1| - 2\gamma |g_{22}|^2 \quad (\text{E.47})$$

De igual modo, despejando $f(\cdot)$ de (E.41) obtenemos

$$\frac{\partial^2 \psi_2}{\partial g_{21} \partial g_{21}^*} = -f'(|g_{22}|^2) - 2\gamma |g_{11}|^2 = 2|g_{22}|^2 |\rho_2| - 2\gamma |g_{11}|^2 \quad (\text{E.48})$$

Utilizando los dos resultados anteriores y (E.30) podemos escribir la matriz \mathbf{H}_{1221} como

$$\mathbf{H}_{1221} = \begin{bmatrix} -2|g_{11}|^2 |\rho_1| - 2\gamma |g_{22}|^2 & -2\gamma g_{11}^* g_{22} \\ -2\gamma g_{11} g_{22} & -2|g_{22}|^2 |\rho_2| - 2\gamma |g_{11}|^2 \end{bmatrix}$$

La condición $\gamma > 0$ es suficiente para asegurar que los términos de la diagonal sean negativos. Además, inspeccionando la forma del determinante de esta matriz

$$\det(\mathbf{H}_{1221}) = 4|g_{11}|^2 |g_{22}|^2 |\rho_1| |\rho_2| + 4\gamma |g_{11}|^4 |\rho_1| + 4\gamma |g_{22}|^4 |\rho_2|$$

se obtiene que la condición $\gamma > 0$ garantiza que el determinante es positivo y, por ello, esta matriz es definida negativa.

Como ya se ha indicado, al ser las submatrices \mathbf{H}_{1111} , \mathbf{H}_{1221} y \mathbf{H}_{2222} definidas positivas se cumple que \mathbf{H} es definida positiva y el punto considerado es un máximo de ψ_2 .

- *Solución indeseada tipo I:* Los combinadores cancelan las dos fuentes, esto es

$$g_{11} = 0, \quad g_{12} = 0, \quad g_{21} = 0, \quad g_{22} = 0$$

Evaluando las segundas derivadas en este punto, se obtiene que los únicos términos distintos de cero de \mathbf{H} están en la diagonal y tienen la forma

$$\frac{\partial^2 \psi_2}{\partial g_{ij} \partial g_{ij}^*} = -f'(0)$$

Por definición $p_j(x) = |\rho_j|x^2 - f(x)$, $j = 1, \dots, L$ y, por ello, se cumple

$$\frac{\partial^2 \psi_2}{\partial g_{ij} \partial g_{ij}^*} = p_j'(0)$$

Al tener $p_j(x)$ un único máximo en $x > 0$, entonces $p_j'(0) > 0$. Por tanto, el hessiano es definido positivo y el punto es un mínimo de la función de coste.

- *Solución indeseada tipo II:* Uno de los combinadores cancela las dos fuentes y el otro extrae una. Nótese que en este caso sólo uno de los combinadores funciona correctamente. Este grupo engloba las siguientes cuatro soluciones

$$g_{11} \neq 0, \quad g_{12} = 0, \quad g_{21} = 0, \quad g_{22} = 0$$

$$g_{11} = 0, \quad g_{12} \neq 0, \quad g_{21} = 0, \quad g_{22} = 0$$

$$g_{11} = 0, \quad g_{12} = 0, \quad g_{21} \neq 0, \quad g_{22} = 0$$

$$g_{11} = 0, \quad g_{12} = 0, \quad g_{21} = 0, \quad g_{22} \neq 0$$

La naturaleza del hessiano para las cuatro soluciones es la misma por lo que nos limitaremos a presentar el análisis de la primera de ellas. Para estudiar la naturaleza de este punto haremos uso de un resultado que afirma que una matriz definida negativa no puede contener elementos positivos en la diagonal [72]. Consideremos el siguiente término de la diagonal de \mathbf{H}

$$\frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} = -f'(0)$$

Siguiendo un análisis similar al presentado en la solución indeseada tipo I se obtiene que

$$\frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} = p_2'(0) > 0$$

de lo que se deduce que \mathbf{H} es no definida positiva y el punto no es un máximo de ψ_2 .

- *Solución indeseada tipo III:* Uno de los combinadores extrae una mezcla de las dos fuentes y el otro cancela las dos fuentes. Dentro de este grupo existen dos soluciones

$$\begin{aligned} g_{11} \neq 0, \quad g_{12} \neq 0, \quad g_{21} = 0, \quad g_{22} = 0 \\ g_{11} = 0, \quad g_{12} = 0, \quad g_{21} \neq 0, \quad g_{22} \neq 0 \end{aligned}$$

El análisis de ambas es similar por lo que sólo presentaremos el primero de ellos. Para esta solución las primeras derivadas implican

$$\begin{aligned} \frac{1}{g_{11}^*} \frac{\partial \psi_2}{\partial g_{11}} &= 2|g_{11}|^2 |\rho_1| - f'(\|\mathbf{g}_1\|^2) = 0 \\ \frac{1}{g_{12}^*} \frac{\partial \psi_2}{\partial g_{12}} &= 2|g_{12}|^2 |\rho_2| - f'(\|\mathbf{g}_1\|^2) = 0 \end{aligned}$$

Para conocer el carácter del hessiano en este punto estudiaremos los determinantes de las matrices superiores izquierdas. La primera matriz está formada por un único elemento

$$\frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} = \frac{1}{g_{11}^*} \frac{\partial \psi_2}{\partial g_{11}} + 2|g_{11}|^2 |\rho_1| - |g_{11}|^2 f''(\|\mathbf{g}_1\|^2) = |g_{11}|^2 p_1''(\|\mathbf{g}_1\|^2) \quad (\text{E.49})$$

donde se ha utilizado el resultado $p_i''(x) = 2|\rho_i| - f''(x)$. Si $p_1''(\|\mathbf{g}_1\|^2) > 0$ entonces (E.49) es positivo y el hessiano tiene un término positivo en su diagonal. Como se ha mencionado anteriormente, esto implica que \mathbf{H} no es definida negativa.

