

**MODELOS MATEMÁTICOS Y NUEVOS ALGORITMOS
PARA EL PROCESAMIENTO DE IMÁGENES**

Autor: CESAR F. CAIAFA

Directora: Araceli N. Proto
Co-director: Christian Barbier

Tesis de Doctorado

FACULTAD DE INGENIERIA, UNIVERSIDAD DE BUENOS AIRES
BUENOS AIRES, ARGENTINA

ÍNDICE GENERAL

CAPÍTULO 1. INTRODUCCIÓN	15
CAPÍTULO 2. ASPECTOS TEÓRICOS: RESULTADOS PREVIOS	19
2.1. Métodos estadísticos y Análisis de Componentes Independientes	19
2.1.1. Análisis de Componentes Principales - <i>PCA</i> (o <i>KLT</i>)	19
2.1.2. <i>Projection Pursuit - PP</i>	23
2.1.3. Separación ciega de fuentes independientes (<i>ICA</i>)	24
2.1.4. Resultados teóricos en <i>ICA</i> : conexiones con la Teoría de la Información de Shannon, <i>PP</i> y <i>KLT</i>	25
2.1.5. Algoritmos <i>ICA</i>	29
2.2. Bases ortonormales óptimas en \mathbb{R}^n	31
2.2.1. Aproximación y clasificación de vectores en \mathbb{R}^n	31
2.2.2. Cálculo de las distancias DASP y DESP _j	32
2.2.3. Bases ortonormales	34
2.2.4. Métodos de ortonormalización de vectores	36
2.2.5. Descomposición de la energía de los patrones	37
2.2.6. La <i>KLT</i> como método de ortogonalización óptimo	38
2.3. Modelos de interacción entre píxeles	39
2.3.1. Modelo de Ising	39
2.3.2. Campos aleatorios gaussianos - <i>GRF's</i>	40
CAPÍTULO 3. SEPARACIÓN CIEGA DE FUENTES DEPENDIENTES	43
3.1. Un nuevo enfoque para DCA	44
3.2. La medida NG basada en la distancia en $L^2(\mathbb{R})$	45
3.2.1. Calculo de $\Gamma_{NG}(f_y)$ usando ventanas de Parzen	45
3.3. Separabilidad de fuentes dependientes	46
3.4. El algoritmo MAXNG	48
3.4.1. Filtrado previo de las señales mezcla	49
3.4.2. Parametrización de la matriz de separación \mathring{D}	50
3.4.3. Búsqueda de máximos locales	51
3.4.4. Eliminación de máximos locales	52
3.4.5. Cálculo acelerado de la medida NG y sus derivadas	52
3.5. Ejemplo de separación de dos fuentes dependientes	56
3.6. Ruido aditivo gaussiano y MAXNG	60
3.6.1. FD modificado para el caso con ruido	62
3.7. Determinación de factores de escala con fuentes condicionadas	62
3.8. Resultados experimentales	64
3.8.1. Experimento 1: Máxima NG versus Mínima Información Mutua	65
3.8.2. Experimento 2: Efecto del ruido sobre MAXNG	66
3.8.3. Experimento 3: MAXNG versus algoritmos <i>ICA</i>	68

3.8.4. Experimento 4: Performance de MAXNG	69
3.9. Apéndices	70
3.9.1. Demostraciones de las ecuaciones (3.8) y (3.9)	70
3.9.2. Demostración del Teorema 3	71
3.9.3. Demostraciones de las ecuaciones (3.41) - (3.46)	74
 CAPÍTULO 4. CAMPOS ALEATORIOS GAUSSIANOS CON CORRELACIONES DE LARGO ALCANCE	 75
4.1. Motivación del modelo	75
4.2. El modelo <i>LC-GRF</i>	76
4.3. Propiedades básicas del modelo <i>LC-GRF</i>	78
4.4. Estimación de parámetros	79
4.4.1. Estimación de máxima verosimilitud de α , ρ_x y θ	80
4.4.2. Estimación de las varianzas de la señal útil y el ruido usando el algoritmo <i>EM</i>	82
4.5. Reducción de ruido	84
4.5.1. Filtro de Wiener para el caso <i>LC-GRF</i>	85
4.5.2. Performance teórica del filtro de Wiener	86
4.6. Evaluación experimental de los algoritmos	87
4.6.1. Estimación de α , ρ_x y θ	87
4.6.2. Estimación de las varianzas de la señal y el ruido a partir de una muestra con ruido	89
4.7. Apéndices	91
4.7.1. Demostración de la función de Log-likelihood $\Gamma(\alpha, \rho_x, \theta)$	91
4.7.2. Detalles del algoritmo <i>EM</i>	91
4.7.3. Demostración de la Ganancia en la restauración	92
 CAPÍTULO 5. APLICACIONES AL PROCESAMIENTO DE IMÁGENES	 95
5.1. Clasificación de imágenes de <i>Mirounga Leonina</i>	95
5.1.1. Identificación de individuos	95
5.1.2. Clasificación en harems	97
5.2. Corregistro de imágenes satelitales	98
5.3. Análisis sub-píxel de imágenes satelitales (<i>Spectral Unmixing</i>)	101
5.3.1. Imágenes utilizadas en los experimentos	102
5.3.2. Experimento con imagen simulada	105
5.3.3. Experimentos con imágenes reales	105
5.4. <i>BSS</i> y reducción de ruido en imágenes astrofísicas	109
5.4.1. Modelo matemático	109
5.4.2. El método MINIMAX ENTROPY	110
5.4.3. Otra manera de estimar la imagen <i>CMB</i>	112
5.4.4. Experimentos sobre imágenes sintetizadas	112
 CAPÍTULO 6. CONCLUSIONES	 119
 BIBLIOGRAFÍA	 123

RESUMEN

En esta tesis se desarrollan modelos matemáticos y técnicas para el tratamiento de señales en dos dimensiones (imágenes) y se elaboran nuevos algoritmos para la resolución de problemas importantes en procesamiento de imágenes, como son: la separación ciega de fuentes (*Blind Source Separation - BSS*), la reducción de ruido y el reconocimiento de patrones.

A modo introductorio, en la Sección 2.1, se describen las propiedades del método estadístico de Análisis de Componentes Principales (*Principal Component Analysis - PCA*), también conocido como Transformada Karhunen-Loëve (*Karhunen-Loëve Transformation - KLT*), y los métodos derivados de él como son: el Análisis de Componentes Independientes (*Independent Component Analysis - ICA*) y *Projection Pursuit - PP*. Se presenta un análisis detallado teórico de estos métodos, basado en elementos de la teoría de la información de Shannon, estableciendo las conexiones entre: gaussianidad, entropía, información mutua e independencia de las variables involucradas. En la Sección 2.2, se describen los métodos de clasificación de vectores en \mathbb{R}^n utilizando distancias euclídeas mínimas, donde las bases ortonormales juegan un rol central y los métodos de ortogonalización de vectores son de vital importancia. Se muestra que la *KLT*, sobre la que se basa el método de las *eigenfaces* para la identificación de patrones, constituye un método de ortogonalización óptimo. En la Sección 2.3, se describen las teorías existentes de campos aleatorios que permiten modelar las interacciones entre píxeles en imágenes, incluyendo el modelo de Ising y los Campos Aleatorios Gaussianos (*Gaussian Random Field's - GRF's*).

En los capítulos 3 y 4 se introducen nuevos resultados teóricos y se desarrollan nuevos algoritmos, cuyas aplicaciones al procesamiento de imágenes son presentadas en el Cap. 5. Como una generalización de *ICA*, en el Cap. 3, se aborda el problema de la separación ciega de fuentes estadísticamente dependientes o Análisis de Componentes Dependientes (*Dependent Component Analysis - DCA*), que es un nuevo campo de investigación, con pocos antecedentes bibliográficos e importantes aplicaciones a la ingeniería. Se propone un nuevo algoritmo *DCA*, llamado MAXNG que permite la separación de fuentes dependientes (Sección 3.4). Se presentan resultados teóricos fundamentales para MAXNG, proveyendo una condición suficiente sobre las fuentes, que garantiza la separabilidad a través de la maximización de la no-gaussianidad (NG) (Sección 3.3). Se propone una medida de la no-gaussianidad (NG) basada en la distancia en $L^2(\mathbb{R})$, se utiliza el estimador no-paramétrico de Parzen para aproximar las funciones de densidad de probabilidad de las fuentes (*pdf's*) (Sección 3.2) y se introducen nuevas técnicas de aproximación que permiten calcular la medida NG de manera óptima a través de la transformada rápida de Fourier (*Fast Fourier Transform - FFT*) (Sección 3.4). También se analiza la robustez de MAXNG ante la presencia de ruido aditivo gaussiano (Sección 3.6) y se propone una solución para el problema de la indeterminación de los factores de escala (Sección 3.7).

En el Cap. 4, motivado por el problema del ruido en imágenes de origen

astrofísico, se analiza la reducción de ruido gaussiano blanco aditivo (*Additive Gaussian White Noise - AGWN*) (Sección 4.2) presente en las fuentes estimadas por MAXNG. Se propone un nuevo modelo gaussiano, el *LC-GRF* (*Long Correlation - Gaussian Random Field*), que permite modelar las correlaciones de largo alcance presentes en imágenes del *CMB* (*Cosmic Microwave Background*). A través del análisis detallado de las propiedades de este modelo (Sección 4.3), se elaboran técnicas de estimación de parámetros, basadas en el criterio de máxima verosimilitud (*Maximum Likelihood - ML*) y en el algoritmo *EM* (*Expectation-Maximization*) (Sección 4.4). Además, se construye el filtro óptimo de Wiener que permite restaurar las imágenes del *CMB* distorsionadas con ruido aditivo gaussiano blanco (Sección 4.5).

En el Cap. 5, se presentan las siguientes aplicaciones al procesamiento de imágenes:

- i) Clasificación de imágenes de *Mirounga Leonina* (Sección 5.1);
- ii) Corregistro de imágenes satelitales (Sección 5.2);
- iii) Análisis sub-píxel de imágenes satelitales hiper-espectrales (Sección 5.3);
- iv) Separación de fuentes y reducción de ruido en imágenes de origen astrofísico (Sección 5.4).

A continuación se especifican los aportes originales de esta tesis haciendo referencia a las publicaciones emanadas de la misma:

- Se desarrolló un nuevo algoritmo denominado MAXNG que da solución al problema de la separación ciega de fuentes dependientes (*DCA*) (Cap. 3). Se obtuvieron resultados teóricos que sustentan al método y que permiten la implementación de MAXNG de manera eficiente, en términos de su complejidad computacional, y bajo la influencia de ruido aditivo gaussiano. Además, se analizó experimentalmente el desempeño del algoritmo MAXNG en distintos escenarios, utilizando tanto datos simulados como reales, y comparándolo con técnicas *ICA* tradicionales [29, 30, 32, 33, 35].
- Se introdujo un modelo de campo aleatorio gaussiano (*GRF*) con correlaciones de largo alcance (*LC-GRF*), se desarrollaron nuevas técnicas de estimación de parámetros basados en él y se diseñó un filtro reductor de ruido gaussiano aditivo (Cap. 4) [34].
- Se desarrolló un algoritmo original para la clasificación de especímenes de *Mirounga Leonina* a partir de fotografías de sus rostros explotando las propiedades de la representación de imágenes por *KLT* (Sección 5.1) [26, 27].
- Se desarrolló un nuevo método automático para el corregistro de imágenes satelitales, basado en la representación por *KLT*, que fue aplicado a imágenes de tipo radar (Sección 5.2) [28].

-
- Se utilizó el algoritmo MAXNG como nuevo método de análisis sub-píxel de imágenes hiper-espectrales (*Spectral Unmixing*), mediante el cual es posible estimar los porcentajes de ocupación por píxel de los diversos materiales que componen una imagen (agua, vegetación, ladrillo, etc.) (Sección 5.3). Se realizaron experimentos sobre imágenes simuladas y reales demostrando la efectividad del método y comparándolo con técnicas tradicionales basadas en *ICA* [33, 35].
 - Se desarrolló un nuevo algoritmo llamado MINIMAX ENTROPY basado en MAXNG, especialmente diseñado para la separación ciega de imágenes de origen astrofísico (Sección 5.4). Se realizaron experimentos sobre imágenes simuladas demostrando la efectividad del método y comparándolo con técnicas tradicionales basadas en *ICA* [32].
 - Se utilizó el modelo *LC-GRF* para modelar las correlaciones de largo alcance existentes en imágenes del *Cosmic Microwave Background (CMB)* y para desarrollar un filtro reductor de ruido de Wiener (Sección 5.4) [34].

ABSTRACT

In this thesis, mathematical models and techniques for processing two-dimensional signals (images) are developed and new algorithms are proposed for solving important problems in image processing such as: Blind Source Separation - BSS, noise reduction and pattern recognition. As an introduction, in Section 2.1, main properties of the Principal Component Analysis - PCA method, also known as Karhunen-Loève Transformation - KLT, are described together with the analysis of related methods as: Independent Component Analysis - ICA and Projection Pursuit - PP. A detailed theoretical analysis is made based on elements of the Shannon information theory, stating the connections among: Gaussianity, entropy, mutual information and independence of variables. In Section 2.2, classification methods for vectors in \mathbb{R}^n by using minimum Euclidean distances are described, where orthonormal bases and orthogonalization methods are very important. It is shown that KLT, on which the eigenfaces technique for pattern recognition is based, is the optimal method. In Section 2.3, random field theories are described allowing to model pixel interactions in images including the Ising model and Gaussian Random Field's -GRF's as particular cases.

In chapters 3 and 4 new theoretical results are introduced, and new algorithms are developed with applications to image processing as presented in Chapter 5. As a generalization of ICA, in Chapter 3, blind separation of statistically dependent sources (Dependent Component Analysis - DCA) is approached, which is a new field of research with few previous works in the literature and with important applications to engineering. A new DCA algorithm is proposed, namely MAXNG, which allows the separation of dependent sources (Section 3.4). New fundamental theoretical results are presented, providing a sufficient condition on the sources that guarantees their separability through the maximization of Non-Gaussianity (NG) (Section 3.3). An NG measure based on the $L^2(\mathbb{R})$ distance is proposed, a non-parametric Parzen estimator is used to estimate the source probability density functions (pdf's) (Section 3.2) and new approximation techniques are proposed to compute the NG measure in an optimized way through the Fast Fourier Transform - FFT (Section 3.4). Additionally, the robustness of MAXNG in noisy environments is analysed (Section 3.6) and a solution for the scale factor indeterminacy is provided (Section 3.7).

In Chapter 4, motivated by the noise problem in astrophysical images, the reduction of Additive Gaussian White Noise - AGWN in sources estimated by MAXNG is analyzed (Section 4.2). A new Gaussian model is proposed, namely the Long Correlation - Gaussian Random Field (LC-GRF), which allows to model long range interactions existing in Cosmic Microwave Background - CMB images. Through a detailed analysis of model properties (Section 4.3), new techniques for parameter estimation are developed based on the Maximum Likelihood - ML criterion and the Expectation - Maximization (EM) algorithm (Section 4.4). Additionally, the optimal Wiener filter is derived for the restoration of CMB images distorted with AGWN (Section 4.5).

In Chapter 5, the following applications to image processing are presented:

- i) Classification of *Mirounga Leonina* images (Section 5.1);
- ii) Corregistration of remote-sensed images (Section 5.2);
- iii) Sub-pixel analysis of hyper-spectral remote-sensed images (Section 5.3);
- iv) Source separation and noise reduction in astrophysical images (Section 5.4).

Following, the novel contributions included in this thesis are mentioned, including the corresponding publication references:

- A new algorithm for DCA, called MAXNG, was developed (Chapter 3). Theoretical results are provided justifying the method and providing an efficient implementation of MAXNG, in terms of its computational complexity and under the influence of additive Gaussian noise. Furthermore, an experimental analysis of the MAXNG performance in different scenarios is provided, by using simulated data as well as real data, and comparing it with traditional ICA solutions [29, 30, 32, 33, 35].
- A new model of GRF with long correlations was introduced (LC-GRF) and new techniques of parameter estimation were developed based on it, also a noise reduction filter was designed (Chapter 4) [34].
- A novel algorithm for the classification of face images of *Mirounga Leonina*, using properties of the KLT representation (Section 5.1) [26, 27].
- A new algorithm for automatic corregistration of remote-sensed images was developed by using the KLT representation, which was applied to radar images (Section 5.2) [28].
- A new method for sub-pixel analysis of hyper-spectral images (Spectral Unmixing) based on MAXNG was proposed for the estimation of the percent contributions per pixel of several materials present in an image (water, vegetation, bricks, etc.) (Section 5.3). Experiments on simulated and real images are provided, showing the usefulness of the method and comparing it with traditional techniques based on ICA [33, 35].
- A new algorithm, called MINIMAX ENTROPY, based on MAXNG was developed, which was especially designed for the blind separation of astrophysical images (Section 5.4). Experiments on simulated images are provided showing the usefulness of the method and comparing it with traditional techniques based on ICA [32].
- The LC-GRF model was used for modelling the long correlations existing in Cosmic Microwave Background - CMB images and used for the development of a noise reduction Wiener filter (Section 5.4) [34].

AGRADECIMIENTOS

Deseo expresar mi agradecimiento a mi directora de tesis Dra. Araceli Proto, por haberme brindado la posibilidad de iniciarme en la actividad científica desarrollando mi investigación en el Laboratorio de Sistemas Complejos, Facultad de Ingeniería, UBA. También quiero agradecer al Dr. Nestor Barraza (Facultad de Ingeniería - UBA) por su permanente disposición y valiosos comentarios a los resultados de esta tesis y a mi co-director de tesis, Dr. Christian Barbier (*Centre Spatial de Liège, Université de Liège, Bélgica*) por su colaboración. Además, quiero agradecer a los investigadores Emanuele Salerno y Ercan Kuruoğlu del *Signal & Images Laboratory of ISTI - CNR (Consiglio Nazionale delle Ricerche)*, Pisa, Italia, de quienes he recibido valiosas sugerencias y acertados aportes a la investigación de mi tesis. Finalmente, quiero agradecer a la Facultad de Ingeniería y la Universidad de Buenos Aires (UBA) por haberme permitido dedicarme, de manera exclusiva, al trabajo de investigación a través de la beca doctoral Peruilh (2004 - 2007).

NOTACIÓN

Abreviaturas:

<i>AGWN</i>	<i>Additive Gaussian White Noise</i>
<i>BSS</i>	<i>Blind Source Separation</i>
<i>cdf</i>	<i>Cumulative Distribution Function</i>
<i>CM</i>	<i>Conditional Model</i>
<i>CMB</i>	<i>Cosmic Microwave Background</i>
<i>DASP</i>	Distancia Al Subespacio de Patrones
<i>DCA</i>	<i>Dependent Component Analysis</i>
<i>DESP</i>	Distancia En el Subespacio de Patrones
<i>2D-DFT</i>	<i>Two Dimensional - Discrete Fourier Transform</i>
<i>DFT</i>	<i>Discrete Fourier Transform</i>
<i>DUST</i>	<i>Thermal Dust</i>
<i>EM</i>	<i>Expectation / Maximization</i>
<i>FFT</i>	<i>Fast Fourier Transform</i>
<i>GM</i>	<i>Gaussianity Measure</i>
<i>FD</i>	Filtro Decorrelador
<i>GRF</i>	<i>Gaussian Random Field</i>
<i>ICA</i>	<i>Independent Component Analysis</i>
<i>KLT</i>	<i>Karhunen-Loève Transformation</i>
<i>LC-GRF</i>	<i>Long Correlation - Gaussian Random Field</i>
<i>l.i.</i>	linealmente independientes
<i>ML</i>	<i>Maximum Likelihood</i>
<i>MI</i>	<i>Mutual Information</i>
<i>MISE</i>	<i>Mean Integrated Square Error</i>
<i>MRF</i>	<i>Markov Random Field</i>
<i>MSE</i>	<i>Mean Squared Error</i>
<i>NG</i>	No-gaussianidad
<i>PCA</i>	<i>Principal Component Analysis</i>
<i>pdf</i>	<i>Probability Density Function</i>
<i>PP</i>	<i>Projection Pursuit</i>
<i>SAR</i>	<i>Simultaneous Autorregressive</i>
<i>SE</i>	<i>Shannon Entropy</i>
<i>SIR</i>	<i>Signal to Interference Ratio</i>
<i>SNR</i>	<i>Signal to Noise Ratio</i>
<i>SYN</i>	<i>Synchrotron</i>

Símbolos matemáticos:

$x \in \mathbb{R}$	Variable escalar
$\mathbf{x} \in \mathbb{R}^n$	Vector real de dimensión n
$A \in \mathbb{R}^{m \times p}$	Matriz real de $m \times p$ (filas \times columnas)
$\mathbf{x}^T \in \mathbb{R}^{1 \times n}$	Vector transpuesto (matriz de $1 \times n$)
$U \in \mathbb{C}^{m \times p}$	Matriz compleja de $m \times p$ (filas \times columnas)
$U^H \in \mathbb{C}^{p \times m}$	Matriz transpuesta y conjugada (operador hermítico)
$ D = \det [D]$	Determinante de la matriz D
$\langle \mathbf{x}, \mathbf{y} \rangle$	Producto interno de vectores, en \mathbb{R}^n : $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$
$\ \mathbf{x}\ $	Norma del vector \mathbf{x} , en \mathbb{R}^n : $\ \mathbf{x}\ = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$
$d(\mathbf{x}, \mathbf{y})$	Distancia entre vectores \mathbf{x} e \mathbf{y} , la distancia inducida por una norma es $d(\mathbf{x}, \mathbf{y}) = \ \mathbf{x} - \mathbf{y}\ $
$\mathbf{x} \perp \mathbf{y}$	\mathbf{x} es ortogonal a \mathbf{y} , i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = 0$
M^\perp	Subespacio ortogonal al subespacio M
$f_x(x)$	pdf de la variable x
$f_{x y}(x y)$	pdf de la variable x condicionada a la variable y
$\tilde{f}_y(\omega)$	Transformada de Fourier de la pdf $f_y(y)$
$f_z(z; \alpha, \beta)$	pdf de la variable z con parámetros α y β
$f_{xy}(x, y)$	pdf conjunta de las variables x e y
$f_{\mathbf{x}}(\mathbf{x})$	pdf conjunta del vector \mathbf{x}
$f_{\mathbf{x} \mathbf{y}}(\mathbf{x} \mathbf{y})$	pdf del vector \mathbf{x} condicionado al vector \mathbf{y}
$f_{x_i}(x_i)$	pdf marginal de la variable x_i
$E[\cdot]$	Operador esperanza que puede aplicarse a escalares, vectores o matrices, por ej. $E[x]$, $E[\mathbf{x}]$ o $E[\mathbf{x}\mathbf{x}^T]$
σ	Desviación estándar
$\bar{\mathbf{x}}$	Vector medio, definido por $\bar{\mathbf{x}} = E[\mathbf{x}]$
$\mathring{\mathbf{x}}$	Vector no-correlacionado obtenido con un FD
$H(\mathbf{x})$	Entropía de Shannon del vector \mathbf{x}
$I(\mathbf{x})$	Información mutua del vector \mathbf{x}
$J(\mathbf{x})$	Negentropía del vector \mathbf{x}
$C_{\mathbf{xx}} \in \mathbb{R}^{m \times m}$	Matriz de covarianzas del vector $\mathbf{x} \in \mathbb{R}^m$
$A^\dagger \in \mathbb{R}^{p \times m}$	Matriz pseudo inversa o de Moore-Penrose asociada a A
$s'_i \in \mathbb{R}$, $\mathbf{s}' \in \mathbb{R}^p$	Variables centradas, i.e. $s'_i = s_i - \bar{s}_i$ y $\mathbf{s}' = \mathbf{s} - \bar{\mathbf{s}}$
$\hat{u} \in \mathbb{R}$, $\hat{\mathbf{u}} \in \mathbb{R}^p$	Estimaciones del escalar $u \in \mathbb{R}$ y del vector $\mathbf{u} \in \mathbb{R}^p$
$\{x_{i,j}\}$	Campo de variables aleatorias con $i, j = 0, 1, \dots, m - 1$
$\tilde{\mathbf{x}} = DFT \{ \mathbf{x} \}$	Forma vectorial de la transformada discreta de Fourier en 2D del campo aleatorio $\{x_{i,j}\}$

Capítulo 1

INTRODUCCIÓN

En esta tesis se desarrollan modelos matemáticos para el tratamiento de señales en dos dimensiones (imágenes) y se proponen nuevos algoritmos para la resolución de problemas importantes en procesamiento de imágenes como son: la separación ciega de fuentes (*Blind Source Separation - BSS*), la reducción de ruido y el reconocimiento de patrones.

La estructura matemática más adecuada para modelar imágenes digitales depende del problema específico que se quiera resolver y del modelo físico que interviene en el proceso de captura de las imágenes. Por ejemplo, cuando las imágenes adquiridas por un dispositivo electrónico están compuestas por una combinación lineal de imágenes primarias (o fuentes), es importante diseñar técnicas para la estimación de las fuentes a través del análisis de las propiedades matemáticas de las señales y sus mezclas. Este problema ha sido la motivación principal, durante los últimos años, para el desarrollo de la teoría de del Análisis de Componentes Independientes (*Independent Component Analysis - ICA*) [47, 76].

En *ICA*, a cada señal se le suele asociar una variable $x_i \in \mathbb{R}$ que como tal, tiene asociada una *pdf* $f_{x_i}(x_i)$ (*probability density function*). En el caso de imágenes, cada uno de los píxeles representan muestras de esa variable $x_i(0)$, $x_i(1)$, $x_i(2)$, ..., etc. Este modelo permite explotar los resultados de la teoría de probabilidades y en particular, los resultados de la **Teoría de la Información** de C. Shannon [51, 114]. Dada una imagen o conjunto de imágenes, es posible estimar probabilidades, momentos estadísticos, entropías, información mutua, etc. Existe una conexión natural entre los modelos para *ICA* y las herramientas de análisis estadístico de datos multidimensionales como son el Análisis de Componentes Principales (*Principal Component Analysis - PCA*), la transformada de Karhunen-Loève (*Karhunen-Loève Transformation - KLT*) y *Projection Pursuit - PP*, las cuales se analizan en la Sección 2.1.

Otro ejemplo de modelo matemático aparece cuando se analizan técnicas de reconocimiento de patrones en imágenes 2D (como se ejemplifica en [122]). En este caso las imágenes digitales son elementos del espacio vectorial \mathbb{R}^n . El espacio de las imágenes de $n_1 \times n_2$ píxeles es mapeado a un espacio de vectores de dimensión $n = n_1 \times n_2$ ($\mathbb{R}^{n_1 \times n_2} \longrightarrow \mathbb{R}^n$). A su vez, cuando se trabaja con el espacio \mathbb{R}^n , es posible hacer un análisis determinístico o probabilístico. En el caso de un modelo determinístico en \mathbb{R}^n , es posible aplicar los resultados de la **Teoría de Espacios Métricos con dimensión finita (o Álgebra Lineal clásica)** [68, 100] que implica la utilización de elementos propios de esta teoría, como son: las transformaciones lineales, las proyecciones ortogonales, los subespacios lineales, las distancias, etc. De esta manera, se elaboran técnicas de clasificación de vectores basados en distancias mínimas en \mathbb{R}^n como se discute en la Sección 2.2.

Por otro lado también es posible hacer un análisis probabilístico de imágenes en

\mathbb{R}^n . En este caso, una imagen dada es considerada como una muestra de un vector aleatorio $\mathbf{x} \in \mathbb{R}^n$ que como tal, tiene asociada una *pdf* conjunta $f_{\mathbf{x}}(\mathbf{x})$. La **Teoría de Campos Aleatorios** [17, 25, 66] permite analizar y modelar las interacciones existentes entre píxeles, ya que en las imágenes reales los píxeles vecinos tienden a parecerse. Como ejemplos de estos modelos podemos citar el modelo de Ising [31, 72], los campos aleatorios de Markov (*Markov Random Fields - MRF's*) [25, 62] y los campos aleatorios gaussianos (*Gaussian Random Fields - GRF's*) [15, 25, 42, 102] que son discutidos en la Sección 2.3.

En esta tesis se introducen avances teóricos en los modelos matemáticos mencionados y se proponen nuevos algoritmos para el procesamiento de imágenes. Se analiza teóricamente el problema del reconocimiento de patrones permitiendo la adaptación de la técnica conocida como “Método de las *eigenfaces*” para la clasificación de fotografías de especímenes de *Mirounga Leonina* y para el desarrollo de un nuevo algoritmo de corregistro de imágenes satelitales. En *BSS*, se analiza el caso particular de la separación de señales dependientes derivándose un algoritmo original (*Maximum Non-Gaussianity - MAXNG*) con aplicaciones directas en imágenes satelitales hiper-espectrales e imágenes astrofísicas. Dentro de la teoría de campos aleatorios gaussianos, se propone un modelo con correlaciones de largo alcance, el *LC-GRF* (*Long Correlated - Gaussian Random Field*), se proponen algoritmos para la estimación de los parámetros y se muestra su aplicación en la reducción de ruido en imágenes astrofísicas.

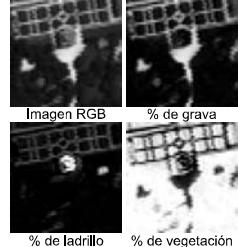
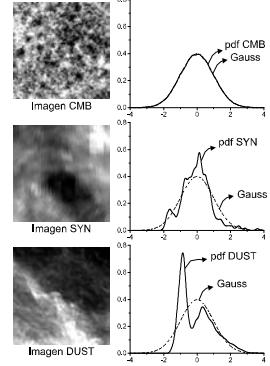
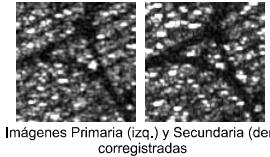
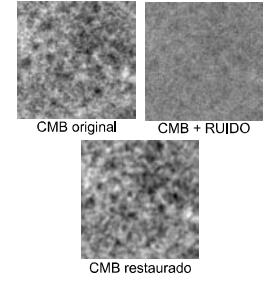
En la TABLA 1.1 se resumen los modelos utilizados en esta tesis, haciendo referencia a los capítulos donde se utilizan. Además, se muestran ejemplos de aplicación originales en procesamiento de imágenes los cuales son desarrollados en detalle en el Cap. 5.

Esta tesis está organizada de la siguiente manera:

En el Cap. 2 se discuten algunas de las herramientas matemáticas existentes para el tratamiento de imágenes. Se presentan los resultados previos más relevantes y se establecen las conexiones entre ellos con demostraciones originales sobre resultados importantes. La Sección 2.1 presenta la teoría de los métodos estadísticos para el análisis de datos multidimensionales, entre los cuales se incluyen: el Análisis de Componentes Principales (*PCA*), la Transformación de Karhunen-Loève (*KLT*), *Projection Pursuit (PP)* y el Análisis de Componentes Independientes (*ICA*). La Sección 2.2 presenta las definiciones y resultados básicos provenientes de la teoría de espacios vectoriales de dimensión finita donde, para un conjunto de vectores patrones de \mathbb{R}^n , se definen subespacios lineales, distancias, proyecciones ortogonales y otros conceptos importantes para el posterior desarrollo de técnicas de reconocimiento de patrones en \mathbb{R}^n . Se analiza también el problema de la ortogonalización de vectores y se demuestra que la *KLT* puede ser interpretada como un método de ortogonalización de vectores óptimo. La Sección 2.3 presenta los modelos de campos aleatorios más relevantes con antecedentes de aplicación al tratamiento de imágenes incluyendo a: los campos aleatorios de Markov (*MRF's*), el modelo de Ising, y los campos aleatorios gaussianos (*GRF's*).

En el Cap. 3 se presenta el problema de la separación ciega de fuentes depen-

TABLA 1.1: Modelos matemáticos para procesamiento de imágenes

Modelo	Ejemplos de Aplicación
<p>Teoría de la Información (Cap. 2, sec. 2.1 y Cap. 3)</p> <p>En una imagen de $n_1 \times n_2$ pixeles (n_1 filas y n_2 columnas), cada pixel representa una muestra de una variable aleatoria $x \in \mathbb{R}$.</p> <p>Por lo tanto, hay $n = n_1 \times n_2$ muestras que pueden utilizarse para estimar la pdf correspondiente, los momentos, la entropía, no-gaussianidad, etc.</p>	<p>Separación Ciega de Fuentes (BSS):</p> <p>a) Imágenes satelitales: La radiancia por pixel capturada por un sensor hiperespectral, es la combinación lineal de las radiancias debidas a cada uno de los materiales presentes en la imagen, multiplicadas por los correspondientes porcentajes de ocupación. La tarea de estimar los porcentajes de ocupación de cada material por pixel, se denomina "Spectral Unmixing" (ver ejemplos en Cap. 5, sec. 5.3)</p> <p>b) Imágenes astrofísicas: Las señales captadas por varios sensores, a distintas frecuencias centrales, pueden modelarse como combinaciones lineales de diferentes fuentes de radiación como son: Cosmic Microwave Background (CMB), Thermal Dust (DUST) y Synchrotron (SYN). La No-Gaussianidad de las imágenes DUST y SYN, junto con la gaussianidad del CMB permite el diseño de un algoritmo BSS para estimar las señales fuentes a partir de sus mezclas. (ver ejemplos en Cap.5, sec. 5.4).</p>  
<p>Espacios vectoriales con dimensión finita (Cap. 2, sec. 2.2)</p> <p>Cada imagen de $n_1 \times n_2$ pixeles es representada por un vector $\mathbf{x} \in \mathbb{R}^n$, siendo $n = n_1 \times n_2$.</p> <p>La teoría de espacios métricos y el álgebra lineal permite utilizar los conceptos de subespacio, distancia, proyección ortogonal, etc.</p>	<p>Reconocimiento de patrones:</p> <p>a) Identificación de Mirounga Leonina: A partir de un conjunto de fotografías de rostros, que definen un subespacio, se puede identificar a un individuo asociándolo con aquella imagen que se encuentra a distancia mínima en el subespacio (Método de las Eigenfacas) (ver ejemplos en Cap. 5, sec. 5.1).</p> <p>b) Corregistro de imágenes satelitales: La capacidad de la representación de Karhunen-Loève para reconocer patrones permite adaptar el método de las Eigenfacas para el corregistro de imágenes. (ver ejemplos en Cap 5, sec. 5.2).</p>  
<p>Teoría de Campos Aleatorios (Cap. 2, sec. 2.3 y Cap. 4)</p> <p>Cada imagen se considera una muestra de un vector aleatorio $\mathbf{x} \in \mathbb{R}^n$ con $n = n_1 \times n_2$ siendo el número de pixeles.</p> <p>La teoría de Campos Aleatorios permite diseñar modelos que reflejen las interacciones entre píxeles usualmente encontrada en imágenes reales.</p>	<p>Reducción de ruido en imágenes: El modelo de Campo Aleatorio Gaussiano LC-GRF (Long Correlated Gaussian Random Field) permite la implementación del filtro de Wiener para la reducción de ruido aditivo gaussiano presente en imágenes astrofísicas CMB (ver ejemplos en Cap. 5, sec. 5.4).</p> 

dientes (*DCA*), que representa un nuevo área de investigación con pocos antecedentes bibliográficos. Se introduce un nuevo algoritmo para la resolución de este problema, llamado MAXNG, que utiliza el criterio de maximización de la no-gaussianidad (NG) de las señales estimadas. Se propone una medida NG basada en la distancia en $L^2(\mathbb{R})$ y se utiliza el estimador no-paramétrico de Parzen para la función de densidad de probabilidad (*pdf*) que se implementa, de manera óptima, a través de la transformada rápida de Fourier (*FFT*). También se analiza la robustez de MAXNG ante la presencia de ruido aditivo gaussiano y se introduce una solución para el problema de la indeterminación de los factores de escala en las estimaciones de las fuentes. A los efectos de demostrar la performance de MAXNG, se presentan resultados experimentales sobre datos simulados comparando los resultados con los obtenidos a través de algoritmos tradicionales *ICA*.

En el Cap. 4 se introduce un modelo gaussiano, denominado *LC-GRF*, que permite modelar interacciones entre píxeles con correlaciones de largo alcance. Se analizan las propiedades del modelo y se elaboran algoritmos para la estimación de los parámetros utilizando el criterio de máxima verosimilitud (*Maximum Likelihood - ML*) y el algoritmo *EM* (*Expectation-Maximization*). Se deriva analíticamente la expresión del filtro de Wiener para la restauración de imágenes *LC-GRF* contaminadas con ruido gaussiano blanco. También se presentan resultados experimentales sobre datos simulados que muestran la efectividad de los algoritmos de estimación de parámetros y del filtro de Wiener.

En el Cap. 5 se presentan ejemplos de aplicación de las teorías y algoritmos desarrollados en los capítulos anteriores, a problemas específicos en procesamiento de imágenes. Se aplica una técnica de reconocimiento de patrones basada en la transformación de Karhunen-Loève (*KLT*) al reconocimiento de individuos de *Mirounga Leonina* y al corregistro de imágenes satelitales. Además, se aplica el algoritmo MAXNG para realizar análisis sub-píxel (*Spectral Unmixing*) de imágenes satelitales hiper-espectrales, también se muestra su aplicación en la separación de fuentes de radiación dependientes en imágenes de origen astrofísico. Se muestran los resultados de la aplicación del filtro de Wiener, basado en el modelo de campos aleatorios gaussianos con correlaciones de largo alcance (*LC-GRF*), para la reducción de ruido gaussiano en imágenes del *Cosmic Microwave Background (CMB)*.

Finalmente, en el Cap. 6, se discuten los resultados de esta tesis, se resaltan los aportes originales de la misma y se discuten las nuevas líneas de investigación abiertas a partir de esta tesis.

Capítulo 2

ASPECTOS TEÓRICOS: RESULTADOS PREVIOS

A continuación se presentan y analizan los resultados previos fundamentales que son la base de los modelos y algoritmos que se desarrollan en los capítulos 3, 4 y 5 de esta tesis.

2.1. Métodos estadísticos y Análisis de Componentes Independientes

En esta sección se presentan las conexiones entre el Análisis de Componentes Principales (*PCA*), la transformación de Karhunen-Loëve (*Karhunen-Loëve Transformation - KLT*), *Projection Pursuit - PP*, y las técnicas de procesamiento de señales de Análisis de Componentes Independientes (*ICA*) utilizadas en la separación ciega de fuentes (*BSS*). Estos métodos, se basan en las propiedades matemáticas que resultan de aplicar una transformación lineal a un vector aleatorio $\mathbf{x} = [x_0 \ x_1..x_{m-1}]^T \in \mathbb{R}^m$ cuya función de densidad de probabilidad conjunta (*joint probability density function - pdf*) es $f_{\mathbf{x}}(\mathbf{x})$.

2.1.1. Análisis de Componentes Principales - *PCA* (o *KLT*)

El Análisis de Componentes Principales (*Principal Component Analysis - PCA*) es una herramienta matemática muy conocida, con aplicaciones en diversas áreas de la ciencia. Esta técnica de análisis estadístico de datos multidimensionales, fue introducida originalmente por Pearson en 1901 en un contexto biológico [108], y desarrollada posteriormente por Hotelling en trabajos de psicometría [71]. La técnica *PCA* es también conocida como la transformación de Karhunen-Loëve (*Karhunen-Loëve Transformation - KLT*) debido a que fue desarrollada de manera independiente por Karhunen [82] y generalizada posteriormente por Loëve [95].

La *KLT* o *PCA* juega un rol fundamental en las técnicas modernas de procesamiento de señales, con aplicaciones a la compresión de datos [56, 87, 126], al reconocimiento de patrones en imágenes [106, 122], a la fusión de imágenes satelitales [40], etc. También es utilizada en algoritmos de redes neuronales [3, 47, 104, 105].

A continuación se introduce una definición formal de la *KLT* [47, 76]:

Transformada de Karhunen-Loëve (*KLT*): Dado un vector aleatorio $\mathbf{x} = [x_0 \ x_1..x_{m-1}]^T \in \mathbb{R}^m$, con momentos de segundo orden finitos, se define la *KLT* como la transformación lineal de \mathbf{x} :

$$\mathbf{y} = V^T \mathbf{x} \quad (2.1)$$

donde $\mathbf{y} = [y_0 \ y_1..y_{p-1}]^T \in \mathbb{R}^p$ ($p \leq m$) es el vector transformado, y la matriz $V = [\mathbf{v}_0 \mathbf{v}_1..\mathbf{v}_{p-1}]^T \in \mathbb{R}^{m \times p}$ se forma colocando en las columnas los primeros p autovectores normalizados de la matriz de covarianzas $C_{\mathbf{xx}} = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$,

ordenados por orden decreciente de sus autovalores no nulos asociados, i.e., $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{p-1} > 0$.

Existe una serie de propiedades muy importantes de la KLT que se resumen a continuación, cuyas demostraciones pueden encontrarse en la literatura [47, 54, 76, 78] pero que, dada su importancia, se incluyen algunas a continuación:

- **Propiedad I:** Las variables aleatorias transformadas y_0, y_1, \dots, y_{p-1} son no-correlacionadas, es decir, poseen covarianzas nulas

$$E[(y_i - \bar{y}_i)(y_j - \bar{y}_j)] = 0 \quad \forall i \neq j \quad (2.2)$$

Dem. Se aplica el operador esperanza a la ecuación (2.1) y se obtiene $\bar{\mathbf{y}} = V^T \bar{\mathbf{x}}$. Usando este resultado junto con (2.1), se calcula la matriz de covarianzas de las variables transformadas

$$C_{\mathbf{yy}} = E[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] = V^T E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T] V = V^T C_{\mathbf{xx}} V \quad (2.3)$$

Debido a que la transformación V^T ha sido elegida de manera que diagonaliza a la matriz de covarianza $C_{\mathbf{xx}}$, se obtiene que $C_{\mathbf{yy}} = \Lambda$ con Λ siendo una matriz diagonal cuyos elementos de la diagonal principal están dados por los autovalores $\lambda_0, \lambda_1, \dots, \lambda_{p-1}$. Por lo tanto queda demostrado que las variables transformadas y_0, y_1, \dots, y_{p-1} son no-correlacionadas. \square

- **Propiedad II:** Las variables y_0, y_1, \dots, y_{p-1} corresponden a proyecciones de las variables originales en las direcciones de máxima (o mínima) varianza. Es decir, de todas las posibles proyecciones $y = \mathbf{v}^T \mathbf{x}$, la varianza de y resulta máxima (o mínima) si se elige la dirección dada por los autovectores de la matriz de covarianzas, i.e., $\mathbf{v} = \mathbf{v}_i$. Además, las varianzas de las variables y_0, y_1, \dots, y_{p-1} están dadas por los autovalores λ_i asociados a los autovectores \mathbf{v}_i .

Dem. Debido a que los vectores de proyección \mathbf{v} deben tener norma unitaria, se utiliza el método de los multiplicadores de Lagrange para maximizar (o minimizar) la varianza de y con la condición $\mathbf{v}^T \mathbf{v} = 1$. Por lo tanto, se desea maximizar (minimizar) la siguiente función del vector \mathbf{v} :

$$\begin{aligned} J(\mathbf{v}) &= E[(y - \bar{y})^2] - \lambda (\mathbf{v}^T \mathbf{v} - 1) \\ &= \mathbf{v}^T C_{\mathbf{xx}} \mathbf{v} - \lambda \mathbf{v}^T \mathbf{v} - \lambda \end{aligned} \quad (2.4)$$

donde λ es el multiplicador de Lagrange asociado. Se plantea el gradiente de la función $J(\mathbf{v})$ y se iguala a cero

$$\nabla_{\mathbf{v}} J(\mathbf{v}) = C_{\mathbf{xx}} \mathbf{v} - \lambda \mathbf{v} = \mathbf{0} \quad (2.5)$$

de donde se obtiene inmediatamente que $C_{\mathbf{xx}} \mathbf{v} = \lambda \mathbf{v}$. Es decir, si se elige a \mathbf{v} como un autovector de $C_{\mathbf{xx}}$, entonces la proyección obtenida es la de

máxima (o mínima) varianza posible, además, el multiplicador de Lagrange λ resulta ser igual al autovalor asociado a \mathbf{v} . Cabe señalar que la matriz $C_{\mathbf{xx}}$ es simétrica y por lo tanto los autovectores asociados a autovalores distintos son ortogonales [100]. Finalmente, se calcula la varianza de cada una de las variables proyectadas en las direcciones de los autovectores y se obtiene:

$$E [(y_i - \bar{y}_i)^2] = \mathbf{v}_i^T C_{\mathbf{xx}} \mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i \quad (2.6)$$

□

- **Propiedad III:** *Dado un vector cualquiera $\mathbf{x}_0 \in \mathbb{R}^m$, la aproximación de orden p dada por el desarrollo en la base ortonormal definida por los p autovectores de $C_{\mathbf{xx}}$ más relevantes, i.e.*

$$\hat{\mathbf{x}}_0 = \bar{\mathbf{x}} + \sum_{i=0}^{p-1} c_i \mathbf{v}_i \quad (2.7)$$

con $c_i = (\mathbf{x}_0 - \bar{\mathbf{x}})^T \mathbf{v}_i$; es una representación óptima en el sentido que minimiza el error cuadrático medio $E [\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|^2]$.

Dem. Dada una base ortonormal arbitraria de \mathbb{R}^m $\{\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{m-1}\}$, un vector cualquiera puede escribirse como [64]

$$\hat{\mathbf{x}}_0 = \bar{\mathbf{x}} + \sum_{i=0}^{p-1} c_i \mathbf{w}_i \quad (2.8)$$

donde $c_i = (\mathbf{x}_0 - \bar{\mathbf{x}})^T \mathbf{w}_i$. Si se realiza un truncamiento del desarrollo conservando los primeros p términos, se observa que el error cuadrático medio queda escrito en función de los últimos $m - p$ vectores de la base, más precisamente:

$$E [\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|^2] = E \left[\left\| \sum_{i=p}^{m-1} c_i \mathbf{w}_i \right\|^2 \right] \quad (2.9)$$

y, usando propiedades básicas de las bases ortonormales, se obtiene:

$$E [\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|^2] = E \left[\sum_{i=p}^{m-1} c_i^2 \right] = \sum_{i=p}^{m-1} E \left[((\mathbf{x}_0 - \bar{\mathbf{x}})^T \mathbf{w}_i)^2 \right] = \sum_{i=p}^{m-1} \mathbf{w}_i^T C_{\mathbf{xx}} \mathbf{w}_i \quad (2.10)$$

En [54, 76] se demuestra que la minimización de la ecuación (2.10), con la condición de ortonormalidad de los vectores \mathbf{w}_i , se obtiene considerando la base ortonormal dada por los primeros p autovectores de la matriz de covarianzas, ordenados de manera decreciente, i.e. $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{p-1}$. Además, el error cuadrático medio está dado por la sumatoria de los autovalores no considerados en la expansión, i.e.,

$$E [\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|^2] = \sum_{i=p}^{m-1} \lambda_i \quad (2.11)$$

□

La Propiedad III muestra claramente la importancia de la *KLT* como técnica para la reducción de la dimensión de un conjunto de datos ya que, usualmente en las aplicaciones, la matriz de covarianzas contiene solo unos pocos autovalores relevantes y puede considerarse $p \ll m$ con un error cuadrático medio despreciable. Esto permite transformar los datos originales de dimensión elevada, en un nuevo conjunto de dimensión reducida donde las características intrínsecas de los datos originales se conservan [78].

Debido a que las variables transformadas son no-correlacionadas, a veces la *KLT* es utilizada como un paso de procesamiento previo y toma el nombre de Filtro Decorrelador - FD (o *whitening filter* en la bibliografía en idioma inglés [76]). En el caso que sea necesario que las variables transformadas, además de ser no-correlacionadas, tengan varianza unitaria $C_{\mathbf{yy}} = I$ se obtiene la siguiente transformación:

$$\mathbf{y} = \Lambda^{-\frac{1}{2}} V^T \mathbf{x} \quad (2.12)$$

donde Λ es la matriz diagonal con los autovalores relevantes definida anteriormente.

Usualmente en las aplicaciones no se conoce directamente la matriz de covarianzas $C_{\mathbf{xx}}$ por lo que se utiliza su estimación a partir de un conjunto de n muestras: $\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(n-1)$; i.e.

$$\hat{C}_{\mathbf{xx}} = \frac{1}{n} \sum_{i=0}^{n-1} (\mathbf{x}(i) - \bar{\mathbf{x}})(\mathbf{x}(i) - \bar{\mathbf{x}})^T \quad (2.13)$$

Esta estimación puede calcularse directamente a partir del conjunto de muestras disponibles e incluso, para aplicaciones en tiempo real, donde las muestras arriban secuencialmente, se puede estimar la matriz de covarianzas de manera recursiva [47], es decir, estimando la matriz de covarianzas para n muestras a partir de la estimación sobre $n-1$ muestras.

Además, si bien existen algoritmos eficaces para el cálculo de autovectores y autovalores de matrices simétricas (por ejemplo, a través del método de las reducciones de Householder [111]), existen algoritmos provenientes del área de redes neuronales, que proponen la búsqueda de las componentes principales a través de algoritmos adaptativos [3, 44, 45]. En estos algoritmos, usualmente se buscan los extremos (máximos o mínimos) de una función a través de aproximaciones sucesivas en lo que se llama un proceso de aprendizaje de la red neuronal [104].

Otro aspecto importante de la *KLT* es que, debido a que es la base ortonormal óptima en el sentido de la Propiedad III, existen numerosos métodos de compresión de datos que han sido basados en ella. Probablemente, el primer trabajo en el que se introduce este concepto data de 1956 [87]. Posteriormente se propusieron aproximaciones a la *KLT* a través de transformaciones trigonométricas rápidas, dando origen a la transformada coseno discreta (*Discrete Cosine Transform - DCT*) en 1974 [2]. Esta última fue la base del estándar de compresión de imágenes

JPEG (Joint Photographic Experts Group) [92]. Más recientemente, el desarrollo de la teoría *wavelets* [99] permitió la elaboración de algoritmos más eficaces para aproximar la *KLT* a través de trasformadas *wavelet* [56, 123, 126] sobre los cuales se basa el nuevo estándar *JPEG 2000* [92].

