



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# Nuevas contribuciones a la teoría y aplicación del procesado de señal sobre grafos

---

Tesis doctoral realizada por Jordi Belda Valls  
Dirigida por Luís Vergara Domínguez y Addison Salazar Afanador

Julio de 2022

# ABSTRACT

Graph signal processing appears as an emerging field of techniques that combine concepts from two highly consolidated areas: signal processing and graph theory. From the perspective of signal processing, it is possible to achieve a more general signal definition by assigning each value of the signal to a vertex of a graph. Conventional signals can be considered particular cases where the sample values are assigned to a uniform (temporal or spatial) grid. From the perspective of graph theory, new transformations of the graph can be defined in such a way that they extend the classical concepts of signal processing such as filtering, prediction and spectral analysis. Furthermore, graph signal processing is finding new applications in detection and classification areas due to its flexibility to model general dependencies between variables.

In this thesis, new contributions are given to graph signal processing. Firstly, it is considered the problem of estimating the Laplacian matrix associated with a graph, which determines the relationship between nodes. Conventional methods are based on the precision matrix, where Gaussianity is implicitly assumed. In this thesis, new methods to estimate the Laplacian matrix from the partial correlations are proposed respectively assuming two different non-Gaussian models in the observation space: Gaussian Mixtures and Independent Component Analysis. The proposed methods have been tested with simulated data and with real data in some selected biomedical applications. It is demonstrated that better estimates of the Laplacian matrix can be obtained with the new proposed methods in cases where Gaussianity is not a correct assumption.

The problem of generating synthetic signal in scenarios where real signals scarcity can be an issue has also been considered. Graph models allow more general pair-wise dependence models between signal samples. Thus a new method based on the Complex Graph Fourier Transform and on the concept of subrogation is proposed. It has been applied in the challenging problem of hand gesture recognition. It has been demonstrated that extending the original training set with graph surrogate replicas, significantly improves the accuracy of the hand gesture classifier.

# RESUMEN

El procesamiento de señal sobre grafos es un campo emergente de técnicas que combinan conceptos de dos áreas muy consolidadas: el procesamiento de señal y la teoría de grafos. Desde la perspectiva del procesamiento de señal puede obtenerse una definición de la señal mucho más general asignando cada valor de la misma a un vértice de un grafo. Las señales convencionales pueden considerarse casos particulares en los que los valores de cada muestra se asignan a una cuadrícula uniforme (temporal o espacial). Desde la perspectiva de la teoría de grafos, se pueden definir nuevas transformaciones del grafo de forma que se extiendan los conceptos clásicos del procesamiento de la señal como el filtrado, la predicción y el análisis espectral. Además, el procesamiento de señales sobre grafos está encontrando nuevas aplicaciones en las áreas de detección y clasificación debido a su flexibilidad para modelar dependencias generales entre variables.

En esta tesis se realizan nuevas contribuciones al procesamiento de señales sobre grafos. En primer lugar, se plantea el problema de estimación de la matriz Laplaciana asociada a un grafo, que determina la relación entre nodos. Los métodos convencionales se basan en la matriz de precisión, donde se asume implícitamente Gaussianidad. En esta tesis se proponen nuevos métodos para estimar la matriz Laplaciana a partir de las correlaciones parciales asumiendo respectivamente dos modelos no Gaussianos diferentes en el espacio de las observaciones: mezclas gaussianas y análisis de componentes independientes. Los métodos propuestos han sido probados con datos simulados y con datos reales en algunas aplicaciones biomédicas seleccionadas. Se demuestra que pueden obtenerse mejores estimaciones

de la matriz Laplaciana con los nuevos métodos propuestos en los casos en que la Gaussianidad no es una suposición correcta.

También se ha considerado la generación de señales sintéticas en escenarios donde la escasez de señales reales puede ser un problema. Los modelos sobre grafos permiten modelos de dependencia por pares más generales entre muestras de señal. Así, se propone un nuevo método basado en la Transformada de Fourier Compleja sobre Grafos y en el concepto de subrogación. Se ha aplicado en el desafiante problema del reconocimiento de gestos con las manos. Se ha demostrado que la extensión del conjunto de entrenamiento original con réplicas sustitutas generadas con los métodos sobre grafos, mejora significativamente la precisión del clasificador de gestos con las manos.