Supongamos ahora que $p_1''(\|\mathbf{g}_1\|^2) < 0$ y estudiemos el determinante de la siguiente matriz superior izquierda de \mathbf{H}

$$\mathbf{H}_{1112} = \begin{bmatrix} \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} & \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{12}^*} \\ \frac{\partial^2 \psi_2}{\partial g_{11}^* \partial g_{12}} & \frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} \end{bmatrix}$$

El segundo término de la diagonal de \mathbf{H} tiene la forma

$$\begin{aligned} \frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} &= \frac{1}{g_{12}^*} \frac{\partial \psi_2}{\partial g_{12}} + 2|g_{12}|^2 |\rho_2| - |g_{12}|^2 f''(\|\mathbf{g}_1\|^2) \\ &= |g_{12}|^2 p_2''(\|\mathbf{g}_1\|^2) \end{aligned} \quad (\text{E.50})$$

Utilizando (E.49), (E.50) y (E.19) la matriz \mathbf{H}_1 puede escribirse como

$$\mathbf{H}_{1112} = \begin{bmatrix} |g_{11}|^2 p_1''(\|\mathbf{g}_1\|^2) & -g_{11}^* g_{12} f''(\|\mathbf{g}_1\|^2) \\ -g_{11} g_{12}^* f''(\|\mathbf{g}_1\|^2) & |g_{12}|^2 p_2''(\|\mathbf{g}_1\|^2) \end{bmatrix}$$

Esta matriz puede descomponerse como una transformación congruente $\mathbf{C}^H \mathbf{B} \mathbf{C}$ siendo las matrices \mathbf{C} y \mathbf{B} las siguientes

$$\mathbf{C} = \begin{bmatrix} g_{11} & 0 \\ 0 & g_{12} \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} p_1''(\|\mathbf{g}_1\|^2) & -f''(\|\mathbf{g}_1\|^2) \\ -f''(\|\mathbf{g}_1\|^2) & p_2''(\|\mathbf{g}_1\|^2) \end{bmatrix}$$

Por la ley de inercia bastará estudiar el carácter de la matriz \mathbf{B} . Las segundas derivadas de $p_i(\cdot)$ evaluadas en el punto $|g_{11}|^2 + |g_{12}|^2$ tienen la forma

$$p_1''(\|\mathbf{g}_1\|^2) = 2|\rho_1| - f''(\|\mathbf{g}_1\|^2) \quad (\text{E.51})$$

$$p_2''(\|\mathbf{g}_1\|^2) = 2|\rho_2| - f''(\|\mathbf{g}_1\|^2) \quad (\text{E.52})$$

Despejando $f''(\cdot)$ de (E.51) se tiene

$$f''(\|\mathbf{g}_1\|^2) = p_1''(\|\mathbf{g}_1\|^2) + 2|\rho_1| \quad (\text{E.53})$$

y sustituyendo en (E.52) resulta

$$p_2''(\|\mathbf{g}_1\|^2) = -p_1''(\|\mathbf{g}_1\|^2) + 2|\rho_1| - 2|\rho_2| \quad (\text{E.54})$$

Utilizando (E.53) y (E.54), la matriz \mathbf{B} puede ser escrita como

$$\mathbf{B} = \begin{bmatrix} p_1''(\|\mathbf{g}_1\|^2) & -p_1''(\|\mathbf{g}_1\|^2) + 2|\rho_1| \\ -p_1''(\|\mathbf{g}_1\|^2) + 2|\rho_1| & -p_1''(\|\mathbf{g}_1\|^2) + 2|\rho_2| - 2|\rho_1| \end{bmatrix}$$

Calculando su determinante se obtiene

$$\det(\mathbf{B}) = 2p_1''(\|\mathbf{g}_1\|^2)(|\rho_1| + |\rho_2|) - 4|\rho_1|^2$$

Recordemos que se está suponiendo que $p_1''(\|\mathbf{g}_1\|^2) < 0$ y, por ello, resulta obvio que el determinante es negativo, $\det(\mathbf{B}) < 0$. Por tanto, el hessiano no es definido negativo y el punto considerado no es máximo de la función ψ_2 .

- *Solución indeseada tipo IV*: Los dos combinadores extraen la misma fuente y ambos cancelan la otra. En este grupo nos encontramos con los siguientes puntos

$$\begin{aligned} g_{11} \neq 0, \quad g_{12} = 0, \quad g_{21} \neq 0, \quad g_{22} = 0 \\ g_{11} = 0, \quad g_{12} \neq 0, \quad g_{21} = 0, \quad g_{22} \neq 0 \end{aligned}$$