2.1.2. *Projection Pursuit - PP*

En problemas de clasificación [57], uno de los principales objetivos es reducir la dimensión del conjunto de datos disponible. El análisis de los datos en bajas dimensiones es más manejable y es más sencillo calcular histogramas o graficar datos en dos o tres dimensiones [79]. De hecho, como hemos visto, *PCA* es una técnica que reduce la dimensión usando las direcciones de proyección, en las cuales los datos proyectados tienen varianza máxima (Propiedad II) y, además, las direcciones resultan ortogonales. Como una alternativa a *PCA*, durante años, se propusieron diferentes criterios para definir cuándo una dirección es más “interesante” que otra definiendo el concepto de índice de proyección y proponiendo varias formas de calcularlo.

Los primeros trabajos con estas ideas aparecieron a fines de los 1960's y en los 1970's [90, 91], sin embargo, la técnica se consolidó con el trabajo de Friedman & Tukey publicado en 1974 [60] quienes la bautizaron con el nombre de *Projection Pursuit - PP*. Recientemente, surgió un renovado interés en estos trabajos debido a la conexión entre *PP* e *ICA* [75, 76].

A continuación, se presenta una definición de *PP* estructuralmente análoga a la definición de *KLT*:

Projection Pursuit - PP: Dado un vector aleatorio $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} = [x_0 \dots x_{m-1}]^T$, *PP* se define como la transformación lineal de \mathbf{x} dada por:

$$\mathbf{y} = V^T \mathbf{x} \quad (2.14)$$

donde $\mathbf{y} = [y_0 \ y_1 \dots y_{p-1}]^T \in \mathbb{R}^p$ ($p \leq m$) es el vector transformado, y la matriz $V = [\mathbf{v}_0 \mathbf{v}_1 \dots \mathbf{v}_{p-1}]^T \in \mathbb{R}^{m \times p}$ se forma colocando en las columnas, los p vectores de proyección cuyos índices de proyección sean máximos.

Para poder completar la definición de *PP*, es necesario precisar de qué manera se mide cuán interesante es una proyección. En la literatura se han introducido, de manera empírica, varios índices de proyección [60, 73].

A diferencia de *PCA*, en *PP* los vectores de proyección pueden ser no ortogonales y no se requiere que la varianza sea máxima. En general, los índices de proyección suelen involucrar a los momentos estadísticos de orden superior, por lo que se dice que, a diferencia de *PCA*, *PP* utiliza información más allá de la matriz de covarianzas.

Matemáticamente, un índice de proyección $q_{\mathbf{v}}$ asociado al vector \mathbf{v} se define como una funcional de la distribución de los datos proyectados:

$$q_{\mathbf{v}} = \Theta [f_y(y)] \quad (2.15)$$

donde $\Theta[\cdot]$ es la funcional y $f_y(y)$ es la *pdf* de los datos proyectados $y = \mathbf{v}^T \mathbf{x}$. Por ejemplo, algunos índices de proyección utilizados en *PP* son: el índice de

Friedman-Tuckey [60, 79]:

$$q_v = \int f_y^2(y) dy ; \quad (2.16)$$

la entropía de Shannon, también conocida como entropía de Boltzman-Gibbs en mecánica estadística [73, 79]:

$$q_v = - \int f_y(y) \log [f_y(y)] dy ; \quad (2.17)$$

el índice de Hall [67], basado en la distancia en L^2 , i.e.:

$$q_v = - \int [f_y(y) - \Phi(y)]^2 dy \quad (2.18)$$

con $\Phi(y)$ siendo la *pdf* de Gauss; además, existen otros índices que están basados en el cálculo de momentos de orden superior a dos, como es el caso del coeficiente de asimetría o momento de 3^{er} orden (*skewness*), o el momento de 4^{to} orden (*kurtosis*) [79].

La elección de la funcional más adecuada no es sencilla, y pueden definirse una serie de características necesarias para que un índice de proyección sea considerado como tal [73]. Sin embargo, es aceptable asumir que las direcciones en las cuales la distribución de los datos es cercana a la gaussiana son definitivamente poco “interesantes” [61, 73]. Esto motivó a los investigadores a proponer diferentes índices de proyección basados en medidas de la no-gaussianidad (NG), lo cual puede hacerse de muchas formas obteniendo propiedades y características diferentes para cada alternativa. Por ejemplo, en [129] se presenta una comparación de distintas medidas de NG basadas en tres diferentes normas definidas sobre el espacio de funciones L^2 [61].

2.1.3. Separación ciega de fuentes independientes (*ICA*)

El Análisis de Componentes Independientes (*Independent Component Analysis - ICA*) es una técnica que surgió en los 1990’s para dar solución a problemas de separación ciega de fuentes (*Blind Source Separation - BSS*), un área relativamente nueva en el procesamiento de señales con infinidad de aplicaciones en la ingeniería actual [47, 76]. El nombre *ICA* proviene de una generalización del término *PCA*. En *PCA* se descomponen las variables observadas x_0, x_1, \dots, x_{m-1} como combinaciones lineales de las variables no-correlacionadas y_0, y_1, \dots, y_{p-1} (componentes principales). La no-correlación de las componentes principales no implica que sean independientes, de hecho existen infinitas descomposiciones en componentes no-correlacionadas de las cuales *PCA* es sólo un caso particular. Puede demostrarse fácilmente que la independencia implica la no-correlación pero la recíproca no es cierta en general [59].

A continuación se introduce una definición formal de *ICA*.

Análisis de Componentes Independientes (*ICA*): *Dado un vector aleatorio $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} = [x_0 \ x_1 \dots x_{m-1}]^T$, *ICA* se define como la transformación lineal de \mathbf{x} dada por:*

$$\mathbf{y} = V^T \mathbf{x} \quad (2.19)$$

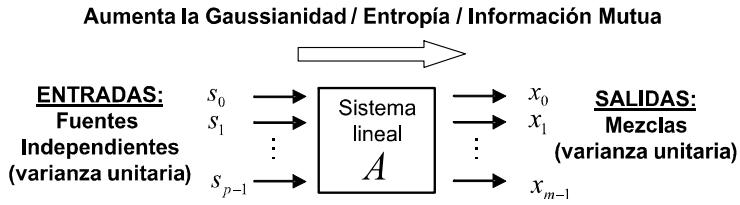


FIGURA 2.1. Modelo ICA

donde $\mathbf{y} = [y_0 \ y_1..y_{p-1}]^T \in \mathbb{R}^p$ ($p \leq m$) es el vector transformado, y la matriz $V = [\mathbf{v}_0 \mathbf{v}_1..\mathbf{v}_{p-1}]^T \in \mathbb{R}^{m \times p}$ se forma colocando en las columnas, los p vectores de proyección que minimizan la dependencia estadística entre las variables proyectadas. Cabe aclarar que las variables aleatorias $y_0, y_1, .., y_p$ son independientes, cuando la pdf conjunta puede escribirse como producto de las pdf's marginales, i.e.,

$$f_{y_0, y_1, .., y_p}(y_0, y_1, .., y_p) = f_{y_0}(y_0) \times f_{y_1}(y_1) \times .. \times f_{y_p}(y_p) \quad (2.20)$$

2.1.4. Resultados teóricos en ICA: conexiones con la Teoría de la Información de Shannon, PP y KLT.

La aplicación principal de *ICA* es la separación ciega de fuentes (*BSS*), donde las señales observadas están compuestas por combinaciones de las señales fuente. *ICA* permite estimar las señales fuente cuando éstas son estadísticamente independientes, que es el caso más común en la literatura ([76] y todas las referencias allí incluidas).

Asumiendo mezclas de tipo lineal, el modelo matemático de *ICA* con m sensores y p fuentes (Fig. 2.1), puede ser expresado matricialmente como:

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

donde $\mathbf{x} = [x_0 \ x_1..x_{m-1}]^T \in \mathbb{R}^m$ es el vector de variables asociadas a observaciones (*señales visibles*), $\mathbf{s} = [s_0 \ s_1..s_{p-1}]^T \in \mathbb{R}^p$ es el vector de variables asociadas a las fuentes independientes (*señales invisibles u ocultas*) y $A \in \mathbb{R}^{m \times p}$ contiene los coeficientes o pesos con los cuales las fuentes son combinadas.

En *ICA*, la estimación de las fuentes $\hat{\mathbf{s}}$ a partir de sus mezclas \mathbf{x} , es ciega en el sentido que la matriz de mezcla A es desconocida. De hecho si se conociera, la estimación de las fuentes $\hat{\mathbf{s}}$ podría calcularse, por ejemplo, utilizando la pseudo inversa o matriz de Moore-Penrose $A^\dagger = (A^T A)^{-1} A^T$ [57] a través de $\hat{\mathbf{s}} = A^\dagger \mathbf{x}$.

En 1994 Comon presentó la teoría que justifica *ICA* como método para la separación ciega de fuentes independientes [49].

Los resultados teóricos de *ICA* más importantes encontrados en la literatura son:

- **No-gaussianidad (NG) de las fuentes:** Las fuentes deben ser no-gaussianas para que *ICA* esté bien planteado. Más precisamente, como máximo, puede existir sólo una fuente con distribución gaussiana. Esta restricción

proviene del hecho que la combinación lineal de gaussianas es a su vez gaussiana, haciendo que los sistemas con entradas gaussianas no sean identificables. Éste es un resultado conocido en el contexto de convolución inversa ciega (*Blind Deconvolution*) [55, 69].

- **Permutación e indeterminación de escala en la estimación de las fuentes:** Las soluciones de *ICA* poseen redundancia de escala y de permutación [49], más precisamente, la solución de *ICA* puede escribirse como:

$$\hat{\mathbf{s}} = \Gamma P \mathbf{s} \quad (2.22)$$

donde Γ es una matriz diagonal y P es una matriz de permutación¹. En general, la permutación de fuentes no es problemática en la mayoría de los casos; sin embargo la indeterminación de escala puede ser crítica en aplicaciones específicas (por ejemplo en el caso de imágenes hiper-espectrales [33, 35]).

- **Información Mutua (*Mutual Information - MI*):** La Teoría de la Información fundada por Shannon [114] provee una forma de medir el grado de dependencia entre variables aleatorias a través del cálculo de la información mutua. Dado un vector de variables aleatorias $\mathbf{y} = [y_0 \ y_1..y_{p-1}]^T \in \mathbb{R}^p$, se define la información mutua $I(\mathbf{y})$ como la entropía relativa [51] entre la *pdf* conjunta y el producto de las correspondientes *pdf*'s marginales:

$$I(\mathbf{y}) = \int f_{\mathbf{y}}(\mathbf{y}) \log \left(f_{\mathbf{y}}(\mathbf{y}) / \prod_{i=0}^{p-1} f_{y_i}(y_i) \right) d\mathbf{y} \quad (2.23)$$

donde $f_{\mathbf{y}}(\mathbf{y})$ es la *pdf* conjunta de las variables $y_0 \ y_1..y_{p-1}$, $f_{y_i}(y_i)$ es la *pdf* marginal de la variable y_i y la integral en (2.23) se toma sobre todo el espacio \mathbb{R}^p . Puede verse fácilmente que $I(\mathbf{y})$ es positiva y se anula solo cuando las variables son independientes, i.e., cuando $f_{\mathbf{y}}(\mathbf{y}) = \prod_{i=0}^{p-1} f_{y_i}(y_i)$ en casi todo punto².

- **Estimador Lineal:** La separación de las fuentes es un problema lineal, es decir, *ICA* se reduce a la búsqueda de una matriz de separación D tal que

$$\mathbf{y} = D\mathbf{x} \quad (2.24)$$

tenga componentes lo más independientes posible. Naturalmente, la mayoría de los algoritmos *ICA* se basan en la búsqueda de las matrices de separación D que hagan mínima la información mutua $I(\mathbf{y})$.

¹Una matriz de permutación P es una matriz cuyas filas contienen un único elemento no nulo igual a +1 o -1.

²En casi todo punto significa que la igualdad podría no cumplirse en un conjunto de medida de Lebesgue nula, lo cual no afecta al cálculo de la integral.

- **KLT como etapa previa de procesamiento:** En [49], Comon observó que la búsqueda de la matriz D que minimiza la información mutua $I(\mathbf{y})$, puede descomponerse en dos etapas. En la primer etapa, se busca una transformación que decorrelacione los datos, usando por ejemplo el FD definido en (2.12). En la segunda etapa, se requiere la búsqueda de una transformación ortogonal que minimice la información mutua, más precisamente:

$$\begin{aligned}\mathbf{\hat{x}} &= \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{V}^T \mathbf{x} && \text{FD (1}^{er} \text{ etapa)} \\ \mathbf{y} &= \mathbf{\hat{D}} \mathbf{\hat{x}} && \text{Transformación ortogonal (2}^{da} \text{ etapa)}\end{aligned}\quad (2.25)$$

La ventaja de esta descomposición es que la búsqueda de la matriz $\mathbf{\hat{D}}$ se restringe al espacio de las matrices ortogonales ya que, si las fuentes son independientes, entonces:

$$C_{\mathbf{yy}} = D C_{\mathbf{\hat{x}}\mathbf{\hat{x}}} D^T = D D^T = I \quad (2.26)$$

- **Minimizar MI equivale a minimizar entropías:** Un resultado ampliamente aceptado en *ICA*, es que la minimización de la información mutua equivale a la minimización de las entropías de las variables. En esta tesis se presenta una demostración nueva y sencilla de este enunciado (una demostración más extensa puede ser encontrada, por ejemplo, en [38]).

Propiedad: *Si las estimaciones de las fuentes (2.24) son no-correlacionadas, i.e., $C_{\mathbf{yy}} = I$; entonces la minimización de MI implica la minimización de la suma de las entropías de las variables $y_0 y_1..y_{p-1}$.*

Dem. Por un lado, se observa de (2.23) que:

$$H(\mathbf{y}) = \sum_{i=0}^{p-1} H(y_i) - I(\mathbf{y}) \quad (2.27)$$

donde

$$H(\mathbf{y}) = - \int f_{\mathbf{y}}(\mathbf{y}) \log(f_{\mathbf{y}}(\mathbf{y})) d\mathbf{y} \quad (2.28)$$

y

$$H(y_i) = \int f_{y_i}(y_i) \log(f_{y_i}(y_i)) dy_i \quad (2.29)$$

son las entropías de Shannon del vector \mathbf{y} y de la variable y_i , respectivamente.

Por otro lado, si se usa una propiedad elemental de la entropía de Shannon [114] en (2.24) se obtiene:

$$H(\mathbf{y}) = H(\mathbf{x}) + \log |D| \quad (2.30)$$

donde $|D|$ es el determinante de D . Igualando (2.27) con (2.30) y usando que $|D| = |C_{\mathbf{yy}}|^{\frac{1}{2}} |C_{\mathbf{xx}}|^{-\frac{1}{2}}$, se obtiene la siguiente ecuación:

$$I(\mathbf{y}) = \sum_{i=0}^{p-1} H(y_i) - H(\mathbf{x}) + \frac{1}{2} \log |C_{\mathbf{xx}}| - \frac{1}{2} \log |C_{\mathbf{yy}}| \quad (2.31)$$

Analizando (2.31) como una función de la matriz de separación D , y teniendo en cuenta que las fuentes deben ser independientes, se observa que el segundo y tercer término ($-H(\mathbf{x}) + \frac{1}{2} \log |C_{\mathbf{xx}}|$) no dependen de D mientras que el último término es nulo (ya que $|I| = 1$). Por lo tanto:

$$I(\mathbf{y}) = \sum_{i=0}^{p-1} H(y_i) + cte \quad (2.32)$$

es decir, minimizar la información mutua o minimizar la suma de las entropías son equivalentes. \square

- **Minimizar la entropía equivale a minimizar la gaussianidad:** En muchas aplicaciones, resulta útil medir distancias entre *pdf's*, y en particular, respecto de la distribución de Gauss.

La teoría de la Geometría de la Información (*Information Geometry*) introducida por Amari [4] en los años 1980's, estudia las propiedades geométricas de familias de *pdf's* midiendo la entropía relativa entre ellas con la definición de Kullback-Leibler [51]. Puede definirse la gaussianidad de una variable aleatoria y como la entropía relativa entre su *pdf* y la *pdf* de Gauss, con igual media μ_y y desvío estándar σ_y :

$$K[f_y(y) \parallel \Phi((y - \mu_y)/\sigma_y)] = \int f_y(y) \log \left(f_y(y) / \Phi \left(\frac{y - \mu_y}{\sigma_y} \right) \right) dy \quad (2.33)$$

donde

$$\Phi(y) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\frac{1}{2}y^2) \quad (2.34)$$

es la *pdf* de Gauss con media cero y varianza unitaria.

Propiedad: *Si se consideran la media y la varianza fijas, entonces minimizar la entropía equivale a minimizar la gaussianidad definida por (2.33).*

Dem. Desarrollando la integral en (2.33) se obtiene:

$$K[f_y(y) \parallel \Phi((y - \mu_y)/\sigma_y)] = H(y) - \frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2 \ln(10)} \quad (2.35)$$

de donde se deduce que, a varianza constante, a la minimización de la entropía es equivalente a la minimización de la gaussianidad. \square

Los resultados recientemente expuestos justifican la idea generalmente aceptada en *ICA*, de que las fuentes tienen *pdf's* que son lo más alejadas posibles a la gaussiana. Este resultado puede interpretarse como una consecuencia natural del Teorema Central del Límite [59] ya que la mezcla de variables tiene siempre una distribución más cercana a la gaussiana (con mayor entropía) que las variables originales. Existen varias demostraciones alternativas a esta idea fundamental que

pueden ser encontradas en la literatura con diversos enfoques y utilizando diferentes herramientas matemáticas (algunos ejemplos son: la demostración de Donoho en un contexto de convolución inversa [55] o la demostración de Cruces et al en un contexto de *ICA* [52]).

Por otro lado, se destaca que *ICA* es esencialmente equivalente a *PP* aplicado a un conjunto de variables que son mezclas de fuentes independientes ya que, si se elige la minimización de la *MI* como criterio para obtener la separación, entonces equivale a encontrar las direcciones ortogonales cuyas proyecciones asociadas maximizan la medida NG de las estimaciones.

Este hecho permitió a los investigadores interesados en *BSS* valerse de herramientas desarrolladas anteriormente para *PP*. Las conexiones entre *PP*, *ICA* y *BSS* fueron remarcadas por varios autores muchos años después que *PP* fuera introducido [52, 76, 112, 129].

2.1.5. Algoritmos *ICA*

Son muchos los algoritmos desarrollados para *ICA* durante los últimos años y sigue siendo un área de creciente actividad por su infinidad de aplicaciones prácticas. Los algoritmos *ICA* existentes pueden clasificarse de acuerdo al criterio utilizado para la determinación de la matriz de separación *D*.

A continuación se presentan los métodos más relevantes en *ICA* agrupados de acuerdo a la clasificación usada en [76]:

- **Maximizando la NG:** Uno de los algoritmos más famosos, FASTICA [74], basa su medida de la NG en el cálculo de la curtosis o momento de orden 4^{to}. En FASTICA, para la búsqueda de máximos se implementa un algoritmo iterativo de punto fijo que garantiza la convergencia a máximos locales con una complejidad de cálculo relativamente baja. La ventaja de FASTICA es su velocidad, aunque es bien conocido que la curtosis, como medida de NG, no es estadísticamente robusta [76]. Una versión mejorada de FASTICA consiste en utilizar la medida de NG basada en una aproximación a la negentropía (detalles en [76]). Se define la negentropía de \mathbf{y} como la diferencia entre la entropía del vector gaussiano, que es la máxima posible, y su entropía:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (2.36)$$

Otras técnicas basadas en la maximización de la NG fueron abordadas en [20, 21].

- **Usando el criterio de máxima verosimilitud (*Maximum Likelihood - ML*):** Uno de los algoritmos *ICA* más antiguos, INFOMAX, tiene su origen en un contexto de redes neuronales, y se deriva planteando el criterio de maximización del flujo de información en una red neuronal (*Info-max principle*) [13]. Puede demostrarse que INFOMAX, al mismo tiempo, maximiza la probabilidad de obtener las observaciones \mathbf{x} (estimador de máxima verosimilitud). Dado que las *pdf*'s de las fuentes no son conocidas,

es necesario imponer un modelo para las mismas. Afortunadamente, puede demostrarse que el estimador es poco sensible a errores en la especificación de las *pdf*'s, y por lo tanto, sólo es necesario seleccionar entre dos familias de *pdf*'s: sub-gaussianas o super-gaussianas. En este algoritmo la maximización se implementa a través de un método de optimización basado en gradientes. También puede utilizarse un algoritmo de punto fijo para máxima verosimilitud arribando a otro algoritmo del tipo FASTICA el cual mejora considerablemente la velocidad de convergencia del algoritmo original [76].

- **Minimizando la *MI*:** Dentro de este grupo de algoritmos se puede mencionar el propuesto originalmente por Comon en [49], que minimiza una estimación de la *MI* a través de la negentropía y usa la aproximación de Edgeworth [83] para las *pdf*'s de las fuentes estimadas. Este algoritmo se reduce al cálculo de *cumulants* [83] de distintos órdenes (detalles en [49]). Otro de los algoritmos en este grupo es el KERNEL ICA [8], para el cual los autores introdujeron una medida de la independencia usando \mathcal{F} -Correlations, una generalización del concepto de correlación para espacios de Hilbert. Puede demostrarse, que este método básicamente minimiza la *MI* [8]. Un algoritmo más reciente es el NON-PARAMETRIC ICA [22], donde se estima la *MI* usando estimadores no-paramétricos de las *pdf*' con la técnica de Parzen [107] usando ventanas gaussianas [115].
- **Utilizando métodos tensoriales:** Otra familia de algoritmos se caracteriza por utilizar tensores de estadísticas de orden superior (*higher-order cumulant tensors*). La matriz de covarianzas es el tensor de estadísticas de segundo orden y su diagonalización, lleva a obtener las componentes principales (*PCA*). Los tensores de estadísticas de orden superior son una generalización de la matriz de covarianzas y por lo tanto, las componentes independientes pueden buscarse como aquellas que hagan mínimas los *cumulant* de orden superior [76]. Usando estas ideas surgieron dos métodos de *ICA*: JADE (*Joint Approximate Diagonalization of Eigenmatrices*) [37] y FOBI (*Fourth-Order Blind Identification*) [36].
- **Utilizando decorrelación no lineal:** Existe un teorema que garantiza que dos variables aleatorias y_1 e y_2 son independientes si y solo si, para todo par de funciones continuas f y g , que se anulan fuera de un intervalo finito, se cumple que $E[f(y_1)g(y_2)] = E[f(y_1)]E[g(y_2)]$ [59]. El primero de los algoritmos publicado en la literatura para *ICA* estuvo basado en este teorema y fue introducido en un contexto de redes neuronales [80].
- **Explotando la estructura temporal de las fuentes:** Finalmente se menciona otro grupo de algoritmos que utilizan sólo estadísticas de segundo orden (matrices de covarianzas) pero explotan la estructura temporal de las señales del mundo real que, usualmente, se encuentran autocorrelacionadas. Algunos ejemplos de estos algoritmos son: AMUSE (*Algorithm for Multiple Unknown Source Extraction*) [119], EVD2 (*Eigen Value Decomposition*) [63],

SOBI (*Second Order Blind Identification*) [14] y otros (en [47], se presenta un resumen completo de este tipo de métodos).

2.2. Bases ortonormales óptimas en \mathbb{R}^n

Una idea fundamental de la teoría de espacios vectoriales utilizada en el procesamiento de señales es la descomposición de una señal como combinación de señales elementales, generalmente más simples. De esta manera, el procesamiento de una señal puede definirse en términos de sus componentes elementales. Como ejemplos clásicos de esta idea, aplicada al procesamiento de imágenes, se pueden mencionar: los métodos de Fourier [110], las descomposiciones en funciones *wavelets* u onditas [99], etc. Existen infinitas maneras de descomponer una señal como combinación lineal de componentes elementales ortogonales o no. Sin embargo, la ortogonalidad permite la aplicación de muchas propiedades muy útiles. Algunas bases ortogonales de \mathbb{R}^n conocidas en el procesamiento de imágenes son las utilizadas por las transformadas de Fourier discreta, Coseno discreta, Walsh, Hadamard, Haar y otras [64].

En esta sección se presentan los resultados teóricos provenientes del álgebra lineal clásica que permiten diseñar bases ortogonales óptimas respecto de un conjunto de datos disponibles. En particular, se demuestra que la transformación *KLT* presentada en la sección anterior en un contexto probabilístico, puede ser derivada utilizando criterios de álgebra lineal que no involucran el uso de probabilidades. La presente formalización matemática fue motivada por el método de las *eigenfaces* utilizado para el reconocimiento de rostros humanos en [106, 122].

2.2.1. Aproximación y clasificación de vectores en \mathbb{R}^n

A continuación, se presentan dos teoremas fundamentales provenientes de la teoría de espacios de Hilbert [50] adaptados para el caso particular de \mathbb{R}^n cuyas demostraciones pueden ser encontradas en la literatura.

Teorema 1 (Proyecciones en \mathbb{R}^n). *Dado un subespacio $M \subset \mathbb{R}^n$, y un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$, existe un único $\mathbf{m}_0 \in M$ tal que $\|\mathbf{x} - \mathbf{m}_0\| \leq \|\mathbf{x} - \mathbf{m}\|$ para todo $\mathbf{m} \in M$; además vale $(\mathbf{x} - \mathbf{m}_0) \perp M$.*

Este teorema indica que, dado un vector cualquiera de $\mathbf{x} \in \mathbb{R}^n$ y un subespacio $M \subset \mathbb{R}^n$, existe un único vector \mathbf{m}_0 del subespacio que minimiza la distancia al vector \mathbf{x} . Además, ese vector puede hallarse como la proyección ortogonal de \mathbf{x} sobre M . El siguiente teorema permitirá descomponer cualquier vector del espacio \mathbb{R}^n en subespacios ortogonales.

Teorema 2 (Descomposición ortogonal de \mathbb{R}^n). *Dado un subespacio $M \subset \mathbb{R}^n$, cualquier vector $\mathbf{x} \in \mathbb{R}^n$ puede expresarse como la suma de dos vectores*

$$\mathbf{x} = \mathbf{x}_M + \mathbf{x}_{M^\perp}$$

donde $\mathbf{x}_M \in M$ y $\mathbf{x}_{M^\perp} \in M^\perp$, además, esta descomposición es única. Otra forma de escribir este resultado es la siguiente $\mathbb{R}^n = M \oplus M^\perp$. Los vectores \mathbf{x}_M y \mathbf{x}_{M^\perp}

se denominan las *proyecciones ortogonales* de \mathbf{x} en los subespacios M y M^\perp respectivamente.

Subespacio de patrones: Supongamos que se disponen de m vectores de \mathbb{R}^n , $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}$, linealmente independientes (l.i.), a los que denominaremos **patrones**, en la práctica estos vectores son representaciones vectoriales de imágenes u otro tipo de datos multidimensionales. Se define al **subespacio de patrones** M como aquel subespacio generado a través de las combinaciones lineales de los patrones, es decir:

$$M = \text{gen} \{ \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1} \} \quad (2.37)$$

La idea subyacente es que, dado un vector cualquiera del espacio $\mathbf{x} \in \mathbb{R}^n$, nos interesa construir una aproximación $\hat{\mathbf{x}}$ de ese vector, usando una combinación lineal de vectores que estén en el subespacio de patrones M . También estaremos interesados en medir distancias de vectores genéricos $\mathbf{x} \in \mathbb{R}^n$ y sus proyecciones en M a cada uno de nuestros patrones $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ para desarrollar mecanismos de clasificación basados en distancias mínimas. Aplicando el teorema de la proyección, se concluye que la mejor aproximación $\hat{\mathbf{x}} \in M$ a un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$ es la proyección ortogonal de \mathbf{x} sobre el subespacio M ($\hat{\mathbf{x}} = \mathbf{x}_M$). A continuación se definen algunas distancias importantes utilizadas en métodos de clasificación de vectores [106, 122].

Distancia Al Subespacio de Patrones (DASP): Es la distancia euclídea entre un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$ y su mejor aproximación $\hat{\mathbf{x}} \in M$, es decir:

$$\text{DASP}(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\| \quad (2.38)$$

Distancia a y_j En el Subespacio de Patrones (DESP_j): Es la distancia euclídea entre la proyección ortogonal $\hat{\mathbf{x}} \in M$ de un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$ y el patrón \mathbf{y}_j , es decir

$$\text{DESP}_j(\mathbf{x}) = \|\mathbf{y}_j - \hat{\mathbf{x}}\| \quad (2.39)$$

En la Fig. 2.2, se hace una interpretación geométrica de estas medidas en \mathbb{R}^3 , en la cual los vectores \mathbf{y}_1 e \mathbf{y}_2 definen un subespacio de dos dimensiones (plano M). El complemento ortogonal M^\perp es una recta que cruza al plano M ortogonalmente por el origen y se cumple que $\mathbb{R}^3 = M \oplus M^\perp$.

La siguiente es una propiedad básica de espacios vectoriales euclídeos cuya demostración es directa y surge a partir de la ortogonalidad de los subespacios M y M^\perp .

Propiedad (Pitágoras): *Dado un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$, la distancia a cada uno de los patrones \mathbf{y}_j puede escribirse en términos de las distancias DASP(\mathbf{x}) y DESP_j(\mathbf{x}) de la siguiente forma:*

$$\|\mathbf{x} - \mathbf{y}_j\|^2 = \text{DASP}^2(\mathbf{x}) + \text{DESP}_j^2(\mathbf{x}) \quad (2.40)$$

2.2.2. Cálculo de las distancias DASP y DESP_j

Puede demostrarse que la distancia mínima de \mathbf{x} al subespacio M (DASP(\mathbf{x})) puede calcularse usando la siguiente relación [97]:

$$\text{DASP}^2(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \frac{\det [G(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}, \mathbf{x})]}{\det [G(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1})]} \quad (2.41)$$

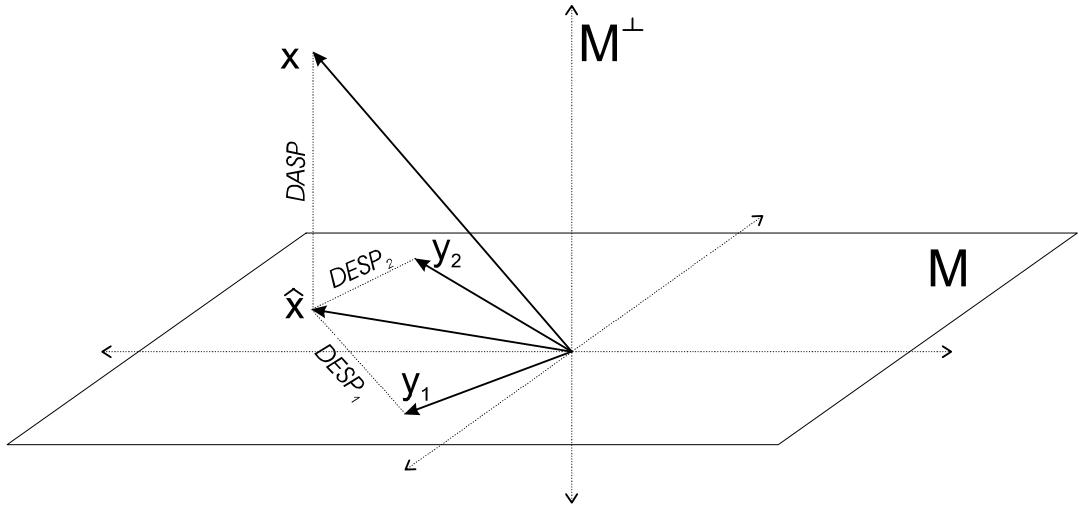


FIGURA 2.2. Interpretación geométrica de $DASP(\mathbf{x})$ y $DESP_j(\mathbf{x})$ en \mathbb{R}^3 .

donde G es la matriz de Gram de dimensiones $m \times m$ asociada a los vectores $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$:

$$G(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}) = \begin{bmatrix} \langle \mathbf{y}_0, \mathbf{y}_0 \rangle & \langle \mathbf{y}_0, \mathbf{y}_1 \rangle & \dots & \langle \mathbf{y}_0, \mathbf{y}_{m-1} \rangle \\ \langle \mathbf{y}_1, \mathbf{y}_0 \rangle & \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \dots & \langle \mathbf{y}_1, \mathbf{y}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{y}_{m-1}, \mathbf{y}_0 \rangle & \langle \mathbf{y}_{m-1}, \mathbf{y}_1 \rangle & \dots & \langle \mathbf{y}_{m-1}, \mathbf{y}_{m-1} \rangle \end{bmatrix} \quad (2.42)$$

Dado que $\hat{\mathbf{x}} \in M$ puede ser escrita como combinación lineal de los patrones de manera única:

$$\hat{\mathbf{x}} = \alpha_0 \mathbf{y}_0 + \alpha_1 \mathbf{y}_1 + \dots + \alpha_{m-1} \mathbf{y}_{m-1}$$

Si se encuentran los valores $\alpha_0, \alpha_1, \dots, \alpha_{m-1}$ se habrá determinado únicamente a $\hat{\mathbf{x}}$ y podremos calcular las distancias $DESP_j$. Para ello se plantea la condición de ortogonalidad de \mathbf{x}_{M^\perp} con cada uno de los patrones \mathbf{y}_j que garantiza la ortogonalidad con cualquier vector de M . Por lo tanto surgen las siguientes m ecuaciones conocidas como **ecuaciones normales** [97]:

$$\langle \mathbf{x}_{M^\perp}, \mathbf{y}_i \rangle = \langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{y}_i \rangle = 0 \quad (2.43)$$

$$\langle \alpha_0 \mathbf{y}_0 + \alpha_1 \mathbf{y}_1 + \dots + \alpha_{m-1} \mathbf{y}_{m-1}, \mathbf{y}_i \rangle = \langle \mathbf{x}, \mathbf{y}_i \rangle \quad (2.44)$$

para $i = 0, 1, \dots, m-1$.

Aplicando la linealidad del producto interno en las m ecuaciones (2.44) se arriba a la siguiente expresión matricial:

$$G\boldsymbol{\alpha} = Y\mathbf{x} \quad (2.45)$$

donde $G = G(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1})$ es la matriz de Gram, $\boldsymbol{\alpha} = [\alpha_0 \ \alpha_1 \ \dots \alpha_{m-1}]^T \in \mathbb{R}^m$, $Y = [\mathbf{y}_0 \ \mathbf{y}_1 \ \dots \ \mathbf{y}_{m-1}]^T \in \mathbb{R}^{m \times n}$ se forma ubicando a los sucesivos patrones en las filas.

La ecuación matricial (2.45) tiene solución única, como era de esperarse por el teorema de la proyección. Los coeficientes $\alpha_0, \alpha_1, \dots, \alpha_{m-1}$ pueden ser calculados de la siguiente manera:

$$\boldsymbol{\alpha} = G^{-1}Y\mathbf{x} \quad (2.46)$$

donde es importante señalar que, dado a que los patrones son linealmente independientes, la matriz G es invertible [97].

Para calcular las distancias $\text{DESP}_j(\mathbf{x})$ entre la aproximación $\hat{\mathbf{x}}$ y cada uno de los patrones $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}$, se utilizan los coeficientes calculados en (2.46) y se obtiene:

$$\text{DESP}_j^2(\mathbf{x}) = \|\mathbf{y}_j - \hat{\mathbf{x}}\|^2 = \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \beta_k^j \beta_l^j \langle \mathbf{y}_k, \mathbf{y}_l \rangle \quad (2.47)$$

donde los coeficientes son

$$\beta_i^j = \begin{cases} \alpha_i & \text{si } i \neq j \\ \alpha_i - 1 & \text{si } i = j \end{cases} \quad (2.48)$$

De las expresiones obtenidas para las distancias $\text{DASP}(\mathbf{x})$ y $\text{DESP}_j(\mathbf{x})$ (ecuaciones (2.41) y (2.47)) se observa que desde el punto de vista computacional no son atractivas, ya que exigen costosos cálculos causados por la complejidad de los determinantes de las matrices de Gram. En la siguiente subsección se muestra que estos cálculos son más sencillos si se elige un sistema de vectores de referencia ortonormal.

2.2.3. Bases ortonormales

Un conjunto de vectores $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ es ortonormal si se cumple que

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad (2.49)$$

Supongamos que existe una base ortonormal de vectores en \mathbb{R}^n que generan exactamente el mismo subespacio M de patrones. De hecho, puede demostrarse que pueden encontrarse muchas bases ortonormales para un mismo subespacio M .

Las bases ortonormales poseen una serie de propiedades muy importantes que las hacen atractivas para representar datos multidimensionales como es el caso de las imágenes. Dado un conjunto de vectores l.i. (patrones) $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}\}$ y una base ortonormal $\{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{m-1}\}$ del subespacio que generan, es decir, tales que

$$M = \text{gen} \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{m-1}\} = \text{gen} \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}\} \quad (2.50)$$

entonces se cumplen las siguientes propiedades.

Propiedad I: Cualquier vector $\mathbf{x}_M \in M$ puede escribirse de manera única como combinación lineal de los vectores de la base ortonormal de la siguiente forma:

$$\mathbf{x}_M = \sum_{j=0}^{m-1} a_j \mathbf{u}_j \quad \text{con } a_j = \langle \mathbf{x}_M, \mathbf{u}_j \rangle \quad (2.51)$$

Dem. (Puede ser encontrada, por ejemplo, en [68]) □

Aplicando este resultado para los patrones \mathbf{y}_i , se obtiene:

$$\mathbf{y}_i = \sum_{j=0}^{m-1} c_{ij} \mathbf{u}_j \quad \text{con} \quad c_{ij} = \langle \mathbf{y}_i, \mathbf{u}_j \rangle \quad (2.52)$$

Energía de un vector: como es usual en la literatura se define la “**energía**” de un vector (señal) $\mathbf{x} \in \mathbb{R}^n$ como su norma al cuadrado, es decir, $\|\mathbf{x}\|^2$.

Propiedad II: *La energía de cualquier vector $\mathbf{x}_M \in M$ puede calcularse a partir de sus coeficientes en la base ortonormal (igualdad de Parseval) mediante*

$$\|\mathbf{x}_M\|^2 = \sum_{j=0}^{m-1} a_j^2 \quad (2.53)$$

y en particular para los patrones se obtiene:

$$\|\mathbf{y}_i\|^2 = \sum_{j=0}^{m-1} c_{ij}^2 \quad (2.54)$$

Dem. (Puede ser encontrada, por ejemplo, en [68]) □

La interpretación de la ecuación (2.53) es la siguiente: la energía total de un vector puede descomponerse en la suma de las energías asociadas a cada vector ortonormal de la base. En otras palabras, el valor de a_j^2 da una idea de la importancia del vector \mathbf{u}_j para la reconstrucción de \mathbf{y}_i . Este concepto puede utilizarse para diseñar diferentes estrategias de búsqueda de bases ortonormales óptimas (Subsección 2.2.4).

Propiedad III: *Dado un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$, su distancia al subespacio de patrones (DASP) puede calcularse con la siguiente fórmula:*

$$\text{DASP}^2(\mathbf{x}) = \|\mathbf{x}\|^2 - \sum_{j=0}^{m-1} a_j^2 \quad \text{donde} \quad a_j = \langle \mathbf{x}, \mathbf{u}_j \rangle \quad (2.55)$$

Dem. El vector $\mathbf{x} \in \mathbb{R}^n$ puede descomponerse ortogonalmente de la siguiente forma: $\mathbf{x} = \hat{\mathbf{x}} + \mathbf{x}_{M^\perp}$ donde $\hat{\mathbf{x}} \in M$ y $\mathbf{x}_{M^\perp} \in M^\perp$. Por lo tanto, usando la igualdad de Pitágoras (2.40) se obtiene una descomposición de la energía de \mathbf{x} como suma de dos energías, la energía en M y la energía en M^\perp :

$$\|\mathbf{x}\|^2 = \|\hat{\mathbf{x}}\|^2 + \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (2.56)$$

y, usando la definición de $\text{DASP}^2(\mathbf{x})$ (ecuación (2.38)), junto con la proposición anterior, se obtiene (2.55). □

Propiedad IV: *Dado un vector cualquiera $\mathbf{x} \in \mathbb{R}^n$, su distancia en el subespacio de patrones (DESP_j) a cada uno de los patrones \mathbf{y}_j , puede calcularse con la siguiente fórmula:*

$$\text{DESP}_j^2(\mathbf{x}) = \sum_{j=0}^{m-1} (a_j - c_{ij})^2 \quad (2.57)$$

Dem. Por definición $\text{DESP}_j^2(\mathbf{x}) = \|\mathbf{y}_j - \hat{\mathbf{x}}\|^2$ y usando las representaciones en la base ortonormal de \mathbf{y}_j y $\hat{\mathbf{x}}$ (ecuaciones (2.52) y (2.51)) se obtiene:

$$\text{DESP}_j^2(\mathbf{x}) = \|\mathbf{y}_j - \hat{\mathbf{x}}\|^2 = \left\| \sum_{j=0}^{m-1} (c_{ij} - a_j) \mathbf{u}_j \right\|^2 = \sum_{j=0}^{m-1} (a_j - c_{ij})^2 \quad (2.58)$$

□

2.2.4. Métodos de ortonormalización de vectores

Los resultados anteriores muestran claramente la ventaja de disponer de bases ortonormales para el subespacio de patrones. El cálculo de los coeficientes en una base ortonormal es tan sencillo como realizar un producto interno (ecuaciones (2.51) o (2.52)) mientras que en el caso no ortogonal se requeriría del cálculo adicional de la inversa de una matriz de Gram (2.46). Otra ventaja es que las expresiones de las distancias DASP(\mathbf{x}) y DESP_j(\mathbf{x}) (ecuaciones (2.55) y (2.57)) son sencillas comparadas con las expresiones en el caso no ortogonal (ecuaciones (2.41) y (2.47)).

Es importante disponer de métodos para generar bases ortonormales a partir de un conjunto de patrones. Si bien existen infinitas elecciones de bases ortonormales del subespacio de patrones, aquí nos concentraremos en las bases ortonormales que resulten óptimas para conjunto de datos disponibles. Siguiendo parcialmente los lineamientos de los trabajos [39, 113, 116], el problema de la ortogonalización de vectores, puede ser planteado de una manera general que contiene a los métodos conocidos como casos particulares.

Para ello se utiliza la matriz $Y = [\mathbf{y}_0 \ \mathbf{y}_1 \ \dots \ \mathbf{y}_{m-1}] \in \mathbb{R}^{n \times m}$ cuyas columnas son los vectores patrones $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}$ que se asumen l.i. De manera equivalente se define la matriz $U = [\mathbf{u}_0 \ \mathbf{u}_1 \ \dots \ \mathbf{u}_{m-1}] \in \mathbb{R}^{n \times m}$ cuyas columnas son los vectores ortonormales $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{m-1}$ buscados.

Como ambos conjuntos de vectores deben generar el mismo subespacio M , se puede escribir a cada vector \mathbf{u}_i como combinación lineal de los $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}$ en forma única, que matricialmente es:

$$U = YA \quad (2.59)$$

donde $A \in \mathbb{R}^{m \times m}$ es invertible.

Como debe cumplirse la condición de ortonormalidad $U^T U = I$, resulta:

$$U^T U = A^T Y^T Y A = A^T G A = I \quad (2.60)$$

siendo G la matriz de Gram introducida anteriormente en la ecuación (2.42). Por lo tanto, una forma generalizada de buscar métodos de ortonormalización, consiste en buscar matrices de transformación invertibles A , que hagan cumplir la condición (2.60).

Un método ampliamente conocido es el de Gram Schmidt o factorización QR, en el que la construcción de la base ortonormal, es de tipo algorítmica y consiste

en elegir secuencialmente los vectores \mathbf{y}_i para obtener uno a uno los vectores \mathbf{u}_i ortonormales (detalles en [68]). La base ortonormal que se obtiene con este método no es la misma si se altera el orden en que van tomando los vectores \mathbf{y}_i en el algoritmo. Por lo tanto, surge la pregunta de si existe algún método en el cual la base obtenida no dependa del orden que se consideren los vectores \mathbf{y}_i , y en particular, uno puede preguntarse cual es la base óptima en el sentido que concentre la energía de los patrones en pocos elementos de esa base. En las próximas subsecciones se presenta la respuesta a esta pregunta y se demuestra que la base óptima resulta equivalente a la *KLT* definida anteriormente.

2.2.5. Descomposición de la energía de los patrones

Siguiendo los lineamientos de [116], en esta subsección se analiza la distribución de las energías de los patrones en una base ortonormal cualquiera. Para ello, se construye la matriz de energías E que contiene las energías de las proyecciones de los patrones $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}\}$ en una base ortonormal genérica $\{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{m-1}\}$, donde el elemento de la fila i , columna j se calcula como $E(i, j) = c_{ij}^2 = \langle \mathbf{y}_i, \mathbf{u}_j \rangle^2$, es decir:

$$E = \begin{bmatrix} \langle \mathbf{y}_0, \mathbf{u}_0 \rangle^2 & \langle \mathbf{y}_0, \mathbf{u}_1 \rangle^2 & \dots & \langle \mathbf{y}_0, \mathbf{u}_{m-1} \rangle^2 \\ \langle \mathbf{y}_1, \mathbf{u}_0 \rangle^2 & \langle \mathbf{y}_1, \mathbf{u}_1 \rangle^2 & \dots & \langle \mathbf{y}_1, \mathbf{u}_{m-1} \rangle^2 \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{y}_{m-1}, \mathbf{u}_0 \rangle^2 & \langle \mathbf{y}_{m-1}, \mathbf{u}_1 \rangle^2 & \dots & \langle \mathbf{y}_{m-1}, \mathbf{u}_{m-1} \rangle^2 \end{bmatrix} \quad (2.61)$$

Se observa que la suma de los elementos de la fila i resulta igual a la energía de cada uno de los patrones, es decir:

$$\sum_{j=0}^{m-1} E(i, j) = \sum_{j=0}^{m-1} \langle \mathbf{y}_i, \mathbf{u}_j \rangle^2 = \sum_{j=0}^{m-1} c_{ij}^2 = \|\mathbf{y}_i\|^2 \quad (2.62)$$

Por otro lado la suma de los elementos de la columna j es la suma de las energías de los patrones en la componente j de la base ortonormal e_j :

$$e_j = \sum_{i=0}^{m-1} E(i, j) = \sum_{i=0}^{m-1} \langle \mathbf{y}_i, \mathbf{u}_j \rangle^2 = \sum_{i=0}^{m-1} c_{ij}^2 \quad (2.63)$$

Se destaca que, la suma sobre las filas es independiente de la base ortonormal que se elija, mientras que, la suma sobre las columnas depende de la base. Por otro lado se cumple la siguiente condición:

$$\sum_{j=0}^{m-1} e_j = \sum_{i=0}^{m-1} \|\mathbf{y}_i\|^2 \quad (2.64)$$

donde el miembro de la derecha de (2.64) es la energía total de los patrones.

Con este enfoque surgen diferentes estrategias para la búsqueda de bases ortonormales ya que pueden plantearse distribuciones arbitrarias de las energías de los patrones en la base [39, 96, 113, 116].

2.2.6. La KLT como método de ortogonalización óptimo

Supongamos que se desea encontrar los vectores ortonormales $\{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_k\}$ uno a uno de manera que, en cada paso, la energía de los patrones proyectada en ellos sea máxima. Para ello, se define la siguiente magnitud que mide la concentración de energía promedio de m patrones $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{m-1}\}$ en una base ortonormal de k elementos $\{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{k-1}\}$ ($k \leq m$):

$$J(k) = \frac{1}{m} \sum_{j=0}^{k-1} e_j \quad \text{con } k = 0, 1, \dots, m-1 \quad (2.65)$$

La energía de los patrones concentrada en el primer vector \mathbf{u}_0 de la secuencia es $J(0) = \frac{1}{m} e_0$ y se puede plantear su maximización, a través del método de multiplicadores de Lagrange, ya que el vector \mathbf{u}_0 buscado debe estar normalizado ($\langle \mathbf{u}_0, \mathbf{u}_0 \rangle = 1$). La función a maximizar (respecto de \mathbf{u}_0) resultante es:

$$J(0) = \frac{1}{m} \sum_{i=0}^{m-1} \langle \mathbf{y}_i, \mathbf{u}_0 \rangle^2 - \lambda_0 (\langle \mathbf{u}_0, \mathbf{u}_0 \rangle - 1) \quad (2.66)$$

Desarrollando los productos internos en términos de productos matriciales y reagrupando se obtiene:

$$J(0) = \mathbf{u}_0^T \left(\frac{1}{m} \sum_{i=0}^{m-1} \mathbf{y}_i \mathbf{y}_i^T \right) \mathbf{u}_0 - \lambda_0 (\mathbf{u}_0^T \mathbf{u}_0 - 1) \quad (2.67)$$

Aplicando gradiente respecto de \mathbf{u}_0 e igualando a cero se deduce que los vectores que maximizan la energía, son aquellos que verifican

$$R_{\mathbf{y}\mathbf{y}} \mathbf{u}_0 = \lambda_0 \mathbf{u}_0 \quad (2.68)$$

con

$$R_{\mathbf{y}\mathbf{y}} = \frac{1}{m} \sum_{i=0}^{m-1} \mathbf{y}_i \mathbf{y}_i^T \quad (2.69)$$

es decir, los autovectores de la matriz $R_{\mathbf{y}\mathbf{y}} \in \mathbb{R}^{n \times n}$.