# RESUM

El processament de senyal sobre grafs és un camp emergent de tècniques que combinen conceptes de dues àrees molt consolidades: el processament de senyal i la teoria de grafs. Des de la perspectiva del processament de senyal pot obtindre's una definició del senyal molt més general assignant cada valor de la mateixa a un vèrtex d'un graf. Els senyals convencionals poden considerar-se casos particulars en els quals els valors de la mostra s'assignen a una quadrícula uniforme (temporal o espacial). Des de la perspectiva de la teoria de grafs, es poden definir noves transformacions del graf de manera que s'estenguen els conceptes clàssics del processament del senyal com el filtrat, la predicció i l'anàlisi espectral. A més, el processament de senyals sobre grafs està trobant noves aplicacions en les àrees de detecció i classificació a causa de la seua flexibilitat per a modelar dependències generals entre variables.

En aquesta tesi es donen noves contribucions al processament de senyals sobre grafs. En primer lloc, es planteja el problema d'estimació de la matriu Laplaciana associada a un graf, que determina la relació entre nodes. Els mètodes convencionals es basen en la matriu de precisió, on s'assumeix implícitament la gaussianitat. En aquesta tesi es proposen nous mètodes per a estimar la matriu Laplaciana a partir de les correlacions parcials assumint respectivament dos models no gaussians diferents en l'espai d'observació: mescles gaussianes i anàlisi de components independents. Els mètodes proposats han sigut provats amb dades simulades i amb dades reals en algunes aplicacions biomèdiques seleccionades. Es demostra que poden obtindre's millors estimacions de la matriu Laplaciana amb els nous mètodes proposats en els casos en què la gaussianitat no és una suposició correcta.

També s'ha considerat el problema de generar senyals sintètics en escenaris on l'escassetat de senyals reals pot ser un problema. Els models sobre grafs permeten models de dependència per parells més generals entre mostres de senyal. Així, es proposa un nou mètode basat en la Transformada de Fourier Complexa sobre Grafs i en el concepte de subrogació. S'ha aplicat en el desafiador problema del reconeixement de gestos amb les mans. S'ha demostrat que l'extensió del conjunt d'entrenament original amb rèpliques substituïdes generades amb mètodes sobre grafs, millora significativament la precisió del classificador de gestos amb les mans.

# Tabla de contenido

Capítulo 1. Revisión del procesado de señal sobre grafos .....	13
1.1. Introducción .....	13
1.2. Procesado de señal algebraico.....	14
1.3. Estudio de la matriz Laplaciana .....	17
1.4. Procesado de señal sobre grafos.....	18
1.5. Hipótesis y objetivos de la tesis .....	21
Referencias.....	25
Capítulo 2. Estimación de la matriz Laplaciana en Modelos de Mezclas Gaussianas .....	28
2.1. Introducción .....	28
2.2. Coeficiente de correlación parcial generalizado GPCC .....	30
2.3. Calculando la GPCC de una GMM.....	33
2.4. Experimentos.....	41
2.4.1. Simulaciones.....	41
2.4.2. Un experimento con datos reales .....	46
2.5. Conclusiones .....	47
A. Apéndice: <i>Derivación de la media condicionada</i> .....	49
B. Apéndice: <i>Una medida de separación entre componentes GMM</i> .....	52
Referencias.....	54
Capítulo 3. Estimación de la matriz Laplaciana en Modelos de Análisis de Componentes Independientes .....	57
3.1. Introducción .....	57
3.2. Correlaciones parciales en modelos ICA .....	58
3.2.1. Planteamiento del problema.....	58
3.2.2. Una fórmula general para la covarianza residual.....	60
3.3. Calculando ICA-PCC.....	62
3.4. Experimentos.....	69
3.4.1. Experimentos con datos sintéticos .....	69



3.4.2.	Una aplicación con datos reales .....	72
3.5.	Conclusiones .....	80
	Referencias.....	87
Capítulo 4.	Un nuevo algoritmo de subrogación para la Transformada de Fourier Compleja sobre Grafos (CGFT).....	89
4.1.	Introducción.....	89
4.1.1.	Planteamiento del problema y trabajos relacionados .....	89
4.1.2.	Nuevas contribuciones y organización del capítulo .....	90
4.2.	La matriz Laplaciana Hermítica .....	92
4.2.1.	Definición y propiedades .....	92
4.2.2.	La transformada de Fourier Compleja sobre grafos (CGFT).....	96
4.3.	Invariantes de la Amplitud del Espectro sobre grafos.....	97
4.4.	Algoritmos de subrogación.....	101
4.4.1.	Algoritmo iterativo de subrogación de señales sobre grafos.....	101
4.4.2.	Seleccionando la matriz Laplaciana Hermítica .....	103
4.5.	Experimentos.....	105
4.5.1.	Clasificación de instancias aleatorias.....	107
4.5.2.	Reconocimiento automático de gestos con las manos.....	110
4.6.	Conclusiones .....	117
	Referencias.....	121
Capítulo 5.	Conclusiones generales y líneas futuras .....	127
5.1.	Conclusiones generales .....	127
5.2.	Líneas futuras .....	128
5.2.1.	“Sparsity” .....	129
5.2.2.	“Smoothness” .....	129
5.2.3.	Aplicación en detección/clasificación.....	130
5.2.4.	Definición de la matriz Laplaciana en los métodos subrogación	130
5.3.	Publicaciones generadas a partir de los resultados de la tesis .....	131