El análisis de ambas soluciones es similar por lo que sólo presentaremos el correspondiente a la primera de ellas. En este punto las primeras derivadas implican

$$\begin{aligned} \frac{1}{g_{11}^*} \frac{\partial \psi_2}{\partial g_{11}} &= 2|g_{11}|^2 |\rho_1| - f'(|g_{11}|^2) - 2\gamma |g_{21}|^2 = 0 \\ &\Leftrightarrow f'(|g_{11}|^2) = 2|g_{11}|^2 |\rho_1| - 2\gamma |g_{21}|^2 \end{aligned} \quad (\text{E.55})$$

$$\begin{aligned} \frac{1}{g_{21}^*} \frac{\partial \psi_2}{\partial g_{21}} &= 2|g_{21}|^2 |\rho_1| - f'(|g_{21}|^2) - 2\gamma |g_{21}|^2 = 0 \\ &\Leftrightarrow f'(|g_{11}|^2) = 2|g_{21}|^2 |\rho_1| - 2\gamma |g_{11}|^2 \end{aligned} \quad (\text{E.56})$$

Utilizando estos resultados obtenemos las siguientes expresiones

$$\frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} = -f'(|g_{11}|^2) = 2|g_{11}|^2 |\rho_1| - 2\gamma |g_{21}|^2 \quad (\text{E.57})$$

$$\frac{\partial^2 \psi_2}{\partial g_{22} \partial g_{22}^*} = -f'(|g_{21}|^2) = 2|g_{21}|^2 |\rho_1| - 2\gamma |g_{11}|^2 \quad (\text{E.58})$$

Es inmediato demostrar que la condición $\gamma > |\rho_1|$ es suficiente para que (E.57) ó (E.58) sean positivos. Bajo esta condición el hessiano de ψ_2 tiene elementos positivos en su diagonal y el punto bajo estudio no es un máximo de la función de coste.

- *Solución indeseada tipo V:* Ambos combinadores extraen una combinación de las fuentes, esto es

$$g_{11} \neq 0 \quad g_{12} \neq 0 \quad g_{21} \neq 0 \quad g_{22} \neq 0$$

Para determinar si el hessiano es definido negativo estudiaremos la siguiente submatriz superior izquierda de \mathbf{H}

$$\mathbf{H}_{1112} = \begin{bmatrix} \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} & \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{12}^*} \\ \frac{\partial^2 \psi_2}{\partial g_{12}^* \partial g_{11}} & \frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} \end{bmatrix}$$

Utilizando las segundas derivadas y $p_i''(x) = 2|\rho_i| - f''(x)$ podemos escribir los términos de la diagonal como

$$\begin{aligned} \frac{\partial^2 \psi_2}{\partial g_{11} \partial g_{11}^*} &= \frac{1}{g_{11}^*} \frac{\partial \psi_2}{\partial g_{11}} + 2|g_{11}|^2 |\rho_1| - |g_{11}|^2 f''(|\mathbf{g}_1|^2) + 2\gamma \frac{g_{11} g_{12}^* g_{21}^* g_{22}}{|g_{11}|^2} \\ &= |g_{11}|^2 p_1''(|\mathbf{g}_1|^2) + 2\gamma \frac{g_{11} g_{12}^* g_{21}^* g_{22}}{|g_{11}|^2} \end{aligned} \quad (\text{E.59})$$

$$\begin{aligned} \frac{\partial^2 \psi_2}{\partial g_{12} \partial g_{12}^*} &= \frac{1}{g_{12}^*} \frac{\partial \psi_2}{\partial g_{12}} + 2|g_{12}|^2 |\rho_2| - |g_{12}|^2 f''(|\mathbf{g}_1|^2) + 2\gamma \frac{g_{11}^* g_{12} g_{21} g_{22}^*}{|g_{12}|^2} \\ &= |g_{12}|^2 p_2''(|\mathbf{g}_1|^2) + 2\gamma \frac{g_{11}^* g_{12} g_{21} g_{22}^*}{|g_{12}|^2} \end{aligned} \quad (\text{E.60})$$

Los términos de la diagonal de una matriz hessiana son reales [72] por tanto $g_{11}g_{12}^*g_{21}^*g_{22} = g_{11}^*g_{12}g_{21}g_{22}^* = |g_{11}||g_{12}||g_{21}||g_{22}|$ es real. Ahora, utilizando (E.59), (E.60) y (E.19), y realizando algunas operaciones sencillas, la matriz \mathbf{H}_{1112} puede ser escrita de la siguiente forma

$$\mathbf{H}_{1112} = \begin{bmatrix} |g_{11}|^2 p_1''(|\mathbf{g}_1|^2) & -g_{11}^* g_{12} f''(|\mathbf{g}_1|^2) \\ -g_{11} g_{12}^* f''(|\mathbf{g}_1|^2) & |g_{12}|^2 p_2''(|\mathbf{g}_1|^2) \end{bmatrix} \\ + 2\gamma \begin{bmatrix} \frac{|g_{11}||g_{12}||g_{21}||g_{22}|}{|g_{11}|^2} & -g_{21}^* g_{22} \\ -g_{21} g_{22}^* & \frac{|g_{11}||g_{12}||g_{21}||g_{22}|}{|g_{12}|^2} \end{bmatrix}$$

La primera matriz a la derecha de la igualdad es igual a la presentada en el estudio de la solución indeseada tipo III donde se ha probado que es no definida negativa. Por otro lado, la condición $\gamma > 0$ garantiza que la segunda matriz es definida positiva ya que contiene elementos negativos en su diagonal y su determinante es igual a cero. Como consecuencia, la matriz \mathbf{H}_{1112} no es definida negativa y el punto que estamos analizando no es máximo de ψ_2 .

Del análisis anterior se obtiene que los máximos de la función de coste ψ_2 coinciden con los estados en los que cada combinador del sistema de separación extrae una fuente diferente. Además, también se deduce que $\gamma > \max(\rho_i)$, $i = 1, \dots, L$ garantiza que la función no contiene máximos no deseados.