Además, puede verse fácilmente que la energía de los patrones concentrada en un autovector de $C_{\mathbf{y}\mathbf{y}}$ está dada por el autovalor asociado, que además es siempre real y no-negativo:

$$J(0) = \mathbf{u}_0^T C_{\mathbf{y}\mathbf{y}} \mathbf{u}_0 = \lambda_0 \quad (2.70)$$

Si se continúa con el proceso de selección para elegir los siguientes vectores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{m-1}$, se observa que para cada uno se obtiene una ecuación equivalente (2.70), por lo tanto los vectores se irán eligiendo como autovectores de $R_{\mathbf{y}\mathbf{y}}$ asociados a los autovalores mayores (no elegidos anteriormente) y que además, resultan ortogonales gracias a que $R_{\mathbf{y}\mathbf{y}}$ es hermitiana [100].

Se señala que, utilizar esta base ortonormal como sistema de referencia es equivalente a aplicar la *KLT*, de hecho, si se redefine un nuevo conjunto de patrones centrados $\{\mathbf{y}'_0, \mathbf{y}'_1, \dots, \mathbf{y}'_{m-1}\}$ tales que:

$$\mathbf{y}'_i = \mathbf{y}_i - \bar{\mathbf{y}} \quad \text{con} \quad \bar{\mathbf{y}} = \frac{1}{m} \sum_{j=0}^{m-1} \mathbf{y}_j \quad (2.71)$$

Entonces la matriz

$$R_{\mathbf{y}'\mathbf{y}'} = \frac{1}{m} \sum_{i=0}^{m-1} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^t = C_{\mathbf{yy}} \quad (2.72)$$

coincide con la estimación de la matriz de covarianzas sobre la cual se basa la *KLT* como método probabilístico.

2.3. Modelos de interacción entre píxeles

En esta sección se presenta un resumen de teorías importantes encontradas en la bibliografía a partir de las cuales, se desarrollan los resultados incluidos en los siguientes capítulos de esta tesis. Generalmente en las imágenes reales existen relaciones entre los valores asociados a píxeles vecinos. Para poder modelar este tipo de interacciones desde el punto de vista estadístico, se han propuesto numerosos modelos, la mayoría de los cuales fueron a su vez utilizados en la Física ya que puede establecerse una analogía entre partículas en un sistema físico con píxeles en una imagen.

Los campos aleatorios (*random fields*) [66] proveen una estructura matemática para modelar imágenes capturando las dependencias entre píxeles. Una de las propiedades usualmente requeridas para un modelo matemático de imágenes es proporcionar probabilidades de los valores de los píxeles condicionados a los píxeles vecinos. Por esta razón los campos aleatorios markovianos (*Markov Random Fields* - *MRF*'s) han sido usados frecuentemente en la literatura [18, 23, 25, 62].

La teoría de *MRF*'s tiene su origen en la mecánica estadística y tuvo gran auge en la aplicación a imágenes desde la publicación de los trabajos de Besag [17, 18], Moran [102] y Geman & Geman [62]. Dentro de los *MRF*'s existe un modelo muy sencillo y antiguo: el Modelo de Ising [72, 77, 117], que se presenta a continuación.

2.3.1. Modelo de Ising

Este modelo tuvo su origen en la física para explicar el comportamiento de materiales ferromagnéticos [72, 77, 117]. Dado un sistema de n variables binarias o *spins* $\{s_i\}$, que pueden tomar dos estados opuestos $s_i = -1$ o $s_i = +1$, la energía del sistema $E_I\{s_i\}$ está dada por los estados de los *spins* de la siguiente manera:

$$E_I\{s_i\} = -\sum_{\langle ij \rangle} \epsilon_{ij} s_i s_j - H \sum_{i=1}^n s_i \quad (2.73)$$

donde el símbolo $\langle ij \rangle$ indica pares de vecinos de primer orden (o vecinos más cercanos), ϵ_{ij} indica el potencial de interacción entre los *spins* s_i y s_j , y H es un posible campo externo que induce a los *spins* a polarizarse, ya sea positivamente $H > 0$ o negativamente $H < 0$ [72].

Este modelo es un caso particular de *MRF* y la probabilidad de una determinada configuración se calcula a partir de la fórmula de Gibbs [66]:

$$P(\{s_i\}) = \frac{1}{z} \exp \left(-\frac{1}{kT} E_I \{s_i\} \right) \quad (2.74)$$

donde z es una constante de normalización conocida como función de partición en mecánica estadística [72], k es la constante de Boltzmann y T es la temperatura del sistema [72]. Se observa que, para el caso de ausencia de campo externo, y considerando interacciones homogéneas de primeros vecinos ($\epsilon_{ij} = 1$ si s_i y s_j son vecinos de primer orden), la probabilidad de una configuración está dada por:

$$P(\{s_i\}) = \frac{1}{z} \exp \left(\beta \sum_{\langle ij \rangle} s_i s_j \right) \quad (2.75)$$

con $\beta = \frac{1}{kT}$ siendo el único parámetro que define el modelo y que es inversamente proporcional a la temperatura del sistema.

Este sencillo modelo permite representar imágenes blanco y negro, donde el parámetro β regula las interacciones entre píxeles [23, 25, 62]. Puede demostrarse que, para valores por debajo de un umbral crítico $|\beta| < \beta_c \approx 0,44$ el modelo de Ising en 2D no presenta magnetización mientras que para valores de β por encima del umbral sí [72].

En aplicaciones de procesamiento de imágenes es fundamental contar con herramientas para ajustar los modelos a datos disponibles lo que exige el desarrollo de técnicas de estimación de parámetros basados en criterios estadísticos. Como ejemplo, en [31] se han propuesto dos nuevas técnicas para la estimación del parámetro β del modelo de Ising a partir de realizaciones del campo aleatorio (imágenes binarias).

2.3.2. Campos aleatorios gaussianos - *GRF*'s

Otro caso particular de *MRF* es el caso de variables son gaussianas dando lugar a los campos aleatorios Gaussianos (*Gaussian Random Fields* - *GRF*'s) [17, 81]. En un *GRF* con media cero, el modelo queda completamente determinado por su matriz de covarianzas [81]. Por definición, dada una imagen de $l \times l$ píxeles, las variables aleatorias $\{x_{i,j}\}$ con $i, j = 0, 1, \dots, l-1$, el vector $\mathbf{x} \in \mathbb{R}^{l^2}$ de píxeles ordenados lexicográficamente forma un *GRF*, con media nula $E[\mathbf{x}] = \mathbf{0}$ y matriz de covarianzas $C_{\mathbf{xx}} = E[\mathbf{x}\mathbf{x}^T]$, cuando su *pdf* conjunta está dada por la distribución de Gauss multivariada [120]:

$$f_{\mathbf{x}}(\mathbf{x}; C_{\mathbf{xx}}) = \frac{1}{\sqrt{(2\pi)^{l^2} |C_{\mathbf{xx}}|}} \exp \left(-\frac{1}{2} \mathbf{x}^T C_{\mathbf{xx}}^{-1} \mathbf{x} \right) \quad (2.76)$$

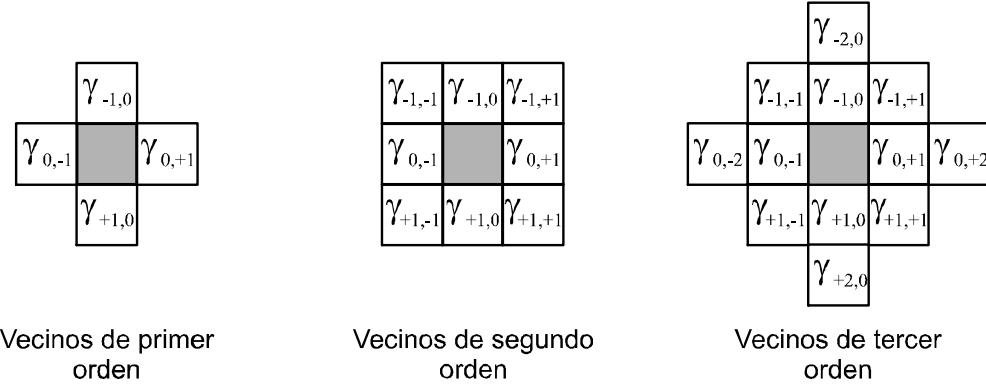


FIGURA 2.3. Sistemas de vecinos de interacción. El orden de interacción define el conjunto de índices ∂ considerado en (2.77)

donde $|C_{\mathbf{xx}}|$ indica el determinante de la matriz $C_{\mathbf{xx}}$. En la bibliografía, existen distintas variantes del modelo con diferentes esquemas de interacción entre vecinos obteniendo distintas matrices de covarianzas [9, 17, 25, 42, 81].

Como es usual en la bibliografía, se pueden analizar los *GRF*'s $\{x_{i,j}\}$ generados por el siguiente conjunto de ecuaciones lineales [15, 17, 25, 81]

$$x_{i,j} = \sum_{(r,s) \in \partial} \gamma_{r,s} \cdot x_{i-r,j-s} + \sqrt{\rho_{\mathbf{x}}} \cdot \epsilon_{i,j} \quad (2.77)$$

donde: $i, j = 0, 1, \dots, l - 1$; ∂ es el conjunto de vecinos del píxel (i, j) (Fig. 2.3); $\gamma_{r,s}$ son los coeficientes de interacción con propiedades de simetría, i.e., $\gamma_{-1,0} = \gamma_{+1,0}$, $\gamma_{0,-1} = \gamma_{0,+1}$, etc.; $\epsilon_{i,j}$ son variables gaussianas con media cero, varianza unitaria con una matriz de covarianzas con estructura conocida, y $\rho_{\mathbf{x}}$ es un escalar positivo. Además, se asume una estructura toroidal del campo lo cual significa que:

$$x_{l-1,j} \equiv x_{-1,j} \quad (2.78)$$

$$x_{i,l-1} \equiv x_{i,-1} \quad (2.79)$$

La ecuación (2.77) puede ser escrita en forma matricial de la siguiente manera:

$$B\mathbf{x} = \sqrt{\rho_{\mathbf{x}}}\boldsymbol{\epsilon} \quad (2.80)$$

donde \mathbf{x} y $\boldsymbol{\epsilon}$ son vectores de \mathbb{R}^{l^2} y la matriz $B \in \mathbb{R}^{l^2 \times l^2}$ tiene una estructura muy particular ya analizada en la bibliografía en el contexto de los *GRF*'s. Aquí, se utilizará la notación de [81]. En el caso más simple, donde sólo se consideran cuatro vecinos por píxel (sistema de vecinos de primer orden), los coeficientes de interacción quedan definidos por dos parámetros $\theta_1 = \gamma_{-1,0} = \gamma_{+1,0}$ (interacciones horizontales), y $\theta_2 = \gamma_{0,-1} = \gamma_{0,+1}$ (interacciones verticales), y la matriz B toma

la forma [81]:

$$B = \begin{bmatrix} B_{1,1} & B_{1,2} & 0 & .. & 0 & B_{1,l} \\ B_{1,l} & B_{1,1} & B_{1,2} & .. & 0 & 0 \\ .. & .. & .. & .. & .. & .. \\ B_{1,2} & .. & .. & .. & B_{1,l} & B_{1,1} \end{bmatrix} \quad (2.81)$$

la cual es simétrica, circulante por bloques ($l \times l$ bloques), sus bloques son a su vez circulantes ($l \times l$) y están definidos por:

$$B_{1,1} = \begin{bmatrix} 1 & -\theta_1 & 0 & .. & -\theta_1 \\ -\theta_1 & 1 & -\theta_1 & 0 & .. \\ 0 & -\theta_1 & 1 & -\theta_1 & 0 \\ .. & .. & .. & .. & .. \\ -\theta_1 & 0 & .. & -\theta_1 & 1 \end{bmatrix}; \quad B_{1,l} = B_{1,2} = \begin{bmatrix} -\theta_2 & 0 & .. & 0 \\ 0 & -\theta_2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & 0 & .. & -\theta_2 \end{bmatrix} \quad (2.82)$$

Desde luego éste modelo puede generalizarse a ordenes de interacción superiores (más de cuatro vecinos) introduciendo más parámetros en la definición de la matriz B (por ejemplo, en [15, 81] se han aplicado modelos de orden superior para sintetizar y modelar texturas). De todas maneras, sin importar el orden del sistema de vecinos considerado, la estructura de la matriz B siempre es la misma: contiene mayoritariamente ceros (*sparse*), es circulante por bloques, sus bloques son también circulantes y por lo tanto pueden ser diagonalizadas a través de la operación *2D-DFT* [65].

La matriz de covarianzas del *GRF* definida en (2.80) es:

$$C_{\mathbf{x}\mathbf{x}} = \rho_{\mathbf{x}} B^{-1} C_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} B^{-1} \quad (2.83)$$

Existen dos casos muy particulares que se corresponden con modelos existentes en la literatura:

1. ***Conditional Model - CM*** [18, 25, 81]: donde $C_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} = B$ y por lo tanto $C_{\mathbf{x}\mathbf{x}} = \rho_{\mathbf{x}} B^{-1}$;
2. ***Simultaneous Autoregressive Model - SAR*** [25, 81]: donde $C_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} = I$ y entonces resulta que $C_{\mathbf{x}\mathbf{x}} = \rho_{\mathbf{x}} B^{-2}$.

Capítulo 3

SEPARACIÓN CIEGA DE FUENTES DEPENDIENTES

En el capítulo anterior se presentaron las técnicas de Análisis de Componentes Independientes (*Independent Component Analysis - ICA*) para la Separación Ciega de Fuentes (*Blind Source Separation - BSS*) en el caso de fuentes estadísticamente independientes. Sin embargo, existen muchos casos en los cuales la condición de independencia no es válida y se requiere de la elaboración de nuevos métodos que generalicen *ICA*. Existen pocos antecedentes bibliográficos de este problema al que algunos autores le han dado el nombre de *Dependent Component Analysis - DCA* [5, 10, 11, 12, 46]. Esta tesis aporta nuevos resultados teóricos y experimentales en *DCA*: se propone un nuevo algoritmo llamado MAXNG para la separación de fuentes dependientes o independientes [30], se analiza teóricamente la separabilidad de las fuentes en función del tipo de dependencia entre ellas (Teorema 3) [35], se desarrollan nuevas técnicas que permiten implementar MAXNG de manera eficiente en términos de su complejidad computacional y bajo la influencia de ruido aditivo gaussiano [35], también se analiza experimentalmente el desempeño del algoritmo en distintos escenarios comparándolo con técnicas *ICA* tradicionales utilizando datos simulados y reales [29, 30, 32, 33, 35].

El modelo matemático de *DCA* es el siguiente: dadas p fuentes dependientes, existen m señales mezcla (sensores) producidas a partir del siguiente modelo lineal:

$$\mathbf{x}(t) = A\mathbf{s}(t) ; \quad (3.1)$$

donde: $\mathbf{x}(t) = [x_0(t) \ x_1(t) \dots x_{m-1}(t)]^T \in \mathbb{R}^m$ es el vector de las señales observadas, $\mathbf{s}(t) = [s_0(t) \ s_1(t) \dots s_{p-1}(t)]^T \in \mathbb{R}^p$ es el vector de las señales fuente no independientes y $A \in \mathbb{R}^{m \times p}$ es la matriz de mezcla. La variable t es un índice que recorre las muestras disponibles de las señales, cuando esta variable representa el tiempo, el modelo (3.1) es conocido como mezcla lineal instantánea ya que la matriz de mezcla A es constante en el tiempo. En lo sucesivo, se evitará la variable t en la notación para hacerla más sencilla.

La separación de las fuentes se obtiene identificando la matriz de separación D tal que

$$\mathbf{y} = D\mathbf{x} \quad (3.2)$$

contenga una copia de las señales fuente \mathbf{s} posiblemente escaladas y/o permutadas.

En esta tesis nos restringiremos al caso sobredeterminado, es decir $m \geq p$ (número de sensores mayor o igual que el número de fuentes). También se asumen señales estacionarias con medias nulas ($\bar{\mathbf{s}} = E[\mathbf{s}] = 0$) con excepción de la Sección 3.7 donde se considera un caso particular de señales con medias no nulas.

Al igual que en *ICA*, estamos interesados en desarrollar técnicas de *DCA* “ciegas” en las cuales no se tiene información sobre la matriz A más allá de sus

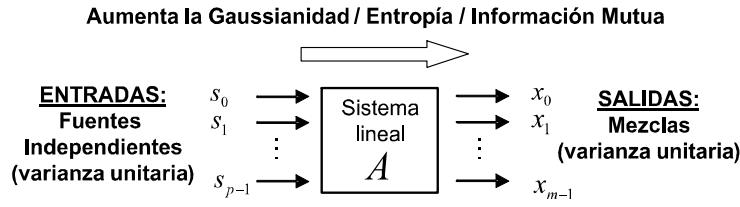


FIGURA 3.1. Modelo ICA

dimensiones (m y p). Está claro que, de conocerse la matriz de mezcla A , el proceso de separación sería tan sencillo como considerar la matriz de separación D igual a la pseudo-inversa o matriz de Moore-Penrose [57]:

$$D = A^\dagger = (A^T A)^{-1} A^T \quad (3.3)$$

Se destaca que el enfoque tradicional de *DCA* encontrado en la bibliografía consiste en utilizar información de la estructura temporal de las señales fuente asumiendo que las señales buscadas están auto-correlacionadas en el tiempo y explotando estas correlaciones [5, 11, 12, 46]. En esta tesis, se presentan nuevos resultados teóricos y se introduce un nuevo algoritmo para *DCA*, llamado MAXNG (Maximum Non-Gaussianity) que no requiere la utilización de la estructura temporal de las señales, por lo tanto puede ser utilizada inclusive cuando las señales son procesos blancos (no auto-correlacionados), ya sean señales dependientes o independientes.

3.1. Un nuevo enfoque para DCA

La distribución de Gauss tiene la máxima entropía (máxima incertidumbre) entre todas las posibles distribuciones de probabilidad definidas en la recta real con igual varianza [114] y existe una relación directa entre la gaussianidad y la entropía (ecuación (2.33)). Cuanto más gaussiana (mayor entropía) es una distribución, menos estructura tienen los datos asociados con ella.

En *ICA*, la utilización de la gaussianidad de las señales se basa en la siguiente idea fundamental:

Las salidas de un sistema lineal, que preserva las varianzas (finitas), tienen siempre entropías mayores que las de las entradas (no-gaussianas) (Fig. 3.1)

Este resultado puede ser expresado en términos matemáticos rigurosos cuando las señales son mutuamente independientes, siendo una consecuencia natural del Teorema Central del Límite, que establece que la distribución de una mezcla lineal de p señales independientes con varianzas finitas tiende asintóticamente a la distribución de Gauss (cuando $p \rightarrow \infty$) [59]. Más aún, puede demostrarse que la combinación lineal finita de señales independientes no-gaussianas tienen entropía mayor que las señales originales, cuando se mantienen las varianzas fijas (ejemplos

de este resultado son las demostraciones de Donoho [55] en su trabajo clásico sobre deconvolución (*deconvolution*) y la demostración de Cruces et al. [52]).

Además, como establecieron Comon en [49] y Donoho en [55], las señales gaussianas no son admisibles en *BSS* ya que sus combinaciones lineales (también gaussianas) conservan la entropía. Por lo tanto surge como criterio natural, para la búsqueda de las estimaciones de las fuentes en *ICA*, el criterio de Máxima NG (o mínima entropía) (Fig. 3.1).

Cuando las señales son dependientes entre sí, el Teorema Central del Límite clásico no se cumple en general. Sin embargo, en [55], Donoho mostró la utilidad del criterio de máxima NG para ejemplos particulares de señales dependientes. En esta tesis, se presenta una teoría que da sustento a este criterio estableciendo una condición suficiente sobre las fuentes, que garantiza su separabilidad. A continuación, se presenta una definición formal del criterio de máxima NG que se utiliza en esta tesis:

Criterio de máxima NG (mínima entropía): *Consiste en buscar la matriz de separación D que produzca estimaciones de las fuentes de máxima NG (mínima entropía), manteniendo la varianza unitaria. Más específicamente, las fuentes son estimadas a través de la parametrización (3.2) y el espacio de búsqueda es el de las matrices D que hacen que las señales y_0, y_1, \dots, y_{M-1} tengan varianza unitaria (lo cual es equivalente a imponer que la matriz de covarianzas $C_{yy} = DC_{xx}D^T$ tenga 1's en su diagonal principal).*

3.2. La medida NG basada en la distancia en $L^2(\mathbb{R})$

En esta tesis se propone una forma natural de medir la NG de una variable utilizando la distancia en el espacio de funciones de cuadrado integrable Lebesgue $L^2(\mathbb{R})$ tal como se define a continuación [30].

Medida NG: *Dada una variable aleatoria y con media cero $E[y] = 0$ y varianza unitaria $E[y^2] = 1$, cuya pdf es $f_y(y)$, se define la medida NG como el cuadrado de la distancia entre $f_y(y)$ y la pdf de Gauss $\Phi(y)$ en $L^2(\mathbb{R})$, es decir:*

$$\Gamma_{NG}(f_y) = \int_{-\infty}^{+\infty} [\Phi(y) - f_y(y)]^2 dy \quad (3.4)$$

donde $\Phi(y)$ es la pdf de Gauss (distribución normal) con media cero y varianza unitaria definida por:

$$\Phi(y) = N(0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \quad (3.5)$$

Se observa que $\Gamma_{NG}(f_y) \geq 0$ para cualquier pdf $f_y(y)$ y además $\Gamma_{NG}(f_y) = 0$ si y solo si $\Phi(y) = f_y(y)$ en casi todo punto.

3.2.1. Calculo de $\Gamma_{NG}(f_y)$ usando ventanas de Parzen

En la práctica no se conoce de una expresión para la pdf $f_y(y)$ sino que se dispone de un conjunto de n muestras de la variable y : $y(0), y(1), y(2), \dots, y(n-1)$.

Estas muestras, pueden ser utilizadas para estimar la *pdf* $\hat{f}_y(y)$ y así poder evaluar la medida NG de (3.4).

En esta tesis se propone utilizar la técnica de estimación no paramétrica denominada “ventanas de Parzen”, que fuera propuesta y analizada originalmente por Parzen en [107]. Este método establece que, si el número de muestras es suficientemente grande, se obtiene una buena estimación $\hat{f}_y(y)$ de la *pdf* desconocida a través de la siguiente fórmula:

$$\hat{f}_y(y) = \frac{1}{nh} \sum_{i=0}^{n-1} \Phi\left(\frac{y - y(i)}{h}\right) \quad (3.6)$$

donde n es el número de muestras, h es el parámetro de apertura que determina el ancho y altura de las funciones ventana utilizadas $\Phi(y)$ (también conocidas como *kernels*). En nuestro caso nos restringimos al caso en el que la función $\Phi(y)$ es la *pdf* de Gauss (definida en (3.5)), ya que, como se verá más adelante, se obtienen ecuaciones sencillas y útiles. En su trabajo original [107], Parzen analizó las condiciones mínimas que garantizan la convergencia de (3.6) a la *pdf* deseada y en [115] se demostró que el valor óptimo para h , que provee un mínimo MISE (*Mean Integrated Square Error*), está dado por $h = 1,06 \times n^{-1/5}$. A partir de estos resultados a continuación se desarrolla un nuevo método para el cálculo de la medida NG.

Se observa que la ecuación (3.4) puede dividirse en tres partes:

$$\Gamma_{NG}(f_y) = \underbrace{\int \Phi^2(y) dy}_{\frac{1}{2\sqrt{\pi}}} - 2 \underbrace{\int \Phi(y) f_y(y) dy}_{\Gamma_1(f_y)} + \underbrace{\int f_y^2(y) dy}_{\Gamma_2(f_y)} \quad (3.7)$$

El primer término del miembro derecho en (3.7) puede calcularse analíticamente y vale $\frac{1}{2\sqrt{\pi}}$. El segundo y tercer término ($\Gamma_1(f_y)$ y $\Gamma_2(f_y)$) dependen de la *pdf* $f_y(y)$, por lo tanto, si se la reemplaza por su estimación de Parzen $\hat{f}_y(y)$ (3.6) se obtiene (demostración en Apéndice 3.9.1):

$$\Gamma_1(\hat{f}_y) = \frac{-2}{n\sqrt{h^2+1}} \sum_{i=0}^{n-1} \Phi\left(\frac{y(i)}{\sqrt{h^2+1}}\right) \quad (3.8)$$

$$\Gamma_2(\hat{f}_y) = \frac{1}{n^2 h \sqrt{2}} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \Phi\left(\frac{y(j) - y(i)}{\sqrt{2}h}\right) \quad (3.9)$$

Las ecuaciones (3.7), (3.8) y (3.9) proveen una forma práctica de calcular la medida NG a partir de un conjunto de muestras de la señal y serán utilizadas luego para el desarrollo del algoritmo MAXNG (Sección 3.4).

3.3. Separabilidad de fuentes dependientes

En esta tesis se presenta una condición suficiente para la separación de fuentes dependientes a través de la medida NG [35]. Para simplificar el tratamiento mate-

mático se considera el caso con $p = 2$ fuentes de media nula $E[s_0] = E[s_1] = 0$ y se definen las fuentes normalizadas u_0 y u_1 de la siguiente manera:

$$u_0 = \frac{s_0}{\sigma_0}; \quad u_1 = \frac{s_1}{\sigma_1} \quad (3.10)$$

donde $\sigma_0^2 = E[s_0^2]$ y $\sigma_1^2 = E[s_1^2]$ son las varianzas de las variables s_0 y s_1 , respectivamente.

Usando (3.1) y (3.2) se obtiene la estimación del vector de las fuentes normalizadas $\hat{\mathbf{u}}$ en términos del vector de las fuentes \mathbf{s} , la matriz de mezcla A y la matriz de separación D :

$$\hat{\mathbf{u}} = DAs \quad (3.11)$$

De (3.11) se deduce que la estimación de las fuentes normalizadas es una combinación lineal de las fuentes s_0 y s_1 :

$$\hat{u}_i = \alpha_{i,0}s_0 + \alpha_{i,1}s_1 \quad (3.12)$$

donde los parámetros $\alpha_{i,j}$ son los elementos de la matriz DA , i.e., $\alpha_{i,j} = [DA]_{i,j}$. En términos de las fuentes normalizadas u_0 , u_1 , la ecuación (3.12) se convierte en:

$$\hat{u}_i = \alpha u_0 + \beta u_1 \quad (3.13)$$

donde $\alpha = \alpha_{i,0}\sigma_0$ y $\beta = \alpha_{i,1}\sigma_1$. Por lo tanto, la reconstrucción de las fuentes se logra cuando los parámetros cumplen la condición $(\alpha, \beta) = (\pm 1, 0)$ para u_0 y $(\alpha, \beta) = (0, \pm 1)$ para u_1 .

Por otro lado se observa que, forzando a la varianza del estimador a ser unitaria, se obtiene:

$$E[\hat{u}_i^2] = \alpha^2 + \beta^2 + 2\alpha\beta\rho = 1 \quad (3.14)$$

donde $\rho = E[u_0u_1]$ es el coeficiente de correlación entre las variables u_0 y u_1 . En otras palabras, forzar la varianza del estimador a ser unitaria es equivalente a restringir a los parámetros (α, β) a pertenecer a la elipse que se muestra en la Fig. 3.2. El objetivo entonces, es encontrar la condición que garantice que la medida NG tendrá máximos locales en los puntos

$$(\alpha, \beta) = \{(1, 0), (0, 1), (-1, 0), (0, -1)\} \quad (3.15)$$

cuando los parámetros están restringidos a pertenecer a la elipse. En consecuencia, se introduce la definición de *separabilidad* como sigue:

Separabilidad de fuentes: *Dadas dos fuentes normalizadas u_0 y u_1 , se dice que la fuente u_1 es separable de u_0 si la medida $\Gamma_{NG}(f_y; \alpha, \beta)$ de la combinación lineal $y = \alpha u_0 + \beta u_1$, tiene un máximo local en $(\alpha, \beta) = (0, 1)$ cuando los parámetros están restringidos a la elipse $\alpha^2 + \beta^2 + 2\alpha\beta\rho = 1$.*

El siguiente teorema provee una condición suficiente para la separabilidad de las fuentes.

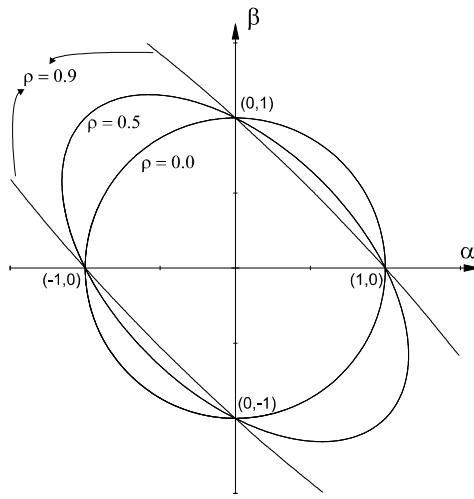


FIGURA 3.2. Conjunto de valores permitidos para los parámetros α y β para $\rho = 0.0, 0.5$ y 0.9 .

Teorema 3. Dadas dos fuentes dependientes s_0 y s_1 , si la esperanza condicional $E[s_0|s_1]$ es lineal en s_1 , es decir, si existen parámetros a y b tales que $E[s_0|s_1] = as_1 + b$, entonces se cumplen las siguientes propiedades:

I) Los parámetros a y b pueden ser calculados como:

$$a = \frac{\sigma_0}{\sigma_1} \rho \quad (3.16)$$

$$b = \bar{s}_0 - \bar{s}_1 \frac{\sigma_0}{\sigma_1} \quad (3.17)$$

donde ρ es el coeficiente de correlación $\rho = E[u_0 u_1]$, σ_0 y σ_1 son las desviaciones estándar, $\bar{s}_0 = E[s_0]$, $\bar{s}_1 = E[s_1]$ son los valores medios (que consideraremos nulos).

II) La esperanza condicional de las fuentes normalizadas es: $E[u_0|u_1] = \rho u_1$

III) La fuente normalizada u_1 es separable de u_0

Dem. Apéndice 3.9.2 □

La principal consecuencia de este teorema es que cuando las fuentes tienen esperanzas condicionadas lineales, entonces son separables y pueden ser extraídas encontrando los máximos locales de la medida NG. Se destaca que, la condición del teorema siempre se cumple para el caso independiente ya que $E[s_0|s_1] = E[s_0] = 0$, es decir que, como corolario, se demostró que las señales independientes son separables con la medida NG.

3.4. El algoritmo MaxNG

En esta tesis, se propone el algoritmo MAXNG (*Maximum Non-Gaussianity*) basado en el criterio de máxima NG definido anteriormente para separar fuentes

dependientes o independientes [29, 30]. En esta sección se presentan los detalles de su implementación para el caso general de m mezclas (sensores) y p fuentes con $m \geq p$ (caso sobredeterminado) que es donde se puede garantizar la existencia de una matriz de separación D tal que $DA = I$.

3.4.1. Filtrado previo de las señales mezcla

Como es usual en *ICA*, la aplicación de un Filtro Decorrelador - FD (Cap 2, Subsección 2.1.4) de las señales mezclas facilita el posterior procesamiento de los datos. En particular, en MAXNG este filtrado previo tiene dos objetivos:

- 1) Reducir la dimensión de los datos a procesar, especialmente cuando $m \gg p$.
- 2) Simplificar la estructura de la matriz de separación buscada forzándola a tener filas de norma unitaria.

El FD se basa en aplicar *PCA* (o *KLT*) a los datos disponibles \mathbf{x} definiendo un nuevo conjunto de mezclas no-correlacionadas $\mathbf{\hat{x}}$ de la siguiente forma:

$$\mathbf{\hat{x}} = \Lambda^{-1/2} V^T \mathbf{x} \quad (3.18)$$

donde V es una matriz de $m \times q$ ($q \leq p \leq m$) con los q autovectores dominantes (cuyos autovalores son no nulos) de la matriz de covarianzas $C_{\mathbf{xx}} = E[\mathbf{xx}^T]$, y Λ es una matriz diagonal de $q \times q$ con los autovalores correspondientes en su diagonal principal. Cabe señalar que, en el caso particular en que $m = p = q$, también puede utilizarse el FD propuesto por Hyvärinen et al en *ICA* [75]:

$$\mathbf{\hat{x}} = V \Lambda^{-1/2} V^T \mathbf{x} \quad (3.19)$$

A través de esta operación lineal, las mezclas no-correlacionadas resultan tener una matriz de covarianzas igual a la identidad como puede ser verificado directamente a partir de (3.18), i.e.

$$C_{\mathbf{\hat{x}}\mathbf{\hat{x}}} = E[\mathbf{\hat{x}}\mathbf{\hat{x}}^T] = I \quad (3.20)$$

Si se define la separación en términos de las mezclas no-correlacionadas se obtiene:

$$\mathbf{y} = \mathbf{\hat{s}} = \mathring{D} \mathbf{\hat{x}} \quad (3.21)$$

donde la nueva matriz de separación \mathring{D} tiene la propiedad de poseer filas de norma unitaria ya que

$$C_{\mathbf{yy}} = E[\mathbf{yy}^T] = \mathring{D} \mathring{D}^T \quad (3.22)$$

y se sabe que las señales estimadas están restringidas a tener varianzas unitarias $E[y_i^2] = 1$.

Por otro lado la matriz de separación real D está relacionada con \mathring{D} de la siguiente manera:

$$D = \mathring{D} \Lambda^{-1/2} V^T \quad (3.23)$$

En la Fig. 3.3 se muestra la secuencia de transformaciones lineales aplicadas sobre las fuentes \mathbf{s} .

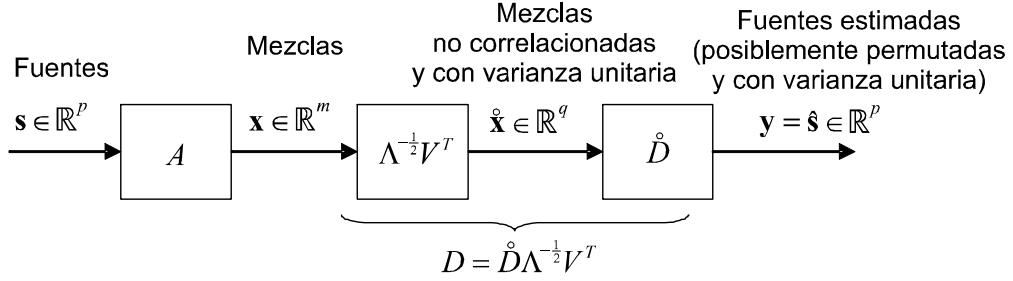


FIGURA 3.3. Secuencia de transformaciones lineales aplicadas sobre las fuentes \mathbf{s} donde $m \geq p \geq q$

3.4.2. Parametrización de la matriz de separación \mathring{D}

Una vez aplicado el FD (3.18), el algoritmo MAXNG se reduce a la búsqueda de p máximos locales de la medida $\Gamma_{NG}(f_y)$, donde y es una combinación lineal de las mezclas no-correlacionadas, i.e.

$$y = \mathring{\mathbf{d}}^T \mathring{\mathbf{x}} = \mathring{d}_0 \mathring{x}_0 + \mathring{d}_1 \mathring{x}_1 + \dots + \mathring{d}_{q-1} \mathring{x}_{q-1} \quad (3.24)$$

con la restricción

$$\left\| \mathring{\mathbf{d}} \right\|^2 = \sum_{i=0}^{q-1} \mathring{d}_i^2 = 1 \quad (3.25)$$

ya que la varianza de y es unitaria y las variables $\mathring{x}_0, \mathring{x}_1, \dots, \mathring{x}_{q-1}$ son no-correlacionadas (ecuación (3.20)). Por lo tanto el espacio de los parámetros $\mathring{d}_0, \mathring{d}_1, \dots, \mathring{d}_{q-1}$ está determinado por una hiper-esfera de radio unitario en \mathbb{R}^q . Cada vector $\mathring{\mathbf{d}} \in \mathbb{R}^q$ para el cual la medida $\Gamma_{NG}(f_y)$ tiene un máximo, es una posible fila de la matriz de separación \mathring{D} .

Para simplificar la búsqueda de los máximos locales, se propone utilizar coordenadas hiper-esféricas las cuales requieren de $q - 1$ ángulos para definir una fila genérica de la matriz \mathring{D} indicada por $\mathring{\mathbf{d}} = [\mathring{d}_0 \mathring{d}_1 \dots \mathring{d}_{q-1}]$, y que está relacionada con el vector de ángulos (coordenadas hiper-esféricas) $\boldsymbol{\theta}^T = [\theta_0 \theta_1 \dots \theta_{q-2}]$ de la siguiente manera [70]:

$$\mathring{\mathbf{d}}^T = \begin{bmatrix} \mathring{d}_0 \\ \mathring{d}_1 \\ \vdots \\ \mathring{d}_{q-2} \\ \mathring{d}_{q-1} \end{bmatrix} = \begin{bmatrix} \cos \theta_0 \\ \sin \theta_0 \cos \theta_1 \\ \vdots \\ \left(\prod_{n=0}^{q-3} \sin \theta_n \right) \cos \theta_{q-2} \\ \prod_{n=0}^{q-2} \sin \theta_n \end{bmatrix} \quad (3.26)$$

Existe una relación interesante entre la matriz de separación buscada \mathring{D} y la matriz de covarianzas de las fuentes C_{ss} que se obtiene reemplazando $\mathbf{y} = \mathbf{s}$ en (3.22) y resulta

$$C_{ss} = \mathring{D} \mathring{D}^T \quad (3.27)$$

Por ejemplo, si las fuentes son independientes, la matriz de separación buscada es una matriz ortogonal $\hat{D}\hat{D}^T = I$. Por otro lado cuanto más correlacionadas sean las fuentes, más cercanos serán los vectores asociados (menor ángulo entre ellos).

3.4.3. Búsqueda de máximos locales

Para garantizar la convergencia a máximos locales se propone utilizar una búsqueda iterativa basada en el gradiente (*gradient ascend technique*) [57] donde el vector de ángulos en la iteración $(k+1)$, $\boldsymbol{\theta}^{(k+1)} = [\theta_0^{(k+1)} \ \theta_1^{(k+1)} \dots \theta_{q-2}^{(k+1)}]^T$ se actualiza de la siguiente manera:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \varepsilon^{(k)} \nabla_{\boldsymbol{\theta}} \Gamma_{NG}(f_y) \quad (3.28)$$

con $\varepsilon^{(k)}$ siendo un escalar que puede ajustarse a medida que se avanza en las iteraciones y $\nabla_{\boldsymbol{\theta}} \Gamma_{NG}(f_y)$ es el gradiente de la medida NG correspondiente a los parámetros actuales.

En el algoritmo MAXNG se seleccionan al azar un número arbitrario de puntos iniciales de búsqueda y, para evitar caer en los mismos máximos locales repetidas veces, una vez que un máximo local es detectado, se procede a eliminarlo (Subsección 3.4.4).

La ventaja de disponer de una fórmula de la medida NG basada en las muestras de la señal a través de las ventanas de Parzen, es que permiten calcular explícitamente sus derivadas respecto a los parámetros de búsqueda (coordenadas hiper-esféricas). Usando la regla de la cadena de las derivadas, se calcula el gradiente de la medida NG respecto del vector de ángulos $\boldsymbol{\theta}^T = [\theta_0 \ \theta_1 \dots \theta_{q-2}]$ de la siguiente manera:

$$\nabla_{\boldsymbol{\theta}} \Gamma_{NG}(f_y) = \left[\frac{\partial \Gamma_{NG}}{\partial \theta_0} \ \frac{\partial \Gamma_{NG}}{\partial \theta_1} \ \dots \ \frac{\partial \Gamma_{NG}}{\partial \theta_{q-2}} \right] = \nabla_{\dot{\mathbf{d}}} \Gamma_{NG}(f_y) \cdot \frac{\partial \dot{\mathbf{d}}}{\partial \boldsymbol{\theta}} \quad (3.29)$$

donde

$$\frac{\partial \dot{\mathbf{d}}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \dot{d}_0}{\partial \theta_0} & \frac{\partial \dot{d}_0}{\partial \theta_1} & \dots & \frac{\partial \dot{d}_0}{\partial \theta_{q-2}} \\ \frac{\partial \dot{d}_1}{\partial \theta_0} & \frac{\partial \dot{d}_1}{\partial \theta_1} & \dots & \frac{\partial \dot{d}_1}{\partial \theta_{q-2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \dot{d}_{q-1}}{\partial \theta_0} & \frac{\partial \dot{d}_{q-1}}{\partial \theta_1} & \dots & \frac{\partial \dot{d}_{q-1}}{\partial \theta_{q-2}} \end{bmatrix} \quad (3.30)$$

La ecuación (3.30) es la derivada matricial del vector $\dot{\mathbf{d}}$ respecto de $\boldsymbol{\theta}$ y puede ser calculada analíticamente a partir de (3.26). Por otro lado, el gradiente de la medida NG con respecto a $\dot{\mathbf{d}}$ puede ser obtenido a partir de (3.8), (3.9) y (3.24) arribando a:

$$\nabla_{\dot{\mathbf{d}}}^T \Gamma_{NG}(\hat{f}_y) = \nabla_{\dot{\mathbf{d}}}^T \Gamma_1(\hat{f}_y) + \nabla_{\dot{\mathbf{d}}}^T \Gamma_2(\hat{f}_y) \quad (3.31)$$

$$\nabla_{\dot{\mathbf{d}}}^T \Gamma_1(\hat{f}_y) = \frac{2}{n\sqrt{h^2 + 1}} \sum_{i=0}^{n-1} y(i) \Phi \left(\frac{y(i)}{\sqrt{h^2 + 1}} \right) \dot{\mathbf{x}}(i) \quad (3.32)$$

$$\nabla_{\dot{\mathbf{d}}}^T \Gamma_2(\hat{f}_y) = \frac{\sqrt{2}}{n^2 h} \sum_{i=0}^{n-1} \frac{\dot{\mathbf{x}}(i)}{\sqrt{2} h} \sum_{j=0}^{n-1} \left(\frac{y(j) - y(i)}{\sqrt{2} h} \right) \Phi \left(\frac{y(j) - y(i)}{\sqrt{2} h} \right) \quad (3.33)$$

3.4.4. Eliminación de máximos locales

Cada vez que MAXNG detecta una máximo local se debe proceder a eliminarlo para evitar la convergencia a un mismo máximo en diferentes instancias de la búsqueda. Esta idea, conocida en inglés como *deflation*, ha sido propuesta originalmente por Friedman et al en *Projection Pursuit* [61], y consiste en aplicar una transformación no lineal a los datos haciéndolos gaussianos en la dirección de proyección donde la estructura no-gaussianas ha sido detectada (máximo local).

Supongamos que se encuentra un máximo local en la proyección dada por el vector \mathbf{d}^* , es decir que $y^* = \mathbf{d}^{*T} \mathbf{\hat{x}}$ es una variable no-gaussianas. En [61] se propone la siguiente transformación:

$$\mathbf{w} = U^T \Theta (U \mathbf{\hat{x}}) \quad (3.34)$$

donde U^T es una matriz ortogonal cuya primer fila está dada por el vector \mathbf{d}^{*T} y Θ es un operador no-lineal dado definido por:

$$\Theta (\mathbf{u}^T) = \Theta \left([u_0 \ u_1 \dots u_{q-1}]^T \right) = \left[G^{-1} (F_{u_0}(u_0)) \ u_1 \dots u_{q-1} \right]^T \quad (3.35)$$

con G^{-1} siendo la inversa de función de probabilidad acumulativa (*cumulative distribution function - cdf*) de Gauss y F_{y^*} es la *cdf* de y^* , i.e.:

$$G(x) = \int_{-\infty}^x \Phi(u) du \quad (3.36)$$

$$F_{y^*}(x) = \int_{-\infty}^x f_{y^*}(u) du \quad (3.37)$$

Friedman demostró que la transformación (3.34) produce un nuevo vector \mathbf{w} cuya distribución es lo más cercana posible a la distribución original de $\mathbf{\hat{x}}$, con la condición que la distribución marginal en la dirección de proyección dada por \mathbf{d}^* es gaussiana. En esta tesis se introduce una nueva técnica para implementar la transformación de Friedman utilizando ventanas de Parzen. En esta tesis se propone estimar la *cdf* $F_{y^*}(y^*)$ integrando el estimador de Parzen [35]:

$$F_{y^*}(y^*) \approx \frac{1}{n} \sum_{i=0}^{n-1} G \left(\frac{y - y(i)}{h} \right) \quad (3.38)$$

y de esta manera se obtiene una forma práctica de remover las estructuras no-gaussianas en la dirección de proyección dada por \mathbf{d}^* .

3.4.5. Cálculo acelerado de la medida NG y sus derivadas

Las ecuaciones (3.8)-(3.9) y (3.32)-(3.33) proveen un cálculo directo de la medida NG y sus derivadas a partir de las muestras de la señal, pero son computacionalmente muy costosas. Se requiere un total de $(n + n^2)$ evaluaciones de la función de Gauss para el cálculo de la medida NG y de su gradiente. La complejidad computacional es $\mathcal{O}(n^2)$ y esto representa un problema cuando el tamaño de los datos n es elevado, lo cual es muy habitual en las aplicaciones.

Es sabido que los estimadores basados en ventanas de Parzen pueden ser implementados de una manera optimizada si las muestras están ubicadas sobre una grilla regular, ya que las sumatorias en las ecuaciones se convierten en convoluciones, que pueden ser calculadas rápidamente a través de la transformada rápida de Fourier (*Fast Fourier Transform - FFT*) [115]. En esta tesis se desarrolla una nueva aproximación basada en esta propiedad [35], como se explica a continuación.

Un estimador general de Parzen basado en un conjunto de n muestras: $y(0), y(1), \dots, y(n-1)$, tiene la siguiente forma:

$$S(t) = \sum_{i=0}^{n-1} \Psi(t - y(i)) \quad (3.39)$$

donde Ψ es una función continua (una ventana de Gauss o su derivada por ejemplo) que tiende a cero en infinito ($|\Psi(t)| \rightarrow 0$ cuando $t \rightarrow \pm\infty$). Se divide el rango continuo de la variable y en n_Δ segmentos, es decir, si los valores de las muestras están en el segmento $[a, b]$, se define la grilla formada por n_Δ segmentos de ancho $\Delta = (b - a) / n_\Delta$ y cada segmento está definido por $\delta(k) = [a + k\Delta, a + (k + 1)\Delta)$ para $k = 0, 1, \dots, n_\Delta - 1$.

La aproximación de la suma $S(t)$ que se propone es:

$$S(t) = \hat{S}(t) \approx \sum_{k=0}^{n_\Delta-1} f(k) \Psi(t - \bar{\delta}(k)) \quad (3.40)$$

donde $f(k)$ es el número de muestras en el segmento k -ésimo, y $\bar{\delta}(n) = a + (k + \frac{1}{2})\Delta$ corresponde a la posición central del segmento k -ésimo. En la Fig. 3.4 se ilustra esta aproximación a través de un ejemplo sencillo. En la Fig. 3.4. a) se muestra la estimación resultante de la *pdf* a partir de $n = 6$ muestras a través del estimador de Parzen (3.6), en las abcisas se indican las ubicaciones de las muestras y las ventanas gaussianas asociadas a cada una de ellas. En la Fig. 3.4. b) se muestran las ventanas gaussianas centradas en cada uno de los segmentos multiplicadas por el número de muestras pertenecientes a cada segmento $f(k-1), f(k), f(k+1)$, etc. En la Fig. 3.4. c) se muestran las sumas Original y Aproximada superpuestas las cuales demuestran ser prácticamente coincidentes.

Es obvio que la aproximación será más precisa a medida que se aumente el número de segmentos considerados. En la práctica, se obtienen muy buenos resultados para un número de segmentos $n_\Delta \ll n$ (por ejemplo para $n_\Delta = 1024$ y $n = 65536$). Sin embargo la gran ventaja de la aproximación (3.40) es que representa una convolución y, como tal, puede calcularse utilizando el algoritmo *FFT* reduciendo más aún su complejidad.