# LISTA DE ACRÓNIMOS

GSP	Graph Signal Processing
ASP	Algebraic Signal Processing
FFT	Fast Fourier Transform
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSP	Discrete Signal Processing
pdf	Probability density function
LLMSE	Linear Least Mean-Square Error
LMSE	Least Mean-Square Error
GMM	Gaussian Mixture Model
ICA	Independent Component Analysis
GPCC	Generalized Partial Correlation Coefficient
aGPCC	Approximated Generalized Partial Correlation Coefficient
DAG	Directed Acyclic Graph
BN	Bayesian Network
EM	Expectation-Maximization
ASI	Alfa-Slow-Index
EEG	Electroencefalografía
MSE	Mean Squared Error
i.i.d.	Independientes e idénticamente distribuidas
FT	Fourier Transform
GFT	Graph Fourier Transform
CGFT	Complex Graph Fourier Transform

GWSS	Graph Wide-Sense Stationary
GSA	Graph Spectrum Amplitude
AAFT	Amplitude Adjusted Fourier Transform surrogating algorithm
IAAFT	Iterative AAFT surrogating algorithm
ICGFT	Iterative CGFT surrogating algorithm
OTS	Original Training Set
ETS	Enlarged Training Set
OITS	Original Imbalanced Training Set
EITS	Enlarged Imbalanced Training Set
HMM	Hidden Markov Models
nIGFT	non-Iterative GFT surrogating algorithm
I	Imbalanced number

# LISTA DE FIGURAS

Figura 2.1. DAG de la BN asociada al GMM. Los círculos sombreados corresponden a las variables latentes, los círculos blancos a las variables observadas, los rectángulos corresponden a los parámetros que definen las probabilidades a priori o las pdfs condicional. a) GMM compacto b) GMM dividido implícito en la estimación de las correlaciones parciales.....	38
Figura 2.2. Error en la estimación de la matriz Laplaciana para varias separaciones de los dos componentes de la GMM, $N=20$ , 5000 de tamaño del conjunto de entrenamiento.....	42
Figura 2.3. Error de estimación de la matriz Laplaciana según la variación del tamaño del conjunto de entrenamiento, $N=20$ , 2 componentes GMM, $b=5$ .....	43
Figura 2.4. Error de estimación de la matriz Laplaciana para diferentes números de componentes $K$ de GMM, tamaño del conjunto de entrenamiento de 5000, $N=20$ .....	44
Figura 2.5. Tiempo computacional necesario para los diferentes métodos de estimación de la matriz Laplaciana variando la dimensión $N$ del vector de observación, 2 componentes GMM, $b=5$ , tamaño del conjunto de entrenamiento de 1000.....	45
Figura 3.1. (continua) y (discontinua) a) Caso sub-gaussiano b) Caso súper-gaussiano c) Caso sub/súper-gaussiano mezclado d) Caso gaussiano.....	71
Figura 3.2. Matrices de adyacencia correspondientes a la característica Amp.....	78
Figura 3.3. Matrices de adyacencia correspondientes a la característica Así.....	79
Figure 4.1. Tasa de error de clasificación para un número creciente de instancias de entrenamiento.....	109
Figura 4.2. Tasa de error de clasificación para un número creciente de instancias de entrenamiento.....	114
Figura 4.3. Tasa de error de clasificación para el aumento del número de desequilibrio.....	116