Apéndice F

Puntos estacionarios del criterio I generalizado

En este apéndice estudiaremos la naturaleza de los puntos estacionarios indeseados de la función de coste (5.8) propuesta en el capítulo 5 para separación ciega de fuentes a partir de mezclas convolutivas. Previamente, en la sección 5.4, hemos expresado la función de coste $J_{1G}(\mathbf{W})$ en términos de las ganancias \mathbf{G} y hemos determinado las primeras derivadas.

Las soluciones indeseadas del sistema de ecuaciones (5.30) pueden ser clasificadas en tres grupos análogos a los estudiados para el caso de mezclas lineales instantáneas

- *Grupo 1:* Alguna salida cancela todas las fuentes y alguna fuente es cancelada por todas las salidas. Consideraremos que la salida y_p cancela todas las fuentes y que la fuente s_l es cancelada por todas las salidas, i.e.,

$$g_{pi}(n) = 0 \quad i = 1, \dots, N, \quad n \in (-\infty, \infty) \quad g_{il}(n) = 0 \quad i = 1, \dots, N, \quad n \in (-\infty, \infty)$$

Evaluando las segundas derivadas de la función de coste (5.33) se obtiene el siguiente término de la diagonal de la matriz hessiana

$$\frac{\partial^2 \psi_G}{\partial g_{pl}(n) \partial g_{pl}^*(n)} = -f'(0) \quad (\text{F.1})$$

Recordemos que una de las premisas de nuestro criterio es la existencia de un único máximo $x > 0$ en las funciones $p_i(x) = |\rho_i|x^2 - f'(x)$. Por tanto, en $x = 0$ se cumple que $p'_i(0) = -f'(0) > 0$. Como consecuencia la matriz hessiana tiene elementos positivos en su diagonal y los puntos en este grupo no son máximos de la función de coste.

- *Grupo 2:* Todas las salidas extraen una única fuente y varias salidas extraen la misma fuente. Consideraremos que la fuente s_l es extraída por las salidas y_p y y_m , i.e.,

$$g_{pl}(n_1) \neq 0, \quad g_{ml}(n_2) \neq 0$$

y

$$\begin{aligned} g_{pj}(n) &= 0, \quad j = 1, \dots, N, \quad n \in (-\infty, \infty) \quad j \neq l \\ g_{mj}(n) &= 0, \quad j = 1, \dots, N, \quad n \in (-\infty, \infty) \quad j \neq l \end{aligned}$$

La diferencia entre este grupo de soluciones y el *grupo 2* del análisis presentado en el apéndice E radica en que ahora puede ocurrir que las salidas extraigan una combinación lineal de los símbolos de las fuentes.

Evaluando el sistema de ecuaciones (5.30) en este punto obtenemos lo siguiente

$$\begin{aligned} \frac{\partial \psi_G}{\partial g_{pl}(n_1)} &= 2|\rho_l|g_{pl}^*(n_1)|g_{pl}(n_1)|^2 - g_{pl}^*(n_1)f'(\|\mathbf{g}_p\|^2) \\ &- 2\gamma|\rho_l|g_{pl}^*(n_1)\sum_{j=1}^N\sum_{k_1} |g_{jl}(k_1)|^2 = 0 \end{aligned} \quad (\text{F.2})$$

$$\begin{aligned} \frac{\partial \psi_G}{\partial g_{ml}(n_2)} &= 2|\rho_l|g_{ml}^*(n_2)|g_{ml}(n_2)|^2 - g_{ml}^*(n_2)f'(\|\mathbf{g}_m\|^2) \\ &- 2\gamma|\rho_l|g_{ml}^*(n_2)\sum_{j=1}^N\sum_{k_2} |g_{jl}(k_2)|^2 = 0 \end{aligned} \quad (\text{F.3})$$

donde el sumatorio con subíndice k_1 en (F.2) engloba a todos los retardos con que las salidas y_j extraen la fuente s_l excepto $y_p(n_1)$. Por otro lado el sumatorio con subíndice k_2 engloba a todos los retardos con que la fuente s_l es extraída por las salidas y_j excepto $g_{ml}(n_2)$.

Obviamente para que las soluciones en este grupo sean puntos estacionarios de la función de coste debe cumplirse lo siguiente

$$\begin{aligned} f'(\|\mathbf{g}_p\|^2) &= 2|\rho_l|(|g_{pl}(n_1)|^2 - \gamma\sum_{j=1}^N\sum_{k_1} |g_{jl}(k_1)|^2) \\ f'(\|\mathbf{g}_m\|^2) &= 2|\rho_l|(|g_{ml}(n_1)|^2 - \gamma\sum_{j=1}^N\sum_{k_2} |g_{jl}(k_2)|^2) \end{aligned} \quad (\text{F.4})$$