Finalmente, aplicando la aproximación (3.40) se obtienen las siguientes ecua-

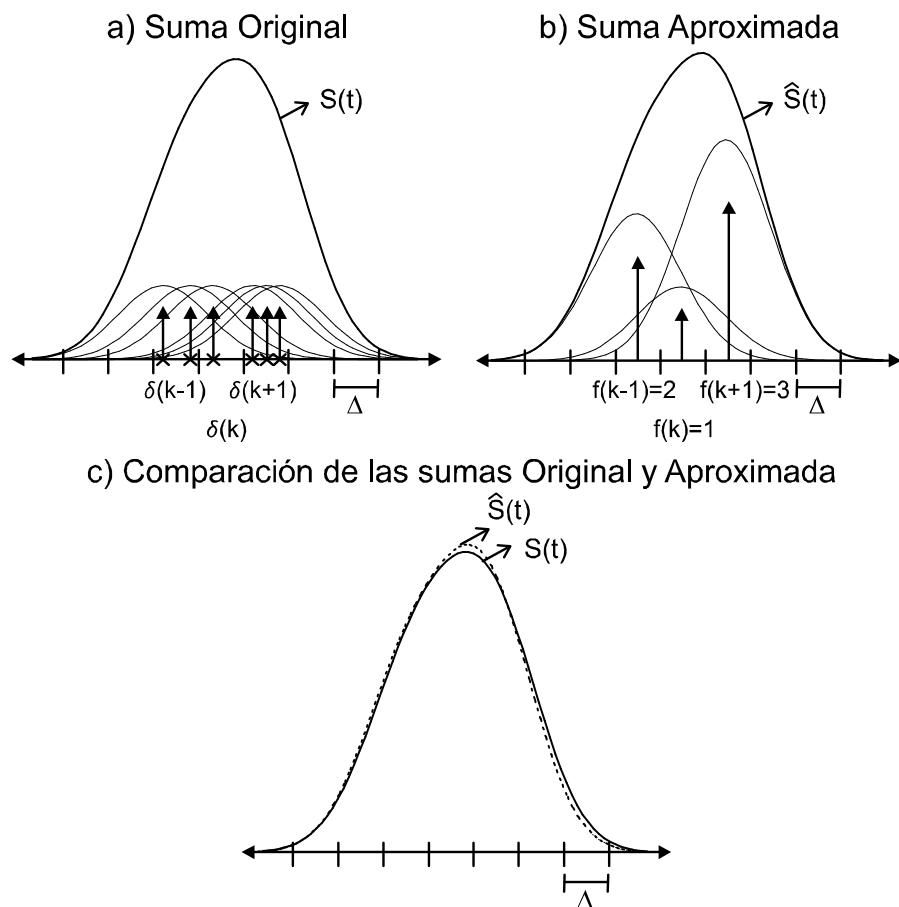


FIGURA 3.4. Ejemplo de aproximación de un estimador basado en ventanas de Parzen.

ciones para la medida NG (demonstración en el Apéndice 3.9.3):

$$\Gamma_{NG}(\hat{f}_y) = \frac{1}{2\sqrt{\pi}} + \Gamma_1(\hat{f}_y) + \Gamma_2(\hat{f}_y) \quad (3.41)$$

$$\Gamma_1(\hat{f}_y) \approx \frac{1}{n} \sum_{k=0}^{n_\Delta-1} f(k) \Phi_1(k) \quad (3.42)$$

$$\Gamma_2(\hat{f}_y) \approx \frac{1}{n} \sum_{k=0}^{n_\Delta-1} f(k) (f * \Phi_2)(k) \quad (3.43)$$

$$\Phi_1(k) = \frac{-2}{\sqrt{h^2 + 1}} \Phi \left(\frac{\bar{\delta}(k)}{\sqrt{h^2 + 1}} \right) \quad (3.44)$$

$$\Phi_2(k) = \frac{1}{n\sqrt{2}h} \Phi \left(\frac{k\Delta}{\sqrt{2}h} \right) \quad (3.45)$$

donde $(f * \Phi_2)(k)$ es la convolución de las funciones f y Φ_2 . Equivalentemente, para las derivadas de la medida NG se obtiene (demonstración en el Apéndice 3.9.3):

$$\frac{\partial \Gamma_{NG}(\hat{f}_y)}{\partial d_l} = \frac{\partial \Gamma_1(\hat{f}_y)}{\partial d_l} + \frac{\partial \Gamma_2(\hat{f}_y)}{\partial d_l} \quad (3.46)$$

$$\frac{\partial \Gamma_1(\hat{f}_y)}{\partial d_l} \approx \frac{2}{n} \sum_{k=0}^{n_\Delta-1} g_l(k) \Phi_3(k) \quad (3.47)$$

$$\frac{\partial \Gamma_2(\hat{f}_y)}{\partial d_l} \approx \frac{-1}{n} \sum_{k=0}^{n_\Delta-1} g_l(k) (f * \Phi_4)(k) \quad (3.48)$$

$$\Phi_3(k) = \frac{\bar{\delta}(k)}{n(h^2 + 1)} \Phi \left(\frac{\bar{\delta}(k)}{\sqrt{h^2 + 1}} \right) \quad (3.49)$$

$$\Phi_4(k) = \frac{k\Delta}{nh^2} \Phi \left(\frac{k\Delta}{\sqrt{2}h} \right) \quad (3.50)$$

donde $g_l(k)$ es proporcional a la suma de las muestras no-correlacionadas $\dot{x}_l(i)$ tal que su correspondiente proyección $y(i)$ (ecuación (3.24)) pertenezca al segmento $\delta(k)$:

$$g_l(k) = \frac{1}{\sqrt{h^2 + 1}} \sum_{i/y(i) \in \delta(k)} \dot{x}_l(i) \quad (3.51)$$

Para analizar la complejidad computacional final de estas aproximaciones, se observa que el cálculo de las funciones $f(k)$ y $g_l(k)$ tienen complejidad $\mathcal{O}(n)$ y la parte más pesada de los cálculos está asociada a las convoluciones las cuales tienen una complejidad de $\mathcal{O}(n_\Delta \log n_\Delta)$. En la práctica se obtienen aproximaciones muy precisas con un número de segmentos $n_\Delta = 1024$ y $n = 65536$ lo cual representa una ahorro significativo en los cálculos teniendo en cuenta que la complejidad original es $\mathcal{O}(n^2)$.

3.5. Ejemplo de separación de dos fuentes dependientes

Con el objetivo de ilustrar sobre el funcionamiento del algoritmo MAXNG, en esta sección se presenta un ejemplo sencillo de separación de dos fuentes dependientes. Las señales utilizadas son reales y han sido extraídas de una imagen satelital cubriendo un área de la superficie de la tierra caracterizada por estructuras geológicas reales. Las señales fuente se han obtenido eligiendo dos columnas de $n = 512$ píxeles de longitud y normalizándolas para forzarlas a tener media cero y varianza unitaria.

En este ejemplo, las fuentes s_0 y s_1 corresponden a columnas contiguas en la imagen (a distancia de un píxel), y por lo tanto tienen un alto nivel de dependencia que se ve reflejado en su coeficiente de correlación $\rho = E[s_0 s_1] \simeq 0,81$ (Fig. 3.5).

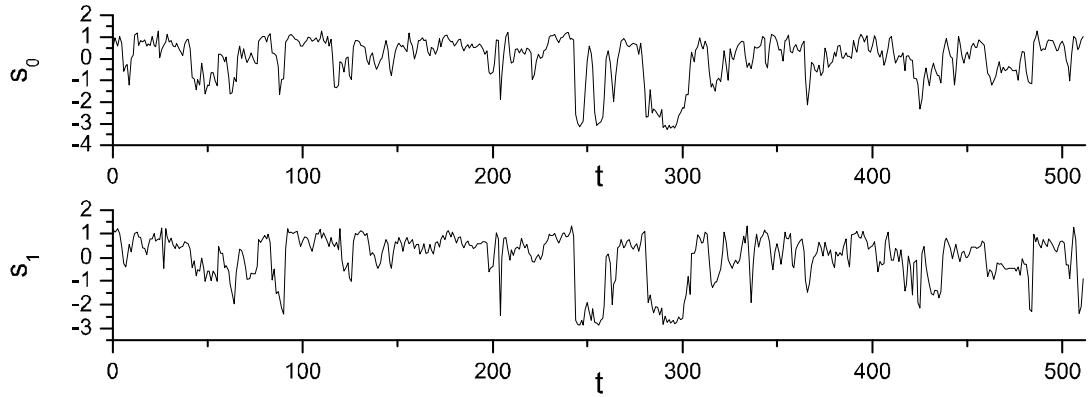


FIGURA 3.5. Señales fuentes dependientes ($\rho = E[s_0 s_1] \simeq 0,81$) correspondientes a dos columnas contiguas de una imagen satelital

Se generan de manera artificial las mezclas a través de (3.1) con una matriz de mezcla arbitraria dada por

$$A = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} \quad (3.52)$$

cuya inversa (matriz de separación) correspondiente es

$$D = A^{-1} = \frac{1}{5} \begin{bmatrix} -\sqrt{10} & 3\sqrt{10} \\ 2\sqrt{10} & -\sqrt{10} \end{bmatrix} \quad (3.53)$$

Aplicando el FD dado por (3.19), se obtienen las mezclas no-correlacionadas \hat{x}_0 y \hat{x}_1 . En la Fig. 3.6, se muestran las *pdf*'s estimadas para las señales s_0 , s_1 , \hat{x}_0 y \hat{x}_1 con sus correspondientes medidas NG.

La matriz de separación objetivo, que puede obtenerse a partir de (3.23), es:

$$\hat{D} = DV\Lambda^{1/2}V^T \simeq \begin{bmatrix} 0,41 & 0,91 \\ 0,87 & 0,50 \end{bmatrix} \quad (3.54)$$

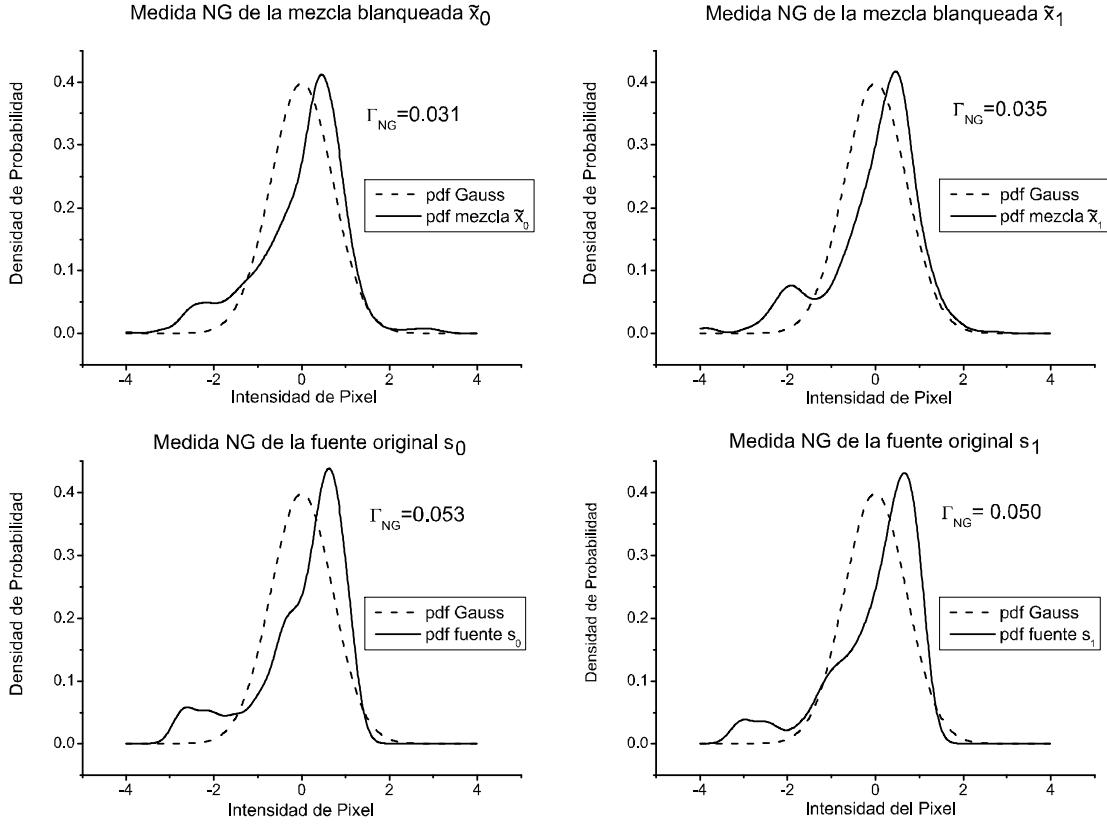


FIGURA 3.6. Estimación de las pdf's de las señales s_0 , s_1 , \tilde{x}_0 , \tilde{x}_1 y sus correspondientes medidas NG.

Usando coordenadas polares θ_0 y θ_1 para las filas de la matriz de separación \hat{D} se obtiene:

$$\hat{D} = \begin{bmatrix} \cos(\theta_0) & \sin(\theta_0) \\ \cos(\theta_1) & \sin(\theta_1) \end{bmatrix} \quad (3.55)$$

La variable proyectada, sobre la que se mide la NG, queda parametrizada con un único parámetro angular θ :

$$y(\theta) = \cos(\theta)\hat{x}_0 + \sin(\theta)\hat{x}_1 \quad (3.56)$$

donde \hat{x}_0 y \hat{x}_1 son las mezclas no-correlacionadas. Además, la medida $\Gamma_{NG}(f_{y(\theta)})$ puede ser calculada en términos de θ usando (3.7), (3.8), (3.9) y (3.56).

Según el algoritmo MAXNG, las estimaciones $\hat{\theta}_0$ y $\hat{\theta}_1$ que permiten estimar las fuentes, se obtienen como los máximos locales de $\Gamma_{NG}(f_{y(\theta)})$. En la Fig. 3.7, se muestra la curva de $\Gamma_{NG}(f_{y(\theta)})$ versus el ángulo θ para este ejemplo. Los máximos locales están ubicados en $\hat{\theta}_0 = 30,0^\circ$ ($\Gamma_{NG} = 0,050$) y $\hat{\theta}_1 = 64,0^\circ$ ($\Gamma_{NG} = 0,053$) que se corresponden con las posiciones teóricas de los mismos: $\theta_0 = \arctan(0,50/0,87) = 29,89^\circ$ y $\theta_1 = \arctan(0,91/0,41) = 65,75^\circ$.

Adicionalmente, se muestra la existencia de un máximo local espurio (solución falsa) en $\theta \simeq 128^\circ$ ($\Gamma = 0,034$), sin embargo los experimentos realizados (Sección

3.8) muestran que la aparición de soluciones falsas está relacionada con el hecho de que el tamaño del conjunto de datos es pequeño, es decir, cuando se considera un mayor número de muestras ($n > 512$) las soluciones falsas tienden a desaparecer como muestran los resultados de la Fig. 3.11.

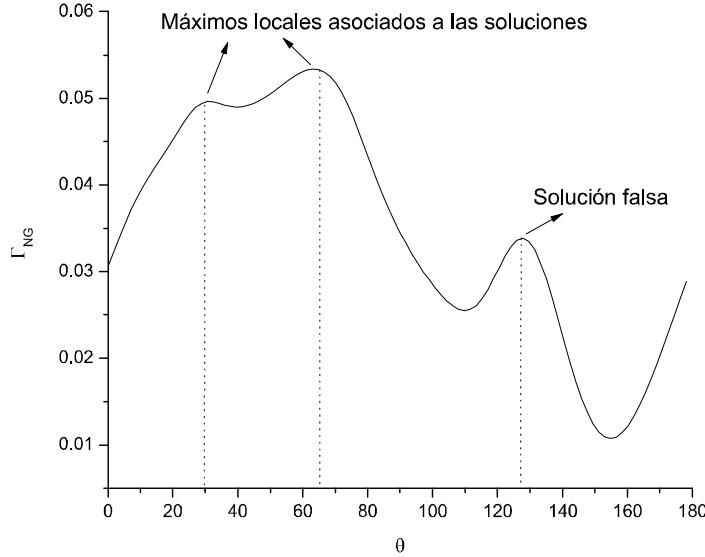


FIGURA 3.7. Medida NG versus parámetro angular θ

A continuación se muestra la relación entre la matriz de covarianzas de las fuentes C_{ss} y las ubicaciones de los máximos locales θ_0 y θ_1 es la siguiente. De las ecuaciones (3.55) y (3.56) se obtiene:

$$E [s_0 s_1] = \cos(\theta_0) \cos(\theta_1) + \sin(\theta_0) \sin(\theta_1) = \cos(\Delta\theta) \quad (3.57)$$

con $\Delta\theta = |\theta_0 - \theta_1|$. Se observa que, cuando $\rho = E [s_0 s_1] \rightarrow 0$ (caso de fuentes no-correlacionadas) entonces $\Delta\theta \rightarrow \pm 90^\circ$ que corresponde al caso en que las fuentes son independientes (*ICA*) y la matriz \tilde{D} es ortogonal. Por otro lado, cuando $\rho = E [s_0 s_1] \rightarrow 1$ (completamente correlacionadas) entonces $\Delta\theta \rightarrow 0^\circ$ lo cual significa que los ángulos θ_0 y θ_1 tienden a ser coincidentes. Esto sugiere que, cuando las fuentes son fuertemente correlacionadas, la detección de los máximos locales podría hacerse más dificultosa.

En la práctica la búsqueda de los máximos se realiza en forma secuencial (un máximo local por vez) a través de una búsqueda guiada por el gradiente y aplicando la técnica de eliminación de máximos explicada anteriormente. En la Fig. 3.8 se muestra un ejemplo particular donde la búsqueda de los máximos se hace a partir de puntos iniciales seleccionados aleatoriamente y, cada vez que un máximo es encontrado, se procede a eliminarlo. En la Fig. 3.8. a) se muestra como el primer máximo es encontrado a partir de una búsqueda guiada por el gradiente (pendiente). En la Fig. 3.8. b), luego de eliminar el primer máximo detectado, se procede a la búsqueda, guiada por el gradiente, del segundo máximo. En la Fig.

3.8. c), luego de eliminar el segundo máximo detectado, se procede a la búsqueda del tercer máximo. Debido a que la eliminación de máximos puede introducir errores, haciendo que los máximos se muevan de sus posiciones, en la Fig. 3.8. d) se muestra la etapa final del proceso, que consiste en realizar nuevamente la búsqueda de los máximos pero esta vez sin eliminación de los mismos, usando como puntos iniciales aquellas posiciones detectadas anteriormente. Cabe aclarar que el orden en que se obtienen los máximos dependerá de los puntos iniciales seleccionados al azar.

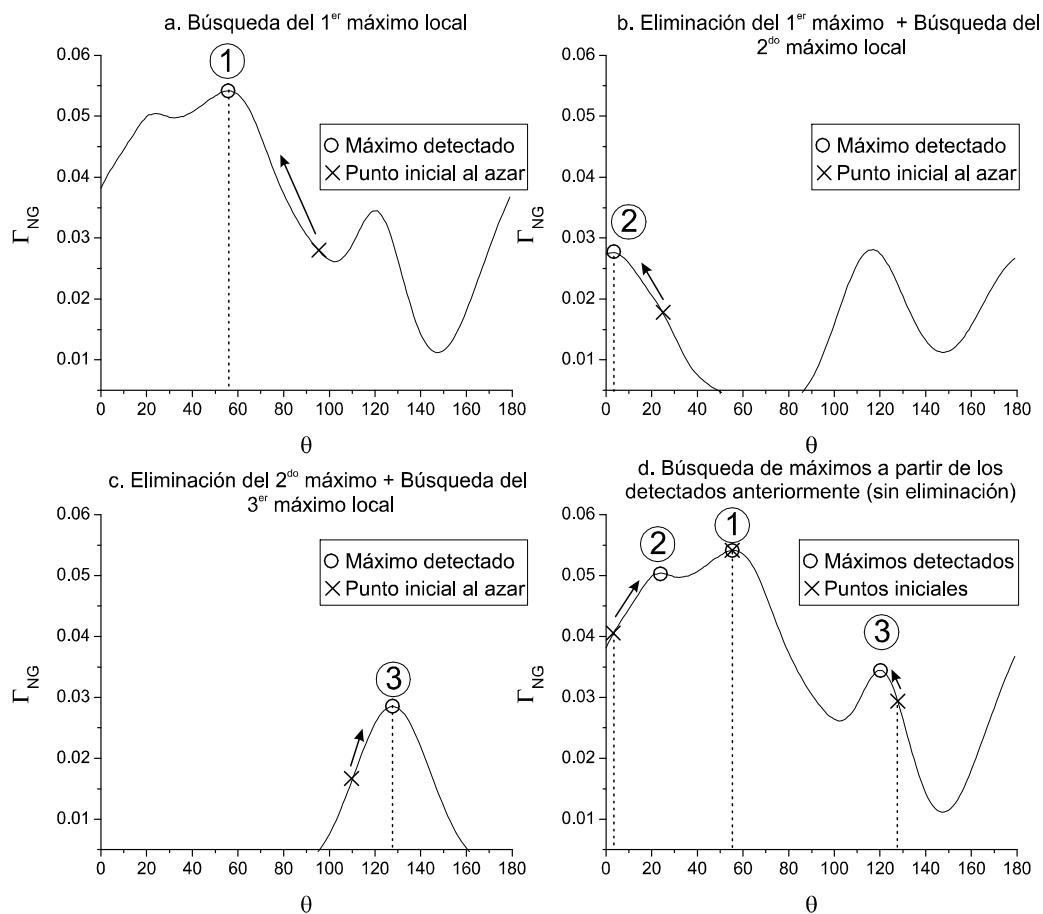


FIGURA 3.8. Etapas en la búsqueda de máximos locales: 1) Se buscan secuencialmente los máximos a partir de puntos iniciales aleatorios (a. b. y c.) y 2) Se buscan máximos sin deflación usando las posiciones antes detectadas como puntos iniciales (d.)

A los efectos de realizar una comparación de *ICA* con *MAXNG*, a continuación se propone evaluar la Información Mutua definida en el Cap. 2 (ecuación (2.23)) de todas las posibles fuentes estimadas. La información mutua puede ser estimada utilizando la técnica de las ventanas de Parzen y, para el caso de dos fuentes, la

formula resultante es [22]:

$$I(\mathbf{y}) = -\frac{1}{n} \sum_{i=0}^{n-1} \log \left\{ \frac{1}{n^2 h^2} \left[\sum_{j=0}^{n-1} \Phi \left(\frac{\Delta^{i,j} y_0}{h} \right) \right] \left[\sum_{j=0}^{n-1} \Phi \left(\frac{\Delta^{i,j} y_1}{h} \right) \right] \right\} - \log |\cos(\theta_0) \sin(\theta_1) - \sin(\theta_0) \cos(\theta_1)| \quad (3.58)$$

donde

$$\begin{aligned} \Delta^{i,j} y_m &= y_m(i) - y_m(j) \\ &= \cos(\theta_m) [\dot{x}_0(i) - \dot{x}_0(j)] - \sin(\theta_m) [\dot{x}_1(i) - \dot{x}_1(j)] \text{ for } m = 0, 1. \end{aligned} \quad (3.59)$$

En la Fig. 3.9 se muestran los valores de $I(\mathbf{y}(\theta_0, \theta_1))$ en el espacio de los parámetros (θ_0, θ_1) para el ejemplo presentado. La ubicación del mínimo global determina la solución de Mínima Información Mutua (solución *ICA*). En este ejemplo, el mínimo se ubica en $\theta_0 = 37,8^\circ$ y $\theta_1 = 127,8^\circ$ (o sus versiones permutadas o invertidas) lo cual representa una estimación equivocada de las posiciones reales de las fuentes ($\theta_0 = 29,89^\circ$ y $\theta_1 = 65,74^\circ$). Este es un claro ejemplo que muestra que, cuando las fuentes son fuertemente correlacionadas, la solución de mínima Información Mutua es errónea.

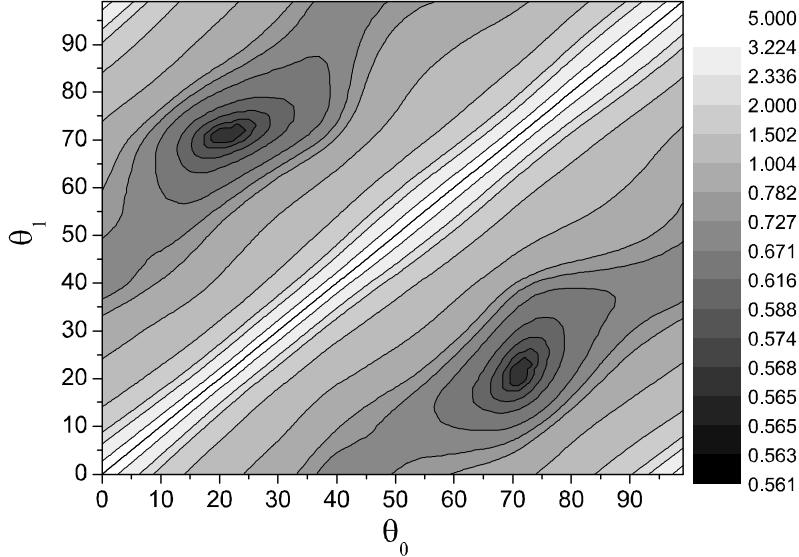


FIGURA 3.9. Información Mutua $I(\mathbf{y})$ versus ángulos θ_0 y θ_1 .

3.6. Ruido aditivo gaussiano y MaxNG

En esta sección, se analiza el efecto del ruido aditivo gaussiano en el algoritmo MAXNG [35]. Es decir, se aplica el algoritmo a un modelo de mezclas dado por:

$$\mathbf{x} = A\mathbf{s} + \mathbf{n} \quad (3.60)$$

donde \mathbf{x} (vector de m mezclas), A (matriz de mezcla) y \mathbf{s} (vector de p fuentes) son los elementos utilizados anteriormente, y \mathbf{n} es un vector de m señales de ruido que, por simplicidad, se asume gaussiano, independiente de las fuentes y con una matriz de covarianzas $C_{\mathbf{nn}} = E[\mathbf{nn}^T]$. Aplicando (3.24) a los datos con ruido (3.60) se obtiene:

$$y(\mathbf{d}) = \mathbf{d}^T A \mathbf{s} + \mathbf{d}^T \mathbf{n} = y_1 + v_1 \quad (3.61)$$

donde y_1 es combinación lineal de las fuentes s_i , y v_1 es gaussiana ya que es combinación lineal de ruidos gaussianos. Dado que y_1 y v_1 son independientes, la *pdf* de y puede escribirse como una convolución:

$$f_y(y) = (f_{y_1} * f_{v_1})(y) \quad (3.62)$$

A la luz de estas observaciones, se identifican dos estrategias diferentes para aplicar MAXNG en el caso con ruido:

Método 1: Maximizar la NG de la *pdf* asociada a la señal con ruido $f_y(y)$: En este caso, se asume que el efecto del ruido, traducido en la convolución (3.62), no afecta demasiado a las posiciones de los máximos locales. En esta tesis se presentan resultados experimentales que demuestran que, para un amplio rango de valores de *SNR* (*Signal to Noise Ratio*), el ruido no afecta significativamente a los máximos locales (Subsección 3.8.2), lo que permite aplicar MAXNG aún cuando no se dispone de información sobre la matriz de covarianzas del ruido $C_{\mathbf{nn}}$.

Método 2: Maximizar la NG de la *pdf* $f_{y_1}(y_1)$: En este otro caso, se asume que se conoce exactamente el modelo de ruido (matriz de covarianzas $C_{\mathbf{nn}}$) y se propone estimar la *pdf* $f_{y_1}(y_1)$ a partir de $f_y(y)$ a través de la convolución inversa (o deconvolución). Asumiendo que existen las transformadas de Fourier de las *pdf*'s, se aplica la transformada de Fourier a la ecuación (3.62):

$$\tilde{f}_{y_1}(\omega) = \frac{\tilde{f}_y(\omega)}{\tilde{f}_{v_1}(\omega)} \quad (3.63)$$

Esta ecuación es válida para todo valor ω (frecuencia) para el cual la transformada de Fourier de la *pdf* del ruido no se anule $\tilde{f}_{v_1}(\omega) \neq 0$. La ecuación (3.63) indica que la función *pdf* buscada $f_{y_1}(y_1)$ puede obtenerse a partir de la transformada inversa de Fourier. Por otro lado, la *pdf* correspondiente al ruido v_1 es

$$f_{v_1}(v_1) = \frac{1}{\sigma_{v_1}} \Phi \left(\frac{v_1}{\sigma_{v_1}} \right) \quad (3.64)$$

donde la varianza está determinada por $\sigma_{v_1}^2 = \mathbf{d}^T C_{\mathbf{nn}} \mathbf{d}$.

Introduciendo la fórmula de Parzen (3.6) en la operación de deconvolución (3.63) se obtiene [35]:

$$f_{y_1}(y_1) \approx \frac{1}{nh^*} \sum_{i=0}^{n-1} \Phi \left(\frac{y - y(i)}{h^*} \right) \quad (3.65)$$

donde

$$h^* = \sqrt{h^2 - \sigma_v^2} = \sqrt{h^2 - \mathbf{d}^T C_{\mathbf{nn}} \mathbf{d}} \quad (3.66)$$

Comparando las ecuaciones (3.65) con (3.6) se concluye que el efecto de considerar la deconvolución es modificar el parámetro de apertura de las ventanas de Parzen, reemplazando h por h^* . Por lo tanto, cuando existe ruido aditivo gaussiano, una forma de disminuir su influencia, es reduciendo el parámetro de apertura de ventana. Sin embargo, existe un límite de aplicación de (3.66) ya que, cuando el ruido es demasiado grande ($\sigma_v^2 > h^2$), esta ecuación ya no tiene sentido.

En la práctica este último método (Método 2) tiene la desventaja de que requiere el conocimiento de la matriz de covarianzas del ruido $C_{\mathbf{nn}}$. Además, el parámetro resultante h^* depende del vector \mathbf{d} , y por lo tanto las derivadas de la medida NG en términos de \mathbf{d} resultan más complicadas de calcular si la comparamos con el método anterior (Método 1), donde el parámetro h es fijo para todo \mathbf{d} .

3.6.1. FD modificado para el caso con ruido

En la Subsección 3.4.1 se presentó el FD cuyo objetivo es mantener la varianza de la variable proyectada constante e igual a uno. En el caso con ruido, cuando se aplica el Método 1, el FD es idéntico al utilizado en el caso libre de ruido quedando definido por la matriz de covarianzas de las muestras $C_{\mathbf{xx}}$ a través de sus autovalores y autovectores.

Sin embargo, en el Método 2, es necesario alterar el FD ya que lo que se necesita es mantener la varianza de la variable deconvolucionada y_1 constante e igual a uno (ecuación (3.61)). En este último caso, el FD debe calcularse a partir de la matriz de covarianzas $E[(\mathbf{As})(\mathbf{As})^T]$ (en lugar de $C_{\mathbf{xx}}$), la cual es:

$$E[(\mathbf{As})(\mathbf{As})^T] = E[(\mathbf{x} - \mathbf{n})(\mathbf{x} - \mathbf{n})^T] = C_{\mathbf{xx}} - C_{\mathbf{nn}} \quad (3.67)$$

donde se observa que es necesario conocer el modelo del ruido (matriz de covarianzas $C_{\mathbf{nn}}$). Por lo tanto, en la práctica, donde difícilmente se conoce con precisión el modelo de ruido, resulta más efectiva la aplicación del Método 1 (Maximizar la NG de la pdf asociada a la señal con ruido $f_y(y)$).

3.7. Determinación de factores de escala con fuentes condicionadas

Como fue demostrado por Comon [49], existe una indeterminación de escala en las fuentes detectadas por los algoritmos *BSS*. De hecho, con MAXNG las estimaciones de las fuentes están forzadas a tener varianza unitaria. En esta tesis se analiza un caso muy particular de fuentes dependientes que aparecen naturalmente en aplicaciones de imágenes hiper-espectrales, para las cuales es posible determinar los factores de escala apropiados. Estos resultados fueron presentados

originalmente en [33] y adaptados al caso con ruido en [35]. Se considera el caso en que las señales fuente satisfacen la siguiente condición:

$$\sum_{i=0}^{P-1} s_i = 1 \quad (3.68)$$

la cual aparece, por ejemplo, cuando las fuentes están asociadas a porcentajes de ocupación de materiales en píxeles. En estos casos, las fuentes están condicionadas a satisfacer $0 \leq s_i \leq 1$ y tienen medias no nulas. Además, la condición (3.68) hace que las señales sean necesariamente dependientes, de hecho cualquier señal puede escribirse como combinación lineal de las restantes.

A continuación, se presenta la notación y definiciones básicas que se usarán en este modelo:

- $\bar{s}_i = E[s_i]$ y $\bar{x}_i = E[x_i]$ son los *valores medios* de la fuente s_i y la mezcla x_i (vectorialmente son $\bar{\mathbf{s}} = E[\mathbf{s}]$ y $\bar{\mathbf{x}} = E[\mathbf{x}]$);
- $s'_i = s_i - \bar{s}_i$ y $x'_i = x_i - \bar{x}_i$ son las *señales centradas* (vectorialmente son $\mathbf{s}' = \mathbf{s} - \bar{\mathbf{s}}$ y $\mathbf{x}' = \mathbf{x} - \bar{\mathbf{x}}$).
- $u_i = \frac{s'_i}{\sigma_i} = \frac{s_i - \bar{s}_i}{\sigma_i}$ son las *fuentes normalizadas*, donde $\sigma_i = \sqrt{E[(s_i - \bar{s}_i)^2]}$ es el desvío estándar de la fuente s_i . Notar que $E[u_i] = 0$ y $E[u_i^2] = 1$.

Dado que el vector de las fuentes tiene media $\bar{\mathbf{s}}$ no nula, se aplica la separación a las mezclas centradas \mathbf{x}' , estimándose entonces las fuentes centradas \mathbf{s}' . La estimación de las fuentes reales se obtiene sumando a las estimaciones centradas sus correspondientes medias, es decir:

$$\hat{\mathbf{s}} = D\mathbf{x}' + D\bar{\mathbf{x}} \quad (3.69)$$

Por otro lado, se observa que, la condición (3.68) puede ser expresada en términos de los valores medios y las fuentes centradas de la siguiente manera:

$$\sum_{i=0}^{p-1} \bar{s}_i = 1 \quad (3.70)$$

$$\sum_{i=0}^{p-1} s'_i = 0 \quad (3.71)$$

A continuación, se demostrará que las condiciones (3.70) y (3.71) pueden ser utilizadas para remover toda indeterminación en los factores de escala. Supongamos que se ha encontrado la matriz de separación D que separara las fuentes para el caso con ruido (3.60). A partir de (3.61) podemos escribir:

$$y'_i = h_i s'_i + v_i \quad (3.72)$$

donde v_i es una señal gaussiana. Las constantes h_0, h_1, \dots, h_{P-1} son entonces los factores de escala de los cuales no se tiene información. Por conveniencia se define $q_i = 1/h_i$ y se observa que la condición sobre las fuentes centradas (3.71) puede escribirse de la siguiente manera:

$$\sum_{i=0}^{p-1} q_i(y'_i - v_i) = 0 \quad (3.73)$$

Si bien no es posible encontrar los valores de q_i que hagan cumplir esta igualdad idénticamente, se puede minimizar el error cuadrático medio (*Mean Squared Error* - *MSE*) definido por:

$$\varepsilon^2 = E \left[\left(\sum_{i=0}^{p-1} q_i(y'_i - v_i) \right)^2 \right] = \sum_{i=0}^{p-1} \sum_{j=0}^{p-1} q_i (E[y'_i y'_j] - E[v_i v_j]) q_j \quad (3.74)$$

donde se ha utilizado que $E[z'_i v_j] = E[z'_j v_i] = E[v_i v_j]$ por ser s'_i y n_i independientes.

Matricialmente, la fórmula anterior es:

$$\varepsilon^2 = \mathbf{q}^T (C_{\mathbf{y}'\mathbf{y}'} - C_{\mathbf{v}\mathbf{v}}) \mathbf{q} = \mathbf{q}^T R \mathbf{q} \quad (3.75)$$

donde $\mathbf{q}^T = [q_0 \ q_1 \dots q_{p-1}]$, $C_{\mathbf{y}'\mathbf{y}'} = E[\mathbf{y}'\mathbf{y}'^T]$, $C_{\mathbf{v}\mathbf{v}} = E[\mathbf{v}\mathbf{v}^T]$ y $R = C_{\mathbf{y}'\mathbf{y}'} - C_{\mathbf{v}\mathbf{v}}$.

Es fácil ver que la minimización de (3.75) con la condición $\|\mathbf{q}\| \neq 0$ se consigue eligiendo $\mathbf{q} = \gamma \mathbf{e}$ con \mathbf{e} siendo el autovector de la matriz R asociado con el mínimo autovalor λ_{\min} ($R\mathbf{e} = \lambda_{\min}\mathbf{e}$ y $\|\mathbf{e}\| = 1$) y γ es una constante global a determinar. Al mismo tiempo, se indica que el *MSE* obtenido está determinado por el mínimo autovalor y la constante global, i.e.,

$$\varepsilon^2 = \mathbf{q}^T R \mathbf{q} = \gamma^2 \mathbf{e}^T R \mathbf{e} = \gamma^2 \lambda_{\min}$$

Para determinar la constante global γ se utiliza la condición sobre los valores medios (3.70), arribando a

$$\bar{y}_i = (D\bar{\mathbf{x}})_i = h_i \bar{s}_i = \frac{\bar{s}_i}{\gamma e_i} \quad (3.76)$$

y por lo tanto, considerando la suma (3.71) se obtiene que:

$$q_i = \gamma e_i \quad \text{con } \gamma = \frac{1}{\sum_{i=0}^{p-1} e_i (D\bar{\mathbf{x}})_i} \quad (3.77)$$

3.8. Resultados experimentales

En esta sección se presentan resultados experimentales que permiten analizar la performance del algoritmo MAXNG como técnica para separar fuentes dependientes. En estos experimentos, se han utilizado los siguientes tipos de fuentes:

a) Señales sintetizadas (Tipo A): En este caso, con la idea de generar señales que cumplan con la condición (3.68) y simular las señales encontradas en

imágenes hiper-espectrales, se procede de la siguiente manera: se generan primero p señales independientes: w_0, w_1, \dots, w_{P-1} utilizando una *pdf* común a todas ellas $f_w(w)$; luego, se obtienen las señales fuente de la siguiente manera:

$$s_k = \frac{w_k}{\sum_{i=0}^{p-1} w_i} \quad (3.78)$$

Se consideran las variables w_0, w_1, \dots, w_{P-1} uniformemente distribuidas en el segmento $[0, 1]$. Es fácil ver que las fuentes s_k generadas con (3.78) tienen idénticas *pdf*'s, además, las esperanzas condicionadas y los coeficientes de correlación están dados por:

$$\rho = E[u_i u_j] = -\frac{1}{p-1} \quad (3.79)$$

$$E[u_i | u_j] = -\frac{1}{p-1} u_j = \rho u_j \quad (3.80)$$

donde u_i y u_j son las fuentes normalizadas. Se señala que la ecuación (3.80) garantiza la separabilidad de las fuentes ya que verifica la condición suficiente del Teorema 3.

b) Señales extraídas de imágenes satelitales (Tipo B): Se usan columnas de píxeles en imágenes satelitales reales monocromáticas. Estas señales tienen *pdf*'s que no responden a un modelo sencillo y permiten regular el grado de dependencia entre señales eligiendo la separación (en píxeles) entre columnas seleccionadas.

3.8.1. Experimento 1: Máxima NG versus Mínima Información Mutua

Se presenta una comparación de las separaciones obtenidas con el criterio de Máxima NG (algoritmo MAXNG) y el criterio de mínima Información Mutua que es la base de la mayoría de los algoritmos *ICA*.

Este experimento se realizó sobre señales del Tipo B (señales extraídas de imágenes satelitales) normalizadas (media cero y varianza unitaria) con diferentes grados de dependencia medidos por el coeficiente de correlación ρ . En total, se han realizado 300 simulaciones para diferentes fuentes con diversos grados de dependencia y con $n = 512$. Se utilizó una matriz de mezcla aleatoria y las matrices de separación fueron estimadas usando ambos criterios: MAXNG y mínima Información Mutua.

Con el objeto de medir la precisión de las estimaciones, dadas las señales fuente s_i y sus estimaciones \hat{s}_i , se define el vector error $\mathbf{e} = \hat{\mathbf{s}} - \mathbf{s}$ y se utiliza la medida *Signal to Interference Ratio - SIR* comúnmente utilizada en algoritmos *BSS*:

$$SIR_i = -10 \log_{10}(var(e_i)) = -10 \log_{10}(C_{\mathbf{ee}}[i, i]) \quad (3.81)$$

En general, niveles de *SIR* por debajo de 8dB, 10dB o 12dB indican un error en la separación [22], por lo tanto se dice que una separación es exitosa si $SIR_i > 8$ dB, 10dB o 12dB para todas las fuentes $i = 0, 1, \dots, p-1$. Al mismo tiempo se define la eficiencia de la separación

$$\text{Eficiencia (\%)} = \frac{n_A}{n_T} \quad (3.82)$$

donde n_A es el número de casos donde los SIR 's superaron el umbral y n_T es el número total de casos analizados.

En la Fig. 3.10, se muestran los niveles de eficiencia para el criterio de mínima Información Mutua y MaxNG como función del coeficiente de correlación de las fuentes $\rho = E[s_0 s_1]$ para dos umbrales diferentes: 10dB y 12dB. Se señala que para niveles de correlación $\rho > 0,3$, la eficiencia del criterio MinMI comienza a decaer y llega a cero para un coeficiente de correlación $\rho > 0,7$. Para niveles de correlación cercanos a cero, ambos criterios proveen niveles de eficiencia equivalentes lo cual concuerda con la teoría. Se observa además que el criterio MinMI ha demostrado ser ligeramente superior en términos de eficiencia para coeficientes de correlación pequeños (0,98 versus 0,81), sin embargo este efecto sólo es observado para n pequeño ($n = 512$), ya que, si se aumenta el número de muestras, la eficiencia de MaxNG se incrementa considerablemente, como muestra la Fig 3.11 donde la eficiencia pasa de 0,87 ($n = 512$) a 0,99 ($n = 5500$).

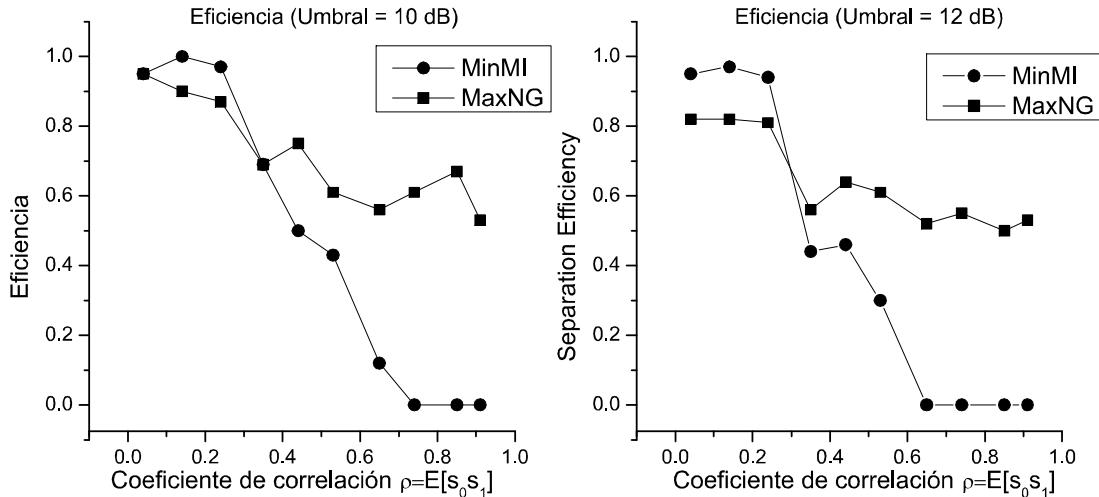


FIGURA 3.10. Eficiencia versus coeficiente de correlación para el criterio de Mínima Información Mutua (MinMI) y Máxima No-Gaussianidad (MaxNG)

3.8.2. Experimento 2: Efecto del ruido sobre MaxNG

En esta subsección se analiza la robustez al ruido de MAXNG (Método 1, presentado en la Sección 3.6). Más precisamente, se analiza el efecto del ruido sobre las ubicaciones teóricas de los máximos locales de la medida NG. En este experimento se consideran $p = 3$ fuentes dependientes sintetizadas (Tipo A) y se fija el número de muestras en $n = 4096$. Se utiliza también un número de mezclas (sensores) $m = 102$ para simular el tipo de señales observadas en el caso de imágenes hiper-espectrales (Sección 5.3). La matriz de mezcla A es generada aleatoriamente y las mezclas x_i son generadas con el modelo lineal (3.60) utilizando ruido blanco gaussiano cuyas varianzas han sido seleccionadas de manera de garantizar un nivel SNR constante para todos los canales (sensores).

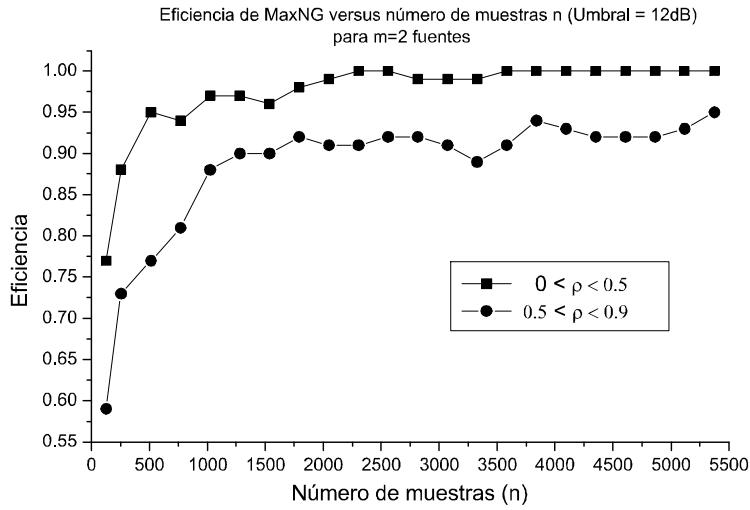


FIGURA 3.11. Eficiencia de MAXNG. 600 casos de separación fueron considerados para cada valor de n . Se consideraron dos categorías de fuentes: $0 < \rho < 0,5$ y $0,5 < \rho < 0,9$

Se realizaron simulaciones para un amplio rango de SNR (-15dB a 50dB) y se calculó la medida NG de la pdf de la señal con ruido (método 1) para el espacio de parámetros de separación. En todos los casos se han considerado los primeros $q = 2$ autovalores como dominantes ya que la dimensión del espacio de las fuentes, causado por la condición (3.68), es $q = 2 = p - 1$. Luego de aplicar el FD correspondiente, se evalúa la medida NG de la variable

$$y(\ddot{\mathbf{d}}) = \ddot{\mathbf{d}}^T \dot{\mathbf{x}} \quad (3.83)$$

donde $\dot{\mathbf{x}}$ son las señales no-correlacionadas (dimensión $q = 2$) y $\ddot{\mathbf{d}}$ es un vector con norma unitaria que puede ser parametrizado a partir de un único parámetro angular θ :

$$\ddot{\mathbf{d}}^T = [\cos \theta \quad \sin \theta] \quad (3.84)$$

Se calculó la medida NG de la variable y con θ en el rango $0^\circ < \theta < 180^\circ$ ya que en el rango $180^\circ < \theta < 360^\circ$ las señales estimadas están invertidas, es decir, multiplicadas por -1. Se utilizó $h = 1,06 \times n^{-0,2} = 0,2$ que minimiza el *Mean Integrated Square Error - MISE* [115] para ventanas y pdf 's gaussianas. En todos los casos, en la Fig. 3.12 se comparan los máximos locales obtenidos con las posiciones teóricas esperadas. Se destaca que, los máximos estimados coinciden con los máximos locales esperados en el rango de SNR de -5dB a 50dB. Inclusive, para un nivel muy bajo de SNR (-10dB) se obtuvieron máximos locales no muy lejanos de sus posiciones teóricas.

La conclusión más importante de este análisis es que se puede aplicar el algoritmo MAXNG directamente sobre los datos con ruido, sin introducir errores importantes y con la ventaja de que este método no requiere información alguna sobre el modelo de ruido.

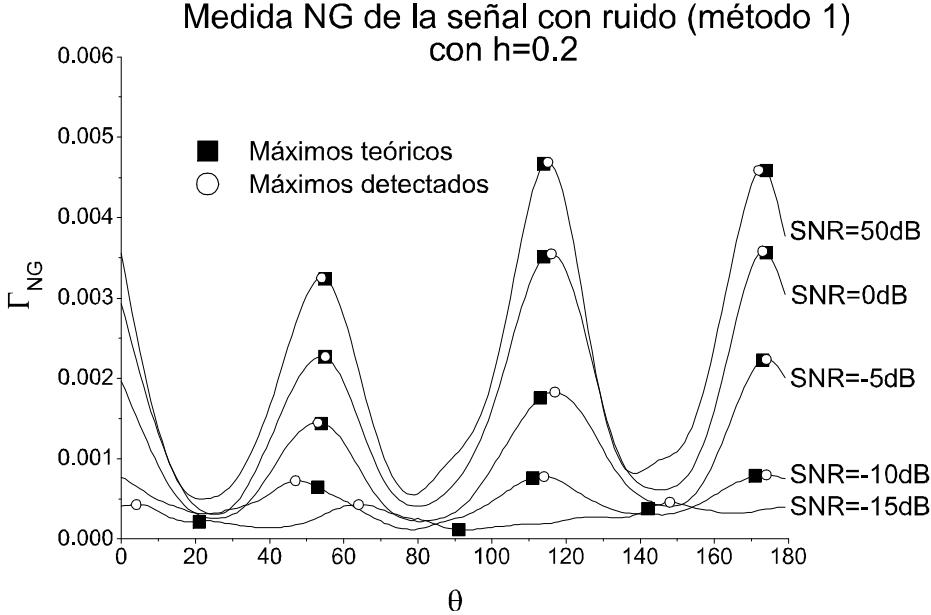


FIGURA 3.12. Máximos locales detectados y teóricos para señales Tipo A con $n = 4096$, $m = 102$ and $p = 3$.

3.8.3. Experimento 3: MaxNG versus algoritmos ICA

El objeto de este experimento es comparar el algoritmo MAXNG con algunos de los algoritmos *ICA* más populares como son el FASTICA [74, 76] y JADE [37, 76] para la separación de fuentes dependientes. Se han considerado casos de separación con $p = 3$ señales sintetizadas Tipo A. Dada la estimación de la fuente normalizada \hat{u}_i , se evalúa el *SIR* correspondiente respecto de la fuente normalizada original u_i , i.e.

$$SIR_i = -10 \log_{10}(var(\hat{u}_i - u_i)) \quad (3.85)$$

Asimismo se define el *SIR* medio como la media aritmética de los SIR_i obtenidos

$$SIR \text{ medio} = \frac{1}{p} \sum_{i=0}^{p-1} SIR_i \quad (3.86)$$

En la Fig. 3.13 se muestra una comparación de los *SIR*'s medios obtenidos con la aplicación de MAXNG, FASTICAY JADE. MAXNG ha sido aplicado sin utilizar la información del modelo de ruido, es decir, usando el Método 1 (medida NG de la *pdf* de la señal con ruido). Los valores de *SIR* medios fueron calculados en el rango de *SNR* de -25dB a 60dB mostrándose el promedio de los valores obtenidos sobre un total de 60 estimaciones por cada nivel de *SNR*. Para la implementación de los algoritmos *ICA*, se ha utilizado el paquete de software de Matlab ICALAB [48]. Los resultados muestran que, aproximadamente para niveles $SNR < -5\text{dB}$, los tres métodos fallan en recuperar las fuentes dado el fuerte contenido de ruido en las mezclas.