# LISTA DE TABLAS

Tabla 2.1. Complejidad computacional en términos de N, K y L.....	40
Tabla 2.2. Errores normalizados obtenidos en el experimento con datos reales.....	47
Tabla 3.1. Resultados correspondientes a la característica Amp.....	76
Tabla 3.2. Resultados correspondientes a la característica Asi.....	76
Tabla 4.1. Invariantes GSA.....	101
Tabla 4.2. Tasa de error promediada sobre el número de instancias de entrenamiento.....	110
Tabla 4.3. Tasas de error para diferentes conjuntos de entrenamiento originales y ampliados.....	115
Tabla 4.4. Tasa de error sobre el número de desequilibrio para los conjuntos de entrena-miento originales y ampliados.....	117

# Capítulo 1

## Revisión del procesamiento de señal sobre grafos

### 1.1. Introducción

El procesamiento de señal sobre grafos (GSP, siglas en inglés de *Graph Signal Processing*) parte de dos áreas principales que han evolucionado en paralelo y que han generado una nueva herramienta de trabajo de procesamiento de señal. Por una parte, se fue desarrollando una aproximación algebraica a los conceptos clásicos del procesamiento de señal, denominada procesamiento de señal algebraico (ASP, siglas en inglés de *Algebraic Signal Processing*) junto a aproximaciones algebraicas a las transformaciones discretas trigonométricas como la transformada discreta de Fourier, para proporcionar nuevos algoritmos más eficientes. Por otra parte, se fue progresando en el estudio de la matriz Laplaciana de la teoría de grafos a partir de la representación sobre ésta de datos provenientes de distintas aplicaciones, para generar nuevas herramientas de procesamiento de datos a través del análisis espectral de la misma. Este contexto histórico y la convergencia en el desarrollo de herramientas de GSP se

expone en [1]. Este artículo, además de situar el estado actual del GSP, plantea los retos a superar, los conceptos clave y algunas de las aplicaciones de este campo de estudio.

En este capítulo se hace una revisión de algunos de los principales trabajos que han originado el GSP. Además, se definen los principales elementos, se revisan los retos a abordar y algunas de las aplicaciones más relevantes que han sido clave para su desarrollo.

## **1.2. Procesado de señal algebraico**

El ASP es una aproximación general y axiomática del procesado de señal lineal convencional. Para ello se define un modelo de señal en el que los conceptos clásicos tales como los filtros temporales y espaciales o el espacio de la señal, tienen sus respectivos conceptos algebraicos equivalentes. Esta aproximación se introdujo principalmente en [2] y [3] a partir de [4] para el caso de señales discretas.

En [2] se define un modelo lineal de la señal mediante una estructura algebraica genérica de filtros, un espacio de señales y un mapeo lineal biyectivo desde un espacio vectorial coordinado sobre el espacio de señales, generalizando la transformada  $Z$ . Este modelo genérico puede definir los conceptos convencionales del procesado de señal a partir de la selección de las bases, las dimensiones y el mapeo lineal para un espacio vectorial dado. En principio, cualquier modelo de señal puede ser válido, por lo que una de las principales cuestiones es determinar cuál es el más adecuado teniendo en cuenta alguno de los siguientes aspectos: los modelos de las aplicaciones más comunes del procesado de señal, los modelos que estén asociados a una transformación lineal dada y el desarrollo de un modelo que refleje propiedades deseadas, según por ejemplo, la aplicación. En este artículo se respon-

de a estas cuestiones desde el dominio temporal eligiendo un operador de translación temporal capaz de generar el álgebra de filtros y mostrando propiedades estándares del procesamiento de señal que puedan ser llevadas a los correspondientes requisitos del modelo de la señal algebraico.

Como se explica en [2]-[5], el operador de translación es un filtro especial y en el caso del procesamiento de señal unidimensional o 1- $D$  sólo es posible un tipo de translación, en el caso  $k$ - $D$  son necesarios  $k$  tipos de translación, según explican los autores. En este artículo se focaliza el caso de un operador de translación y se identifican las álgebras conmutativas posibles. Por ejemplo, forzando invariancia a la translación se asegura que el álgebra sea conmutativa. Además se muestra que los polinomios son clave en la estructura en ASP.

A partir de estas ideas base se presentan las equivalencias en ASP del filtrado, análisis espectral, transformada de Fourier, respuesta en frecuencia y otros conceptos comunes en el procesamiento de señal convencional para el caso discreto en donde las señales son secuencias de números finitas o infinitas de una base que puede ser números reales o complejos sobre un rango de índices. En general se muestra que entendiendo y explotando los beneficios de la relación entre conceptos del procesamiento de señal y su álgebra equivalente, se pueden desarrollar nuevas herramientas de procesamiento de señal, así como modelos de señal diferentes del procesamiento de señal temporal clásico. Desde un punto de vista práctico, esto significa que se puede construir un gran número de herramientas de procesamiento de señal distintas con diferentes nociones de filtrado y espectro. También que imponiendo invariancia a la translación se puede identificar el álgebra polinómica como una estructura clave del procesamiento de señal. Además, esto permite la derivación de nuevos modelos de la señal por otros operadores de translación no estándares.