Adicionalmente, como estamos suponiendo que el número de fuentes es igual al número de salidas debe cumplirse que alguna fuente s_q sea cancelada por todas las salidas, i.e.,

$$g_{iq}(n) = 0, \quad i = 1, \dots, N \quad n \in (-\infty, \infty)$$

Estudiemos los siguientes términos de la diagonal de la matriz hessiana

$$\frac{\partial^2 \psi_G}{\partial g_{pq}(m_1)\partial g_{pq}^*(m_1)} = -f'(\|\mathbf{g}_p\|^2) = -2|\rho_l|(|g_{pl}(n_1)|^2 - \gamma\sum_{j=1}^N\sum_{k_1} |g_{jl}(k_1)|^2)$$

$$= -2|\rho_l|(|g_{pl}(n_1)|^2 - \gamma|g_{ml}(n_2)|^2) + 2\gamma|\rho_l| \sum_{j=1}^N \sum_{k_3} |g_{jl}(k_3)|^2 \quad (\text{F.5})$$

$$\begin{aligned} \frac{\partial^2 \psi_G}{\partial g_{mq}(n_2) \partial g_{mq}^*(n_2)} &= -f'(\|\mathbf{g}_m\|^2) = -2|\rho_l|(|g_{ml}(n_2)|^2 - \gamma \sum_{j=1}^N \sum_{k_2} |g_{jl}(k_2)|^2) \\ &= -2|\rho_l|(|g_{ml}(n_2)|^2 - \gamma|g_{pl}(n_1)|^2) + 2\gamma|\rho_l| \sum_{j=1}^N \sum_{k_3} |g_{jl}(k_3)|^2 \end{aligned} \quad (\text{F.6})$$

Donde el sumatorio con índice k_3 abarca a todas las ganancias con las que es extraída la fuente s_l excepto $g_{pl}(n_1)$ y $g_{ml}(n_2)$. Inspeccionando estas expresiones, podemos concluir que la condición $\gamma > 1$ asegura que alguno de los términos (F.5) o (F.6) en la diagonal del hessiano es positivo. Como consecuencia, el punto no es un máximo de la función de coste.

- *Grupo 3:* Alguna salida extrae una combinación lineal de varias fuentes (o de la misma fuente). En particular, asumiremos que la salida y_p extrae una combinación lineal de las fuentes s_l y s_m ,

$$\text{Existe } n_1, n_2 \text{ tal que } g_{pl}(n_1) \neq 0, \quad g_{pm}(n_2) \neq 0$$

A diferencia del caso equivalente para mezclas instantáneas, ahora puede ocurrir que las fuentes sean extraídas por y_p con diferentes retardos.

El sistema de ecuaciones (5.30) nos indica que los puntos estacionarios deben cumplir las siguientes igualdades

$$\begin{aligned} \frac{\partial \psi_G}{\partial g_{pl}(n_1)} &= 2|\rho_l|g_{pl}^*(n_1)|g_{pl}(n_1)|^2 - g_{pl}^*(n_1)f'(\|\mathbf{g}_p\|^2) \\ &- 2\gamma|\rho_l|g_{pl}^*(n_1) \sum_{j=1}^N \sum_{k_1} |g_{jl}(k_1)|^2 = 0 \end{aligned} \quad (\text{F.7})$$

$$\begin{aligned} \frac{\partial \psi_G}{\partial g_{pm}(n_2)} &= 2|\rho_m|g_{pm}^*(n_2)|g_{pm}(n_2)|^2 - g_{pm}^*(n_2)f'(\|\mathbf{g}_p\|^2) \\ &- 2\gamma|\rho_m|g_{pm}^*(n_2) \sum_{j=1}^N \sum_{k_2} |g_{jn}(k_2)|^2 = 0 \end{aligned} \quad (\text{F.8})$$

donde el sumatorio con índice k_1 en (F.7) abarca a todas las ganancias con que las salidas $y_j, j = 1, \dots, N$, extraen la fuente s_l excepto $g_{pl}(n_1)$ y el sumatorio con índice k_2 en (F.8) abarca a todas las ganancias con que las salidas $y_j, j = 1, \dots, N$, extraen la fuente s_m excepto $g_{pl}(n_2)$.

Para demostrar que este punto no es un máximo de la función de coste es suficiente con estudiar la siguiente matriz 2×2

$$\mathbf{H}_{plpn}(n_1, n_2) = \begin{bmatrix} \frac{\partial^2 \psi_G}{\partial g_{pl}(n_1) \partial g_{pl}^*(n_1)} & \frac{\partial^2 \psi_G}{\partial g_{pl}(n_1) \partial g_{pm}^*(n_2)} \\ \frac{\partial^2 \psi_G}{\partial g_{pl}^*(n_1) \partial g_{pm}(n_2)} & \frac{\partial^2 \psi_G}{\partial g_{pm}(n_2) \partial g_{pm}^*(n_2)} \end{bmatrix} \quad (\text{F.9})$$

cuyos elementos se obtienen evaluando (5.33) en el punto que estamos analizando

$$\begin{aligned} \frac{\partial^2 \psi_G}{\partial g_{pl}(n_1) \partial g_{pl}^*(n_1)} &= \frac{\partial \psi_G}{\partial g_{pl}(n_1)} \frac{1}{g_{pl}^*(n_1)} + |g_{pl}(n_1)|^2 (2|\rho_l| - f''(\|\mathbf{g}_p\|^2)) \\ &= |g_{pl}(n_1)|^2 (2|\rho_l| - f''(\|\mathbf{g}_p\|^2)) \\ \frac{\partial^2 \psi_G}{\partial g_{pl}(n_1) \partial g_{pm}^*(n_2)} &= -g_{pl}^*(n_1) g_{pm}(n_2) f''(\|\mathbf{g}_p\|^2) \\ \frac{\partial^2 \psi_G}{\partial g_{pm}(n_2) \partial g_{pm}^*(n_2)} &= \frac{\partial \psi_G}{\partial g_{pm}^*(n_2)} \frac{1}{g_{pm}^*(n_2)} + |g_{pm}(n_2)|^2 (2|\rho_m| - f''(\|\mathbf{g}_p\|^2)) \\ &= |g_{pm}(n_2)|^2 (2|\rho_m| - f''(\|\mathbf{g}_p\|^2)) \end{aligned} \quad (\text{F.10})$$