Como referencia se comparan estos resultados con los que se obtendrían con una matriz de separación perfecta, es decir, si D es tal que $DA = I$. En este caso, el error en la estimación es causado sólo por el ruido (no existen interferencias cruzadas entre las fuentes). Fácilmente puede verse que, para el caso de tener $C_{nn} = \sigma_n^2 I$ y una matriz de separación perfecta, el SIR obtenido está dado por:

$$SIR_i = -10 \log_{10}(E [(D\Lambda^{-1/2}V^T \mathbf{n})^2]) \quad (3.87)$$

que a su vez puede ser escrita en términos del nivel de SNR y los autovalores λ_i

$$SIR_i = SNR - 10 \log_{10} \left[\sum_{i=0}^{q-1} \frac{1}{\lambda_i} \right] \quad (3.88)$$

Se observa que la performance de MAXNG, en términos de SIR medio es prácticamente equivalente a la utilización de la matriz de separación perfecta para el rango de SNR de $-25dB$ a $25dB$. Para niveles SNR superiores a $25dB$ el SIR medio obtenido por MAXNG tiende a un valor constante de $38dB$ que es un valor muy alto si se lo compara con el valor mínimo aceptable de $12dB$ (Subsección 3.8.1). Por otro lado, se observa claramente, que para los algoritmos *ICA* considerados, el SIR medio obtenido no supera los $15dB$.

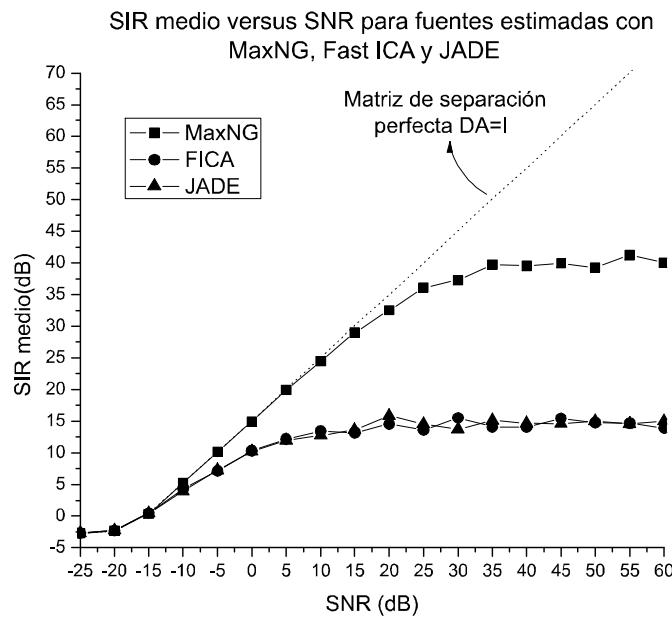


FIGURA 3.13. Valores de SIR medio para algoritmos MaxNG, Fast ICA y JADE para señales sintetizadas Tipo A con $n = 4096$, y $m = 102$.

3.8.4. Experimento 4: Performance de MaxNG

Existen varios factores que afectan la performance de MAXNG. En esta subsección se analiza experimentalmente cómo inciden el número de sensores m (canales

o mezclas), el número de muestras n y el parámetro de apertura de las ventanas de Parzen h en el valor de SIR medio obtenido para el caso de $p = 3$ señales sintetizadas Tipo A.

La Fig. 3.14 muestra los SIR medios obtenidos en función de los parámetros m , n y h para diferentes niveles SNR . En la Fig. 3.14 (izq.) puede observarse que, para mayor número de muestras n , el SIR medio obtenido resulta incrementado (para un SNR fijo) destacándose que, para $n > 2000$ la mejora en la performance es imperceptible, especialmente para un $SNR = 0\text{dB}$ y $SNR = 10\text{dB}$. La Fig. 3.14 (centro) muestra que, el efecto de incrementar el número de canales m (para un SNR fijo), es mejorar la eficiencia en términos del SIR medio resultante. Con respecto al parámetro de apertura h , en la Fig. 3.14 (der.) se verifica que el valor óptimo (en términos de MISE) dado por $h = 1,06 \times n^{-0,2} = 0,2$ para $n = 4096$ [115], provee también una buena solución en términos del SIR medio resultante.

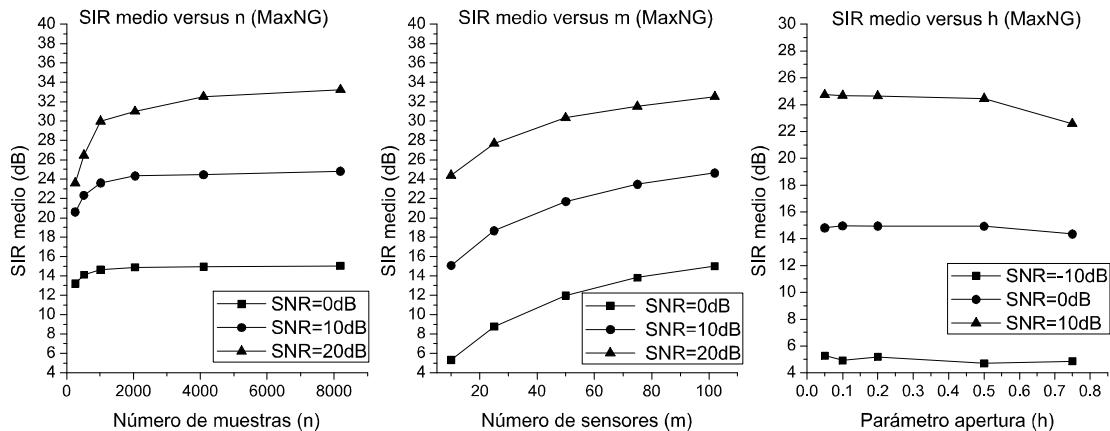


FIGURA 3.14. SIR medio versus n para el caso $m = 102$ y $h = 1,06 \times n^{-0,2}$ (izq.); SIR medio versus m con $n = 4096$ y $h = 1,06 \times n^{-0,2} = 0,2$ (centro); y SIR medio versus h con $n = 4096$ y $m = 102$ (der.)

3.9. Apéndices

3.9.1. Demostraciones de las ecuaciones (3.8) y (3.9)

Sustituyendo (3.6) en la integral de (3.7) e intercambiando el orden de la suma y la integral, se obtiene:

$$\Gamma_1(\hat{f}_y) = \frac{-2}{nh} \sum_{i=0}^{n-1} \int_{-\infty}^{+\infty} \Phi(y) \Phi\left(\frac{y - y(i)}{h}\right) dy \quad (3.89)$$

Reemplazando (3.5) en la ecuación anterior se arriba a:

$$\Gamma_1(\hat{f}_y) = \frac{-2}{n\sqrt{h^2 + 1}} \sum_{i=0}^{n-1} \Phi\left(\frac{y(i)}{\sqrt{h^2 + 1}}\right) \underbrace{\int_{-\infty}^{+\infty} \frac{\sqrt{h^2 + 1}}{h} \Phi\left(\frac{y - \frac{y(i)}{\sqrt{h^2 + 1}}}{\frac{h}{\sqrt{h^2 + 1}}}\right) dy}_{=1} \quad (3.90)$$

donde la última integral vale 1 resultando entonces la ecuación(3.8).

Para demostrar (3.9), se usa nuevamente (3.6) dentro de la integral de (3.7) arribando a:

$$\Gamma_2(\hat{f}_y) = \frac{1}{n^2 h^2} \int_{-\infty}^{+\infty} \left(\sum_{i=0}^{n-1} \Phi\left(\frac{y - y(i)}{h}\right) \right)^2 dy \quad (3.91)$$

Distribuyendo productos e intercambiando orden de integración y sumatorias se obtiene:

$$\Gamma_2(\hat{f}_y) = \frac{1}{n^2 h^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \int_{-\infty}^{+\infty} \Phi\left(\frac{y - y(i)}{h}\right) \Phi\left(\frac{y - y(j)}{h}\right) dy \quad (3.92)$$

y, a través del cambio de variables $v = \frac{y-y(i)}{h}$, se llega a:

$$\begin{aligned} \Gamma_2(\hat{f}_y) &= \frac{1}{n^2 h} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \underbrace{\int_{-\infty}^{+\infty} \Phi(v) \Phi\left(\frac{y(j) - y(i)}{h} - v\right) dy}_{=(\Phi * \Phi)\left(\frac{y(j) - y(i)}{h}\right)} \end{aligned} \quad (3.93)$$

donde la integral es la convolución entre dos funciones de Gauss valiendo:

$$(\Phi * \Phi)(x) = \frac{1}{\sqrt{2}} \Phi\left(\frac{x}{\sqrt{2}}\right) \quad (3.94)$$

por lo que se obtiene finalmente (3.9).

3.9.2. Demostración del Teorema 3

I) En primer lugar, se calcula la esperanza condicional de las fuentes normalizadas $E[u_0|u_1]$ usando la hipótesis de linealidad $E[s_0|s_1] = as_1 + b$. Aplicando propiedades básicas de la esperanza y la definición de fuentes normalizadas, se obtiene:

$$E[u_0|u_1] = \frac{1}{\sigma_0} E[s_0 - \bar{s}_0|u_1] = \frac{\sigma_1}{\sigma_0} au_1 + \frac{a\bar{s}_1 + b - \bar{s}_0}{\sigma_0} = \frac{\sigma_1}{\sigma_0} au_1 + \kappa \quad (3.95)$$

con $\kappa = \frac{a\bar{s}_1 + b - \bar{s}_0}{\sigma_0}$.

Ahora, usando la propiedad $E_y [E[x|y]] = E[x]$ en (3.95), se deduce que:

$$E_{u_1} [E[u_0|u_1]] = \frac{\sigma_1}{\sigma_0} a E[u_1] + \kappa = E[u_0] \quad (3.96)$$

y, sabiendo que $E[u_0] = E[u_1] = 0$, se concluye que $\kappa = 0$ además:

$$b = \bar{s}_0 - a\bar{s}_1 \quad (3.97)$$

$$E[u_0|u_1] = \frac{\sigma_1}{\sigma_0} au_1 \quad (3.98)$$

A continuación, para simplificar la notación, se asigna momentáneamente $x = u_0$ y $y = u_1$. Usando la definición de coeficiente de correlación y la definición de función de densidad condicional, se puede demostrar que:

$$\rho = E[xy] = \int \int xy f_{xy}(x,y) dx dy = \int y f_y(y) \left[\int x f_{x|y}(x|y) dx \right] dy \quad (3.99)$$

donde las integrales son tomadas sobre todo el rango de las variables x e y ; y $f_{xy}(x,y)$, $f_{x|y}(x|y)$ y $f_y(y)$ son las funciones de densidad conjunta, condicional y marginal, respectivamente.

Considerando que $E[x|y] = \int x f_{x|y}(x|y) dx$, introduciendo (3.98) en (3.99), y usando el hecho que $E[u_1^2] = 1$, finalmente se obtiene:

$$\rho = \frac{\sigma_1}{\sigma_0} a \int y^2 f_y(y) = \frac{\sigma_1}{\sigma_0} a \quad (3.100)$$

por lo que los parámetros a y b están relacionados con los momentos de las fuentes de la siguiente manera:

$$a = \frac{\sigma_0}{\sigma_1} \rho \quad (3.101)$$

$$b = \bar{s}_0 - \frac{\sigma_0}{\sigma_1} \rho \bar{s}_1 \quad (3.102)$$

II) Introduciendo (3.101) en (3.98), se obtiene directamente el resultado deseado, es decir, $E[u_0|u_1] = \rho u_1$

III) A continuación se demostrará que la derivada de la medida NG es cero en el punto $(\alpha, \beta) = (0, 1)$, con el par (α, β) restringido a la elipse (3.14).

Se considera la siguiente parametrización de la elipse, donde los coeficientes α y β son escritos en términos de un parámetro común t :

$$\begin{aligned} \alpha(t) &= t \\ \beta(t) &= -t\rho \pm \sqrt{t^2(\rho^2 - 1) + 1} \end{aligned} \quad (3.103)$$

Para simplificar la notación, consideraremos nuevamente $x = u_0$, $y = u_1$ y sus combinaciones lineales $z = \alpha x + \beta y$.

La *pdf* de la variable mezcla z puede ser escrita en forma integral en términos de la *pdf* conjunta $f_{xy}(x,y)$ y los coeficientes de mezcla α y β (asumimos $\beta \neq 0$):

$$f_z(z; \alpha, \beta) = \frac{1}{\beta} \int f_{xy} \left(x, \frac{z - \alpha x}{\beta} \right) dx \quad (3.104)$$

donde, está claro que, en el caso de variables independientes, (3.104) se reduce a la convolución de las *pdf*'s marginales.

Primero, se demuestra que la derivada de la *pdf* de la mezcla z es idénticamente nula para el valor del parámetro $t = 0$, i.e., $f'_z(z; \alpha(0), \beta(0)) = 0 \forall z$. Se observa que, usando la regla de la cadena de las derivadas, vale que:

$$f'_z(z; \alpha(t), \beta(t)) = \frac{\partial}{\partial \alpha} f_z(z; \alpha, \beta) \alpha'(t) + \frac{\partial}{\partial \beta} f_z(z; \alpha, \beta) \beta'(t) \quad (3.105)$$

Asumiendo que se pueden calcular las derivadas respecto de α y β introduciendo la operación de derivada dentro de la integral¹ en (3.104), y luego de manejos algebraicos, se obtiene (demostración más abajo):

$$\frac{\partial}{\partial \alpha} f_z(z = y; \alpha = 0, \beta = 1) = -\frac{d}{dy} (f_y(y) E[x|y]) \quad (3.106)$$

$$\frac{\partial}{\partial \beta} f_z(z = y; \alpha = 0, \beta = 1) = -\frac{d}{dy} (y f_y(y)) \quad (3.107)$$

Las derivadas de los parámetros α y β en $t = 0$ se obtienen fácilmente de (3.103):

$$\begin{aligned} \alpha'(0) &= 1 \\ \beta'(0) &= -\rho \end{aligned} \quad (3.108)$$

Sustituyendo (3.106), (3.107) y (3.108) en (3.105), se obtiene:

$$f'_z(z = y; \alpha(0), \beta(0)) = -\frac{d}{dy} (f_y(y) E[x|y]) + \frac{d}{dy} (y f_y(y)) \rho \quad (3.109)$$

donde, usando la condición $E[x|y] = \rho y$ es obvio que el término de la derecha, en la ecuación (3.109), es idénticamente cero.

Para finalizar la demostración se debe mostrar que la derivada de la medida NG es también cero en $(\alpha(0), \beta(0))$, lo cual puede verificarse considerando la derivada de la expresión (3.4) respecto de t :

$$\Gamma'_z(t) = \int_{-\infty}^{+\infty} 2 [f_z(z; \alpha(t), \beta(t)) - \Phi(z)] f'_z(z; \alpha(t), \beta(t)) dz \quad (3.110)$$

donde, evaluando en $t = 0$ y usando el hecho que $f'_z(z; \alpha(0), \beta(0)) = 0$, permite arribar al resultado deseado.

Demostraciones de las ecuaciones (3.106) y (3.107):

Se calcula primero la derivada de la ecuación (3.104) respecto de α :

$$\frac{\partial}{\partial \alpha} f_z(z; \alpha, \beta) = \frac{-1}{\beta^2} \int \frac{\partial}{\partial y} f_{xy} \left(x, \frac{z - \alpha x}{\beta} \right) x dx \quad (3.111)$$

y, cuando evaluamos en $\alpha = 0$, $\beta = 1$ y $z = y$, se obtiene:

$$\begin{aligned} \frac{\partial}{\partial \alpha} f_z(z = y; \alpha = 0, \beta = 1) &= - \int \frac{\partial}{\partial y} f_{xy}(x, y) x dx = \\ &- \frac{d}{dy} \int f_{x|y}(x|y) f_y(y) x dx = -\frac{d}{dy} (f_y(y) E[x|y]) \end{aligned} \quad (3.112)$$

¹Para garantizar que la operación de diferenciación puede pasarse dentro de la integral es necesario usar resultados del Análisis Funcional y teoría de integral de Lebesgue [85] restringiendo el espacio de las pdfs admisibles. En este trabajo, no se hace un análisis detallado matemático, sino que se asume que las *pdf's* verifican esas condiciones.

En forma similar, se calcula la derivada de la ecuación (3.104) respecto de β :

$$\begin{aligned} \frac{\partial}{\partial \beta} f_z(z; \alpha, \beta) = \\ -\frac{1}{\beta^2} \int f_{xy} \left(x, \frac{z - \alpha x}{\beta} \right) dx - \frac{1}{\beta^3} \int \frac{\partial}{\partial y} f_{xy} \left(x, \frac{z - \alpha x}{\beta} \right) (z - \alpha x) dx \end{aligned} \quad (3.113)$$

y, cuando se evalúa en $\alpha = 0$, $\beta = 1$ y $z = y$, se obtiene:

$$\begin{aligned} \frac{\partial}{\partial \beta} f_z(z = y; \alpha = 0, \beta = 1) = \\ - \int f_{xy}(x, y) dx - y \int \frac{d}{dy} f_{xy}(x, y) dx = \\ -f_y(y) dx - y \frac{d}{dy} f_y(y) = -\frac{d}{dy} (y f_y(y)) \end{aligned} \quad (3.114)$$

3.9.3. Demostraciones de las ecuaciones (3.41) - (3.46)

Asignado $\Psi = \Phi$, tomando $t = 0$ y $z(i) = \frac{y(i)}{\sqrt{h^2 + 1}}$ en (3.39), es posible aplicar la aproximación en (3.40) obteniendo:

$$\Gamma_1 \approx \frac{1}{n} \sum_{k=0}^{n_\Delta-1} f(k) \Phi_1(k) \quad (3.115)$$

con

$$\Phi_1(k) = \frac{-2}{\sqrt{h^2 + 1}} \Phi \left(\frac{\bar{\delta}(k)}{\sqrt{h^2 + 1}} \right) \quad (3.116)$$

Para la aproximación de (3.9), primero se asigna $\Psi = \Phi$, $t = \frac{y(j)}{\sqrt{2}h}$ y $z(i) = \frac{y(i)}{\sqrt{2}h}$, que produce la siguiente aproximación:

$$\Gamma_2 \approx \frac{1}{n^2 h \sqrt{2}} \sum_{i=0}^{n-1} \sum_{k=0}^{n_\Delta-1} f(k) \Phi \left(\frac{\bar{\delta}(k) - y(i)}{\sqrt{2}h} \right) \quad (3.117)$$

y, aplicando nuevamente la aproximación (3.40), es que finalmente se obtiene:

$$\Gamma_2 \approx \frac{1}{n^2 h \sqrt{2}} \sum_{l=0}^{n_\Delta-1} \sum_{k=0}^{n_\Delta-1} f(k) \Phi \left(\frac{k - l}{\sqrt{2}h} \right) = \frac{1}{n} \sum_{k=0}^{n_\Delta-1} f(k) (f * \Phi_2)(k) \quad (3.118)$$

con

$$\Phi_2(l) = \frac{1}{n \sqrt{2}h} \Phi \left(\frac{l \Delta}{\sqrt{2}h} \right) \quad (3.119)$$

y donde $(f * \Phi_2)(k)$ es la convolución de las funciones f y Φ_2 .

Siguiendo un procedimiento equivalente, se obtiene la aproximación (3.46).

Capítulo 4

CAMPOS ALEATORIOS GAUSSIANOS CON CORRELACIONES DE LARGO ALCANCE

4.1. Motivación del modelo

El algoritmo MAXNG, como la mayoría de los algoritmos *ICA*, permite estimar las fuentes a partir de una operación lineal sobre las mezclas (ecuación 3.2). Puede verse fácilmente que, cuando se aplica MAXNG al modelo con ruido, existirá un remanente de ruido gaussiano en las estimaciones, es decir,

$$\hat{\mathbf{s}} = D\mathbf{x} = \mathbf{s} + D\mathbf{n} \quad (4.1)$$

donde, siguiendo la notación del capítulo anterior, \mathbf{s} y $\hat{\mathbf{s}}$ son las señales fuente y sus estimaciones; \mathbf{x} es el vector de señales mezcla; D es la matriz de separación perfecta ($DA = I$) y \mathbf{n} es el vector de ruido gaussiano aditivo presente en las mezclas. Además, el ruido remanente definido por:

$$\mathbf{v} = D\mathbf{n} \quad (4.2)$$

resulta gaussiano por tratarse de combinaciones lineales de variables gaussianas.

Existen algunos antecedentes de tratamiento de ruido en *ICA*, por ejemplo en [6], se propone un algoritmo para *Independent Factor Analysis - IFA*, donde el estimador de las señales fuente es un operador no-lineal que se aplica a las mezclas. *IFA*, solamente es aplicable para el caso de señales independientes y su implementación es muy costosa en términos del volumen de cálculos requeridos. A pesar de ello, éste algoritmo, ha sido aplicado a *BSS* en imágenes hiper-espectrales [103] e imágenes de origen astrofísico [88].

En esta tesis se propone tratar el ruido en *BSS* aprovechando la robustez de MAXNG para estimar la matriz de separación aún en entornos con ruido. La idea es diseñar una técnica de reducción de ruido y aplicarla a las estimaciones provistas por MAXNG. En términos matemáticos, esta técnica deberá ser capaz de reducir el nivel (varianza) de ruido aditivo gaussiano que contamina las señales útiles no-gaussianas. El problema de la reducción de ruido aditivo es clásico en el área del procesamiento de señales, y requiere de modelos probabilísticos de la señal útil y del ruido. Como ejemplo, el filtro de Wiener clásico [127], se asumen al ruido y a la señal útil ambas gaussianas, permitiendo obtener una expresión del filtro óptimo en el sentido que, el error cuadrático medio (*Mean Squared Error - MSE*) es minimizado [124].

En general se puede decir que, cuanto más rico y preciso sea el modelo probabilístico de las señales involucradas, mejor será la performance del filtro de Wiener. En el caso de señales 2D (imágenes), un modelo sencillo es asumir a todos los píxeles (variables) independientes entre sí, lo cual es una burda aproximación al

comportamiento usual de las imágenes [23, 62]. Para mejorar el modelo es necesario introducir, de alguna forma, las dependencias existentes entre las variables modelando las interacciones entre píxeles.

En esta tesis se introduce un nuevo modelo de campo aleatorio gaussiano con correlaciones de largo alcance (*Long Correlation Gaussian Random Fields - LC-GRF's*) para el cual se analizan las propiedades y se derivan nuevos algoritmos de estimación de parámetros [34]. La motivación principal de este modelo, es permitir el diseño un filtro de Wiener reductor de ruido para la restauración de imágenes fuente separadas con MAXNG. En particular estamos interesados en modelar un tipo especial de imágenes astrofísicas producidas por el *Cosmic Microwave Background - CMB*, para las cuales se sabe que su función de densidad de probabilidad es gaussiana y además posee correlaciones espaciales de largo alcance [101].

4.2. El modelo *LC-GRF*

A continuación se presenta la notación que se utilizará a lo largo del presente capítulo como complemento a la utilizada en los capítulos anteriores:

- Un campo aleatorio de $l \times l$ (dos dimensiones) o imagen se denota con $\{x_{i,j}\}$ donde $i, j = 0, 1, \dots, l - 1$ son los índices para filas y columnas, respectivamente.
- El vector columna $\mathbf{x} \in \mathbb{R}^{l^2}$ contiene a todas las variables (píxeles) $x_{i,j}$ de un campo aleatorio, ordenadas fila por fila (orden lexicográfico).
- La transformada discreta de Fourier en dos dimensiones (*2D-Discrete Fourier Transform - 2D-DFT*) de una imagen $\{x_{i,j}\}$ está definida por:

$$\tilde{x}_{u,v} = \frac{1}{l} \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} x_{i,j} \exp \left(-\sqrt{-1} \frac{2\pi}{l} (ui + vj) \right) \quad (4.3)$$

donde $u, v = 0, 1, \dots, l - 1$ son los índices en el espacio transformado. Como es usual en la bibliografía, esta transformación también puede ser escrita en forma matricial:

$$\tilde{\mathbf{x}} = U^H \mathbf{x} \quad (4.4)$$

donde U^H es la operación hermítica (transposición y conjugación) sobre la matriz $U \in \mathbb{R}^{l^2 \times l^2}$ cuyos elementos son:

$$[U]_{il+j, ul+v} = \exp \left(\sqrt{-1} \frac{2\pi}{l} (ui + vj) \right) \quad (4.5)$$

Se observa que la matriz U es unitaria ($U^H U = I$) y por lo tanto, la inversa de la operación 2D-DFT se obtiene haciendo

$$\mathbf{x} = U \tilde{\mathbf{x}} \quad (4.6)$$

Otro resultado muy importante que se utilizará a menudo es que, la matriz U diagonaliza cualquier matriz circulante por bloques [65]. Para simplificar la notación, a menudo se hará referencia a la operación *2D-DFT* con: $\tilde{\mathbf{x}} = DFT\{\mathbf{x}\}$ o $\tilde{x}_{u,v} = DFT\{x_{i,j}\}$.

- Por definición, un *GRF* $\{x_{i,j}\}$ con $i, j = 0, 1, \dots, l-1$, con media nula $E[\mathbf{x}] = \mathbf{0}$ y matriz de covarianzas $C_{\mathbf{xx}} = E[\mathbf{xx}^T]$, posee una *pdf* conjunta definida por la distribución de Gauss multivariada [120], es decir:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^l |C_{\mathbf{xx}}|}} \exp\left(-\frac{1}{2} \mathbf{x}^T C_{\mathbf{xx}}^{-1} \mathbf{x}\right) \quad (4.7)$$

donde $|C_{\mathbf{xx}}|$ indica el determinante de la matriz $C_{\mathbf{xx}}$.

En esta tesis se propone una generalización de los modelos *CM* y *SAR* discutidos en la Sección 2.3.2 asumiendo una matriz de covarianzas dada por

$$C_{\mathbf{xx}} = \rho_{\mathbf{x}} B^{-\alpha}$$

donde α es un parámetro escalar y la matriz B es la definida anteriormente en la Sección 2.3.2 (ecuaciones (2.81) y (2.82)). Es obvio que para $\alpha = 1$ y $\alpha = 2$ se obtienen los casos *CM* y *SAR* conocidos.

Cabe señalar que para utilizar potencias no enteras de matrices, se usa una definición basada en la descomposición canónica de la matriz B , es decir:

$$B^\alpha \stackrel{Def}{=} U \Lambda^\alpha U^H \quad (4.8)$$

donde U^H es la *2D-DFT* que diagonaliza a B [65], Λ es la matriz diagonal cuyos elementos de la diagonal principal son los autovalores de B que depende de los coeficientes de interacción considerados $\theta_1, \theta_2, \dots, \theta_o$ [15, 34]. Es fácil ver que, si se asume un *GRF* no degenerado, i.e. con una matriz de covarianzas definida positiva, entonces todos los autovalores son positivos y (4.8) queda perfectamente definida para cualquier $\alpha \in \mathbb{R}$.

En consecuencia la definición formal de un Campo Aleatorio gaussiano con Correlaciones de Largo alcance (*LC-GRF*) es la siguiente [34]:

LC-GRF: Un campo aleatorio $\{x_{i,j}\}$ de $l \times l$ variables (imagen), es un *LC-GRF* (Long Correlation - GRF) si su *pdf* conjunta obedece a la siguiente forma:

$$f(\mathbf{x} ; \alpha, \rho_{\mathbf{x}}, \boldsymbol{\theta}) = \frac{|B|^{\alpha/2}}{(2\pi\rho_{\mathbf{x}})^{l^2/2}} \exp\left(-\frac{1}{2\rho_{\mathbf{x}}} \mathbf{x}^T B^\alpha \mathbf{x}\right) \quad (4.9)$$

donde $\mathbf{x} \in \mathbb{R}^{l^2}$ es un vector de media nula cuyos elementos son $x_{i,j}$ con $i, j = 0, 1, \dots, l-1$, $\rho_{\mathbf{x}}$ y α son escalares y la matriz $B \in \mathbb{R}^{l^2 \times l^2}$ está definida a partir de los coeficientes de interacción $\boldsymbol{\theta}^T = [\theta_1 \ \theta_2 \dots \theta_o]$ como se explicó anteriormente (Subsección 2.3.2).

Puede demostrarse que el presente modelo resulta equivalente al presentado por Bennet et al. en [15] en un contexto totalmente diferente [34].

4.3. Propiedades básicas del modelo *LC-GRF*

Cálculo de autovalores: Una de las propiedades más importantes del modelo, sobre la que se basan los resultados siguientes, es que la matriz de covarianzas puede ser diagonalizada utilizando la operación *2D-DFT* (ecuación (4.8)). Más aún, los autovalores $\lambda_{i,j}$ de la matriz B pueden ser calculados explícitamente [15, 81] y valen:

$$\lambda_{i,j} = 1 - 2\boldsymbol{\theta}^T \boldsymbol{\psi}_{i,j} \quad (4.10)$$

donde, para el caso de orden $o = 2$ (sistemas de vecinos de primero orden), los vectores $\boldsymbol{\theta}$ y $\boldsymbol{\psi}_{uv}$ son:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \text{ y } \boldsymbol{\psi}_{ij}^T = \begin{bmatrix} \cos(2\pi i/l) \\ \cos(2\pi j/l) \end{bmatrix} \text{ with } i, j = 0, 1, \dots, l-1 \quad (4.11)$$

Puede demostrarse que si se asume $C_{\mathbf{xx}}$ definida positiva, existe una restricción para el conjunto de parámetros, por ejemplo para orden $o = 2$ los parámetros θ_1 y θ_2 están sujetos a [15, 25, 81]:

$$|\theta_1| + |\theta_2| < 0,5 \quad (4.12)$$

Función de Autocorrelación: Como es usual la función de autocorrelación en 2D se define por [64]:

$$r_{s,t} = E[x_{i,j}x_{i-s,j-t}] \quad (4.13)$$

Se observa que cada fila (y columna) de la matriz $C_{\mathbf{xx}}$ está formada por versiones circulantes de la función de autocorrelación $r_{s,t}$ [65]. Se puede calcular directamente $r_{s,t}$ a partir de la descomposición canónica (4.8) de $C_{\mathbf{xx}}$. De esta manera, se obtiene la función de autocorrelación en términos de la operación *2D-DFT*:

$$r_{s,t} = \frac{\rho_{\mathbf{x}}}{l} DFT \left\{ \lambda_{i,j}^{-\alpha} \right\} \quad (4.14)$$

donde $\lambda_{i,j}$ está definido por (4.10).

Varianza: Como corolario del resultado anterior, si se evalúa la función de autocorrelación en $(0, 0)$, se obtiene la varianza de las variables del *GRF*:

$$\sigma_{\mathbf{x}}^2 = r_{0,0} = \frac{\rho_{\mathbf{x}}}{l^2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \lambda_{u,v}^{-\alpha} \quad (4.15)$$

Se señala que, dados los parámetros α y $\boldsymbol{\theta}$, la varianza del *GRF* es proporcional al parámetro $\rho_{\mathbf{x}}$.

Densidad espectral: La densidad espectral (*spectral density* o *power spectrum*) del *LC-GRF* \mathbf{x} , definida como la *2D-DFT* de la función de autocorrelación $r_{s,t}$ [64], toma la siguiente forma:

$$S_{u,v}^{\mathbf{x}} = \frac{\rho_{\mathbf{x}}}{\lambda_{u,v}^{\alpha}} \quad (4.16)$$

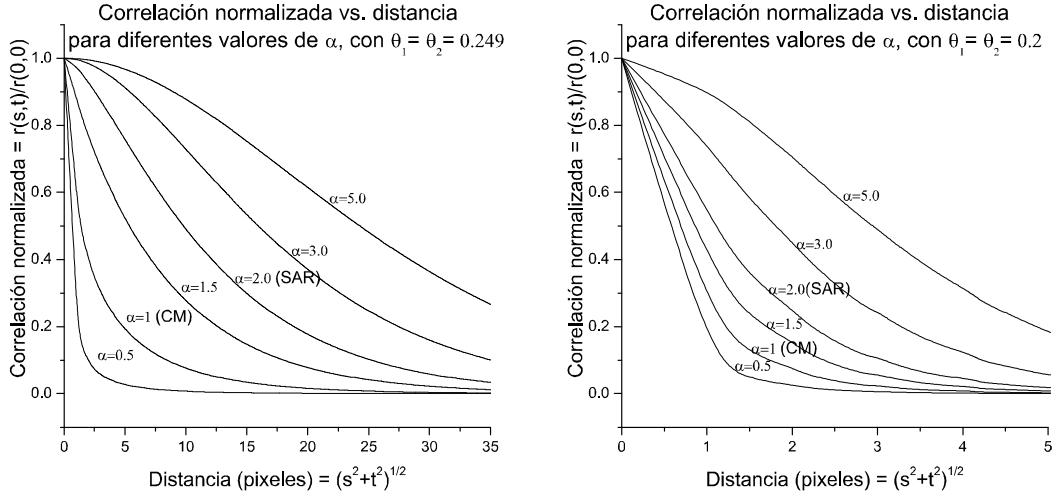


FIGURA 4.1. Correlación normalizada ($r_{s,t}/r_{0,0}$). Los casos para los modelos CM y SAR corresponden a $\alpha = 1$ y $\alpha = 2$ respectivamente.

Correlaciones de largo alcance: Una de las características más interesantes del modelo *LC-GRF*, es que permite modelar fuertes correlaciones entre píxeles lejanos dentro de una imagen utilizando pocos parámetros. Puede verse que el parámetro α está directamente relacionado con esta característica. Con el objeto de ilustrar esta propiedad, en la Fig. 4.1 se muestra el valor teórico del coeficiente de correlación normalizado entre píxeles ($r_{s,t}/r_{0,0}$) versus la distancia, para $\alpha = 0,5, 1,0$ (CM), $1,5, 2,0$ (SAR), $3,0$ y $5,0$, con $\theta_1 = \theta_2 = 0,249$ (izquierda) y $\theta_1 = \theta_2 = 0,2$ (derecha). Se observa que la característica de correlación de largo alcance está también determinada por los parámetros de interacción θ_1 y θ_2 .

Además, en la Fig. 4.2, se muestran imágenes generadas aleatoriamente correspondientes al modelo *LC-GRF* con diferentes valores de los parámetros, donde puede verse la influencia de los mismos en las correlaciones entre píxeles, es decir, a mayores valores de θ_1 , θ_2 y α , los píxeles vecinos tienden a parecerse más entre sí agrupándose en forma de “manchas”. Las muestras fueron generadas utilizando un algoritmo equivalente al diseñado en [25] para el caso de los modelos *CM* y *SAR*.

4.4. Estimación de parámetros

Para poder hacer un uso efectivo del modelo *LC-GRF* se debe contar con métodos de estimación de parámetros que permitan ajustar el modelo a datos disponibles. En esta tesis se presentan las soluciones para los siguientes problemas de estimación de parámetros con el modelo *LC-GRF* [34]:

1. Estimación de los parámetros α , ρ_x y θ que mejor se ajusten a una imagen de muestra

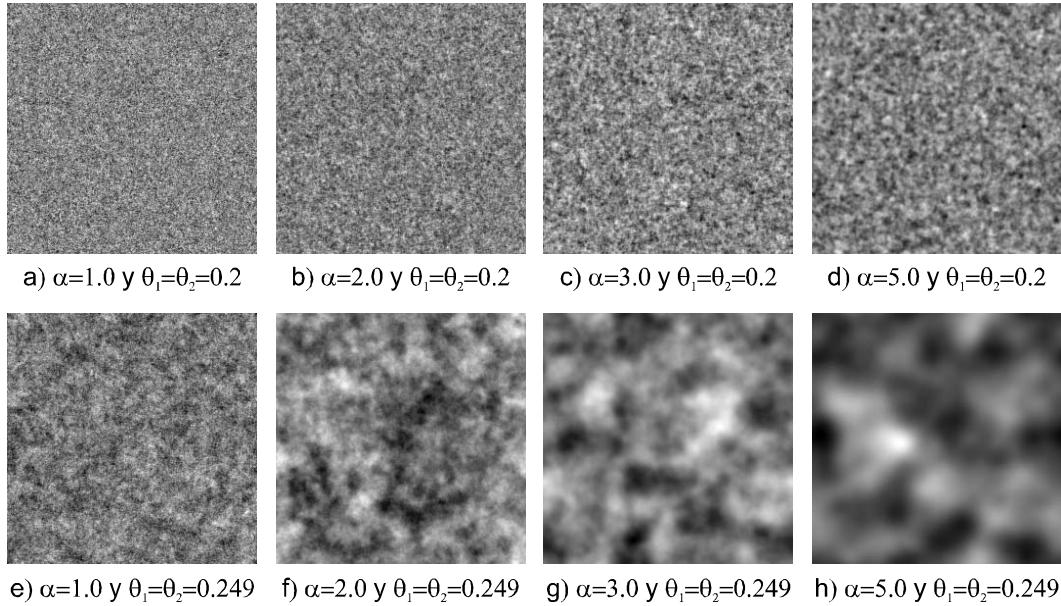


FIGURA 4.2. Muestras de imágenes LC-GRF generadas aleatoriamente para distintos valores de los parámetros θ_1 , θ_2 y α . El parámetro ρ_x se eligió de manera tal que asegure una varianza unitaria

2. Estimación de las varianzas correspondientes a una imagen *LC-GRF* y al ruido aditivo gaussiano no-correlacionado (*Additive White Gaussian Noise - WGN*), a partir de una muestra de la imagen con ruido).

A continuación se presentan los algoritmos especialmente diseñados para estos casos [34].

4.4.1. Estimación de máxima verosimilitud de α , ρ_x y θ

Se propone un algoritmo iterativo para la estimación de los parámetros escalares α , ρ_x y el vector $\boldsymbol{\theta}^T = [\theta_1 \ \theta_2 .. \theta_o]$ a partir de una muestra de la imagen. El algoritmo está basado en el método de máxima verosimilitud (*Maximum Likelihood - ML*) y se obtiene de la siguiente manera. Aplicando el logaritmo a la ecuación (4.9), se obtiene la función *log-likelihood* a maximizar

$$\Gamma(\alpha, \rho_x, \boldsymbol{\theta}) = \log(f_{\mathbf{x}}(\mathbf{x}; \alpha, \rho_x, \boldsymbol{\theta})) \quad (4.17)$$

la cual puede ser escrita de la siguiente manera (Apéndice 4.8.1):

$$\Gamma(\alpha, \rho_x, \boldsymbol{\theta}) = \frac{\alpha}{2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \log(\lambda_{u,v}) - \frac{l^2}{2} \log(2\pi\rho_x) - \frac{1}{2\rho} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} |\tilde{x}_{u,v}|^2 \lambda_{u,v}^\alpha \quad (4.18)$$

donde $\lambda_{u,v}$ son los autovalores de B (ecuación (4.10)) y \tilde{x}_{uv} es la *2D-DFT* de la muestra evaluada en las frecuencias (u, v) .

Una forma de determinar las estimaciones *ML* de los parámetros es buscar los valores $\alpha = \hat{\alpha}$, $\rho_{\mathbf{x}} = \hat{\rho}_{\mathbf{x}}$ y $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ tales que la función *log-likelihood* tenga derivadas nulas, i.e.:

$$\frac{\partial \Gamma}{\partial \alpha}(\hat{\alpha}, \hat{\rho}_{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = 0, \quad (4.19)$$

$$\frac{\partial \Gamma}{\partial \rho_{\mathbf{x}}}(\hat{\alpha}, \hat{\rho}_{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = 0, \quad (4.20)$$

$$\nabla_{\boldsymbol{\theta}} \Gamma(\hat{\alpha}, \hat{\rho}_{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \mathbf{0}, \quad (4.21)$$

donde $\nabla_{\boldsymbol{\theta}} \Gamma$ es el gradiente de Γ respecto del vector $\boldsymbol{\theta}$.

Se observa que, dadas las estimaciones $\hat{\boldsymbol{\theta}}$ y $\hat{\alpha}$, es posible calcular $\hat{\rho}_{\mathbf{x}}$ directamente ya que, usando ecuación (4.21) conjuntamente con (4.18) y (4.10), se obtiene:

$$\hat{\rho}_{\mathbf{x}} = \frac{1}{l^2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} |\tilde{x}_{u,v}|^2 \left[1 - 2\hat{\boldsymbol{\theta}}^T \boldsymbol{\psi}_{uv} \right]^{\hat{\alpha}} \quad (4.22)$$

Para el resto de los parámetros ($\hat{\alpha}$ y $\hat{\boldsymbol{\theta}}$) no existen fórmulas cerradas, por lo que se propone una búsqueda iterativa de los máximos locales a través de las derivadas, las cuales pueden obtenerse explícitamente de (4.18) y son:

$$\nabla_{\boldsymbol{\theta}} \Gamma(\alpha, \rho_{\mathbf{x}}, \boldsymbol{\theta}) = \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \frac{\alpha \boldsymbol{\psi}_{uv}}{\lambda_{u,v}} \varphi_{uv}(\alpha, \rho_{\mathbf{x}}, \boldsymbol{\theta}) \quad (4.23)$$

$$\frac{\partial \Gamma}{\partial \alpha}(\alpha, \rho_{\mathbf{x}}, \boldsymbol{\theta}) = -\frac{1}{2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \log(\lambda_{u,v}) \varphi_{uv}(\alpha, \rho_{\mathbf{x}}, \boldsymbol{\theta}) \quad (4.24)$$

donde

$$\varphi_{uv}(\alpha, \rho_{\mathbf{x}}, \boldsymbol{\theta}) = \frac{1}{\rho_{\mathbf{x}}} |\tilde{x}_{u,v}|^2 \lambda_{u,v}^{\alpha} - 1 \quad (4.25)$$

En consecuencia se deriva el siguiente algoritmo que puede ser interpretado como una modificación del algoritmo clásico de Newton de búsqueda de máximos a través del gradiente [57]:

Algoritmo 4. *Estimación ML de parámetros en un LC-GRF:*

- Inicializar las estimaciones $\alpha^{(0)}$, $\rho_{\mathbf{x}}^{(0)}$ y $\boldsymbol{\theta}^{(0)}$.
- Iniciar LOOP ($i = 0, 1, 2, \dots$)
- Actualizar $\hat{\boldsymbol{\theta}}$ (usando ecuación (4.23)):

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \hat{\boldsymbol{\theta}}^{(i)} + D^{(i)} \nabla_{\boldsymbol{\theta}} \Gamma \left(\alpha^{(i)}, \rho_{\mathbf{x}}^{(i)}, \boldsymbol{\theta}^{(i)} \right) \quad (4.26)$$

- Actualizar $\hat{\alpha}$ (usando ecuación (4.24)):

$$\hat{\alpha}^{(i+1)} = \hat{\alpha}^{(i)} + h^{(i)} \frac{\partial \Gamma}{\partial \alpha} \left(\hat{\alpha}^{(i)}, \hat{\rho}_{\mathbf{x}}^{(i)}, \hat{\boldsymbol{\theta}}^{(i)} \right) \quad (4.27)$$

-
- Calcular $\hat{\rho}_{\mathbf{x}}^{(i+1)}$ en función de $\hat{\boldsymbol{\theta}}^{(i)}$ y $\hat{\alpha}^{(i)}$ usando (4.22).
 - Fin LOOP.

En este algoritmo, $D^{(i)} \in \mathbb{R}^{o \times o}$, $h^{(i)}$ es un escalar y ambos son parámetros que especifican el paso en la actualización de los parámetros. En el algoritmo clásico de Newton la matriz $D^{(i)}$ y $h^{(i)}$ se actualizan a través del cálculo de las matrices hessianas [57]. En el presente algoritmo, se evitan los costosos cálculos de derivadas segundas proponiéndose una regla más simple (y no óptima) de actualización. Además, aquí se considera una matriz $D^{(i)}$ diagonal, cuyos elementos de la diagonal principal: $d_1^{(i)}, d_2^{(i)}, \dots, d_o^{(i)}$ controlan la variación de los parámetros $\hat{\theta}_1^{(i)}, \hat{\theta}_2^{(i)}, \dots, \hat{\theta}_o^{(i)}$, en la iteración i . Para la actualización de los valores de paso $d_1^{(i)}, d_2^{(i)}, \dots, d_o^{(i)}$, $h^{(i)}$ se utiliza un criterio empírico: se usan valores pequeños iniciales y, luego de cada iteración, se compara el valor de las derivadas con sus valores en la iteración anterior. Más precisamente, para cada parámetro, si la nueva derivada no cambia su signo entonces se incrementa el paso correspondiente usando un factor fijo de “aceleración” (*speed-up*) $\mu_{su} > 1,0$. Por otro lado, si la derivada cambia de signo, significa que la estimación del parámetro cruzó la posición del máximo y, para refinar la búsqueda, se reduce el valor del paso correspondiente dividiéndolo por un factor fijo de “reducción” (*slow-down*) $\mu_{sd} > 1,0$.

Otro aspecto que tiene que ser tenido en cuenta durante la búsqueda del máximo, es que existe una restricción sobre los parámetros $\hat{\theta}_1^{(i)}, \hat{\theta}_2^{(i)}, \dots, \hat{\theta}_o^{(i)}$ (por ejemplo, para el caso de orden $o = 2$, la condición es $|\theta_1| + |\theta_2| < 0,5$ determinando el área indicada en la Fig. 4.3). Es más, puede verse que esta condición es necesaria para evitar singularidades en el cálculo de las derivadas (ecuaciones (4.23) y (4.24)). Por lo tanto, para evitar que los parámetros escapen de la región permitida, sólo se acepta la actualización del vector $\hat{\boldsymbol{\theta}}$ si este no viola la condición de admisibilidad, de lo contrario, los pasos $d_1^{(i)}, d_2^{(i)}, \dots, d_o^{(i)}$ son reducidos dividiéndolos por μ_{sd} tantas veces como sea necesario hasta que el parámetro $\hat{\boldsymbol{\theta}}$ sea aceptable.

4.4.2. Estimación de las varianzas de la señal útil y el ruido usando el algoritmo *EM*

Dada una imagen con ruido

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (4.28)$$

donde \mathbf{x} es una imagen que responde al modelo *LC-GRF*, \mathbf{n} es ruido gaussiano blanco aditivo (*AWGN*) independiente de \mathbf{x} y los parámetros α y $\boldsymbol{\theta}$ de la señal útil son conocidos. El objetivo es obtener las estimaciones $\hat{\rho}_{\mathbf{x}}$ y $\hat{\rho}_{\mathbf{n}}$, lo que equivale a determinar las varianzas de \mathbf{x} y \mathbf{n} . Éste es un problema de estimación de máxima verosimilitud con un conjunto de datos incompletos (*incomplete dataset*) que puede resolverse a través del algoritmo *EM* (*Expectation / Maximization*) [53] ya que, en lugar de tener acceso a los datos completos (*complete dataset*), es decir las imágenes \mathbf{x} y \mathbf{n} por separado, se tiene solo acceso a la suma (4.28).

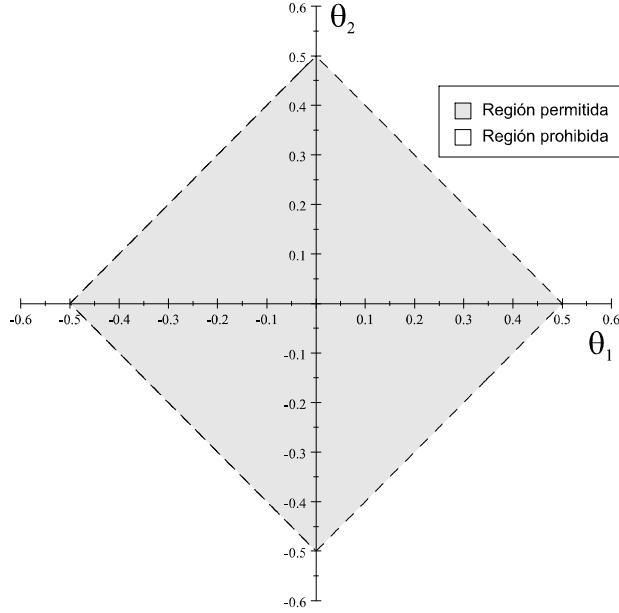


FIGURA 4.3. Regiones de valores permitidos y prohibidos para los parámetros θ_1 y θ_2 para el caso de $s = 2$ (sistema de vecinos más cercanos)

El algoritmo *EM* provee una técnica iterativa para la maximización del *log-likelihood* de los datos completos usando los datos incompletos [53]. Más formalmente, se definen:

Datos Completos: $\mathcal{C} = \{\mathbf{x}, \mathbf{n}\}$

Datos Incompletos: $\mathcal{I} = \{\mathbf{x} + \mathbf{n}\}$

Parámetros a estimar: $\hat{\rho}_{\mathbf{x}}$ y $\hat{\rho}_{\mathbf{n}}$

El algoritmo *EM* consiste en la repetición de dos pasos principales llamados *E-step (Expectation)* y *M-step (Maximization)*:

E-step (Expectation): Dadas las estimaciones correspondientes al paso i ($\hat{\rho}_{\mathbf{x}}^{(i)}$, $\hat{\rho}_{\mathbf{n}}^{(i)}$) y los datos incompletos $\{\mathbf{x} + \mathbf{n}\}$, se calcula la esperanza del *log-likelihood* que es:

$$V(\rho'_{\mathbf{x}}, \rho'_{\mathbf{n}} | \hat{\rho}_{\mathbf{x}}^{(i)}, \hat{\rho}_{\mathbf{n}}^{(i)}) = E \left\{ \log f(C | \rho'_{\mathbf{x}}, \rho'_{\mathbf{n}}) \mid I, \hat{\rho}_{\mathbf{x}}^{(i)}, \hat{\rho}_{\mathbf{n}}^{(i)} \right\} \quad (4.29)$$

M-step (Maximization): Se asignan, como nuevos valores de las estimaciones (paso $i + 1$) aquellos que maximizan la función $V(\rho'_{\mathbf{x}}, \rho'_{\mathbf{n}} | \hat{\rho}_{\mathbf{x}}^{(i)}, \hat{\rho}_{\mathbf{n}}^{(i)})$, es decir:

$$V(\hat{\rho}_{\mathbf{x}}^{(i+1)}, \hat{\rho}_{\mathbf{n}}^{(i+1)} | \hat{\rho}_{\mathbf{x}}^{(i)}, \hat{\rho}_{\mathbf{n}}^{(i)}) = \underset{(\rho'_{\mathbf{x}}, \rho'_{\mathbf{n}})}{\operatorname{Max}} \left(V(\rho'_{\mathbf{x}}, \rho'_{\mathbf{n}} | \hat{\rho}_{\mathbf{x}}^{(i)}, \hat{\rho}_{\mathbf{n}}^{(i)}) \right) \quad (4.30)$$

La convergencia al máximo de la función de *log-likelihood* está garantizada por la demostración teórica provista en [53, 57]. Para nuestro caso particular de imágenes *LC-GRF* contaminadas con ruido gaussiano, luego de varias operaciones algebraicas se obtienen las fórmulas correspondientes a los pasos *E-step* (4.29) y *M-step* (4.30) que definen el siguiente algoritmo (Apéndice 4.8.2):

Algoritmo 5. *Estimación de los parámetros ρ_x y ρ_n*

- Inicializar las estimaciones $\hat{\rho}_x^{(0)}$ y $\hat{\rho}_n^{(0)}$

- Iniciar LOOP ($i = 0, 1, 2, \dots$)

- Actualizar las estimaciones:

$$\hat{\rho}_x^{(i+1)} = \frac{1}{l^2} \Phi_1 \left(\hat{\rho}_x^{(i)}, \hat{\rho}_n^{(i)}, \mathbf{y} \right) \quad (4.31)$$

$$\hat{\rho}_n^{(i+1)} = \frac{1}{l^2} \Phi_2 \left(\hat{\rho}_x^{(i)}, \hat{\rho}_n^{(i)}, \mathbf{y} \right) \quad (4.32)$$

- END LOOP.

donde $\Phi_1 \left(\hat{\rho}_x^{(i)}, \hat{\rho}_n^{(i)}, \mathbf{y} \right)$ y $\Phi_2 \left(\hat{\rho}_x^{(i)}, \hat{\rho}_n^{(i)}, \mathbf{y} \right)$ están dados por las siguientes expresiones (demostraciones en el Apéndice 4.8.2):

$$\Phi_1 \left(\hat{\rho}_x, \hat{\rho}_n, \mathbf{y} \right) = \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \lambda_{uv} \left(\frac{1}{\hat{\rho}_n} + \frac{\lambda_{uv}}{\hat{\rho}_x} \right)^{-1} \left[1 + \frac{|\tilde{y}_{uv}|^2}{\hat{\rho}_n^2} \left(\frac{1}{\hat{\rho}_n} + \frac{\lambda_{uv}}{\hat{\rho}_x} \right)^{-1} \right] \quad (4.33)$$

$$\Phi_2 \left(\hat{\rho}_x, \hat{\rho}_n, \mathbf{y} \right) = \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \left(\frac{1}{\hat{\rho}_n} + \frac{\lambda_{uv}}{\hat{\rho}_x} \right)^{-1} \left[1 + \frac{|\tilde{y}_{uv}|^2 \lambda_{uv}^2}{\hat{\rho}_x^2} \left(\frac{1}{\hat{\rho}_n} + \frac{\lambda_{uv}}{\hat{\rho}_x} \right)^{-1} \right] \quad (4.34)$$

siendo \tilde{y}_{uv} la *2D-DFT* de la imagen con ruido \mathbf{y} evaluada en las frecuencias (u, v) .