Finalmente, cabe mencionar que en este artículo también se hace un repaso histórico del uso del álgebra en teoría de señal en dos direcciones principales: la transformada rápida de Fourier (FFT, siglas en inglés de *Fast Fourier Transform*) y el análisis de Fourier en grupos.

En otros trabajos como [5], [6], [7] o [22] se desarrollaron aproximaciones algebraicas para proponer algoritmos de transformaciones trigonométricas más eficientes y rápidos.

En [3] se utiliza la teoría del álgebra para desarrollar modelos espaciales de señales finitas e infinitas. En este caso se analiza el dominio espacial, por lo que el concepto de traslación es espacial, no dirigido o simétrico y se complementa al artículo anterior en el que el análisis era en el dominio temporal.

Junto a estos trabajos, se siguió desarrollando el procesamiento de señal algebraico en [8] y [9] dónde se presentan herramientas algebraicas para analizar señales discretas a partir de combinaciones lineales de polinomios ortogonales y especialmente en muestras de señales distribuidas.

Adicionalmente, el ASP a partir de una definición apropiada de un modelo de espacio de la señal en forma de grafo línea, se puede utilizar para mostrar que la transformada discreta del coseno (DCT, siglas en inglés de *Discrete Cosine Transform*) desempeña el mismo papel para ese modelo de señal como el que juega la transformada discreta de Fourier (DFT, siglas en inglés de *Discrete Fourier Transform*) para el modelo temporal cíclico. A partir de estas definiciones se introduce la posibilidad de generar matrices de adyacencia de grafos con pesos como traslaciones para generar el modelo de señal basado en grafos para señales indexadas sobre nodos de un grafo arbitrario dirigido o no dirigido. Por lo tanto, estos trabajos sientan unas primeras bases para el desarrollo del GSP.

### **1.3. Estudio de la matriz Laplaciana**

En paralelo al desarrollo del procesado de señal algebraico aparecieron diferentes trabajos enfocados en el estudio de la representación de datos de altas dimensiones en estructuras de menor dimensión a través de proyecciones sobre un subespacio de bases propias de la matriz Laplaciana del grafo asociado.

En [10] y [11] se desarrollan representaciones de dimensiones bajas para datos de gran dimensión a partir de los autovectores de la matriz Laplaciana de un grafo y del operador de Laplace-Beltrami de la geometría diferencial.

En [12] y [13] se eligen aproximaciones discretas para otros operadores continuos, como por ejemplo una conjugación de un operador elíptico de Schrödinger y obtener otras bases espectrales para la caracterización de la geometría subyacente a los datos.

En [14] se describen diferentes grafos Laplacianos y sus propiedades básicas y se discute sobre las ventajas y desventajas de los diferentes algoritmos de agrupación espectral utilizados hasta ese momento.

Además, a partir de la recolección de datos provenientes de redes de sensores situados espacialmente de forma irregular, varios autores [15], [16] desarrollaron diferentes algoritmos de regresión, filtrado, eliminación de ruido y esquemas de compresión a partir de la matriz Laplaciana asociada a un grafo. Otros trabajos se centraban en el procesado de señales sobre grafos indexando las muestras de las señales sobre un grafo no dirigido con pesos reales y no negativos. Con esta aproximación se adoptaba la matriz Laplaciana del grafo como un bloque básico en construcción a partir del cual desarrollar una teoría de procesado de señal sobre grafos no dirigidos.

En este camino hacia el GSP también han sido muy relevantes los trabajos realizados desde el procesado de imagen mediante el enfoque de la teoría de grafos. Así por ejemplo, varios autores [17] representaban las imágenes segmentándolas para generar un grafo a partir del cual desarrollar métodos de procesado, filtrado y representación [18].

Aplicaciones más actuales [23] han usado el GSP para análisis de datos recogidos por los sensores de la red eléctrica y detectar ataques de inyección de datos falsos como los de lectura de contadores o errores de lectura en general.