Sustituyendo en (F.9) resulta una matriz que tiene la siguiente forma

$$\mathbf{H}_{plpm}(n_1, n_2) = \begin{bmatrix} |g_{pl}(n_1)|^2 (2|\rho_l| - f''(\|\mathbf{g}_p\|^2)) & -g_{pl}^*(n_1) g_{pm}(n_2) f''(\|\mathbf{g}_p\|^2) \\ -g_{pl}(n_1) g_{pm}^*(n_2) f''(\|\mathbf{g}_p\|^2) & |g_{pm}(n_2)|^2 (2|\rho_m| - f''(\|\mathbf{g}_p\|^2)) \end{bmatrix}$$

Aquellos puntos donde se cumpla

$$f''(\|\mathbf{g}_p\|^2) < 2|\rho_l| \quad \text{o} \quad f''(\|\mathbf{g}_p\|^2) < 2|\rho_m| \quad (\text{F.11})$$

dan lugar a una matriz $\mathbf{H}_{plpm}(n_1, n_2)$ con términos positivos en su diagonal y, por tanto, no es definida negativa. Para determinar la naturaleza de $\mathbf{H}_{plpm}(n_1, n_2)$ en los otros casos, estudiaremos su determinante

$$\begin{aligned} \det(\mathbf{H}_{plpm}(n_1, n_2)) &= 2|g_{pl}(n_1)|^2 |g_{pm}(n_2)|^2 (2|\rho_l| |\rho_m| - |\rho_l| f''(\|\mathbf{g}_p\|^2) - |\rho_l| f''(\|\mathbf{g}_p\|^2)) \\ &= 2|\rho_l| |\rho_m| |g_{pl}(n_1)|^2 |g_{pm}(n_2)|^2 \left(2 - \frac{f''(\|\mathbf{g}_p\|^2)}{|\rho_l|} - \frac{f''(\|\mathbf{g}_p\|^2)}{|\rho_m|} \right) \end{aligned} \quad (\text{F.12})$$

Es evidente que el determinante (F.12) es negativo cuando

$$f''(\|\mathbf{g}_p\|^2) \geq 2|\rho_l| \quad \text{y} \quad f''(\|\mathbf{g}_p\|^2) \geq 2|\rho_m| \quad (\text{F.13})$$

Como consecuencia, \mathbf{H} no es definida negativa y los puntos en este grupos no son máximos de la función de coste.

Apéndice G

Notación y abreviaturas empleadas

G.1 Notación

Representación	Significado	Ejemplos
Letra minúscula normal	Variable escalar o entera	n, t, i, λ
Letra minúscula negrilla	Vector columna	$\mathbf{x}, \mathbf{s}, \mathbf{y}$
Letra mayúscula normal	Constante entera	M, N
Letra mayúscula negrilla	Matriz	$\mathbf{A}, \mathbf{\Delta}$
Subíndices	Elemento de un vector o columna de una matriz	s_i, \mathbf{x}_i
Superíndice *	Conjugado de un complejo	w_i^*
Superíndice T	Traspuesto	$\mathbf{w}^T, \mathbf{W}^T$
Superíndice H	Traspuesto conjugado	$\mathbf{w}^H, \mathbf{W}^H$
Superíndice -1	Inversa de una matriz	\mathbf{A}^{-1}
∇	Operador gradiente	$\nabla_{\mathbf{w}}, \nabla_{\mathbf{W}}$
$\text{tr}(\cdot)$	Traza de una matriz	$\text{tr}(\mathbf{W})$
$\det(\cdot)$	Determinante de una matriz	$\det(\mathbf{W})$
$\text{diag}(\cdot)$	Diagonal de una matriz	$\text{diag}(\mathbf{A})$
$E[\cdot]$	Operador esperanza	$E[s_i]$
$\text{cum}(\cdot)$	Cumulante de una señal	$\text{cum}(y_i, y_i^*, y_i, y_i^*)$

G.2 Abreviaturas

AGC	Adaptive Gain Control
CDMA	Code Division Multiple Access
CM	Constant Modulus
CMA	Constant Modulus Algorithm
f.d.p.	Función densidad de probabilidad
DSSS	Direct-Sequence Spread-Spectrum
EASI	Equivariant Adaptive Separation via Independence
GHA	Generalized Hebbian Algorithm
HOS	Higher Order Statistics
ICA	Independent Component Analysis
IIR	Infinite Impulse Response
LMS	Least Mean Squares
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MMSE	Minimum Mean Square Error
ODE	Ordinary Differential Equation
PCA	Principal Component Analysis
QAM	Quadrature Amplitude Modulation
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio

Bibliografía

- [1] S.-I. Amari. Gradient learning in structured parameter spaces: Adaptive blind separation of signal sources. In *Proc. WCNN'96*, pages 951–956. San Diego, 1996.
- [2] S.-I. Amari, T.-P. Chen, and A. Cichocki. Stability analysis of learning algorithms for blind source separation. *Neural Networks*, vol. 10(no. 8):pp. 1345–1351, August 1997.
- [3] R. Barbolla and P. Sanz. *Algebra Lineal y Teoría de Matrices*. Prentice Hall, Madrid, 1998.
- [4] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, vol. 7(no. 6):pp. 1129–1159, November 1995.
- [5] A. Belouchrani, K. Abed-Meraim, and J.-F. Cardoso. A blind source separation technique using second-order statistics. *IEEE Trans. Signal Processing*, vol. 45(no. 2):pp. 434–444, February 1997.
- [6] A. Benveniste, M. Métivier, and P. Priourent. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, New York, 1990.
- [7] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proc. F and G .*, vol. 130:pp. 11–16, February 1983.
- [8] J. Cadzow. Blind deconvolution via cumulant extrema. *IEEE Signal Processing Magazine*, pages pp. 24–42, May 1996.
- [9] X. R. Cao and R. W. Liu. General approach to blind source separation. *IEEE Transactions on Signal Processing*, vol. 44(no. 3):pp. 562–571, March 1996.
- [10] J.-F. Cardoso. Statistical principles of source separation. In *Proc. SYSID'97, Fukuoka, Japan*, pages 1837–1840, 1997.
- [11] J.-F. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, vol. 4(no. 4):pp. 112–114, April 1997.