4.5. Reducción de ruido

En esta subsección se aborda el problema de la reducción del ruido aditivo gaussiano contenido en una imagen \mathbf{y} . Más precisamente, dada una imagen con ruido \mathbf{y} (ecuación (4.28)), se busca la mejor estimación $\hat{\mathbf{x}}$ de la imagen *LC-GRF* \mathbf{x} .

Es bien conocido que, cuando las señales son gaussianas, los estimadores MAP (*Maximum A Posteriori*) y MMSE (*Minimum Mean Squared Error*) son equivalentes [57, 124] y la reducción del ruido se lleva a cabo a través de la aplicación del filtro de Wiener [127], cuya expresión en el dominio de frecuencias de Fourier es [64]:

$$\tilde{\hat{x}}_{uv} = \frac{S_{u,v}^{\mathbf{x}}}{S_{u,v}^{\mathbf{x}} + S_{u,v}^{\mathbf{n}}} \tilde{y}_{uv} \quad (4.35)$$

donde $\tilde{\hat{x}}_{uv}$ y \tilde{y}_{uv} son las transformaciones *2D-DFT* de la imagen \mathbf{x} estimada y la imagen con ruido \mathbf{y} respectivamente, $S_{u,v}^{\mathbf{x}}$ y $S_{u,v}^{\mathbf{n}}$ son los correspondientes densidades espectrales, i.e. las *2D-DFT* de las funciones de autocorrelación de \mathbf{x} y \mathbf{n} . En esta tesis, se obtiene una expresión del filtro de Wiener basada en el modelo *LC-GRF* y se analiza su performance en forma teórica [34].

4.5.1. Filtro de Wiener para el caso *LC-GRF*

Una de las ventajas del modelo *LC-GRF*, es que el filtro de Wiener puede ser implementado directamente ya que la densidad espectral de la señal es conocida y tiene una expresión sencilla (ecuación (4.16)). Usando que el ruido *AWGN* tiene una densidad espectral plana $S_{u,v}^n = \rho_n$ para todo $u, v = 0, 1, \dots, l-1$, con ρ_n siendo la varianza del ruido ($\rho_n = \sigma_n^2$), e introduciendo (4.16) en (4.35), se obtiene:

$$\hat{\tilde{x}}_{uv} = w_{uv} \tilde{y}_{uv} \quad (4.36)$$

donde el filtro de Wiener está dado por:

$$w_{uv} = \left(1 + \frac{\rho_n}{\rho_x} \lambda_{uv}^\alpha \right)^{-1} \quad (4.37)$$

La ecuación (4.36), en su forma matricial es:

$$\hat{\tilde{\mathbf{x}}} = W \tilde{\mathbf{y}} \quad (4.38)$$

donde $W \in \mathbb{R}^{l^2 \times l^2}$ es diagonal y los elementos de la diagonal principal son w_{uv} . Utilizando la definición de la operación *2D-DFT* se puede escribir:

$$\hat{\tilde{\mathbf{x}}} = U^H \hat{\mathbf{x}}$$

y

$$\tilde{\mathbf{y}} = U^H \mathbf{y}$$

por lo tanto, la ecuación (4.38) se transforma en:

$$\hat{\mathbf{x}} = UWU^H \mathbf{y} = T \mathbf{y} \quad (4.39)$$

donde queda escrito de manera explícita la forma lineal del filtro de Wiener:

$$T = UWU^H \quad (4.40)$$

Por otro lado es fácil ver que la densidad espectral del modelo *LC-GRF* tiene predominio de bajas frecuencias (largas correlaciones) por lo tanto el filtro de Wiener resulta ser un filtro pasa bajos atenuando las componentes de alta frecuencia.

En la Fig. 4.4 se ilustra el efecto del filtro de Wiener en el espacio de las imágenes transformadas por Fourier, donde u representa a la frecuencia vertical y v a la frecuencia horizontal. En la Fig. 4.4. a) se muestra el módulo de la transformada de Fourier discreta en 2D (*2D-DFT*) de la señal útil \mathbf{x} , que corresponde a un *LC-GRF* con parámetros $\alpha = 3$, $\rho_x = 1,0$ y $\theta = \theta_1 = \theta_2 = 0,249$, puede observarse que el contenido de esta imagen está concentrado en las bajas frecuencias, es decir, en un entorno de $(u, v) = (0, 0)$. En la Fig. 4.4. b) se muestra el módulo de la *2D-DFT* del ruido aditivo gaussiano \mathbf{n} cuya varianza es igual a la de la señal útil ($SNR = 0\text{dB}$), puede apreciarse que el contenido del ruido es plano en todo el

rango de frecuencias ya que se trata de ruido blanco. En la Fig. 4.4. c) se muestra el módulo de la *2D-DFT* de señal con ruido $\mathbf{y} = \mathbf{x} + \mathbf{n}$ que es la superposición de los espectros de Fourier de la señal útil \mathbf{x} y del ruido aditivo gaussiano \mathbf{n} . En la Fig. 4.4. d) se muestra el módulo del filtro de Wiener w_{uv} obtenido con la fórmula (4.37) el cual toma valores entre 0 y 1. Se observa que el efecto de aplicar el filtro de Wiener es el de atenuar las componentes de frecuencias altas donde la señal útil es más débil y de dejar prácticamente inalteradas las componentes de bajas frecuencias donde la señal útil es preponderante sobre el ruido blanco. Finalmente, en la Fig. 4.4. d) se muestra el módulo de la *2D-DFT* de la señal restaurada $\hat{\mathbf{x}}$ a partir del filtro de Wiener donde se observa claramente que el espectro de la imagen restaurada (Fig. 4.4. d)) es idéntico al de la imagen original (Fig. 4.4. a)). Por otro lado, en este ejemplo se observa que las correlaciones de largo alcance producen un alto contenido de bajas frecuencias.

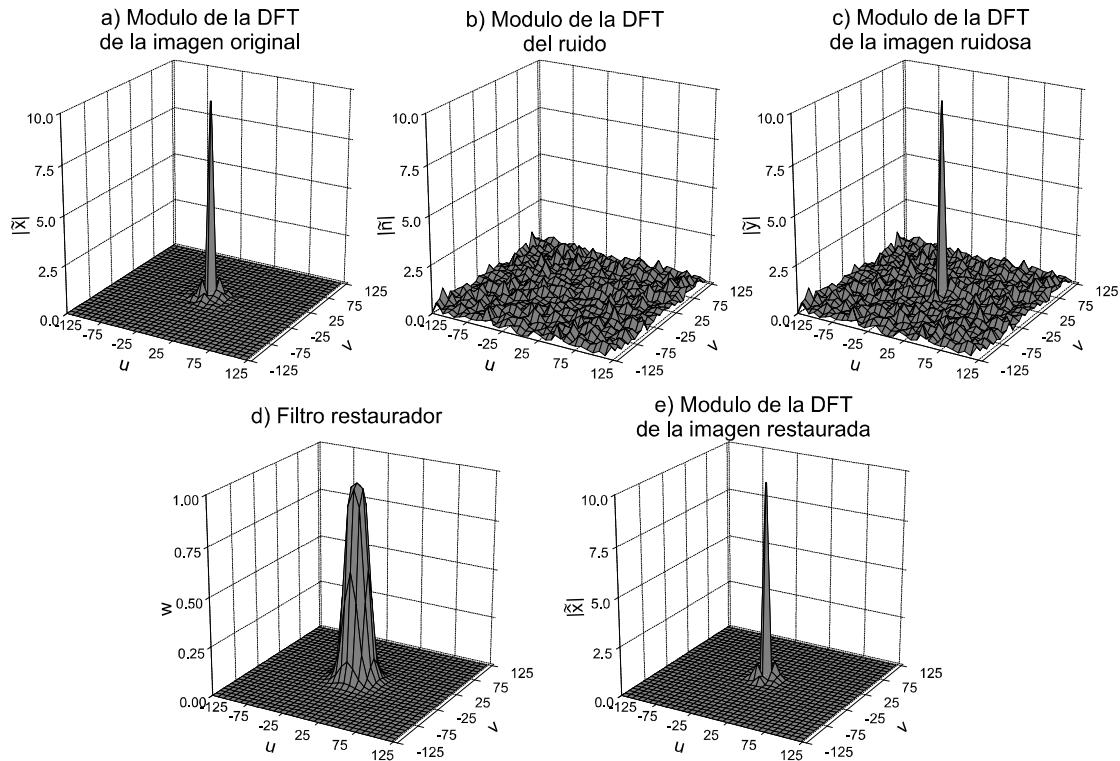


FIGURA 4.4. Interpretación del filtro de Wiener en el espectro de frecuencias de Fourier.

4.5.2. Performance teórica del filtro de Wiener

El filtro de Wiener provee la mejor estimación $\hat{\mathbf{x}}$ de la señal en el sentido que minimiza el error cuadrático medio (*MSE*), pero no elimina completamente el ruido. A continuación se incluye un análisis teórico de la calidad de la restauración producida por el filtro de Wiener a través de la varianza del error remanente que se define como:

$$\mathbf{e} = \widehat{\mathbf{x}} - \mathbf{x} \quad (4.41)$$

Para medir la performance de la restauración, se comparan las relaciones señal a ruido antes y después de aplicar el filtro de Wiener. Es decir, dada la definición habitual de la relación señal a ruido (*Signal to Noise Ratio - SNR*) se definen la SNR_i (de entrada o *input*) y la SNR_o (de salida o *output*) de la siguiente manera:

$$SNR_i(\text{dB}) = 20 \log \left(\frac{\sigma_{\mathbf{x}}}{\sigma_{\mathbf{n}}} \right) \quad (4.42)$$

$$SNR_o(\text{dB}) = 20 \log \left(\frac{\sigma_{\mathbf{x}}}{\sigma_{\mathbf{e}}} \right) \quad (4.43)$$

y la “ganancia” de la restauración es:

$$G(\text{dB}) = SNR_o(\text{dB}) - SNR_i(\text{dB}) = 20 \log \left(\frac{\sigma_{\mathbf{n}}}{\sigma_{\mathbf{e}}} \right) \quad (4.44)$$

donde $\sigma_{\mathbf{n}}$ y $\sigma_{\mathbf{e}}$ son las desviaciones estándar del ruido y el error respectivamente. Un valor de $G = 0\text{dB}$ indica que no se ha mejorado el nivel de ruido remanente mientras que, cuanto mayor sea el valor de G , mayor será la reducción de ruido lograda. Usando (4.37), (4.39) y (4.41), se obtiene (Apéndice 4.8.3):

$$G(\text{dB}) = -10 \log \left[\frac{1}{l^2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \left(1 + \frac{\rho_{\mathbf{n}}}{\rho_{\mathbf{x}}} \lambda_{uv}^{\alpha} \right)^{-1} \right] \quad (4.45)$$

En la Fig. 4.5 se muestra la ganancia en la restauración versus SNR_i para $\alpha = 0,5; 2,5$ y $5,0$ con $\rho_{\mathbf{x}} = 0,0225$ y $\theta_1 = \theta_2 = 0,249$. Se observa que, cuanto más correlacionado es el campo se consiguen mejores restauraciones, por ejemplo, para $SNR=0\text{dB}$ la ganancia obtenida para $\alpha = 0,5$ es $G = 7,55\text{dB}$ mientras que para $\alpha = 5,0$ la ganancia es $G = 51,88\text{dB}$.

4.6. Evaluación experimental de los algoritmos

4.6.1. Estimación de α , $\rho_{\mathbf{x}}$ y θ

En esta subsección se analiza experimentalmente la precisión del algoritmo propuesto en la Subsección 4.4.1 para la estimación de los parámetros del modelo *LC-GRF*. Se han realizado estimaciones sobre un total de $20 \times 9 \times 10 = 1800$ casos (20 repeticiones del experimento para promediar los resultados, 9 casos diferentes de valores en los parámetros θ_1 y θ_2 ; y 10 casos diferentes de valores en el parámetro α).

Se utilizaron imágenes de 256×256 generadas con un algoritmo equivalente a los propuestos en [25, 41] que utiliza la transformada rápida de Fourier *FFT* (*Fast Fourier Transform*). Se utilizó un esquema de vecinos más cercanos (orden $o = 2$) seleccionando los parámetros θ_1 y θ_2 de entre los siguientes valores $-0,24; -0,16; -0,08; 0,08; 0,16; 0,24$ y se consideró $\alpha = 0,5; 1,0; 1,5; \dots; 5,0$. El parámetro

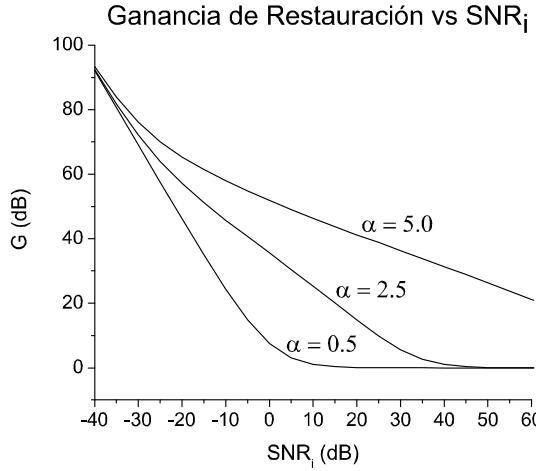


FIGURA 4.5. Ganancia de restauración G vs SNR_i para $\rho_{\mathbf{x}} = 0,0225$ y $\theta_1 = \theta_2 = 0,249$

$\rho_{\mathbf{x}}$ fue calculado a partir de la ecuación (4.15) de manera de asegurar una varianza unitaria ($\sigma_{\mathbf{x}} = 1,0$). Los valores iniciales de los parámetros son $\alpha^{(0)} = 0,1$, $\rho_{\mathbf{x}}^{(0)} = 1,0$ y $\theta^{(0)} = 0,0$. Los parámetros de optimización fueron seleccionados empíricamente como $\mu_{su} = 1,1$ (*speed-up*) y $\mu_{sd} = 2,0$ (*slow-down*); y los pasos iniciales utilizados son: $d_1^{(0)} = d_2^{(0)} = 1,0 \times 10^{-8}$ y $h^{(0)} = 1,0 \times 10^{-7}$.

En la Fig. 4.6 se muestra la precisión de las estimaciones luego de 400 iteraciones en términos del error absoluto ($|\alpha - \hat{\alpha}|$ y $|\theta_1 - \hat{\theta}_1|$) (no se muestran los resultados para θ_2 ya que son análogos a los del parámetro θ_1). Los errores absolutos graficados son los valores promediados sobre las 20 simulaciones hechas para cada conjunto de parámetros.

En la Fig. 4.6-arriba se muestran los valores para los casos simétricos ($\theta_1 = \theta_2$). Se observa que las estimaciones son mejores para los casos en que las correlaciones de largo alcance son más importantes, es decir, para valores mayores de α y parámetros de interacción cercanos al límite permitido, i.e. $|\theta_1| + |\theta_2| \simeq 0,5$.

Otra observación importante es que la estimación *ML* del parámetro α solo es posible cuando θ_1 y θ_2 son no nulos, lo cual puede ser verificado directamente de la ecuación (4.18) ya que, cuando $\theta_1 = \theta_2 = 0$, los autovalores son $\lambda_{u,v} = 1$ y la función de *log-likelihood* resulta:

$$\Gamma(\rho_{\mathbf{x}}, \boldsymbol{\theta}) = -\frac{l^2}{2} \log(2\pi\rho_{\mathbf{x}}) - \frac{1}{2\rho} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} |\tilde{x}_{u,v}|^2 \quad (4.46)$$

que es independiente del parámetro α . Por esta razón cuando los parámetros θ_1 y θ_2 tienden a cero, los errores en la estimación de α tienden a ser grandes (Fig. 4.6-arriba-derecha para $\theta = \pm 0,08$).

El algoritmo presentado también puede ser aplicado a casos no simétricos, es decir, cuando $\theta_1 \neq \theta_2$ tal como se muestra en la Fig. 4.6-abajo, donde se

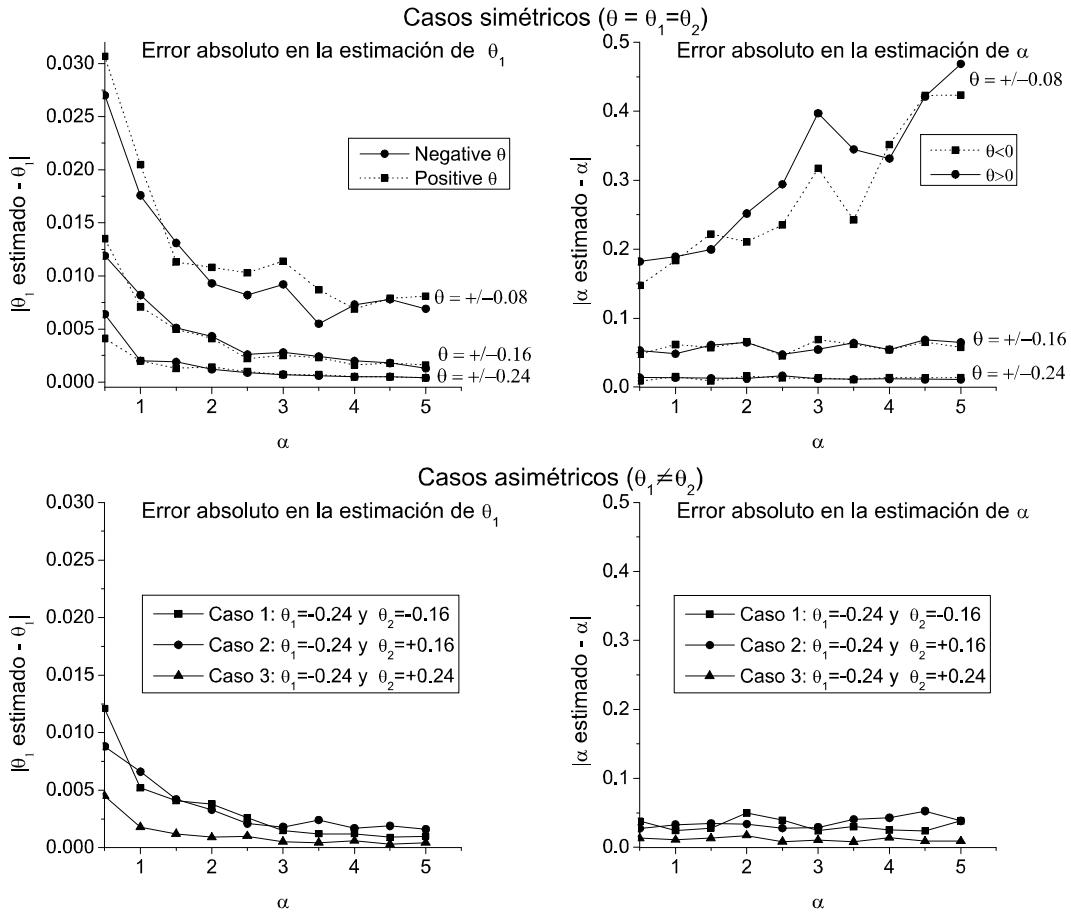


FIGURA 4.6. Precisión en la estimación de α y θ_1 luego de 400 iteraciones para los casos simétricos (arriba) y asimétricos (abajo).

han considerado tres casos diferentes de parámetros no simétricos. En la Fig. 4.7 se muestra la evolución de los parámetros estimados $\hat{\alpha}$, $\hat{\rho}_x$ y $\hat{\theta}_1$ para seis casos diferentes ($\alpha = 0,5; 2,5; 5,0$ y $\theta = \theta_1 = \theta_2 = 0,16; 0,24$). Se observa que el algoritmo requiere de aproximadamente 150 iteraciones para converger a un valor muy cercano del esperado.

4.6.2. Estimación de las varianzas de la señal y el ruido a partir de una muestra con ruido

Aquí se presentan resultados experimentales donde se muestra la eficiencia de la técnica presentada en la Subsección 4.4.2 utiliza el algoritmo *EM* para el caso de imágenes modeladas con un *LC-GRF* contaminadas con ruido aditivo gaussiano blanco (*AWGN*). Para estos experimentos se han generado imágenes de 256×256 píxeles usando el modelo *LC-GRF* con parámetros $\theta_1 = \theta_2 = 0,249$, con α en el rango entre 0,5 y 5,0. A estas imágenes se las ha contaminado con ruido gaussiano aditivo blanco manteniendo una relación señal a ruido constante *SNR*= 0dB, es decir, $\sigma_x = \sigma_n$ (en particular se utilizó $\sigma_x = \sigma_n = 1,0$).

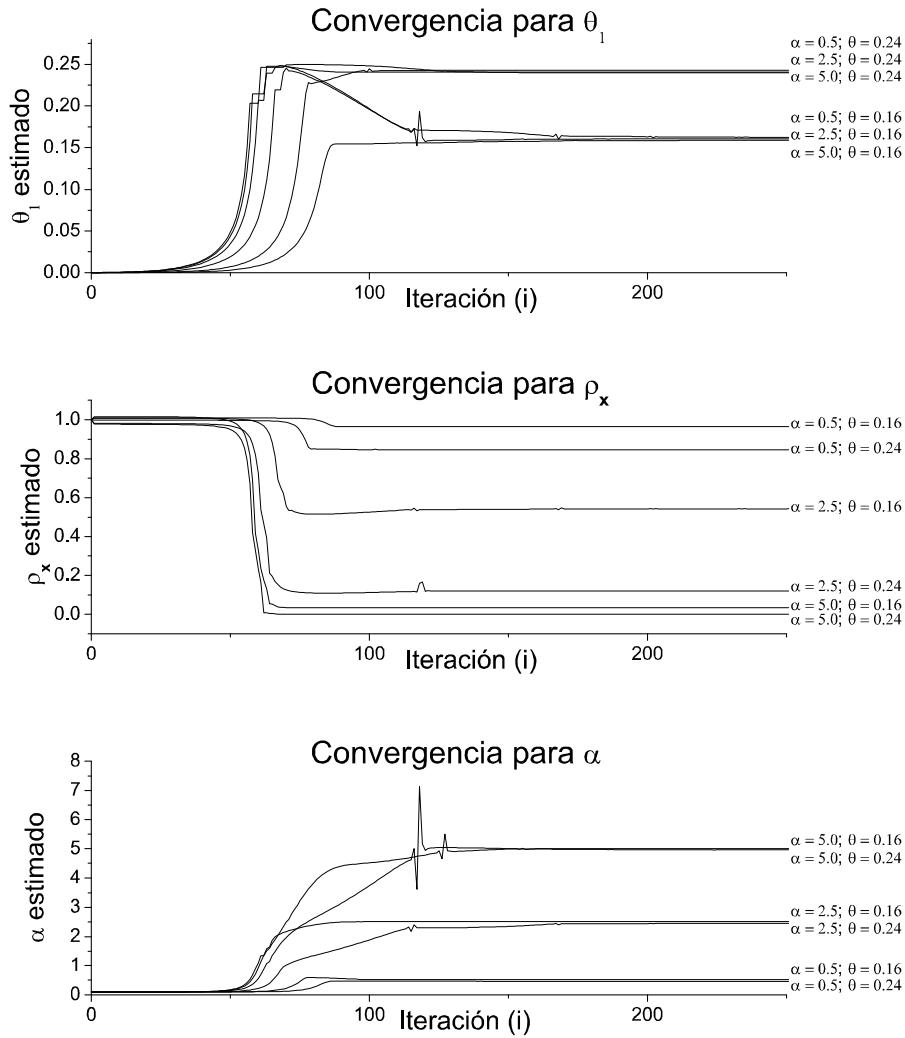


FIGURA 4.7. Evolución de las estimaciones $\hat{\alpha}$, $\hat{\rho}_x$ y $\hat{\theta}_1$ versus el número de iteración i

El algoritmo *EM* fue utilizado para estimar los parámetros ρ_x y ρ_n . Se observa que siempre se cumple que $\rho_n = 1$, mientras que ρ_x está determinado por (4.15). Las condiciones iniciales del algoritmo utilizadas fueron $\hat{\rho}_x^{(0)} = \hat{\rho}_n^{(0)} = 5,0$ y las iteraciones del algoritmo *EM* fueron repetidas hasta lograr que el cambio en los parámetros en una iteración sea suficientemente chico como para asumir que se arribó al punto fijo, en este caso se uso un umbral de $1,0 \times 10^{-7}$. En la Fig. 4.8 (izquierda), se muestran el error absoluto final para ambos parámetros versus α (se han promediado los resultados sobre un total de 20 casos para cada α). Nuevamente, se observa que los resultados son mejores para α mayores. En la Fig. 4.8 (derecha) se muestra la evolución de las estimaciones de los parámetros para el caso $\alpha = 2,5$; y se puede decir que se requieren aproximadamente 100 iteraciones para que las estimaciones converjan a sus valores finales.

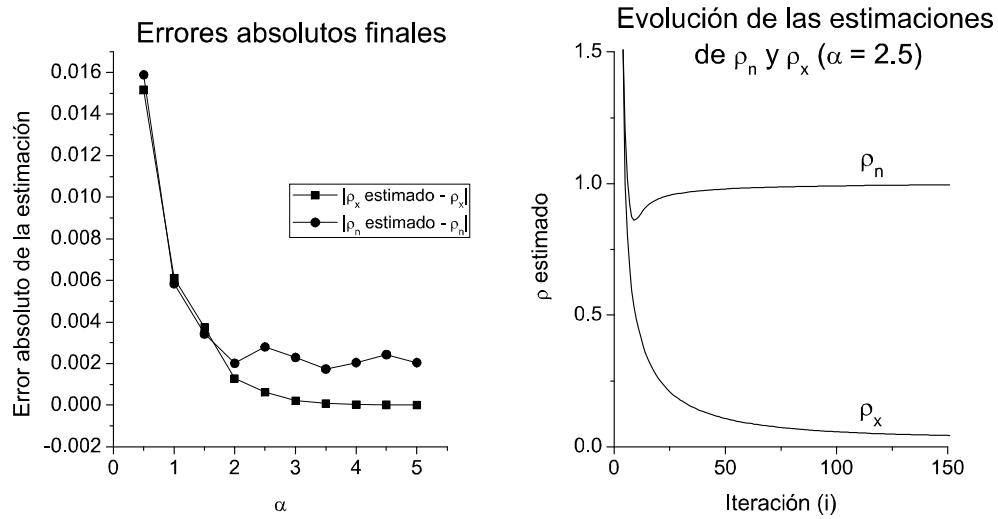


FIGURA 4.8. Eficiencia del algoritmo: errores finales (izquierda) y convergencia (derecha)

4.7. Apéndices

4.7.1. Demostración de la función de Log-likelihood $\Gamma(\alpha, \rho_{\mathbf{x}}, \theta)$

Tomando el logaritmo de la ecuación (4.9), se obtiene:

$$\Gamma(\alpha, \rho_{\mathbf{x}}, \theta) = \frac{\alpha}{2} \log(|B|) - \frac{l^2}{2} \log(2\pi\rho_{\mathbf{x}}) - \frac{1}{2\rho_{\mathbf{x}}} \mathbf{x}^T B^\alpha \mathbf{x} \quad (4.47)$$

Utilizando la descomposición canónica de la matriz B^α (ecuación (4.8)), aplicando propiedades del determinante y usando la definición de la operación $2D-DFT$ ($\tilde{\mathbf{x}} = U^H \mathbf{x}$), se obtiene finalmente la ecuación (4.18).

4.7.2. Detalles del algoritmo *EM*

En el *E-step* se requiere del cálculo de la esperanza condicionada (miembro derecho en la ecuación (4.29)). La función *log-likelihood* de los datos completos condicionados a los parámetros está dada por:

$$\log [f(C; \rho'_{\mathbf{x}}, \rho'_{\mathbf{n}})] = \log [f(\mathbf{x}; \rho'_{\mathbf{x}})] + \log [f(\mathbf{n}; \rho'_{\mathbf{n}})] \quad (4.48)$$

donde se utilizó la independencia entre \mathbf{x} y \mathbf{n} . Además como \mathbf{x} y \mathbf{n} tienen ambos distribuciones gaussianas conocidas, sus *pdf's* conjuntas son:

$$f(\mathbf{x}; \rho'_{\mathbf{x}}) = \frac{|B|^{\alpha/2}}{(2\pi\rho'_{\mathbf{x}})^{l^2/2}} \exp\left(-\frac{1}{2\rho_{\mathbf{x}}} \mathbf{x}^T B^\alpha \mathbf{x}\right) \quad (4.49)$$

$$f(\mathbf{n}; \rho'_{\mathbf{n}}) = \frac{1}{(2\pi\rho'_{\mathbf{n}})^{l^2/2}} \exp\left(-\frac{1}{2\rho_{\mathbf{n}}} \mathbf{n}^T \mathbf{n}\right) \quad (4.50)$$

Introduciendo (4.49) y (4.50) en (4.48) se obtiene la siguiente ecuación:

$$\begin{aligned} \log [f(C; \rho'_{\mathbf{x}}, \rho'_{\mathbf{n}})] &= \frac{\alpha}{2} \log(|B|) - \frac{l^2}{2} \log(2\pi\rho'_{\mathbf{x}}) \\ &\quad - \frac{l^2}{2} \log(2\pi\rho'_{\mathbf{n}}) - \frac{1}{2\rho'_{\mathbf{x}}} \mathbf{x}^T B^\alpha \mathbf{x} - \frac{1}{2\rho'_{\mathbf{n}}} \mathbf{n}^T \mathbf{n} \end{aligned} \quad (4.51)$$

Aplicando la esperanza condicionada a esta ecuación, utilizando la descomposición canónica de la matriz B^α (ecuación (4.8)) y usando propiedades del determinante, se obtiene la siguiente expresión:

$$\begin{aligned} E \{ \log f(C; \rho'_{\mathbf{x}}, \rho'_{\mathbf{n}}) \mid \mathbf{y}; \rho_{\mathbf{x}}, \rho_{\mathbf{n}} \} &= \\ \frac{\alpha}{2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \log(\lambda_{u,v}) - \frac{l^2}{2} \log(\rho'_{\mathbf{x}}) - \frac{l^2}{2} \log(\rho'_{\mathbf{n}}) \\ - \frac{1}{2\rho'_{\mathbf{x}}} \Phi_1(\rho_{\mathbf{x}}, \rho_{\mathbf{n}}, \mathbf{y}) - \frac{1}{2\rho'_{\mathbf{n}}} \Phi_2(\rho_{\mathbf{x}}, \rho_{\mathbf{n}}, \mathbf{y}) \end{aligned} \quad (4.52)$$

donde se definen

$$\Phi_1(\rho_{\mathbf{x}}, \rho_{\mathbf{n}}, \mathbf{y}) = E \{ \mathbf{x}^T B^\alpha \mathbf{x} \mid \mathbf{y}; \rho_{\mathbf{x}}, \rho_{\mathbf{n}} \} \quad (4.53)$$

$$\Phi_2(\rho_{\mathbf{x}}, \rho_{\mathbf{n}}, \mathbf{y}) = E \{ \mathbf{n}^T \mathbf{n} \mid \mathbf{y}; \rho_{\mathbf{x}}, \rho_{\mathbf{n}} \} \quad (4.54)$$

En el *M-step* se necesita maximizar (4.52) con respecto a los parámetros $\rho'_{\mathbf{x}}$ y $\rho'_{\mathbf{n}}$. Calculando sus derivadas e igualándolas a cero, se obtiene finalmente:

$$\rho'_{\mathbf{x}} = \frac{1}{l^2} \Phi_1(\rho_{\mathbf{x}}, \rho_{\mathbf{n}}, \mathbf{y}) \quad (4.55)$$

$$\rho'_{\mathbf{n}} = \frac{1}{l^2} \Phi_2(\rho_{\mathbf{x}}, \rho_{\mathbf{n}}, \mathbf{y}) \quad (4.56)$$

las cuales son exactamente las ecuaciones (4.31) y (4.32) si se reemplaza $\rho'_{\mathbf{x}}$ y $\rho'_{\mathbf{n}}$ por $\hat{\rho}_{\mathbf{x}}^{(i+1)}$ y $\hat{\rho}_{\mathbf{n}}^{(i+1)}$; $\rho_{\mathbf{x}}$ y $\rho_{\mathbf{n}}$ por $\hat{\rho}_{\mathbf{x}}^{(i)}$ y $\hat{\rho}_{\mathbf{n}}^{(i)}$.

A los efectos de finalizar la demostración, debemos encontrar las expresiones de (4.53) y (4.54) las cuales, como son esperanzas condicionadas, requieren la utilización de las *pdf*'s condicionales $f_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}; \rho_{\mathbf{x}}, \rho_{\mathbf{n}})$ y $f_{\mathbf{n}|\mathbf{y}}(\mathbf{n} \mid \mathbf{y}; \rho_{\mathbf{x}}, \rho_{\mathbf{n}})$. Si se utiliza el hecho que \mathbf{x} y \mathbf{n} son independientes, aplicando propiedades básicas de probabilidades como el teorema de Bayes, usando los modelos paramétricos para las *pdf*'s de \mathbf{x} y \mathbf{n} (ecuaciones (4.49) y (4.50)), y utilizando nuevamente la descomposición de la matriz B^α (ecuación (4.8)), se obtienen finalmente las ecuaciones (4.33) y (4.34).

4.7.3. Demostración de la Ganancia en la restauración

Primero se observa que el error remanente \mathbf{e} , luego de aplicar el filtro de restauración, puede ser escrito en términos de la señal útil \mathbf{x} y el ruido \mathbf{n} usando la ecuación (4.39) como sigue:

$$\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x} = T\mathbf{y} - \mathbf{x} = T(\mathbf{x} + \mathbf{n}) - \mathbf{x} = (T - I)\mathbf{x} + T\mathbf{n} \quad (4.57)$$

Por lo tanto podemos calcular la matriz de covarianzas del error directamente:

$$C_{\mathbf{ee}} = E \{ \mathbf{ee}^T \} = (T - I) C_{\mathbf{xx}} (T - I) + T C_{\mathbf{nn}} T \quad (4.58)$$

donde se utilizó la independencia entre \mathbf{x} y \mathbf{n} . Si se utiliza la descomposición canónica de $C_{\mathbf{xx}}$, la ecuación (4.38) y que $U^H U = I$, se obtiene:

$$C_{\mathbf{ee}} = U \left[(W - I) \rho_{\mathbf{x}} \Lambda^{-\alpha} (W - I) + \rho_{\mathbf{n}} W^2 \right] U^H \quad (4.59)$$

La varianza del error puede calcularse de la siguiente forma:

$$\sigma_{\mathbf{e}}^2 = \frac{1}{l^2} \text{Tr} \{ C_{\mathbf{ee}} \} \quad (4.60)$$

Introduciendo (4.59) en (4.60) y aplicando propiedades básicas de la traza de una matriz, se arriba a:

$$\sigma_{\mathbf{e}}^2 = \frac{1}{l^2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} (w_{uv} - 1)^2 \rho_{\mathbf{x}} \lambda_{uv}^{-\alpha} + \rho_{\mathbf{n}} w_{uv} \quad (4.61)$$

y, reemplazando w_{uv} y λ_{uv}^α por sus correspondientes valores, se obtiene una expresión final para la varianza del error remanente:

$$\sigma_{\mathbf{e}}^2 = \frac{1}{l^2} \sum_{u=0}^{l-1} \sum_{v=0}^{l-1} \frac{\rho_{\mathbf{n}}}{1 + \frac{\rho_{\mathbf{n}}}{\rho_{\mathbf{x}}} \lambda_{uv}^\alpha} \quad (4.62)$$

Finalmente usando que $\sigma_{\mathbf{n}}^2 = \rho_{\mathbf{n}}$ se arriba a la ecuación deseada (4.45).

Capítulo 5

APLICACIONES AL PROCESAMIENTO DE IMÁGENES

En esta tesis se presentan aplicaciones originales para el procesamiento de imágenes basadas en las teorías y algoritmos introducidos en los capítulos anteriores. En particular se presentan dos aplicaciones de la representación *KLT* en el reconocimiento de patrones, y dos casos de aplicación del algoritmo MAXNG para la separación de fuentes estadísticamente dependientes en imágenes satelitales e imágenes de origen astrofísico. En éste último caso también se muestra cómo el modelo *LC-GRF* puede ser utilizado para reducir el ruido en imágenes del *CMB* (*Cosmic Microwave Background*).

5.1. Clasificación de imágenes de *Mirounga Leonina*

En 1991, Turk & Pentland [122] introdujeron una técnica de reconocimiento de patrones basada en la *KLT* conocida como el método de las *eigenfaces*. En esta tesis se presenta una aplicación al reconocimiento de especímenes de *Mirounga Leonina*, vulgarmente conocidos como elefantes marinos [26, 27]. La técnica aquí presentada permite a los biólogos trabajar en la clasificación de animales a partir de las fotografías de sus rostros, sin recurrir a métodos invasivos, constituyendo un aporte original a la identificación de especímenes.

Las imágenes digitales utilizadas son monocromáticas, de 256×256 píxeles y a cada píxel se le asigna un nivel de gris en el rango de 0 (negro) a 255 (blanco) (cuantización de 8 bits). Por lo tanto cada imagen será asociada a un vector de \mathbb{R}^n con $n = 256 \times 256 = 65536$.

5.1.1. Identificación de individuos

Dado un conjunto de $m = 56$ imágenes de distintos especímenes que constituyen los patrones $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ (Fig. 5.2-izq.), el objetivo es identificar una imagen alternativa \mathbf{x} de cualquiera de esos individuos. Es natural asociar la imagen \mathbf{x} con aquel patrón con el cual tenga distancia euclídea mínima¹, es decir:

$$\mathbf{x} \text{ pertenece al individuo } j \text{ si } \|\mathbf{x} - \mathbf{y}_j\|^2 < \|\mathbf{x} - \mathbf{y}_i\|^2 \quad \forall i \neq j \quad (5.1)$$

Como fuera demostrado en el Cap. 2, Subsección 2.2.1, estas distancias al cuadrado pueden ser escritas en términos de las distancias en el subespacio de los patrones y en su complemento ortogonal:

$$\|\mathbf{x} - \mathbf{y}_j\|^2 = \text{DASP}^2(\mathbf{x}) + \text{DESP}_j^2(\mathbf{x}) \quad (5.2)$$

¹Existe una conexión directa entre esta clasificación y la definición de los diagramas de Voronoi para \mathbb{R}^n [125].

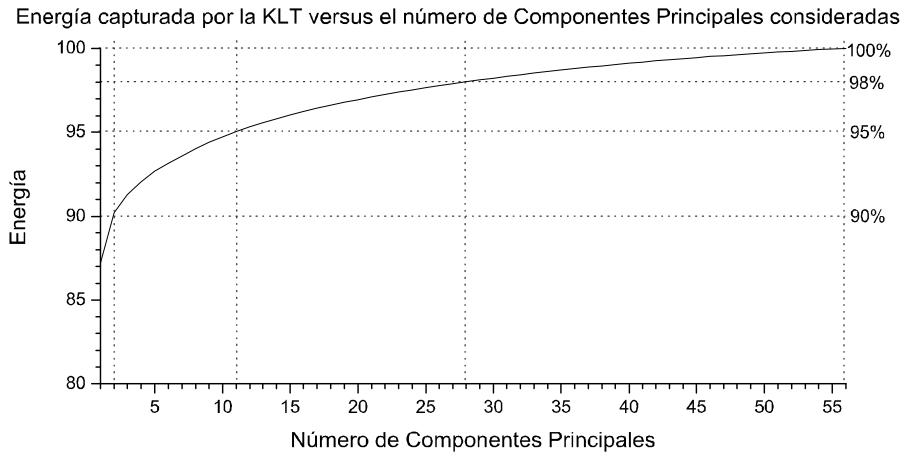


FIGURA 5.1. Con 2 elementos de la base se captura el 90 % de la energía de los patrones; con 11 el 95 % y con 28 el 98 %.

En la práctica la distancia al subespacio de los patrones $DASP(\mathbf{x})$ no aporta información respecto de la similitud con alguno de los patrones. Por el contrario, la distancia de la proyección de \mathbf{x} a cada uno de los patrones, indica qué tan parecido es a cada uno de los patrones conocidos. El método de las *eigenfaces* consiste en asociar a \mathbf{x} con el individuo cuya $DESP_j(\mathbf{x})$ es mínima [122].

Si bien la distancia $DESP_j(\mathbf{x})$ podría ser calculada con cualquier base ortonormal, la ventaja de utilizar la *KLT* es que pueden usarse solamente las componentes principales sin introducir errores considerables. De hecho, y como se discutió en el Cap. 2, los autovalores de la matriz de covarianzas indican la cantidad de energía asociada con cada *eigenface* siendo la *KLT* óptima respecto a la concentración de las energías. En la figura 5.1 se muestra la energía capturada por las componentes *KLT* (*eigenfaces*) donde se observa claramente que la energía total se concentra en unas pocas *eigenfaces* significativas (por ejemplo, las primeras 2 *eigenfaces* concentran el 90 % de la energía y con las primeras 11 se arriba al 95 % de la energía total).

Sobre este conjunto de 56 patrones se calcularon las *eigenfaces* (Fig. 5.2-der.) y se sometieron al algoritmo de clasificación fotografías alternativas de los individuos de las fotos originales. A los efectos de analizar el criterio de clasificación basado en las distancias mínimas, en los siguientes ejemplos se han utilizado la totalidad de las *eigenfaces*, sin embargo es importante tener en cuenta que, si el número de patrones fuera elevado (por ejemplo 10000 o más) la reducción en la dimensión del espacio de las *eigenfaces* es muy importante para reducir el tiempo de cálculo y economizar recursos computacionales. En la Fig. 5.3 se muestran 3 ejemplos de imágenes originales y las correspondientes imágenes alternativas a clasificar. En la Fig. 5.4 se muestran los valores de las distancias $DESP_j(\mathbf{x})$ para cada uno de

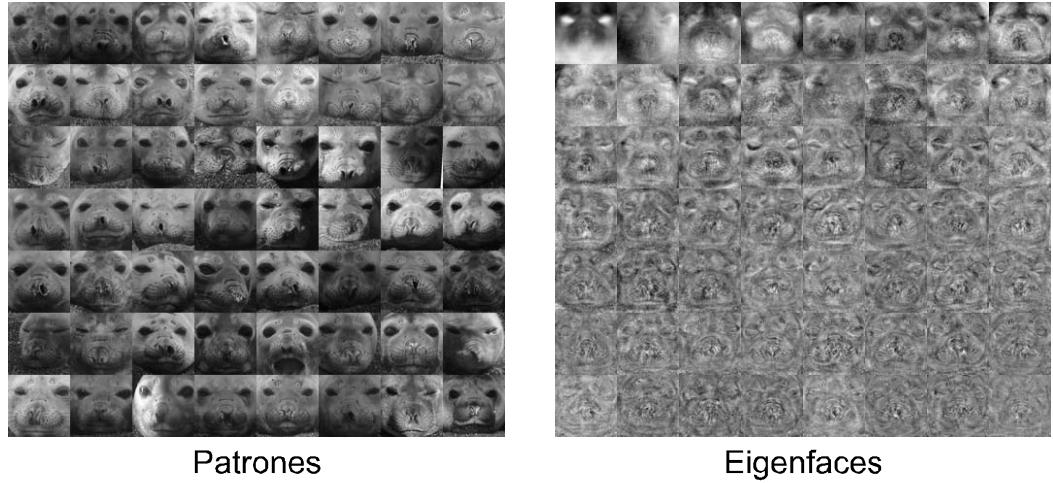


FIGURA 5.2. Las *Eigenfaces* están ordenadas decrecientemente de izquierda a derecha y de arriba hacia abajo.

los tres casos de clasificación y puede observarse que, en todos los casos el mínimo obtenido corresponde con una clasificación correcta.

A los efectos de analizar estadísticamente los resultados se han realizado un total de 40 clasificaciones de imágenes alternativas obteniendo una efectividad del 100 %, además se calculó la media y el desvío estándar de los valores mínimos y no-mínimos de $\text{DESP}_j(\mathbf{x})$ obtenidos en cada caso. En la Fig. 5.4 se muestran las franjas de los valores mínimos y no-mínimos y se observa que las mismas se encuentran suficientemente separadas como para evitar errores de clasificación.

5.1.2. Clasificación en harems

A veces en lugar de identificar una imagen asociándola con un individuo, es suficiente con asociar dicha imagen como perteneciente a un grupo determinado de especímenes o *harem*. En este caso, dados diferentes grupos de patrones o *harems*, resulta más útil medir la distancia de la muestra a clasificar a los distintos subespacios de patrones. Por ejemplo, dados dos grupos: *harem A* y *harem B*, se puede basar la clasificación en el cálculo de las distancias $\text{DASP}_A(\mathbf{x})$ y $\text{DASP}_B(\mathbf{x})$. En [27] se propuso una variante que resultó más efectiva que consiste en calcular el ángulo al subespacio de patrones θ definido por:

$$\sin(\theta(\mathbf{x})) = \frac{\text{DASP}(\mathbf{x})}{\|\mathbf{x} - \mathbf{y}_0\|} \quad (5.3)$$

Por lo tanto surgen naturalmente los siguientes coeficientes de clasificación ρ_A y ρ_B para decidir si una imagen pertenece al *harem A* o *B*:

$$\rho_A = \left[\frac{\sin(\theta_A)}{\sin(\theta_B)} \right]^2 ; \rho_B = \frac{1}{\rho_A} \quad (5.4)$$

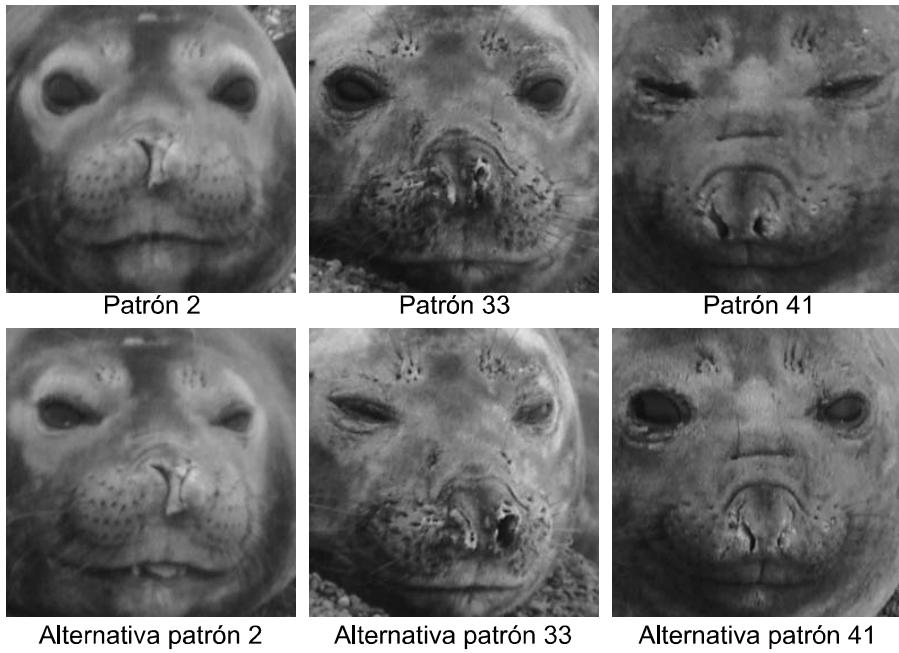


FIGURA 5.3. Imágenes originales de los patrones 2, 33 y 41 (arriba); y sus correspondientes imágenes alternativas sometidas a la clasificación (abajo)

Es decir, si $\rho_A < 1$, la muestra \mathbf{x} es clasificada como perteneciente al *harem A*, por el contrario si $\rho_A > 1$, se clasifica como perteneciente al *harem B*. En la Fig. 5.5 se muestran resultados experimentales para un caso donde dos *harems* han sido definidos arbitrariamente con 28 individuos cada uno, y donde se han sometido a la clasificación 36 fotografías alternativas. Se observó que en todos los casos la clasificación fue correcta.