#### **1.4. Procesado de señal sobre grafos**

Los trabajos anteriores sentaron las bases para la aparición del concepto de GSP. Unos de los artículos que sirve como base de este campo emergente es [19] en donde se hace un repaso de los conceptos desarrollados hasta el momento, los retos a encarar y las aplicaciones más relevantes. En este artículo se explica que el GSP aparece a partir de la unión de los conceptos de procesado de señal y el análisis espectral de grafos con el análisis armónico computacional. Además se revisan los métodos para generalizar operaciones fundamentales como el filtrado o el submuestreo. También se hace un repaso a las transformaciones multi-escala que se han ido proponiendo para extraer información de grafos con datos de grandes dimensiones.

La representación de datos a través de grafos ha sido muy útil para describir la estructura geométrica en numerosas aplicaciones actuales como las relacionadas con medicina, energía, transporte, redes de sensores o tratamiento de datos a gran escala en general. En todas estas aplicaciones en las que los datos tienen una componente temporal, el GSP ha proporcionado nuevos enfoques para tratar problemas

de campos muy distintos dentro de la ingeniería o la ciencia. El artículo mencionado anteriormente hace un repaso de los conceptos a tener en cuenta para el análisis de los grafos asociados a este tipo de problemas. Por ejemplo, explica la similitud entre la transformada de Fourier clásica y su equivalencia a través de los autovectores de la matriz Laplaciana asociada al grafo.

El artículo [20] propone una extensión de la teoría del procesado de señal sobre grafos a partir de aplicaciones de clasificación, predicción lineal y compresión de datos. Para ello se propone nueva teoría de procesado de señal discreta (DSP, siglas en inglés de *Discrete Signal Processing*) en la que las relaciones de dependencia entre los datos son representadas mediante grafos arbitrarios. De este modo se extienden las estructuras y los conceptos fundamentales del DSP viendo los datos como señales indexadas en nodos de un grafo. En este artículo se demuestra que el DSP sobre grafos es una extensión de la teoría clásica de series temporales y que los conceptos de DSP y sus estructuras pueden obtenerse utilizando un grafo como series temporales discretas. También se muestra la eficiencia de esta aproximación en conjuntos de datos de diferente naturaleza.

En [21] se realiza un análisis frecuencial desde el enfoque de DSP sobre grafos. Para ello se estudian los diferentes filtrados frecuenciales y se plantean distintos métodos para el diseño de filtros sobre grafos y se ilustra esta aproximación con dos aplicaciones principales: la detección de fallos en sensores y la clasificación de datos. Así en [24] se muestra una guía del uso del GSP para aplicación de filtros paso bajo en estructuras grandes para obtener información subyacente en aplicaciones financieras, de redes sociales o de energía para detectar entre otras cosas anomalías.

En [1] se recopilan todos los avances realizados en GSP hasta 2018 exponiendo los conceptos clave, las diferentes áreas de aplicación y dando una perspectiva histórica de los trabajos realizados hasta el momento. A modo de resumen se puede decir que a partir del estudio de señales temporales discretas y del rol que juega el operador de traslación se pudo extender el DSP indexando las muestras de las señales sobre nodos de un grafo. A partir de esta extensión se pudo estudiar el modelo de la señal, sus representaciones, así como todo tipo de sistemas de procesado de señales (filtros), las transformaciones como la transformada Z o la transformada de Fourier, el muestro de señales y diferentes conceptos concretos del procesado de señal clásico pero con este nuevo enfoque a través de la teoría de grafos.

En los últimos años [25] se han continuado exponiendo las bases y los avances del GSP en general y en campos en particular como el aprendizaje automático [26] o redes neuronales [27] aplicadas sobre datos biológicos o eliminación de ruido [28].

En línea con lo comentado anteriormente en relación con el ASP, para poder extender el concepto de operador de traslación del DSP al GSP se reescribieron las señales como vectores de números reales o complejos y de este modo se pudo representar un filtro de respuesta al impulso  $h$  como una matriz  $H$ , y en general un sistema de procesado de señal como operaciones algebraicas. Así por ejemplo, se puede considerar una operación de filtrado como una multiplicación matriz-vector. Por ello la matriz de adyacencia del grafo asociado juega un papel muy importante y con el estudio de esta matriz aparecieron diferentes propuestas utilizando esta misma matriz o la matriz Laplaciana asociada.