- [12] J.-F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of IEEE*, vol. 86(no. 10):pp. 2009–2025, October 1998.
- [13] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, vol. 44(no. 12):pp. 3017–3030, December 1996.
- [14] L. Castedo, C. Escudero, and A. Dapena. A blind signal separation method for multiuser communications. *IEEE Transactions on Signal Processing*, vol. 45(no. 5):pp. 1343–1348, May 1997.
- [15] L. Castedo and A. Figueiras-Vidal. An adaptive beamforming technique based on cyclostationary signal properties. *IEEE Trans. on Signal Processing*, vol. 43(no. 7):pp. 1637–1650, July 1995.
- [16] L. Castedo and O. Macchi. Maximizing the information transfer for adaptive unsupervised source separation. In *Proc. SPAWC'97*, pages 65–68. Paris, France, April 1997.
- [17] A. Cichocki and R. Undehauen. Robust neural networks with on-line learning for blind identification and blind separation of source. *IEEE Transactions on Circuits and Systems*, vol. 43(no. 11):pp. 894–906, November 1996.
- [18] P. Comon. Independent component analysis, a new concept? *Signal Processing*, vol. 36:pp. 287–314, 1994.
- [19] P. Comon. Blind separation of sources, part II: Problems statement. *Signal Processing*, vol. 24:pp. 11–20, July 1991.
- [20] P. Comon. Contrasts for multichannel blind deconvolution. *IEEE Signal Processing Letters*, vol. 3(no. 7):pp. 209–211, July 1996.
- [21] A. Dapena. *Algoritmos no Supervisados para la Separación de Señales Mezcladas Linealmente*. Tesis de Licenciatura de la Facultad de Informática. Universidad de La Coruña, February 1997.
- [22] A. Dapena and L. Castedo. Stochastic gradient adaptive algorithms for blind source separation. *To appear in Signal Processing*, 1999.
- [23] A. Dapena, L. Castedo, and C. Escudero. An unconstrained single stage criterion for blind source separation. In *Proc. ICASSP96*, pages 2706–2709. Atlanta, USA, April 1996.
- [24] A. Dapena, D. Iglesia, and L. Castedo. An unconstrained criterion for blind source separation of convolutive mixtures. In *Proc. ICSPC'98*, pages 404–407. Gran Canaria, Spain, February 1998.

- [25] A. Dapena, D. Iglesia, and L. Castedo. Blind separation of cyclostationary signals. In *Proc. EUSIPCO98*, volume III, pages 1813–1816. Rhodes, Greece, September 1998.
- [26] A. Dapena, C. Mejuto, and L. Castedo. Convergencia de algoritmos de separación de fuentes basados en contrastes. In *Proc. URSI97*, volume I, pages 487–490. Bilbao, Spain, September 1997.
- [27] A. Dapena, C. Mejuto, and L. Castedo. Análisis de estabilidad de algoritmos de separación de fuentes. In *Proc. URSI98*, pages 425–426. Pamplona, Spain, September 1998.
- [28] J. L. de la Fuente O'Connor. *Tecnologías Computacionales para Sistemas de Ecuaciones, Optimización Lineal y Entera*. Editorial Rerverté, 1993.
- [29] G. Deco and W. Brauer. Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. *Neural Networks*, vol. 8(no. 4):pp. 525–535, April 1995.
- [30] G. Deco and D. Obradovic. *An Information-Theoretic Approach to Neural Computing*. Springer-Verlang, 1996.
- [31] Y. Deville. A unified stability analysis of the Héroult-Jutten source separation neural networks. *Signal Processing*, vol. 51(no. 3):pp. 229–233, June 1996.
- [32] K. Diamantaras. Second order hebbian neural networks and blind source separation. In *Proc. EUSIPCO'98*, volume III, pages 1317–1320. Rhodes, Greece, September 1998.
- [33] K. I. Diamantaras and S. Y. Kung. *Principal Component Neural Networks Theory and Applications*. John Wiley & Sons Inc., 1996.
- [34] Z. Ding, C. R. Johnson, and R. Kennedy. Global convergence issues with linear blind adaptive equalizers. In *Blind Deconvolution*, pages 60–190. Prentice Hall, Simon Haykin editor, 1994.
- [35] S. C. Douglas and A. Cichocki. Neural networks for blind decorrelation of signals. *IEEE Trans. on Signal Processing*, vol. 45(no. 11):pp. 2829–2842, November 1997.
- [36] C. Escudero. *Algoritmos Adaptativos Ciegos para Igualación y Separación de Señales*. Tesis Doctoral de la Facultad de Informática. Universidad de La Coruña, March 1998.
- [37] W. A. Gardner. An introduction to cyclostationary signals. In *Cyclostationarity: In communications and Signal Processing*, pages 1–90. IEEE Press, 1994.