5.2. Corregistro de imágenes satelitales

En esta tesis se propone un nuevo algoritmo para el corregistro de imágenes que explota la capacidad de la representación *KLT* para realizar reconocimiento de patrones[28]. En aplicaciones de teledetección (*remote sensing*) es muy común disponer de dos imágenes satelitales \mathbf{x} e \mathbf{y} las cuales cubren aproximadamente el mismo área geográfica y que deben ser alineadas una respecto de la otra. Las imágenes \mathbf{x} e \mathbf{y} se denominan imágenes primaria y secundaria, respectivamente. El objetivo de un proceso de corregistro, es proveer una transformación adecuada sobre la imagen primaria $\mathbf{x}' = T(\mathbf{x})$, de manera que, las imágenes \mathbf{x}' e \mathbf{y} cubran exactamente la misma área cuando son superpuestas [24, 93].

Un método tradicional de corregistro consiste en la identificación de *puntos de control* en ambas imágenes, es decir, posiciones que refieren a un mismo objeto generalmente asociados con accidentes geográficos fácilmente identificables por inspección visual (por ej.: edificios, cruce de caminos, etc.). Sin embargo, este tipo de tarea requiere un entrenamiento por parte del analista y los resultados del

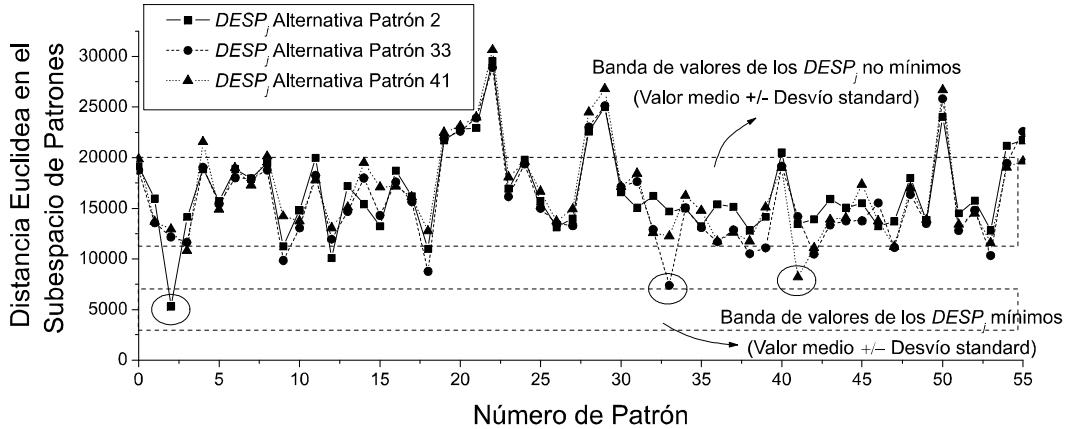


FIGURA 5.4. Distancia en el Subespacio de los Patrones *DESP* para imágenes de los patrones 2, 33 y 41. Se muestran las franjas de valores de las distancias mínimas y no-mínimas.

corregistro son altamente dependientes del nivel de entrenamiento del mismo. En los últimos años, se han desarrollado algoritmos de corregistro automático, tanto para imágenes satelitales como para el caso de aplicaciones en medicina como son las tomografías, ecografías, etc. [24, 109]. Estas técnicas automáticas se basan en la maximización del coeficiente de correlación (*cross correlation coefficient*) [24], en la maximización de la información mutua (*mutual information*) [98, 128] o en la minimización del error cuadrático [118].

En esta tesis se presenta un método original que define *puntos de control* de manera totalmente automática y utiliza la representación *KLT* como técnica de reconocimiento de patrones. Se presentan los resultados publicados en [28] con imágenes de tipo radar de apertura sintética de la ciudad de La Plata, Argentina, capturadas por el satélite ERS 1 de la agencia espacial europea (*European Space Agency - ESA*) los días 14-Oct-1997 y 21-Feb-2000. El algoritmo propuesto consiste en subdividir la imagen primaria en sub-imágenes a través de una grilla regular y asignar a cada sub-imagen un patrón \mathbf{y}_j (Fig. 5.6.a)). El algoritmo realiza la búsqueda de los patrones en la imagen secundaria a través de una ventana deslizante (Fig. 5.6.b)) calculando, para cada posición de la ventana, las distancias al subespacio de patrones. Es decir, si la posición de la ventana deslizante está definida por las coordenadas (k_1, k_2) , se calcula la distancia DASP² como función de las coordenadas (k_1, k_2) usando la siguiente ecuación derivada de la fórmula (2.55):

$$\text{DASP}^2(k_1, k_2) = \|\mathbf{y}'(k_1, k_2)\|^2 - \sum_{j=0}^m c_j^2(k_1, k_2) \quad (5.5)$$

donde $\mathbf{y}'(k_1, k_2)$ corresponde a la imagen centrada ($\mathbf{y}' = \mathbf{y} - \mathbf{y}_0$) determinada por la ventana deslizante en la posición (k_1, k_2) y $c_j(k_1, k_2)$ es el coeficiente asociado a la componente j en la representación de la imagen $\mathbf{y}(k_1, k_2)$.

La figura 5.6.c) muestra la imagen de los valores de $\text{DASP}^2(k_1, k_2)$ en la región

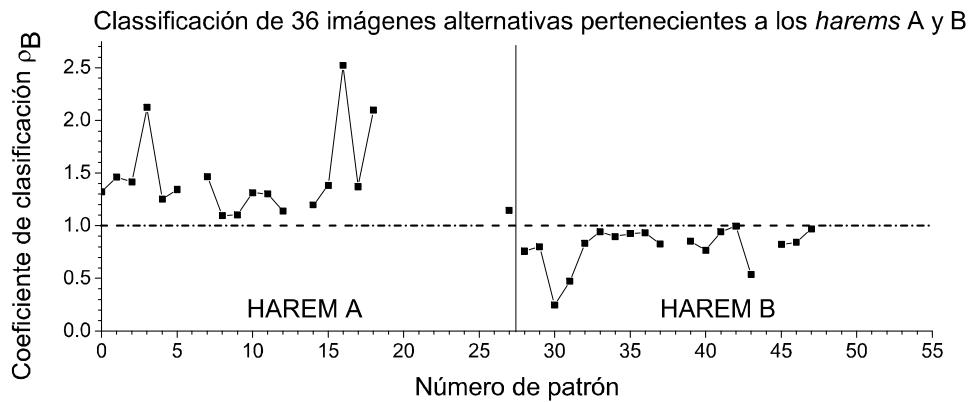


FIGURA 5.5. Cálculo del coeficiente de clasificación ρ_B para fotos alternativas de los *harems A* y *B*. Patrones 0-27 corresponden al *harem A*; y 28-55 al *harem B*.

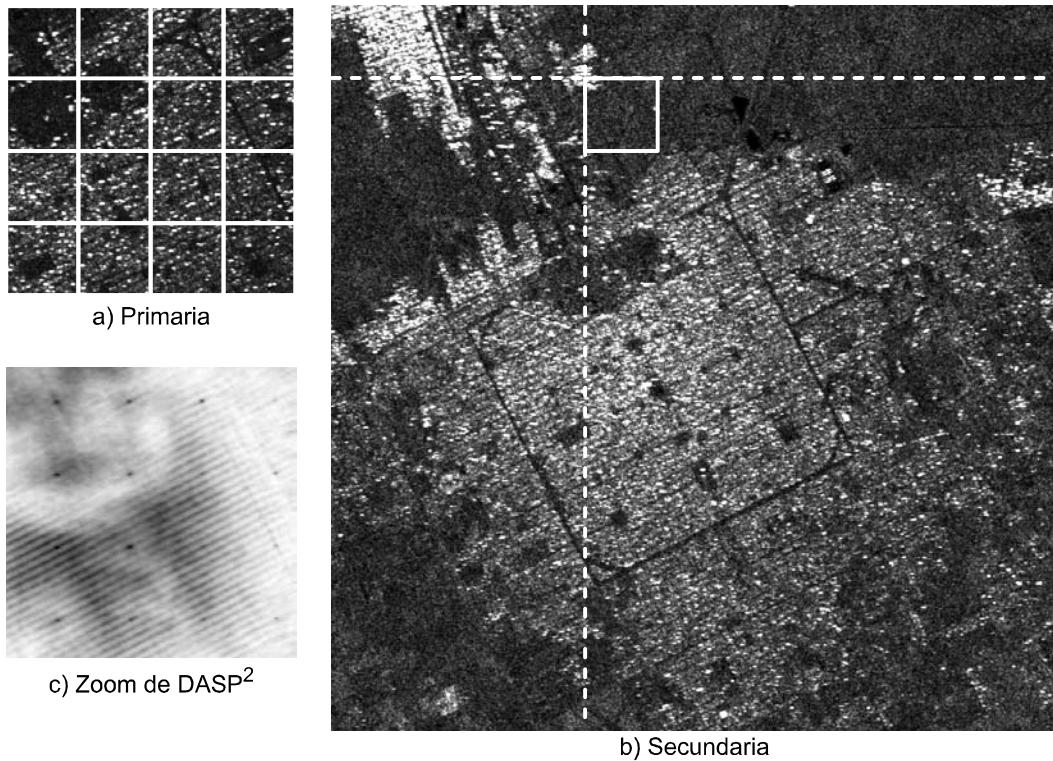


FIGURA 5.6. a) y b): Imágenes radar de la ciudad de La Plata: primaria (400×400) del 14-Oct-1997 y secundaria (1024×1024) del 21-Feb-2000. En la imagen secundaria se muestra la ventana deslizante de 100×100 . c) Mínimos locales de la función $DASP^2(k_1, k_2)$.

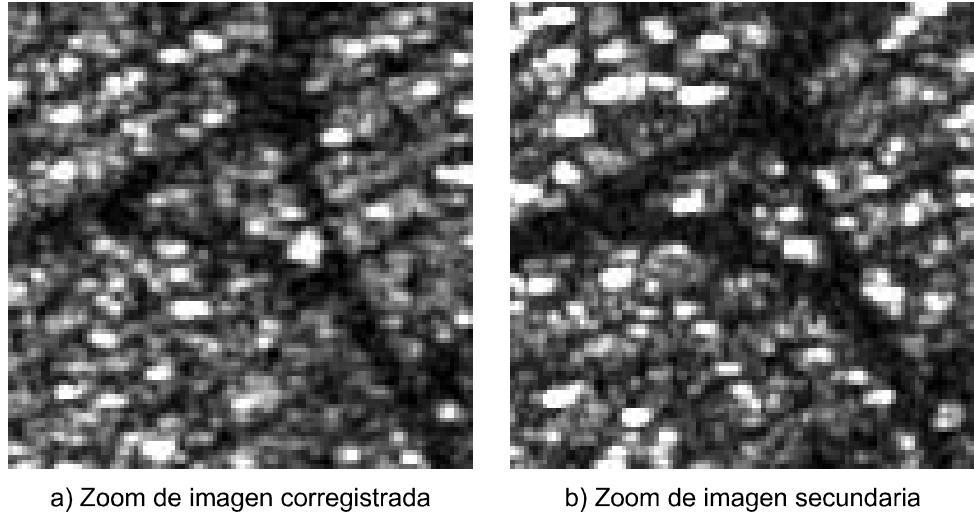


FIGURA 5.7. Zoom de la imagen primaria corregistrada con la imagen secundaria.

de interés de la imagen secundaria. Los puntos oscuros indican valores pequeños de esta magnitud. Como puede verse, la magnitud DASP²(k_1, k_2) es muy sensible a desplazamientos de la ventana, por ello aparecen puntos oscuros que representan mínimos locales pronunciados. Una vez determinadas las posiciones de los patrones, éstas son utilizadas para aplicar una transformación a la imagen primaria para corregistrarla con la secundaria aplicando un método de interpolación [64]. En la Fig. 5.7 se muestra el resultado final luego de la aplicación de una interpolación de tipo bilineal [64].

5.3. Análisis sub-píxel de imágenes satelitales (*Spectral Unmixing*)

Recientemente, con la incorporación de sensores hiper-espectrales en satélites, que proporcionan información espectral de alta resolución, se han posibilitado nuevas aplicaciones en teledetección. Las imágenes multi-espectrales, es decir, aquellas que proveen menos de 10 bandas de medición, han permitido la implementación de algoritmos de clasificación detectando el material preponderante por píxel (con porcentaje de ocupación mayor) [121]. Sin embargo, el análisis sub-píxel se encuentra limitado por el pequeño número de bandas disponibles [84]. Por otro lado, con imágenes hiper-espectrales, donde se disponen de 100 o 200 bandas de frecuencia diferentes, es posible estimar los porcentajes de ocupación de los materiales presentes por píxel. Esta tarea se conoce con el nombre de *Spectral Unmixing* y constituye un nuevo y fascinante área de investigación [84, 103].

Bajo ciertas condiciones ([84, 103] y las referencias allí incluidas), la radiancia recibida en el sensor puede ser aproximada por una combinación lineal de las radiancias de los materiales que componen al píxel, multiplicados por sus correspondientes porcentajes de ocupación. Además, puede existir un componente

de ruido introducido por el instrumento. Por lo tanto, el problema de *Spectral Unmixing* puede ser modelado como un caso particular de *BSS*, es decir:

$$\mathbf{x} = A\mathbf{s} + \mathbf{n} \quad (5.6)$$

donde $\mathbf{x} = [x_0 \ x_1 \dots x_{m-1}]^T$ es el vector de las mediciones en los m canales (mezclas), $A \in \mathbb{R}^{m \times p}$ es una matriz de mezcla que contiene los espectros (*spectral signatures*) de los p materiales, $\mathbf{s} = [s_0 \ s_1 \dots s_{p-1}]^T$ es el vector de los porcentajes de ocupación de los materiales (fuentes) y $\mathbf{n} = [n_0 \ n_1 \dots n_{m-1}]^T$ es el vector de ruido aditivo que, por simplicidad, se asume gaussiano y de media nula.

Se destaca que, *Spectral Unmixing* es un caso especial de *BSS* en el que las señales fuente tienen sus valores acotados $0 \leq s_i \leq 1$, y además están condicionadas a verificar la ecuación (3.68) que aquí se reproduce:

$$\sum_{i=0}^{p-1} s_i = 1 \quad (5.7)$$

Cuando los espectros de los materiales (matriz A) son conocidos, la estimación de los porcentajes de ocupación es un problema inverso que puede ser resuelto, por ejemplo, usando mínimos cuadrados [84] u otros métodos [1]. Sin embargo, en la realidad no se conocen a priori estos espectros, es así que, en los últimos años, se han realizado avances en *Blind Spectral Unmixing* [16, 84, 103].

La técnica más popular en la bibliografía es utilizar los algoritmos *ICA* con los cuales se han obtenido algunos resultados promisorios [19, 43, 86, 94, 103]. Sin embargo, en *ICA* se asume que los porcentajes de ocupación son variables independientes lo cual no es realista por diversas razones [84, 103]. Es fácil ver que la condición (5.7) impone dependencia entre las variables, de hecho, cada fuente puede ser escrita como combinación lineal de las otras haciéndolas explícitamente dependientes. En esta tesis, se presentan resultados de la aplicación del algoritmo MAXNG al problema *Blind Spectral Unmixing* [33, 35]. En los experimentos del Cap. 3, se demostró que MAXNG muestra una performance superior a los algoritmos *ICA* en la separación de fuentes dependientes que obedecen a la condición (5.7) por lo que su aplicación al caso de imágenes hyperspectrales es más adecuado. Además, la condición (5.7) puede ser utilizada para determinar los factores de escala (Sección 3.7) proveyendo así una solución más completa y precisa que las soluciones *ICA* encontradas en la bibliografía.

5.3.1. Imágenes utilizadas en los experimentos

Imagen hiperespectral real: Para los experimentos con datos reales, se han utilizado sub-imágenes de una imagen de la ciudad de Roma provista por el *Airborne Laboratory for Environmental Research at IIA-CNR* en Roma, Italia (Fig. 5.8, arriba). Esta imagen, de 540×337 píxeles, ha sido capturada por el espectrómetro MIVIS de tipo “*sweepbroom*” montado en una aeronave. Este instrumento consta de 102 canales los cuales provienen de 4 sensores independientes en diferentes bandas de frecuencia: canales 1 – 20 ($0,43 - 0,83 \mu\text{m}$); canales 21 – 28 ($1,15 - 1,55 \mu\text{m}$); canales 29 – 92 ($2,00 - 2,50 \mu\text{m}$) y canales 93 – 102 ($8,20 - 12,70 \mu\text{m}$).

Además de la imagen hiper-espectral, se cuenta con una clasificación de los píxeles en 9 clases diferentes (Fig. 5.8, abajo). Esta clasificación fue obtenida a través del método clásico SAM (*spectral angle mapper*) [89] supervisado con inspecciones de campo. Las clases identificadas son: ladrillo (rojo), grava/cemento (amarillo), otras superficies (gris), infraestructuras (marrón), árboles (verde oscuro), arbustos/matas (verde medio), prados (verde claro), agua (azul) y no clasificado (negro). Desde luego, esta clasificación no permite evaluar cuantitativamente la performance de los algoritmos de *spectral unmixing* ya que no proporciona ninguna información sub-píxel, sin embargo permite hacer una evaluación cualitativa de los resultados.

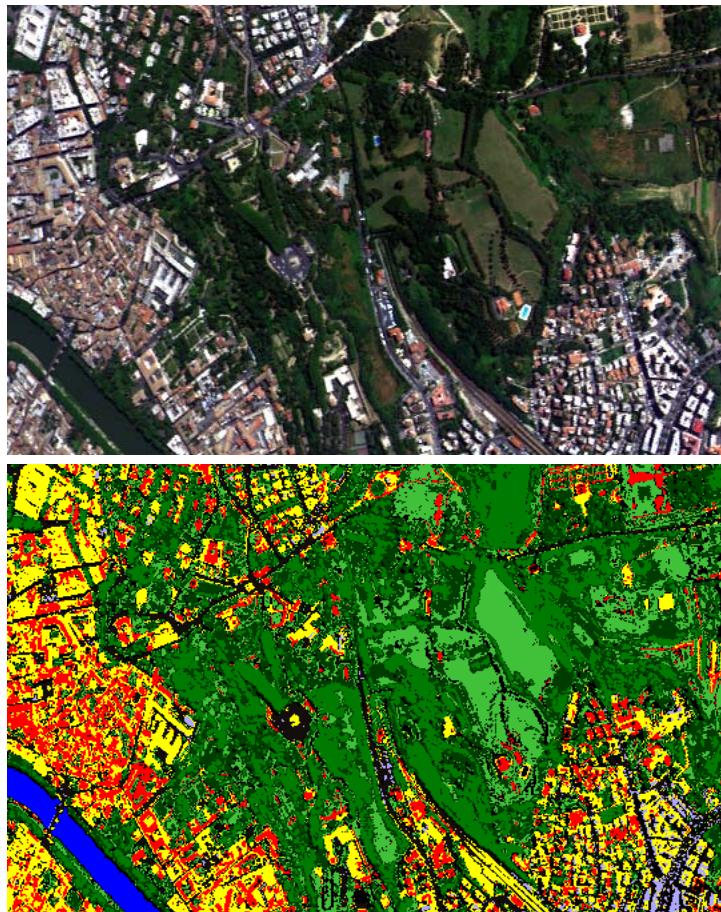


FIGURA 5.8. Arriba: canales RGB de la imagen hiperespectral original (540×337 píxeles). Abajo: clasificación a través del método SAM (*Spectral Angle Mapper*) (9 clases detectadas)

Imagen hiper-espectral simulada: A los efectos de posibilitar una validación más precisa de la separación de fuentes, se ha generado una imagen simulada usando la ecuación 5.6, donde los perfiles espectrales (filas de la matriz A) de cada una de las clases de la Fig. 5.8 - abajo, se han estimado a través de los espectros promedios de los píxeles que pertenecen a ellas.

Para la generación de los porcentajes de ocupación (vector de fuentes \mathbf{s}), se generó una imagen simulada de tamaño inferior a la imagen original real, mezclando los espectros de los píxeles incluidos en sub-áreas de 8×8 píxeles. De esta manera, se puede calcular el porcentaje de ocupación por píxel en la imagen simulada, como el porcentaje de píxeles de ese material incluidos en el sub-área correspondiente. En la Fig. 5.9, se muestran los espectros aproximados (espectros promedio) y sus correspondientes desviaciones estándar. La Fig. 5.10 muestra las imágenes simuladas de 67×42 correspondientes a los porcentajes de ocupación por píxel de cada clase.

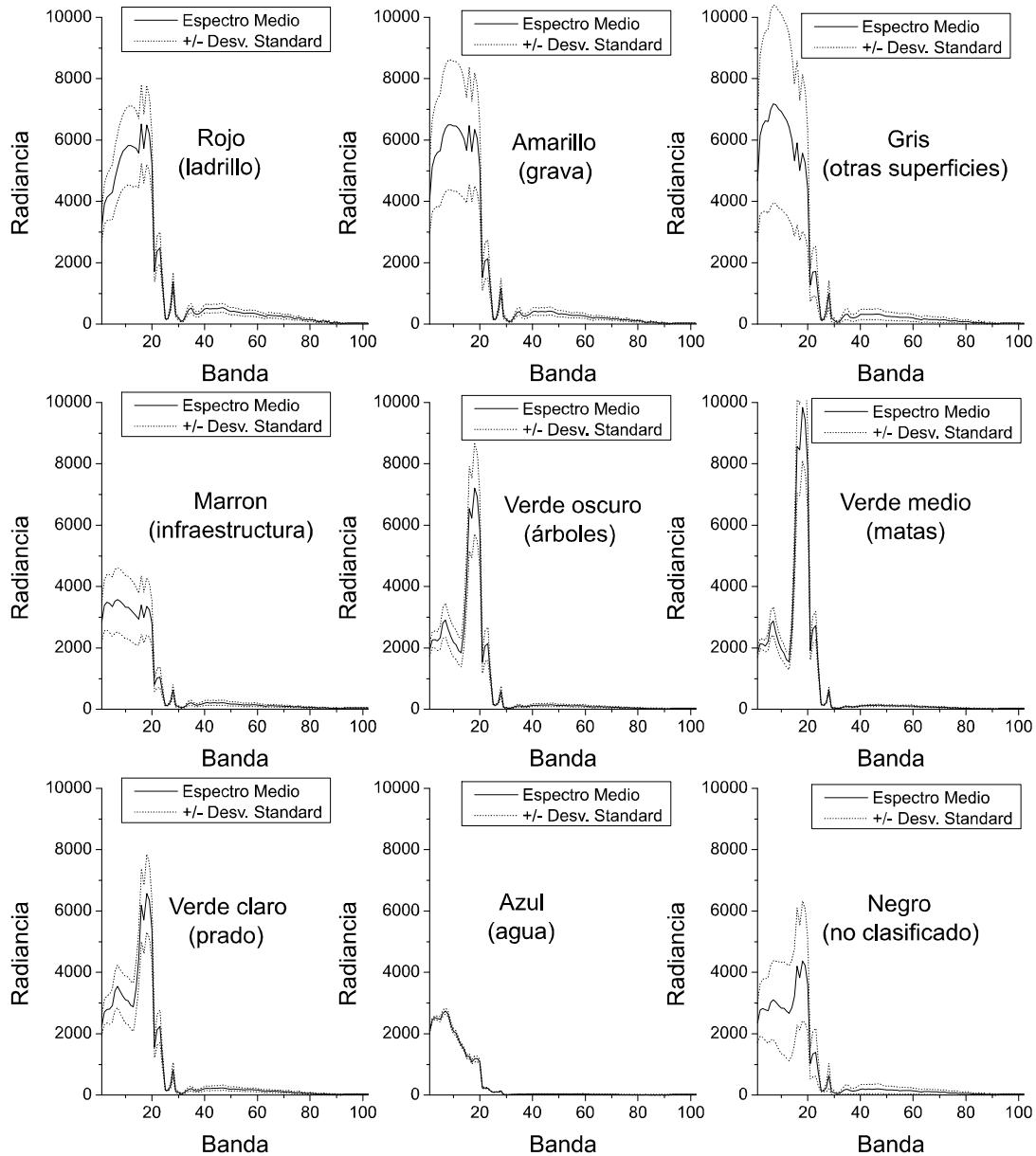


FIGURA 5.9. Espectros aproximados de las 9 clases provistas

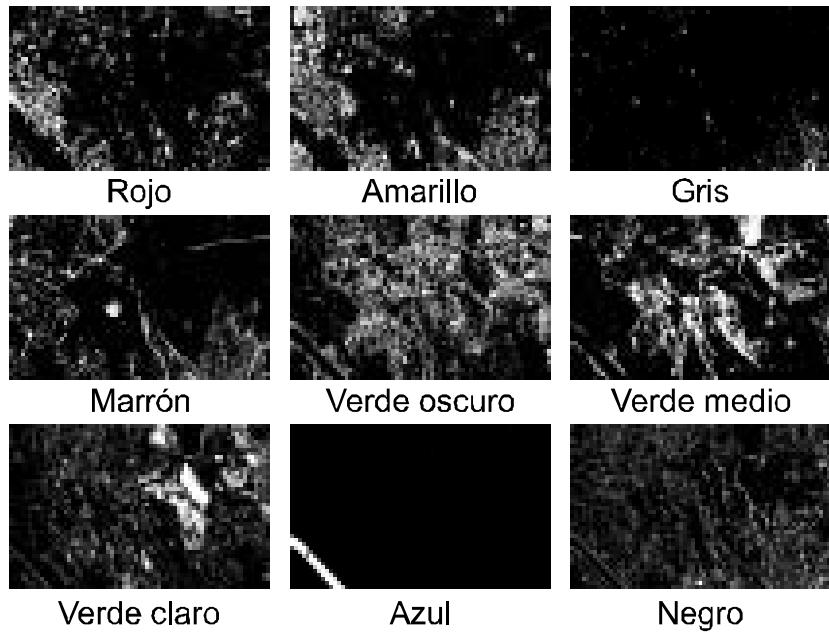


FIGURA 5.10. Imágenes correspondientes a porcentajes de ocupación por pixel de cada clase.

5.3.2. Experimento con imagen simulada

En este experimento se aplica el algoritmo MAXNG a la imagen simulada para el caso de ruido nulo ($\mathbf{n} = \mathbf{0}$) y se comparan las fuentes estimadas con las fuentes reales de la Fig. 5.10. El primer paso es, como siempre, la aplicación del FD que reduce la dimensión de los datos de 102 (número de mezclas) a 8 ya que existen solo 8 autovalores no nulos de la matriz de covarianzas. Puede verse fácilmente que, la condición (5.7) obliga a reducir el espacio de las fuentes en una dimensión, es decir, $p - 1$.

A los efectos de comparar las estimaciones de las fuentes con sus originales, se evalúa el *SIR* (*Signal to Interference Ratio*) sobre las fuentes normalizadas (media cero y varianza nula) teniendo en cuenta que, en general, se puede decir que un valor de *SIR* por debajo de 12dB representa una detección errónea [22]. En la Tabla 5.1, se muestran los valores de *SIR* obtenidos para cada una de las fuentes y se observa que todas fueron perfectamente recuperadas por el algoritmo MAXNG. En la Fig. 5.11 se muestra una comparación visual de las señales fuente y sus estimadas para las clases Verde claro y Azul que corresponden con las señales peor estimada (mínimo *SIR*= 14.6dB) y mejor estimada (máximo *SIR*= 63.8dB), respectivamente.

5.3.3. Experimentos con imágenes reales

En esta subsección se presentan ejemplos de aplicación del algoritmo MAXNG a pequeñas sub-áreas de la imagen hiper-espectral real de la Fig. 5.8 - arriba.

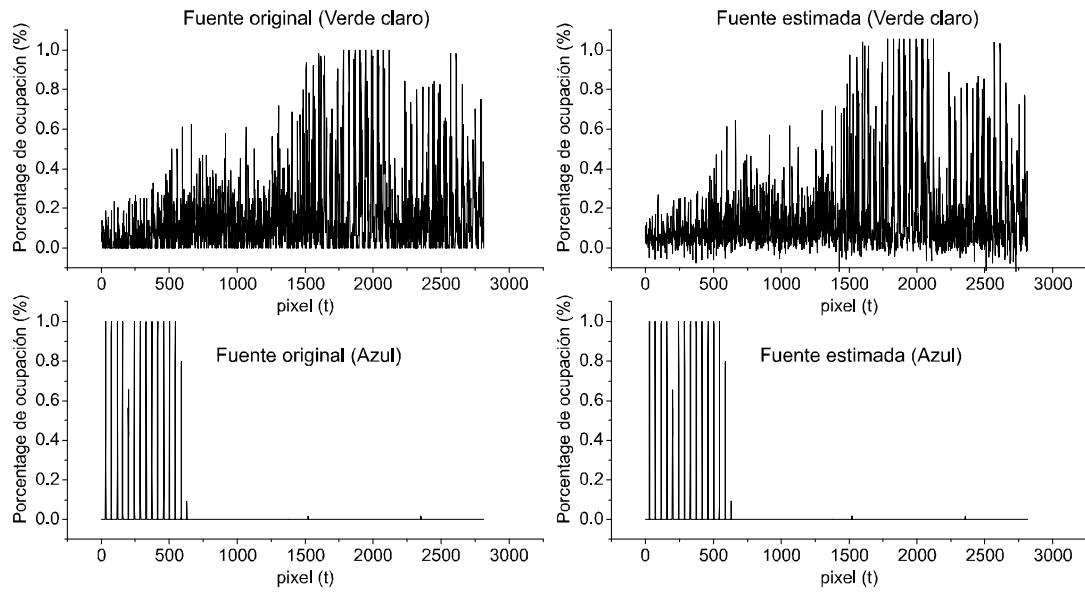


FIGURA 5.11. Comparación de las fuentes originales y estimadas correspondientes a las clases Verde claro (arriba) y Azul (abajo).

TABLA 5.1: Valores SIR de las fuentes estimadas en imagen simulada

	Negro	Marrón	Gris	Amarillo	Rojo	Verde oscuro	Verde medio	Verde claro	Azul	Promedio
SIR (dB)	22.5	17.9	35.7	15.0	26.9	18.4	20.3	14.6	63.8	26.1

En teoría cuantos más píxeles sean considerados (imágenes más grandes) mejor es la eficiencia de MAXNG ya que las estimaciones de la medida NG se vuelve más precisa. Sin embargo, cuando se trabaja con imágenes reales, existe un nuevo fenómeno que debe ser tenido en cuenta: la variabilidad espectral ya que el espectro de un material, no es idéntico para todos los píxeles. Algunos de los factores que provocan este efecto son: la geometría de iluminación-dispersión, la topografía local (efectos 3D) y otros [84, 103]. Por lo tanto, existe una situación de compromiso para la selección del tamaño de la imagen a procesar, por un lado se requiere una imagen suficientemente grande para obtener buenas estadísticas de los datos pero, por otro lado, cuanto más grande sea la imagen, más crítico será el efecto de variabilidad espectral.

En nuestros experimentos se analizaron distintos tamaños de imágenes. Un indicador de la precisión en la estimación de las fuentes, es medir el error cuadrático medio *MSE* (*Mean Square Error*) entre la suma de las fuentes (hipotéticamente igual a uno) y la constante 1.0:

$$MSE = E \left[\left(\left(\sum_{i=0}^{p-1} \hat{s}_i \right) - 1 \right)^2 \right]$$

Si la separación fue exitosa, el error residual será pequeño; por otro lado un valor alto de *MSE* indica un error en la separación.

La Fig. 5.12 (“Casino del bel respiro” en Villa Doria-Pamphilj, Roma), muestra los resultados de aplicar MAXNG a una sub-imagen de 50×50 píxeles (Figs. 5.12.a)-.b)-.c)) junto con la imagen original en formato RGB (Fig. 5.12.a)) y su clasificación (Fig. 5.12. b)). Se observa que las fuentes estimadas 1 y 2 coinciden con las clases Amarillo y Rojo, respectivamente, y la tercera fuente estimada corresponde a las clases Verde (oscuro, medio y claro) mezcladas con las clases minoritarias Gris, Negro, Marrón y Azul. Este ejemplo muestra que MAXNG no es capaz de discriminar las clases minoritarias, ya que el cálculo estadístico se ve afectado por la escasez de píxeles de esas clases. Además se observa que las tres clases de verde, aparecen identificadas como una única clase, lo cual es razonable ya que todas estas clases tienen un perfil espectral muy similar y sufren de altos niveles de variabilidad espectral como muestra la Fig. 5.9.

La Fig.5.13 (“Palazzo Corsini alla Lungara” y Jardín Botánico, Roma), muestra los resultados para una sub-imagen de tamaño mayor (75×75 píxeles). En este caso MAXNG, identificó cuatro clases: las fuentes 1, 2 y 3 se corresponden con las clases Amarillo, Verde (oscuro, medio y claro) y Rojo, respectivamente; la fuente estimada 4 contiene una mezcla de las clases Negro, Marrón y Gris.

Finalmente, en la Tabla 5.2 se comparan los *MSE* obtenidos para diferentes tamaños de sub-imagen (50×50 , 75×75 y 150×150 píxeles). Se observa claramente que, para el primer y segundo caso los errores remanentes son pequeños lo cual sugiere buenos resultados. Por otro lado, para 150×150 se obtiene un error significativo, causado principalmente por el alto nivel de variabilidad espectral, que es consecuencia del tamaño de la imagen.

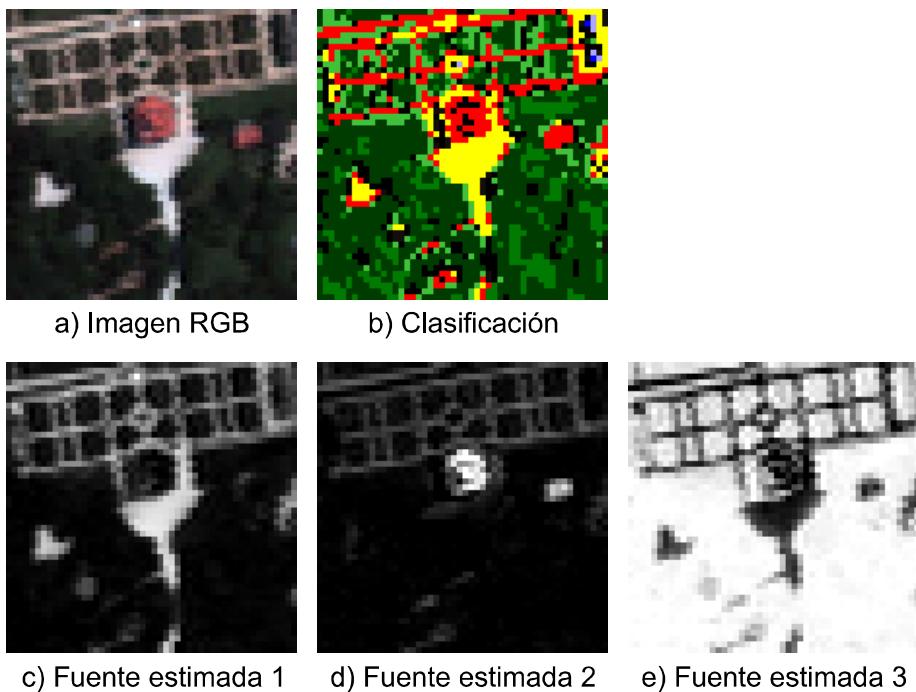


FIGURA 5.12. Imágen hiperespectral de 50×50 píxeles y resultados obtenidos con MAXNG

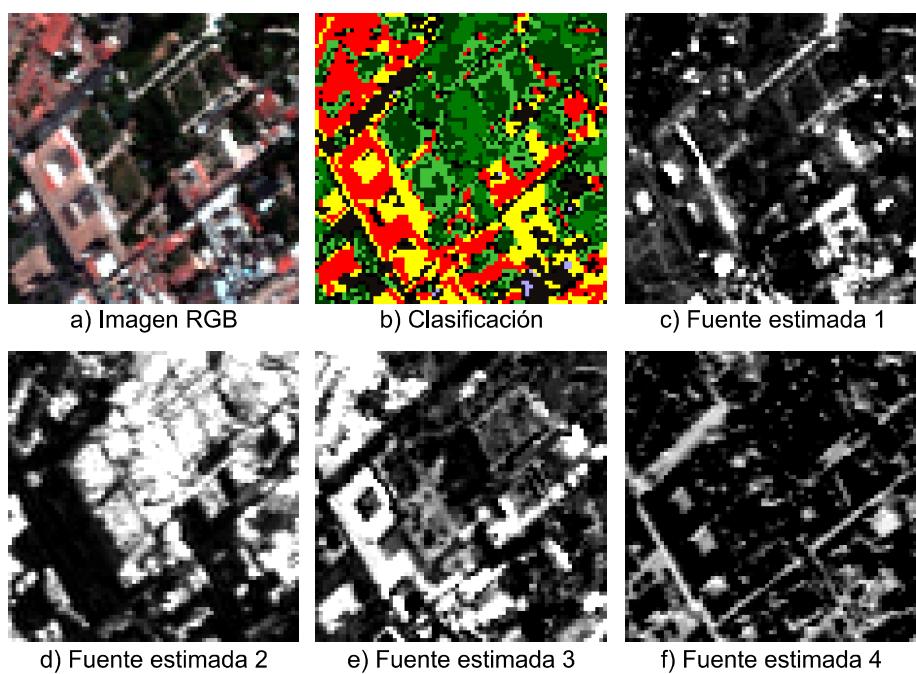


FIGURA 5.13. Imágen hiperespectral de 75×75 píxeles y resultados obtenidos con MAXNG

TABLE 5.2: Error Cuadrático Medio (MSE)

Tamaño Imagen	50x50	75x75	150x150
MSE	8.2E-05	4.2E-03	0.47

5.4. *BSS* y reducción de ruido en imágenes astrofísicas

En Astronomía se presentan muy a menudo problemas de separación de señales y de reducción de ruido para aislar señales de interés. En esta tesis se presenta una técnica para la separación de imágenes astrofísicas correlacionadas (dependientes) [32]. Además, el método propuesto, contiene una etapa final de reducción de ruido para las imágenes extraídas del *CMB* (*Cosmic Microwave Background*) utilizando el modelo *LC-GRF* [34].

Los resultados experimentales se hicieron sobre la base de imágenes astrofísicas sintetizadas provistas por el proyecto *Planck Surveyor Satellite mission* de la Agencia Espacial Europea (ESA) a través del *Signal & Images Laboratory of ISTI - CNR (Consiglio Nazionale delle Ricerche)* de Pisa, Italia. Estas imágenes simulan las mediciones de radiación electromagnética recibidas por un satélite con sensores en distintas bandas de frecuencia, las cuales son combinaciones lineales de las siguientes fuentes de radiación: *Cosmic Microwave Background (CMB)*, *Synchrotron (SYN)* y *Galactic Dust (DUST)* [88].

En algunos trabajos se ha abordado el problema de la separación de fuentes en imágenes astrofísicas a través de los algoritmos *ICA* [7, 88], sin embargo, la dependencia entre las fuentes afecta negativamente los resultados. Es así que algunos autores mejoraron las técnicas admitiendo dependencia o correlación entre las fuentes [12]. En esta tesis, se presenta una solución basada en el algoritmo MAXNG adaptado para este caso particular de señales [32].

5.4.1. Modelo matemático

El modelo matemático que posibilita la aplicación de MAXNG es

$$\mathbf{x} = A\mathbf{s} + \mathbf{n} \quad (5.8)$$

donde $\mathbf{x} = [x_0 \ x_1 \dots x_{m-1}]^T$ es un vector que contiene las mediciones correspondientes a m canales (mezclas), $\mathbf{s} = [s_0 \ s_1 \dots s_{p-1}]^T$ es un vector que contiene las señales asociadas con cada una de las p fuentes de radiación, $A \in \mathbb{R}^{m \times p}$ es una matriz de mezcla que contiene los coeficientes con los que las fuentes son combinadas en cada canal y $\mathbf{n} = [n_0 \ n_1 \dots n_{m-1}]^T$ es el vector de ruido aditivo que, por simplicidad, se asume gaussiano y de media nula.

El caso particular que se analiza aquí posee una serie de características que son explotadas en la adaptación del algoritmo MAXNG y que permite el diseño de un nuevo algoritmo llamado MINIMAX ENTROPY [32]. A continuación se discuten las características del modelo:

- El satélite cuenta con $m = 4$ canales de medición (sensores) que proveen distintas mezclas lineales de diversas fuentes de radiación electromagnética que

se corresponden con las siguientes frecuencias centrales de canal: 100GHz, 70GHz, 44GHz y 30GHz [12, 88];

- Se considera el caso donde existen sólo $p = 3$ fuentes de radiación presentes en las mezclas: el *Cosmic Microwave Background (CMB)*, el *Galactic Synchrotron (SYN)* y el *Galactic Dust (DUST)*;
- Las imágenes *CMB* son gaussianas mientras que las imágenes *SYN* y *DUST* no lo son [88]. Esta suposición, que fuera utilizada previamente por Bedini et al en [12], no es estrictamente necesaria para el algoritmo MINIMAX ENTROPY, sin embargo puede ser utilizada para mejorar la estimación del *CMB* como se muestra luego. En la Fig. 5.14 (arriba), se muestran ejemplos de imágenes *CMB*, *SYN* y *DUST* cuyas *pdf*'s estimadas a través de la técnica de ventanas de Parzen [107] son comparadas con la distribución de Gauss en la Fig. 5.15.
- Los pares de imágenes *CMB-SYN* y *CMB-DUST* son no-correlacionados, es decir $E[s_0 s_1] = E[s_0 s_2] = 0$ con s_0 , s_1 y s_2 correspondiendo a las fuentes *CMB*, *SYN* y *DUST* respectivamente.

5.4.2. El método Minimax Entropy

Este método surge como una adaptación natural del algoritmo MAXNG [30]. Como en cualquier problema de *BSS*, el objetivo es encontrar la matriz de separación D tal que las fuentes puedan ser estimadas a partir de la ecuación lineal

$$\hat{\mathbf{s}} = D\mathbf{x} \quad (5.9)$$

Las filas de la matriz D , que proveen las estimaciones de las fuentes no-gaussianas *SYN* y *DUST* son encontradas a partir de la minimización de la gaussianidad (maximización de la NG) de la variable

$$y = \mathbf{d}^T \mathbf{x}$$

donde el vector \mathbf{d} es tal que la varianza de y sea unitaria. Al mismo tiempo, puede verse que y será siempre no-gaussiana, a menos que se corresponda con la señal fuente *CMB* (gaussiana) con ruido gaussiano aditivo. Por lo tanto, la fila de la matriz D que provee la estimación del *CMB* se puede obtener como el vector \mathbf{d} que hace que la variable y resulte gaussiana.

En consecuencia las fuentes de radicación astrofísicas pueden encontrarse a través de la maximización y minimización de una medida de gaussianidad. La medida NG del algoritmo MAXNG puede generalizarse, usando además la medida basada en la distancia en $L^2(\mathbb{R})$, utilizando la entropía relativa que resulta en maximizar/minimizar la entropía de Shannon (Capítulo 2, Subsección 2.1.4). En esta tesis se analizan dos medidas de gaussianidad, que llamaremos medidas entrópicas: la entropía de Shannon (*Shannon Entropy - SE*) y la medida de gaussianidad

(Gaussianity Measure - GM).

$$H_{SE}(f_y) = - \int f_y(y) \log(f_y(y)) dy \quad (5.10)$$

$$H_{GM}(f_y) = - \int [f_y(y) - \Phi(y)]^2 dy \quad (5.11)$$

donde $\Phi(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)$ es la *pdf* de Gauss y $f_y(y)$ es la *pdf* de y . Se señala que, la medida de gaussianidad $H_{GM}(f_y)$ es exactamente igual a menos la medida $\Gamma_{NG}(f_y)$ definida en MAXNG.

En la implementación del algoritmo de separación, se deberán estimar las medidas entrópicas (5.10) y (5.11) a partir de un conjunto de n muestras de la señal y : $y(0), y(1), \dots, y(n-1)$, para lo cual usamos la técnica de estimación no paramétrica basada en las ventanas de Parzen (Subsección 3.2.1). Para la medida (5.10) puede derivarse una técnica similar como fuera propuesta en [58]. El algoritmo MINIMAX ENTROPY consiste en la aplicación de los siguientes pasos principales:

Decorrelación de las mezclas: Con el objeto de simplificar la parametrización a través del uso de coordenadas esféricas, se aplica un FD basado en la *KLT* (Subsección 3.4.1) es decir, se define al conjunto de mezclas no-correlacionadas $\hat{\mathbf{x}}$ a través de

$$\hat{\mathbf{x}} = T\mathbf{x} \quad (5.12)$$

donde $T \in \mathbb{R}^{q \times m}$ es el FD basado en la *KLT*, $m = 4$ es el número de sensores y q es la dimensión del espacio reducido. Se consideran niveles bajos de ruido aditivo y por lo tanto existirán siempre solo $q = p = 3$ autovalores relevantes de la matriz de covarianzas.

Las mezclas no-correlacionadas \hat{x}_0 , \hat{x}_1 y \hat{x}_2 son no-correlacionadas, es decir

$$E[\hat{x}_0 \hat{x}_1] = E[\hat{x}_0 \hat{x}_2] = E[\hat{x}_1 \hat{x}_2] = 0 \quad (5.13)$$

y se usan coordenadas esféricas con radio unitario para restringir la varianza de las combinaciones lineales a ser unitaria:

$$y(\theta_0, \theta_1) = \hat{x}_0 \cos \theta_0 + \hat{x}_1 \sin \theta_0 \cos \theta_1 + \hat{x}_2 \sin \theta_0 \sin \theta_1 \quad (5.14)$$

El espacio de búsqueda del máximo y los mínimos resulta bidimensional (θ_0, θ_1) . En términos de las mezclas no-correlacionadas, el objetivo es buscar una matriz de separación \hat{D} tal que

$$\hat{\mathbf{s}} = \hat{D}\hat{\mathbf{x}} \quad (5.15)$$

Etapa de mínima entropía: Dado que las fuentes no-gaussianas están asociados con mínimos locales de las medidas entrópicas de la variable $y(\theta_0, \theta_1)$, se procede a buscar los mínimos locales de las medidas *SE* o *GM*. Las fuentes asociadas a estos mínimos locales corresponderán con las fuentes no-gaussianas (*SYN* o *DUST*).

Etapa de máxima entropía: La fuente estimada asociada con la máxima entropía de $y(\theta_0, \theta_1)$ corresponderá con una fuente gaussiana que, en este caso, es el *CMB*.

Etapa de reducción de ruido: Una vez estimada la matriz de separación D la ecuación (5.9) nos provee de las estimaciones de las fuentes *CMB*, *SYN* y *DUST*. Si D es una matriz de separación perfecta ($DA = I$) entonces las estimaciones de las fuentes estarán contaminadas con ruido aditivo gaussiano. La reducción del ruido aditivo gaussiano puede llevarse a cabo a través de un filtro óptimo o filtro de Wiener para lo cual se requiere de un modelo probabilístico de la señal útil que quiere recuperarse. Para el caso de las imágenes *CMB*, el modelo *LC-GRF* permite encontrar una expresión analítica para el filtro de Wiener (Sección 4.6).

5.4.3. Otra manera de estimar la imagen *CMB*

La condición de no-correlación de los pares de señales *CMB-SYN* y *CMB-DUST* puede ser utilizada para mejorar la estimación de la imagen *CMB* ya que, en los experimentos, se observó que la posición del máximo local asociado al *CMB* está ubicado en un valle muy suave siendo mayores los errores de estimación. Supongamos que se han obtenido la segunda (*SYN*) y tercera (*DUST*) columna de la matriz de separación $\hat{D} \in \mathbb{R}^{3 \times 3}$. Fácilmente se observa que, esta matriz \hat{D} se relaciona con la matriz de covarianzas de las fuentes estimadas:

$$C_{\hat{\mathbf{s}}\hat{\mathbf{s}}} = E[\hat{\mathbf{s}}\hat{\mathbf{s}}^T] = \hat{D}\hat{D}^T \quad (5.16)$$

La condición de no-correlación de los pares *CMB-SYN* y *CMB-DUST*:

$$E[\hat{s}_0\hat{s}_1] = E[\hat{s}_0\hat{s}_2] = 0 \quad (5.17)$$

implica que la primera fila de \hat{D} pertenece al subespacio ortogonal del subespacio definido por las filas segunda y tercera, lo cual determina únicamente a la primera fila (*CMB*) de \hat{D} .