Otro de los conceptos relevantes del GSP fue la representación frecuencial de las señales indexadas sobre un grafo. Mediante la definición de filtros como matri-

ces, se mostró que sus autovectores son una equivalencia a las señales propias de un filtro convencional en DSP. Para ello se demostró que los autovectores de la matriz de adyacencia son las componentes espectrales del grafo y que sus autovalores son las frecuencias del grafo. La interpretación de las frecuencias del grafo también se estudió en profundidad para determinar que las frecuencias están directamente relacionadas con el grado de variación de las componentes espectrales del grafo, por ejemplo, las bajas frecuencias corresponden a las componentes espectrales con menor variación. Por último, se propusieron representaciones frecuenciales basadas en la matriz Laplaciana ya que por sus características (semidefinida positiva), la interpretación de las frecuencias equivalentes en DSP quedaba más clara dependiendo del tipo de grafo elegido. La mejor representación frecuencial dependerá de la aplicación del grafo asociado seleccionado. Esto último es todavía un tema de investigación abierto.

### **1.5. Hipótesis y objetivos de la tesis**

En esta tesis se realizan nuevas aportaciones al GSP, que se van a ubicar brevemente en el contexto de todo lo anterior. Como ya se ha mencionado, el procesamiento de señales sobre grafos (GSP) es un área emergente de creciente interés, que permite extender el concepto de una señal definida en una rejilla uniforme de tiempo o espacio a rejillas y dominios más generales. Esencialmente, GSP abre nuevas posibilidades para la comunidad de procesamiento de señales, al establecer un puente entre el procesamiento de señal y el procesamiento de datos. Como se ha visto, muchos esfuerzos se han dirigido y se dirigen actualmente a definir conceptos, perspectivas y aplicaciones para demostrar que GSP tiene su propio interés con respecto a otras áreas relacionadas del procesamiento de datos. Así, por ejemplo, el concepto clásico de filtrado se ha extendido a GSP mediante una definición apro-

piada de operadores de translación y autofunciones del grafo. No menos importantes son los aspectos relacionados con el modelado estadístico, para poder extender el tratamiento estadístico de señales a señales sobre grafos. Así por ejemplo, conceptos básicos, como la estacionaridad, se han extendido de diferentes maneras, siendo todavía un tema abierto de investigación teórica. Son muchos los aspectos en los que aún se puede avanzar en GSP, tanto en nuevos algoritmos como en aplicaciones. Basta ver el listado de posibles líneas de investigación que aparecen habitualmente en los *workshops* específicos sobre DSP y que aquí se reproducen:

- *Sampling and recovery of graph signals*
- *Graph filter and filter bank design*
- *Uncertainty principles and other fundamental limits*
- *Graph signal transforms*
- *Graph filter identification*
- *Graph topology inference*
- *Statistical graph signal processing*
- *Signal processing on in high-order graphs*
- *Non-linear graph signal processing*
- *Prediction and learning in graphs*
- *Geometric deep learning (graph CNNs/RNNs)*
- *Reinforcement learning on graphs*
- *Representation learning on graphs*
- *Applications to image and video processing*
- *Applications to neuroscience and other medical fields*
- *Applications to economics and social networks*
- *Applications to infrastructure networks (e.g., communication, transportation, power networks).*

En concreto esta tesis hace aportaciones en dos aspectos del procesado de señal sobre grafos:

- Estimación de la matriz Laplaciana que caracteriza el grafo, asumiendo modelos probabilísticos no Gaussianos.

- Síntesis de señales a partir de modelos de grafos.

El primero de ellos es un problema clásico en procesado de señal sobre grafos, ya que la matriz Laplaciana caracteriza el grafo al definir todas las conexiones entre sus nodos. Los métodos de definición de Laplacianas propuestos hasta ahora lo hacen a partir de consideraciones específicas de la aplicación o asumiendo un modelo probabilístico de la variable aleatoria multivariante formada por todos los valores de la señal en cada nodo del grafo. Hasta ahora, los modelos probabilísticos asumidos son puramente Gaussianos, lo que puede no ajustarse suficientemente a la realidad. Por ello en esta tesis se aborda el problema de la no-Gaussianidad. Así, los capítulos 2 y 3 proponen métodos para estimar la matriz Laplaciana a partir respectivamente de modelos de mezclas gaussianas y de modelos de Análisis de Componentes Independientes (*Independent Component Analysis*, ICA) no gaussianas respectivamente. Se establece así un puente entre el enfoque clásico de tratamiento estadístico de señales y datos basados en modelos probabilísticos y el GSP basado en modelos de grafos.