- [38] D. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Transactions on Communications*, vol. 28:pp. 1867–1875, November 1980.
- [39] L. J. Griffiths. A simple adaptive algorithm for real-time processing in antenna arrays. *Proceedings IEEE*, vol. 57:pp. 1696–1704, 1969.
- [40] S. Haykin. *Neural Networks A Comprehensive Foundation*. Macmillan College Publishing Company, New York, 1994.
- [41] S. Haykin. *Adaptive Filter Theory*. 3rd. ed., Prentice Hall, 1996.
- [42] M. W. Hirsch and S. Smale. *Ecuaciones diferenciales, sistemas dinámicos y álgebra lineal*. Alianza, 1983.
- [43] A. Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, vol. 9:pp. 1483–1492, 1997.
- [44] A. Hyvarinen and E. Oja. Independent component analysis by general nonlinear hebbian-like learning rules. *Signal Processing*, vol. 64:pp. 301–313, 1998.
- [45] D. Iglesia, A. Dapena, and L. Castedo. Using fractional cyclic moments of CPFSK signals in blind adaptive beamforming. In *Proc. ICASSP97*, pages 3441–3445. Munich, Germany, 1997.
- [46] Y. Inouye and T. Habe. Multichannel blind equalization using second- and fourth-order cumulants. In *Proc. HOS'95*, pages 96–100. Girona, Spain, June 1995.
- [47] Y. Inouye and T. Sato. Iterative algorithm based on multistage criteria for multichannel blind deconvolution. *Submitted to IEEE Trans. on Signal Processing*, 1997.
- [48] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, vol. 24:pp. 1–10, July 1991.
- [49] J. Karhunen. Neural approaches to independent component analysis and source separation. In *Proc. ESANN'96*, pages 249–266. Bruges, Belgium, April 1996.
- [50] J. Karhunen and J. Joutsensalo. Representation and separation of signals using nonlinear PCA type learning. *Neural Networks*, vol. 7(no. 1):pp. 113–127, 1994.
- [51] O. Macchi and C. Jutten. Adaptive unsupervised algorithms. Technical report, Groupement de Recherche Traitement du Signal et Images, CNRS, 1995.
- [52] A. Manseur and C. Jutten. Fourth-order criteria for blind source separation. *IEEE Transactions on Signal Processing*, vol. 43(no. 8):pp. 2022–2025, August 1995.

- [53] C. Mejuto and L. Castedo. A neural network approach to blind source separation. In *Proc. Neural Networks for Signal Processing VII*, pages 486–595. Florida, USA, September 1997.
- [54] E. Moreau and O. Macchi. Complex self-adaptive adaptive algorithms for source separation based on higher order contrasts. *Signal Processing VII: Theories and Applications*, pages 1157–1160, 1994.
- [55] E. Moreau and O. Macchi. High-order contrasts for self-adaptive source separation. *International Journal on Adaptive Control and Signal Processing*, vol. 10:pp. 19–46, 1996.
- [56] E. Moreau and O. Macchi. Self-adaptive source separation, part II: Comparison of direct feedback and mixed neural network. *IEEE Transactions on Signal Processing*, vol. 46(no. 1):pp. 39–50, January 1998.
- [57] E. Moreau and J.-C. Pesquet. Generalized contrasts for multichannel blind deconvolution of linear system. *IEEE Signal Processing Letter*, vol. 4(no. 6):pp. 182–183, June 1997.
- [58] E. Moreau and N. Thirion. Non symmetrical contrasts for source separation. *submitted to IEEE Transactions on Signal Processing*, December 1997.
- [59] J.-P. Nadal and N. Parga. Non linear neurons in the low noise limit: a factorial code maximizes information transfer. *Network*, vol. 5:pp. 565–581, 1994.
- [60] A. Nandi and V. Zarzoso. Fourth-order cumulant based blind source separation. *IEEE Signal Processing Letters*, vol. 3(no. 12):pp. 312–314, December 1996.
- [61] H. L. Nguyen-Thi and C. Jutten. Blind source separation for convolutive mixtures. *Signal Processing*, vol. 45:pp. 209–229, 1995.
- [62] C. Nikias and J. Mendel. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, pages 10–35, July 1993.
- [63] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, San Diego, 1970.
- [64] C. B. Papadias and A. J. Paulraj. A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays. *IEEE Signal Processing Letters*, vol. 4(no. 6):pp. 178–182, June 1997.
- [65] A. Papoulis. *Probability Random Variables and Stochastic Processes*. McGraw-Hill, New York, 1984.
- [66] S. U. Pillai. *Array Signal Processing*. Springer-Verlag, New York, 1989.

- [67] J. G. Proakis. *Digital Communications*. McGraw-Hill International, New York, 1995.
- [68] O. Shalvi and E. Weinstein. Universal methods for blind deconvolution. In *Blind Deconvolution*, pages 121–180. Prentice Hall, Simon Haykin editor, 1994.
- [69] O. Shalvi and E. Weinstein. New criteria for blind deconvolution of nonminimum phase system. *IEEE Transactions on Information Theory*, vol. 36(no. 2):pp. 312–321, March 1990.
- [70] J. J. Shynk. Adaptive filtering. In *The Handbook of Brain Theory and Neural Networks*, pages 74–78. Massachusetts Institute of Technology, Michael A. Arbib editor, 1995.
- [71] E. Sorouchyari. Blind separation of sources, part III: Stability analysis. *Signal Processing*, vol. 24:pp. 21–29, July 1991.
- [72] G. Strang. *Algebra Lineal y sus Aplicaciones*. Addison-Wesley Iberoamericana, 1982.
- [73] C. W. Therrien. *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall, 1992.
- [74] S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998.