5.4.4. Experimentos sobre imágenes sintetizadas

Experimento 1: Ruido nulo

En este caso se han generado mezclas a través de la ecuación (5.8) con ruido nulo ($\mathbf{n} = \mathbf{0}$) y utilizando una matriz de mezcla acorde con [7]. Las imágenes utilizadas para este experimento son imágenes sintetizadas de 256×256 píxeles (65536 muestras) para el *CMB*, *SYN* and *DUST* y fueron provistas por el equipo del proyecto Planck en Italia a través del *Signal & Images Laboratory of ISTI - CNR (Consiglio Nazionale delle Ricerche)* de Pisa, Italia (en [88] se provee una descripción sobre los métodos usados en la generación de las imágenes). Se utilizaron imágenes normalizadas, es decir, con media nula y varianza unitaria. En la Figura 5.14 (arriba y centro), se muestra un ejemplo de imágenes fuente sintetizadas y las mezclas generadas de acuerdo al modelo.

Para el ejemplo de la Fig. 5.14, la matriz \hat{D} y sus estimaciones a través del algoritmo, usando las medidas MG y *SE* son:

$$\hat{D} = \begin{bmatrix} -0,36 & 0,62 & 0,70 \\ 1,00 & -0,05 & 0,02 \\ 0,04 & 0,85 & -0,53 \end{bmatrix} \quad (5.18)$$

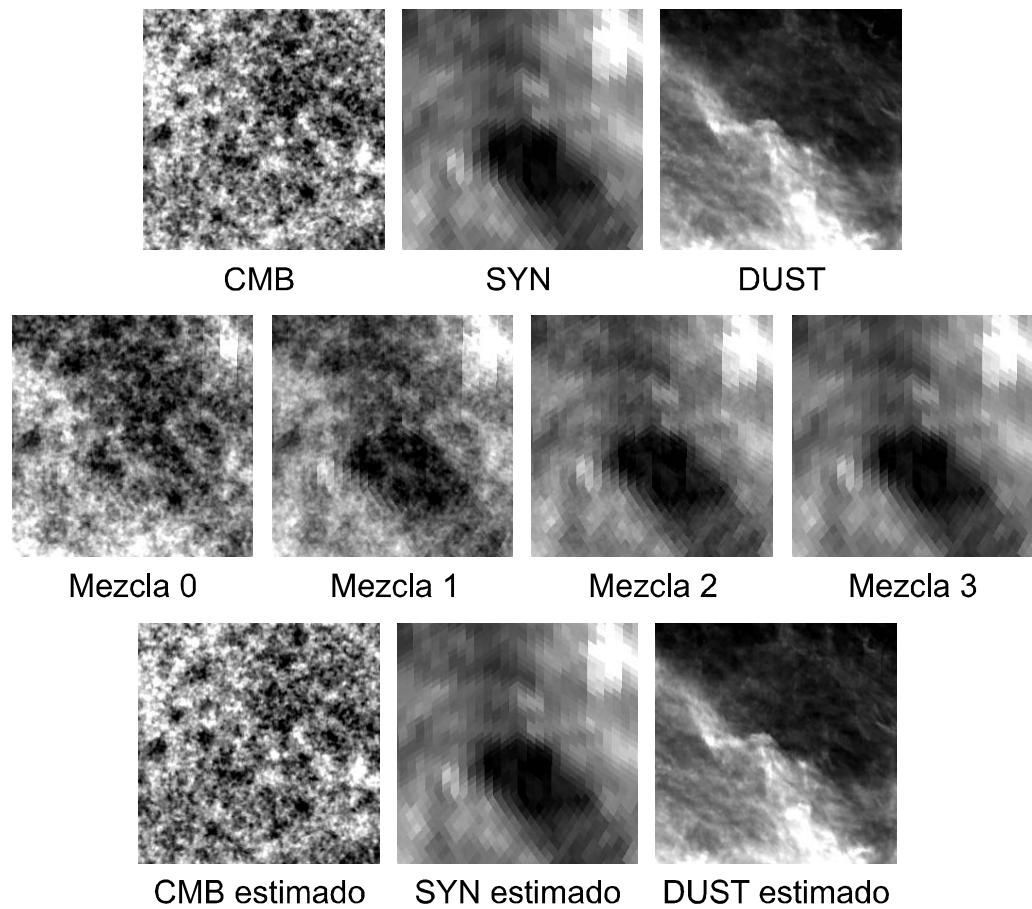


FIGURA 5.14. Ejemplo de imágenes fuentes (arriba), mezclas (centro) y estimaciones de fuentes usando la medida SE (abajo)

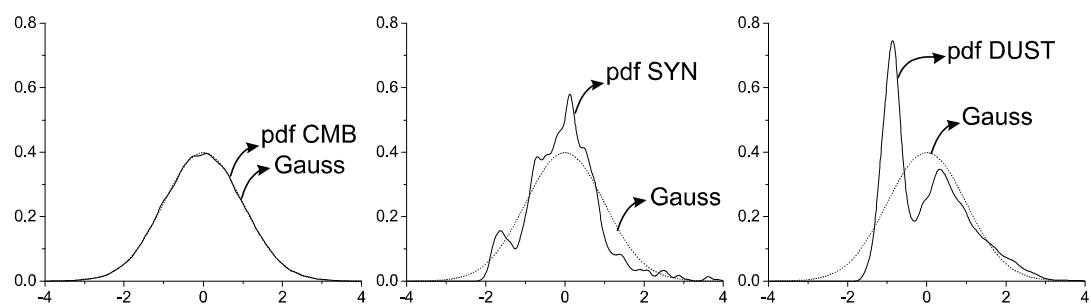


FIGURA 5.15. Funciones de densidad de probabilidad (pdfs) estimadas a partir de las imágenes CMB (izq.), SYN (centro) y DUST (der.) usando la técnica de estimación de las ventanas de Parzen

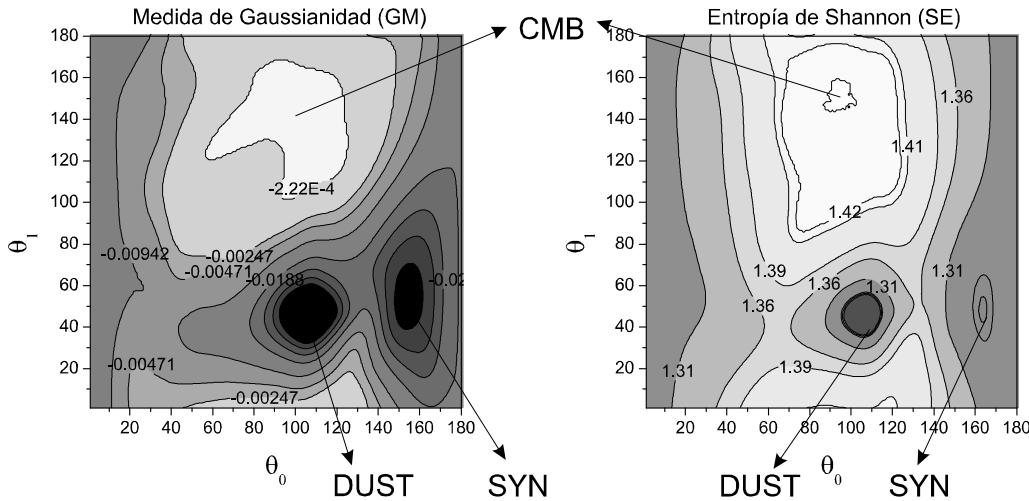


FIGURA 5.16. Curvas de nivel para las superficies de las medidas de Gaussianidad (GM) y entropía de Shannon (SE) versus los ángulos θ_1 y θ_2 .

$$\mathring{D}_{MG} = \begin{bmatrix} -0,29 & 0,66 & 0,69 \\ 0,90 & -0,27 & -0,35 \\ -0,07 & 0,71 & -0,70 \end{bmatrix}; \quad \mathring{D}_{SE} = \begin{bmatrix} -0,28 & 0,67 & 0,69 \\ 1,0 & -0,02 & 0,02 \\ 0,03 & 0,72 & -0,69 \end{bmatrix} \quad (5.19)$$

En la Figura 5.16 se muestran las superficies de las medidas *GM* y *SE* en función de los ángulos θ_1 y θ_2 para el ejemplo de la Fig. 5.14. Se observa que se pueden identificar claramente las posiciones de los mínimos (*SYN* y *DUST*) y el máximo (*CMB*). Además, la forma que toman estas superficies resultan similares entre sí.

El experimento de las Figs. 5.14 y 5.16 fue repetido para 15 conjuntos de imágenes diferentes correspondientes sub-áreas del mapa total simulado y se ha calculado, en cada caso, el correspondiente *SIR* (*Signal to Interference Ratio*) para las estimaciones utilizando las medidas *GM* y *SE*. Además, se han estimado las fuentes utilizando uno de los más famosos algoritmos de *ICA*: FASTICA [74].

En la tabla 5.3 se comparan los resultados obtenidos en los tres casos: medidas *GM*, *SE* y algoritmo FASTICA. La sub-área número 2 corresponde al ejemplo mostrado previamente en el ejemplo de la Fig. 5.14. Se observa que las soluciones provistas por MINIMAX ENTROPY son superiores a las obtenidas por FASTICA, además la medida *SE* muestra ser levemente superior a la medida *GM*.

Experimento 2: Robustez al ruido

Aquí se analiza el comportamiento del algoritmo MINIMAX ENTROPY con la presencia de ruido aditivo gaussiano blanco (*AWGN*) ya que es un componente muy frecuente en imágenes astrofísicas. En este experimento, se analizó cualitativamente la variación en las ubicaciones del máximo y los mínimos para diferentes niveles de *SNR* (*Signal to Noise Ratio*). En la Fig. 5.17 se muestran los resultados para el caso de la medida *SE* para tres niveles diferentes de *SNR*. Se verifica que las posiciones del máximo y los mínimos se mantienen aproximadamente inalteradas para un amplio rango de *SNR*'s ($SNR \geq 20\text{dB}$).

TABLA 5.3: Resultados del algoritmo Entropy Minimax usando las medidas GM y SE. Comparación con algoritmo FastICA. Se resaltan los niveles de SIR por debajo de 8dB. Los porcentajes de imágenes separadas exitosamente ($SIR > 8dB$) son: 91%, 93% y 71% para GM, SE y FastICA respectivamente.

Subareas	SIR (dB) usando GM			SIR (dB) usando SE			SIR (dB) usando FastICA		
	CMB	SYN	DUST	CMB	SYN	DUST	CMB	SYN	DUST
1	15.8	10.0	13.9	15.2	13.2	15.6	22.9	13.0	10.2
2	12.0	7.1	22.1	13.6	31.9	21.4	8.3	22.1	6.4
3	23.3	2.4	17.3	23.1	2.2	17.4	25.4	13.6	9.8
4	9.8	15.4	8.8	10.3	13.8	9.8	8.7	6.9	2.9
5	12.5	15.2	20.6	19.7	23.6	20.1	17.5	17.9	22.8
6	14.9	11.5	14.5	18.4	13.0	16.1	20.8	14.8	10.7
7	16.5	11.7	9.6	16.0	14.5	10.3	16.8	15.6	3.0
8	42.3	6.4	15.6	52.1	6.7	15.3	-1.9	5.6	7.6
9	7.8	14.7	8.6	11.8	24.1	9.3	5.0	2.6	1.5
10	14.2	13.3	13.8	20.8	22.2	19.4	19.8	16.9	27.0
11	15.8	10.1	14.0	15.3	13.2	14.7	22.7	13.5	10.3
12	20.7	26.6	17.8	21.0	22.5	24.4	28.2	19.1	9.6
13	8.7	11.4	17.4	10.6	6.2	12.6	4.1	1.0	4.8
14	11.9	18.6	21.3	13.4	23.6	16.9	13.2	8.5	15.3
15	12.4	15.2	21.0	19.6	23.8	19.9	17.5	17.9	22.8
Promedio	15.9	12.6	15.7	18.7	17.0	16.2	15.3	12.6	11.0

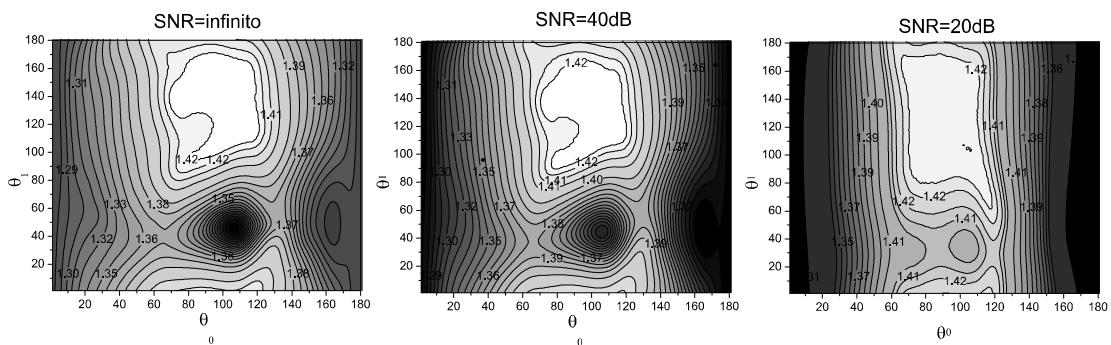


FIGURA 5.17. Curvas de nivel de las superficies de la medida SE para diferentes niveles de SNR: ∞ , 40dB y 20dB

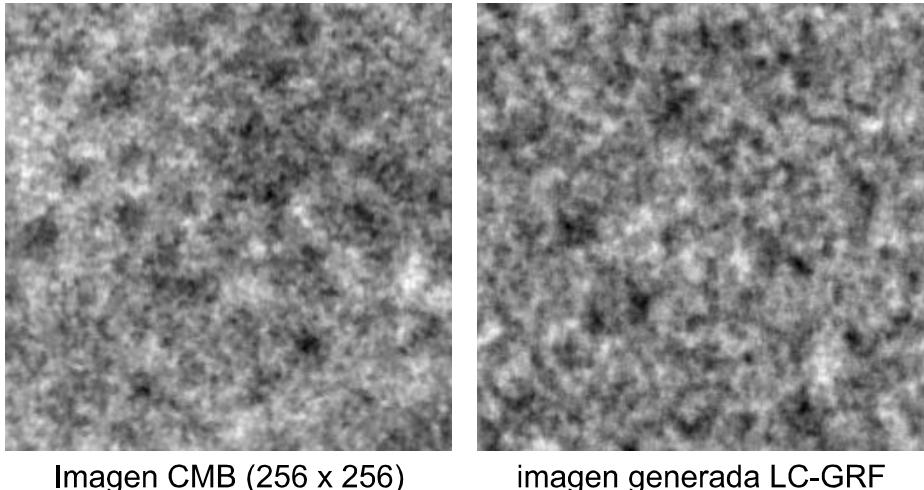


FIGURA 5.18. Imagen CMB original (izquierda) e imagen LC-GRF generada con los parámetros estimados.

Experimento 3: Reducción de ruido gaussiano aditivo en las imágenes *CMB* estimadas

En este experimento se evalúa la performance del filtro de Wiener para el modelo *LC-GRF* como herramienta para la reducción de ruido gaussiano en imágenes *CMB*. En primer lugar, y a los efectos de comparar cualitativamente la capacidad del modelo *LC-GRF* para reproducir imágenes *CMB*, se procedió a estimar los parámetros α , ρ_x , θ_1 y θ_2 del modelo a través del algoritmo de máxima verosimilitud propuesto en la Subsección 4.7.1, usando como muestra, una imagen *CMB* de 256×256 píxeles. Los valores obtenidos de los parámetros fueron: $\hat{\alpha} = 2.48$, $\hat{\rho}_x = 0.028$, $\hat{\theta}_1 = 0.226$ y $\hat{\theta}_2 = 0.257$. A los efectos de proveer una evaluación visual, en la Fig 5.18 (izquierda) se muestra la imagen *CMB* usada de muestra y se la compara con una imagen aleatoria generada a partir del modelo *LC-GRF* con los parámetros estimados. Se observa claramente que las características espaciales de la imagen son preservadas por el modelo.

En la Fig. 5.19 se presentan los resultados de restauración para dos casos con diferente nivel de ruido: en la Fig 5.19 (arriba) se muestran las imágenes para un $SNR=-10\text{dB}$ (nivel de ruido alto), la ganancia de restauración experimental resulta $G = 34,21\text{dB}$ y el valor teórico esperado es de $G = 34,23\text{dB}$. En la Fig 5.19 (abajo) se muestra un caso con $SNR=+10\text{dB}$ (nivel de ruido moderado), la ganancia de restauración experimental es $G = 11,37\text{dB}$ que es muy cercana al valor teórico esperado $G = 11,24\text{dB}$. En ambos casos se muestran las imágenes *CMB* originales (izquierda), las imágenes contaminadas con ruido (centro) y las restauraciones logradas con el filtro de Wiener (derecha).

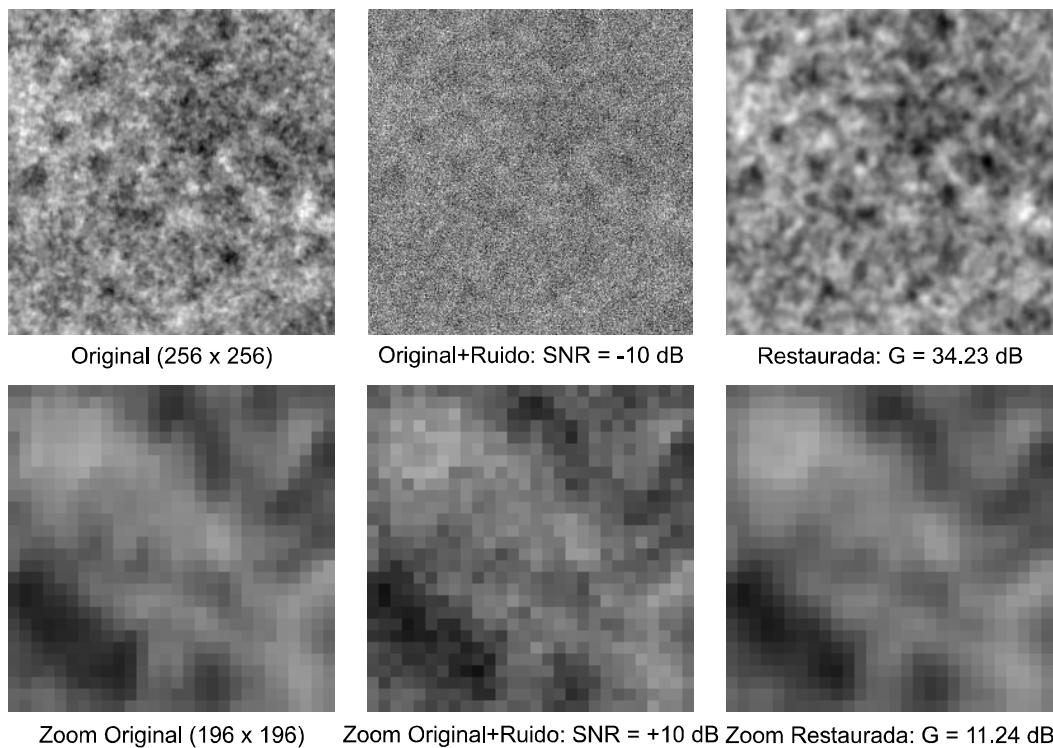


FIGURA 5.19. Imagen CMB original (izquierda), imagen ruidosa (centro) e imagen restaurada (derecha). Se aplicó el filtro de Wiener a los casos: $\text{SNR}=-10\text{dB}$ (arriba) y $\text{SNR}=10\text{dB}$ (abajo)

Capítulo 6

CONCLUSIONES

En esta tesis se desarrollaron modelos matemáticos para el tratamiento de señales en dos dimensiones (imágenes) y se elaboraron nuevos algoritmos para la resolución de problemas importantes en procesamiento de imágenes, como son: la separación ciega de fuentes (*Blind Source Separation - BSS*), la reducción de ruido y el reconocimiento de patrones.

Se analizaron las propiedades matemáticas del método estadístico de Análisis de Componentes Principales (*PCA*) o Transformada Karhunen-Loève (*KLT*), que es una herramienta esencial para el análisis de datos multidimensionales que comprime la “energía” (varianza) de los datos de manera óptima, en un conjunto reducido de componentes principales ortogonales (no-correlacionadas). Asimismo, se analizaron los métodos derivados de *PCA* como es el caso del Análisis de Componentes Independientes (*ICA*), donde se buscan las componentes que, además de ser no-correlacionadas, son lo más independientes posibles, y *Projection Pursuit (PP)*, donde lo que se busca es minimizar la gaussianidad o entropía de las componentes. Se destaca que, tanto en *ICA* como en *PP*, se recurre a medidas provenientes de la teoría de la información como la entropía de Shannon o la información mutua. Un característica importante de *ICA* y *PP* es que utilizan estadísticas de orden superior a dos, a diferencia de *PCA*, que solo utiliza la matriz de covarianzas de las observaciones (estadísticas de orden dos). En esta tesis se analizaron las conexiones profundas entre *PCA*, *ICA* y *PP* haciendo explícitas las relaciones entre gaussianidad, entropía, independencia, información mutua, etc.

Como una generalización de *ICA*, en esta tesis se abordó el problema de la se-paración ciega de fuentes estadísticamente dependientes o Análisis de Componentes Dependientes (*DCA*), que es un área nueva de investigación con pocos antecedentes bibliográficos. Se propuso un nuevo algoritmo *DCA*, llamado MAXNG [30] que permite resolver satisfactoriamente, por ejemplo, el problema de *Spectral Unmixing* en imágenes satelitales hiper-espectrales [33] y la separación ciega de fuentes en imágenes astrofísicas [32], obteniendo mejores resultados que los métodos tradicionales *ICA*. Se presentaron resultados teóricos fundamentales para MAXNG proveyendo, por ejemplo, una condición suficiente sobre las fuentes dependientes que garantiza su separabilidad a través de la maximización de la no-gaussianidad. También se introdujeron resultados que permiten la implementación de MAXNG de manera eficiente incluso en entornos con ruido. Además se propuso un método para la determinación de los factores de escala de las fuentes estimadas que resulta útil en la aplicación de *Spectral Unmixing*.

Por otro lado se analizó la teoría que da sustento a los métodos de clasificación de vectores utilizando distancias euclídeas mínimas en \mathbb{R}^n , donde las bases ortonormales tienen un rol central y los métodos de ortogonalización de vectores son de vital importancia. De todos los posibles métodos de ortogonalización de

vectores, se demostró que la *KLT* es el método óptimo en el sentido que permite concentrar las energías de los patrones. De hecho, la *KLT* es la base del método de las *eigenfaces* utilizado para el reconocimiento de rostros humanos, un método muy conocido en la literatura. Además, se propuso un algoritmo original para la clasificación de especímenes de *Mirounga Leonina* a partir de fotografías de sus rostros, lo cual representa un nuevo método de clasificación en Biología [26, 27]. Además se diseñó un nuevo método automático para el corregistro de imágenes satelitales utilizando métodos de clasificación de vectores en \mathbb{R}^n basados en la *KLT* [28].

Adicionalmente, motivado por el problema del ruido en las imágenes astrofísicas, se estudió la reducción de ruido gaussiano blanco aditivo (*AWGN*) presente en la salidas de MAXNG. Se propuso un modelo gaussiano, el *LC-GRF* (*Long Correlation - Gaussian Random Field*) [34], que permite modelar las correlaciones de largo alcance presentes en imágenes del *Cosmic Microwave Background (CMB)*. Un análisis detallado de las propiedades matemáticas de este modelo, permitió la elaboración de técnicas originales de estimación de parámetros y el diseño de un filtro óptimo de Wiener reductor de ruido.

A continuación se especifican los aportes originales de esta tesis haciendo referencia a las publicaciones producidas:

- Se desarrolló un nuevo algoritmo denominado MAXNG que da solución al problema de la separación ciega de fuentes dependientes (*DCA*) (Cap. 3). Se obtuvieron resultados teóricos que sustentan al método y que permiten la implementación de MAXNG de manera eficiente, en términos de su complejidad computacional, y bajo la influencia de ruido aditivo gaussiano. Además, se analizó experimentalmente el desempeño del algoritmo MAXNG en distintos escenarios, utilizando tanto datos simulados como reales, y comparándolo con técnicas *ICA* tradicionales [29, 30, 32, 33, 35].
- Se introdujo un modelo de campo aleatorio gaussiano (*GRF*) con correlaciones de largo alcance (*LC-GRF*) y se desarrollaron nuevas técnicas de estimación de parámetros basados en él, como así también, se diseñó un filtro reductor de ruido gaussiano aditivo (Cap. 4) [34].
- Se desarrolló un algoritmo original para la clasificación de especímenes de *Mirounga Leonina* a partir de fotografías de sus rostros explotando las propiedades de la representación de imágenes por *KLT* (Sección 5.1) [26, 27].
- Se desarrolló un nuevo método automático para el corregistro de imágenes satelitales, basado en la representación por *KLT*, que fue aplicado a imágenes de tipo radar (Sección 5.2) [28].
- Se utilizó el algoritmo MAXNG como nuevo método de análisis sub-píxel de imágenes hiper-espectrales (*Spectral Unmixing*), mediante el cual es posible estimar los porcentajes de ocupación por píxel de los diversos materiales que componen una imagen (agua, vegetación, ladrillo, etc.) (Sección 5.3). Se

realizaron experimentos sobre imágenes simuladas y reales demostrando la efectividad del método y comparándolo con técnicas tradicionales basadas en *ICA* [33, 35].

- Se desarrolló un nuevo algoritmo llamado MINIMAX ENTROPY basado en MAXNG, especialmente diseñado para la separación ciega de imágenes de origen astrofísico (Sección 5.4). Se realizaron experimentos sobre imágenes simuladas demostrando la efectividad del método y comparándolo con técnicas tradicionales basadas en *ICA* [32].
- Se utilizó el modelo *LC-GRF* para modelar las correlaciones de largo alcance existentes en imágenes del *Cosmic Microwave Background (CMB)* y para desarrollar un filtro reductor de ruido de Wiener (Sección 5.4) [34].

Por otro lado, los resultados aquí presentados, abren nuevas áreas de investigación y permiten plantear problemas que podrían ser explorados en el futuro, como por ejemplo:

- Los métodos *DCA* son muy nuevos en la literatura, y las técnicas *BSS* son un área de creciente actividad en los últimos años ya que existen infinidad de aplicaciones de ingeniería que los requieren. Por ejemplo, en esta tesis se caracterizó un tipo de dependencia entre señales que pueden ser separadas por MAXNG que permiten modelar a las señales presentes en imágenes hiper-espectrales. Sin embargo, es posible que en otras aplicaciones particulares, existan otro tipo de características de las señales, con diferentes tipos de dependencias y para esos casos se podría analizar la aplicabilidad de MAXNG o métodos similares.
- Es interesante analizar diferentes medidas de no-gaussianidad, de hecho existen métodos reportados en la bibliografía que utilizan diferentes medidas de no-gaussianidad, los cuales podrían ser útiles también para *DCA*.
- Un problema importante, encontrado en la aplicación de MAXNG a imágenes hiper-espectrales reales, es la variabilidad espectral, ya que las imágenes reales podrían no ajustarse al modelo de mezcla instantánea considerado en esta tesis, donde la matriz de mezcla A es invariante. Las extensiones adaptativas de MAXNG para mezclas no instantáneas, podrían ser de interés en este caso.
- Otro tema de interés es la separación de fuentes a partir de mezclas no-lineales ya que en muchos problemas reales, la mezcla lineal es un modelo que puede no ajustarse a la realidad existiendo términos no-lineales en las mezclas.

BIBLIOGRAFÍA

- [1] Adams J. B., Smith M. O., “Spectral mixture modelling: A new analysis of rock and soil types at the Viking Lander 1 Site”, *Journal of Geophysical Research*, 91, pags. 8098-8112 (1986).
- [2] Ahmed N., Natarajan T., Rao K. R., “Discrete cosine transform”, *IEEE Trans. Comput.*, C-23 (1), pags. 88-93 (1974).
- [3] Amari S. I., “Neural theory of association and concept formation”, *Biological Cybernetics*, 26, pags. 175–185 (1977).
- [4] Amari S-I., *Differential-geometrical methods in statistics, Lecture notes in statistics*, Springer-Verlag, Berlin (1985).
- [5] Araujo D., Barros A. K., Estombelo-Montesco C., Zhao H., Silva Filho A. R., Baffa O. Wakai R., Ohnoishi N., “Fetal Source Extraction from Magnetocardiographic Recordings by Dependent Component Analysis”. *Physics in Medicine and Biology*, 50 (19), pags. 4457-4464 (2005).
- [6] Attias H., “Independent Factor Analysis”, *Neural Computation*, 11 (4), pags. 803-851 (1999).
- [7] Baccigalupi C., Bedini L., Burigana C., De Zotti G., Farusi A., Maino D., Maris M., Perrotta F., Salerno E., Toffolatti L., Tonazzini A., “Neural Networks and Separation of Background and Foregrounds in Astrophysical Sky Maps”, *Mon. Not. R. Astron. Soc.*, 318, pags. 769-780 (2000).
- [8] Bach F. R., Jordan M. I., “Kernel Independent Component Analysis”, *Journal of Machine Learning Research*, 3, pags. 1-48 (2002).
- [9] Balram N., Moura J., “Noncausal Gauss-Markov random fields: Parameter structure and estimation”, *IEEE Trans. Inform. Theory* 39 (4), pags. 1333–1355 (1993).
- [10] Barros A., “Chapter 4: The Independence Assumption: Dependent Component Analysis”, *Advances in Independent Component Analysis*, Springer (2000).
- [11] Barros A. K., Cichocki A., Ohnishi N., “Extraction of Statistically Dependent Sources with Temporal Structure”. Proc. Brazilian Simposium on Neural Networks, Rio de Janeiro (2000).
- [12] Bedini L., Herranz D., Salerno E., Baccigalupi C., Kuruoglu E., Tonazzini A., “Separation of correlated astrophysical sources using multiple-lag data covariance matrices”, *Eurasip J. on Appl. Sig. Proc.*, 15, pags. 2400-2412 (2005).

- [13] Bell A. J., Sejnowski T. J., “A Non-linear Information Maximisation Algorithm that Performs Blind Separation”, *Advances in Neural Information Processing Systems*, 7, pags. 467-474 (1995).
- [14] Belouchrani A., Abed-Meraim K., Cardoso J. F. , Moulines E., “Second-order blind separation of temporally correlated sources”, Proc. Int. Conf. on Digital Sig. Proc., (Cyprus), pags. 346-351 (1993).
- [15] Bennett J., Khotanzad A., “Modeling Textured Image using Generalized Long Correlation Models”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20 (12), pags. 1365-1370 (1998).
- [16] Berman M., Kiiveri H., Lagerstrom R., Ernst A., Dunne R., Huntington J., “ICE: A statistical approach to identifying endmembers”, *IEEE Transactions on Geoscience and Remote Sensing*, 42 (10), pags. 2085-2095 (2004).
- [17] Besag J., “Spatial Interaction and Statistical Analysis of Lattice Systems”, *Journal of the Royal Statistical Society. Series B (Methodological)*, 36 (2), pags. 192-236 (1974).
- [18] Besag J., “On the statistical analysis of dirty pictures”, *Journal of the Royal Statistical Society. Series B (Methodological)*, 48 (3), pags. 259-302 (1986).
- [19] Bijaoui A., Nuzillard D., Deb Barma T., “BSS, Classification and Pixel Demixing”, in Proc. of 5th International Conference on Independent Component Analysis and Blind Source Separation, University of Granada (Spain), pags. 96-103 (2004).
- [20] Blanco Y., “Aplicación de los Estadísticos de Orden en el Análisis de Componentes Independientes”, Tesis de Doctorado, Universidad Politécnica de Madrid (2001).
- [21] Blanco Y., Zazo S., “New Gaussianity measures based on order statistics: application to ICA”, *Neurocomputing*, 51, pags 303 -320 (2003).
- [22] Boscolo R., Pan H., Roychowdhury V. P., “Independent Component Analysis Based on Nonparametric Density Estimation”, *IEEE Trans. on Neural Networks*, 15 (1), pags. 55-65 (2004).
- [23] Bouman C. A., “Tutorial: Markov Random Fields and Stochastic Image Models”, IEEE International Conference on Image Processing (ICIP '95) (1995).
- [24] Brown L. G., “A survey of image registration techniques”, *ACM Computing Surveys (CSUR)*, 24 (4), pags. 325-376 (1992).
- [25] Caiafa C. F., *Procesos Aleatorios Bidimensionales. Su aplicación al Procesamiento de Imágenes*, Tesis de Ingeniería Electrónica, Facultad de Ingeniería, Universidad de Buenos Aires, Argentina (1996).

- [26] Caiafa C. F., Proto A. N., “Desarrollo de un software para la identificación de Elefantes Marinos por Eigenfaces”, *Technical Papers (ITBA)*, 5, pags. 27-40 (2003).
- [27] Caiafa C. F., Proto A. N., Vergani D., Stanganelli Z., “Development of individual recognition of female southern elephant seals, Mirounga Leonina, applying Principal Component Analysis”, *J. of Biogeography*, 32 (7), pags. 1257-1266 (2005).
- [28] Caiafa C. F., Sassano M. P., Proto A. N., “Wavelet and Karhunen-Loève transformations applied to SAR signals and images”, *Physica A*, 356 (1), pags. 172-177 (2005).
- [29] Caiafa C. F., Proto A. N., “A non-Gaussianity measure for blind source separation”, Proceedings of SPARS05, Nov. 16-18, 2005 – IRISA - Rennes (France).
- [30] Caiafa C. F., Proto A. N., “Separation of statistically dependent sources using an L^2 -distance non-Gaussianity measure”, *Signal Processing*, 86 (11), pags. 3404-3420 (2006).
- [31] Caiafa C. F., Proto A. N., “Temperature estimation in the two dimensional Ising model”, *International Journal on Modern Physics C*, 17 (1), pags. 29-38 (2006).
- [32] Caiafa C. F., Kuruoglu E. E., A. N. Proto, “A Minimax Entropy Method for Blind Separation of Dependent Components in Astrophysical Images”, Proc. of the Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2006), CNRS, Paris, France, July 8-13 (2006).
- [33] Caiafa C. F., Salerno E., Proto A. N., Fiumi L., “Dependent Component Analysis as a Tool for Blind Spectral Unmixing of Remote Sensed Images”, Proc. of 14th European Signal Processing Conference (EUSIPCO06), Florence, Italy, September 4-8 (2006).
- [34] Caiafa C. F., Proto A. N., Kuruoglu E. E., “Long Correlation Gaussian Random Fields: Parameter Estimation and Noise Reduction”, *Digital Signal Processing Elsevier*, 17, pags. 819-835, (2007).
- [35] Caiafa C. F., Salerno E., Proto A. N., Fiumi L., “Blind Spectral Unmixing by Local Maximization of Non-Gaussianity”, *Signal Processing EURASIP* (aceptado Junio 2007).
- [36] Cardoso J. F., “Source separation using higher order moments”, Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP’89), pags. 2109-2112, Glasgow, UK (1989).

- [37] Cardoso J. F., Souloumiac A., “Jacobi angles for simultaneous diagonalization”, *SIAM Journal of Matrix Analysis and Applications*, 17 (1), pags. 161-164 (1996).
- [38] Cardoso J. F., “Dependence, Correlation and Gaussianity in Independent Component Analysis”, *Journal of Machine Learning Research*, 4, pags. 1177-1203 (2003).
- [39] Chaturvedi S., Kapoor A. K., Srinivasan V., “A new orthogonalization procedure with an extremal property”, *J. Phys. A: Math. Gen.*, 31, pags. 367-370 (1998).
- [40] Chavez P. S., Stuart J., Sides C., Anderson J. A., “Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic”, *Photogramm. Eng. Remote Sens.*, 57, pags. 259–303 (1991).
- [41] Chellappa R., Kashyap R. L., “Synthetic Generation and Estimation in Random Field Models of Images”, Proc. IEEE CS Conf. Pattern Recognition and Image Processing, pags. 577-582, Dallas, Texas. (1981).
- [42] Chellappa R. “Two-Dimensional Discrete Gaussian Markov Random Field Models for Image Processing”, *Progress in Pattern Recognition*, 2, pags 79-112 (1985).
- [43] Chiang S-S, Chang C-I, Ginsberg I. W., “Unsupervised Hyperspectral Image Analysis Using Independent Component Analysis”, in Proc. of International Geoscience and Remote Sensing Symposium (IGARSS 2000), Honolulu Hawaii, July 24-28, pags. 3136-3138 (2000).
- [44] Cichocki A., Unbehauen R. “Robust estimation of principal components in real time”, *Electronics Letters*, 29 (21), pags. 1869–1870 (1993).
- [45] Cichocki A., Swiniarski R., Bogner R. E., “Hierarchical neural network for robust PCA of complex-valued signals”. Proc. of the World Congress on Neural Networks (WCNN-96), San Diego, USA, pags. 818–821 (1996).
- [46] Cichocki A., Rutkowski T., Barros A. K., “Blind Extraction of Temporally Correlated but Statistically Dependent Acoustic Signals”. Proc. of IEEE Workshop on Neural Networks for Signal Processing (NNSP ’2000), Sydney (2000).
- [47] Cichocki A., Amari S. I., *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. J. Wiley & Sons, Chichester UK (2002).
- [48] Cichocki A., Amari S., Siwek K., Tanaka T., et al., ICALAB Toolboxes, <http://www.bsp.brain.riken.jp/ICALAB>.

- [49] Comon P., “Independent Component Analysis, a new concept?”, *Signal Processing*, 36 (3) pages. 287-314 (1994).
- [50] Conway J. B., *A course in Functional Analysis*, Springer-Verlag (1985).
- [51] Cover T., Thomas J., *Elements of Information Theory*, John Wiley & Sons, New York (1991).
- [52] Cruces S., Cichocki A., Amari S-I., “The Minimum Entropy and Cumulants Based Contrast Functions for Blind Source Extraction”, IWANN 2001, Granada, Spain, June 13-15, (2001).
- [53] Dempster A. P., Laird N. M., Rubin D. B., “Maximum likelihood from incomplete data via the *EM* algorithm”. *J. of the Royal Statistical Society, Series B*, 34 pags. 1-38 (1977).
- [54] Diamantaras K. I., Kung S. Y., *Principal Component Neural Networks: Theory and Applications*. Wiley (1996).
- [55] Donoho D., “On minimum entropy deconvolution”, *Applied Time Series Analysis II*, pags. 565-609, Academic, New York: (1981).
- [56] Donoho D. L., Vetterli M., De Vore R. A., Daubechies I., “Data Compression and Harmonic Analysis”, *IEEE Trans. on Inform. Theory*, 44 (6), pags. 2435–2476 (1998).
- [57] Duda R. O., Hart P. E., Stork D. G., *Pattern Classification*, J. Wiley, 2^{da} Edición, New York (2000).
- [58] Erdogmus D., Hild K. E., Principe J. C., Lazaro M., Santamaria I., “Adaptive blind deconvolution of linear channels using Renyi’s entropy with Parzen window estimation”, *IEEE Trans. on Sig. Proc.*, 52 (6), pags. 1489-1498 (2004).
- [59] Feller W., *An Introduction to Probability Theory and Its Applications, Volume 1*, J. Wiley (1950).
- [60] Friedman J. H., Tukey J. W., “A Projection Pursuit Algorithm for Exploratory Data Analysis”, *IEEE Transactions on Computers*, C-23 (9), pags. 881-890 (1974).
- [61] Friedman J. H., “Exploratory projection pursuit”, *J. Am. Stat. Assoc.*, 82, pags. 249–266, (1987).
- [62] Geman S., Geman D., “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Trans. Pattern Anal. Machine Intell.*, 6 (6), pags. 721–741 (1984).

- [63] Georgiev P., Cichocki A., “Blind source separation via symmetric eigenvalue decomposition”, Proc. of Sixth International Symposium on Signal Processing and its Applications, Aug. 2001, pags. 17-20, Kuala Lumpur, Malaysia (2001).
- [64] Gonzalez R. C., Woods R. E. *Digital Image Processing.*, Prentice Hall. Upper Saddle River, 2^{da} Edición, New Jersey (2002).
- [65] Gray R. M., “Toeplitz and Circulant Matrices: A review”, *Foundations and Trends in Communications and Information Theory*, 2 (3), pags. 155-239 (2006).
- [66] Greffeath D., *Introduction to Random Fields*, Springer-Verlag, New York (1976).
- [67] Hall P., “Polynomial Projection Pursuit”, *Annals of Statistics*, 17, pags. 589-605 (1989).
- [68] Halmos P. R., *Finite-Dimensional Vector Spaces*, Springer-Verlag (1974).
- [69] Haykin S., *Unsupervised Adaptive Filtering. Volume I: Blind Source Separation*, J. Wiley, New York (2000).
- [70] Hocking J. G., Young G. S., *Topology*. Dover, New York (1988).
- [71] Hotelling H., “Analysis of a complex of statistical variables into principal components”, *Journal of Educational Psychology*, 24, pags. 417-441 (1933).
- [72] Huang K., *Statistical Mechanics*, Wiley, 2^{da} Edición, New Jersey (1987).
- [73] Huber P. J., “Projection Pursuit”, *The Annals of Statistics*, 13, pags. 435-475 (1985).
- [74] Hyvärinen A., Oja E., “A fast fixed-point algorithm for independent component analysis”, *Neural Computation*, 9 (7), pags. 1483–1492 (1997).
- [75] Hyvärinen A., Oja E., “Independent component analysis: algorithms and applications”, *Neural networks*, 13 (4-5), pags. 411-430 (2000).
- [76] Hyvärinen A., Karhunen J., Oja E., *Independent Component Analysis*, J. Wiley & Sons, New York (2001).
- [77] Ising E., “Beitrag zur theorie des ferromagnetismus”, *Z. Physik*, 31, pags. 253-258 (1925).
- [78] Jolliffe I.T., *Principal Component Analysis*, Springer Series in Statistics, Springer-Verlag, New York (2002).
- [79] Jones M. C.,Sibson R., “What is projection pursuit?” (with discussion), *J. R. Statist. Soc. A*, 150, pags. 1-36 (1987).

- [80] Jutten C., Herault J., “Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture”, *Signal Processing*, 24 (1), pags. 1-10 (1991).
- [81] Kashyap R. L., Chellappa R., “Estimation and choice of neighbors in spatial-interaction models of images”, *IEEE Trans. Inform. Theory*, 29 (1), pags. 60-72 (1983).
- [82] Karhunen K., “Über lineare methoden in der Wahrscheinlichkeitsrechnung”, *Annales Academiae Scientiarum Fennicae, Seried A1: Mathematica-Physica*, 37, pags. 3-79 (1947).
- [83] Kendall M., Stuart A., *The advanced theory of statistics*, volume 2. Griffin, London, 4th edition (1979).
- [84] Keshava N., Mustard J., “Spectral unmixing”, *IEEE Signal Process. Mag.*, 19 (1), pags. 44–57 (2002).
- [85] Kolmogorov A. N., Fomin S. V., *Measure, Lebesgue Integrals and Hilbert Space*, Academic Press, New York and London (1961).
- [86] Kosaka N., Uto K., Kosugi Y., “ICA-Aided Mixed-Pixel Analysis of Hyperspectral Data in Agricultural Land”, *IEEE Geoscience and Remote Sensing Letters*, 2, pags. 220-224 (2005).
- [87] Kramer H. P., Mathews M. V., “A linear coding for transmitting a set of correlated signals” *IRE Trans. Inform. Theory*, 23, pags. 41–46 (1956).
- [88] Kuruoglu E. E. , Bedini L., Paratore M. T., Salerno E., Tonazzini A., “Source separation in astrophysical maps using independent factor analysis”, *Neural Networks*, 6 (3-4), pags. 479-491 (2003).
- [89] Kruse F., Lefkoff A., Boardman J., Heidebrecht K., Shapiro A., Barloon P., Goetz A. “The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data”, *Remote Sensing of Environment*, 44, pags. 145-163 (1993).
- [90] Kruskal J. B., “Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new index of condensation”, *Statistical Computation*, pags. 427-440, Academic, New York, 1969.
- [91] Kruskal J. B., “Linear transformation of multivariate data to reveal clustering”, *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, pags. 179-191, Semenir Press, New York-London (1972).
- [92] Lee D.T., “JPEG 2000: retrospective and new developments”, *Proc. of the IEEE*, 93 (1), pags. 32-41 (2005).

- [93] Le Moigne J., Morisette J., Cole-Rhoades A., Netanyahu N. S., Eastman R., Stone H., "Earth science imagery registration", Proc. of International Geoscience and Remote Sensing Symposium (IGARSS'03), Toulouse, pags. 161-163 (2003).
- [94] Lennon M., Mercier G., Mouchot M. C., Hubert-Moy L., "Spectral unmixing of hyperspectral images with the Independent Component Analysis and wavelet packets", in IEEE Proc. of the International Geoscience and Remote Sensing Symposium, Sydney, Australia, July (2001).
- [95] Loéve M., *Probability Theory*, Van Nostrand, New York (1963).
- [96] Löwdin P-O., "A Quantum Mechanical Calculation of the Cohesive Energy, the Interionic Distance, and the Elastic Constants of Some Ionic Crystals. I", *Ark. Mat. Astr. Fys.* 35A, 9, pags. 1-10 (1947).
- [97] Luenberger D. G., *Optimization by Vector Space Methods*, John Wiley & Sons, Inc. (1997).
- [98] Maes F., Vandermeulen D., Suetens P., "Medical image registration using mutual information", *Proc. of the IEEE* , 91 (10), pags. 1699-1722 (2003).
- [99] Mallat S., *A Wavelet Tour of Signal Processing*, Academic Press Elsevier, 2^{da} Edición (1999).
- [100] Meyer C. D., *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2000).
- [101] Movahed M. S., Ghasemi F., Rahvar S., Tabar M. R. R., "New Computational Approaches to Analysis of *CMB* Map: The Statistical Isotropy and Gaussianity", eprint arXiv:astro-ph/0602461 (2006).
- [102] Moran P. A. P., "A Gaussian Markovian Process on a Square Lattice", *J. Appl. Prob.*, 10 (1), pags. 54-62 (1973).
- [103] Nascimento J. M. P., Bioucas Dias J. M., "Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?", *IEEE Transactions on Geoscience and Remote Sensing*, 43 (1), pags. 175-187 (2005).
- [104] Oja E., "Principal components, minor components and linear neural networks", *Neural Networks*, 5, pags. 927–935 (1992).
- [105] Oja E., Karhunen J., Hyvärinen A. "From neural principal components to neural independent components", *Lecture Notes In Computer Science*, 1327, Proc. of the 7th International Conference on Artificial Neural Networks, Lausanne, Switzerland, pags. 519-528 (1997).
- [106] Moghaddam B., Pentland A., "Probabilistic Visual Learning for Object Representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19 (7), pags. 696-710 (1997).

- [107] Parzen E., “On the estimation of a probability density function and mode”, *Annals of Mathematical Statistics*, 33, pags.1065-1076 (1962).
- [108] Pearson K., “On lines and planes of closest fit to systems of points in space”, *Philosophical Magazine*, 2, pags. 559-572 (1901).
- [109] Pluim, J. P. W., Fitzpatrick J. M., “Image registration”, *IEEE Transactions on Medical Imaging*, 22 (11), pags. 1341-1343 (2003).
- [110] Pratt W. K., *Digital Image Processing: PIKS Inside*, John Wiley & Sons, 3^{ra} Edición (2001).
- [111] Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 2^{da} Edición, Cambridge (1992).
- [112] Sarajedini A., Chau P.M., “Blind signal separation with a projection pursuit index”, Proc. of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998 (ICASSP '98), pags. 2125-2128 (1998).
- [113] Schweinler H. C., Wigner E. P., “Orthogonalization Methods”, *J. Math. Phys.*, 11, pags. 1693-1694 (1970).
- [114] Shannon C. E., “A mathematical theory of communication”, *Bell System Tech. J.*, 27, pags. 379-423, (1948).
- [115] Silverman B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York (1985).
- [116] Srivastava V., “A unified view of the orthogonalization methods”, *J. Phys. A: Math. Gen.*, 33, pags. 6219-6222 (2000).
- [117] Stauffer D., Stanley H. E., *From Newton to Mandelbrot*, Springer-Verlag, 2^{da} Edición, New York (1995).
- [118] Thevenaz P., Ruttimann U. E., Unser M., “A pyramid approach to subpixel registration based on intensity”, *IEEE Transactions on Image Processing*, 7 (1), pags. 27-41 (1998).
- [119] Tong L., Soon V., Huang Y. F., Liu R., “Indeterminacy and identifiability of blind identification”, *IEEE Trans. Circuits and Systems*, 38 (5), pags 499-509 (1991).
- [120] Tong, L. *The Multivariate Normal Distribution*. New York: Springer-Verlag, 1990.
- [121] Tso B., Mather P. M., *Classification methods for remotely sensed data*, Taylor and Francis, New York (2001).

- [122] Turk M., Pentland A., “Eigenfaces for Recognition”, *J. Cognitive Neuroscience*, 3 (1), pags. 71-86 (1991).
- [123] Unser M., “Wavelets, Filterbanks, and the Karhunen-Loève Transform”, Proc. of the 9th European Signal Processing Conference (EUSIPCO’98), Rhodes, Greece, Sep. 8-11, pags. 1737-1740 (1998).
- [124] Vaseghi S. V., *Advanced Digital Signal Processing and Noise Reduction*, Wiley & Sons, 3_{ra} Edición, New York (2000).
- [125] Voronoi G., “Nouvelles applications des paramètres continus à la théorie des formes quadratiques” *Journal für die Reine und Angewandte Mathematik*, 133, pags. 97-178 (1907)
- [126] Vetterli M., “Wavelets, Approximation and Compression”, *IEEE Signal Processing Magazine*, 18 (5), pags. 59-73 (2001).
- [127] Wiener N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, New York (1949).
- [128] Xie H., Pierce L. E., Ulaby F. T., “Mutual information based registration of SAR images”, Proc. of International Geoscience and Remote Sensing Symposium (IGARSS’03), Toulouse, pags. 4028-4031 (2003).
- [129] Yokoo T., Knight B. W., Sirovich L., “ L^2 De-Gaussianization and Independent Component Analysis”. Proc. of ICA 2003, pags. 757-762, 4th International Symposium on Independent Component Analysis and Blind Source Separation, Nara, Japan (2003).