En relación con la contribución relativa a la síntesis de señales, cabe destacar que este es un aspecto novedoso en cuanto al potencial del GSP. Asumimos que un modelo de grafo obtenido a partir de una señal temporal convencional incorpora información relevante sobre la misma. El objetivo es aprovechar dicha información para mejorar la síntesis de señales, que permitan llevar a cabo simulaciones de sistemas, entrenamiento de clasificadores,... con señales sintéticas que reproduzcan mejor el comportamiento de las señales reales. Así, en el Capítulo 4 se extienden los métodos clásicos de subrogación de señales a subrogación de señales sobre grafos, permitiendo así implementar métodos de síntesis de señales que incorporan modelos de grafos.



Como se irá comprobando, estas contribuciones encajan en varias líneas del anterior listado tales como: *Graph signal transforms*, *Graph topology inference*, *Statistical graph signal Processing*, *Non-linear graph signal Processing*, *Prediction and learning in graphs*, *Representation learning on graphs*. También se considerarán nuevas aplicaciones con bioseñales y en reconocimiento de signos.

## Referencias

- [1] Antonio Ortega, Pascal Frossard, Jelena Kovacevic, José M. F. Moura, and Pierre Vandergheynst, "Graph Signal Processing: Overview, Challenges, and Applications," *Proceedings of the IEEE*, pp. 808 - 828, 2018.
- [2] Markus Püschel and José M. F. Moura, "Algebraic Signal Processing Theory: Foundation and 1-D Time," *IEEE Transactions on Signal Processing*, pp. 3572-3585, 2008.
- [3] Markus Püschel and José M. F. Moura, "Algebraic Signal Processing Theory: 1-D Space," *IEEE Transactions on Signal Processing*, pp. 3586-3599, 2008.
- [4] Markus Püschel and José M. F. Moura, "Algebraic Signal Processing Theory," *arXiv*, pp. 1-67, 2006.
- [5] Markus Püschel and José M. F. Moura, "The algebraic approach to the discrete cosine and sine transforms and their fast algorithms," *SIAM Journal on Computing*, pp. 1280-1316, 2003.
- [6] M. Püschel and José M.F. Moura, "Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms for DCTs and DSTs," *IEEE Transactions on Signal Processing*, pp. 1502-1521, 2008.
- [7] Y. Voronenko and M. Püschel, "Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms for Real DFTs," *IEEE Transactions on Signal Processing*, pp. 205-222, 2009.
- [8] M. Püschel and M. Rötteler, "Algebraic Signal processing Theory: 2-D Spatial Hexagonal Lattice," *IEEE Transactions On Image Processing*, pp. 1506-1521, 2007.
- [9] A. Snadryhaila, J. Kovacevic, and M. Püschel, "Algebraic Signal Processing Theory: 1-D Nearest Neighbor Models," *IEEE Transactions on Signal Processing*, pp. 2247-2259, 2012.
- [10] P. Niyogi M. Belkin, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computing*, pp. 1373-1396, 2003.
- [11] P. Niyogi M. Belkin, "Using manifold structure for partially labelled classification," *NIPS*, pp. 953-960, 2002.

- [12] R. R. Coifman, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Nat. Acad. Sci. USA*, pp. 7426-7431, 2005.
- [13] R. R. Coifman, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods," *Proc. Nat. Acad. Sci.*, pp. 7432-7437, 2005.
- [14] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, pp. 395-416, 2007.
- [15] P. Bodik, R. Thibaux, M. Paskin, S. Madden C. Guestrin, "Distributed regression: an efficient framework for modeling sensor network data," *IPSN*, pp. 1-10, 2004.
- [16] B. Greenstein, D. Estrin, J. Heideman, R. Govindan D. Ganesan, "Multiresolution storage and search in sensor networks," *ACM Trans. Storage*, pp. 277-315, 2005.
- [17] R. Leahy Z. Wu, "An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1101-1113, 1993.
- [18] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process.*, pp. 106-128, 2013.
- [19] D. I. Shuman, K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs," *IEEE Signal Processing Magazine*, pp. 83-98, 2013.
- [20] A. Sandryhaila and J. M.F. Moura, "Discrete Signal Processing on Graphs," *IEEE Transactions on Signal Processing*, pp. 1644-1656, 2013.
- [21] A. Sandryhaila and J. M.F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," *IEEE Transactions on Signal Processing*, pp. 3042-3054, 2014.
- [22] Y. Voronenko and M. Püschel, "Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms for Real DFTs," *IEEE Transactions on Signal Processing*, pp. 205-222, 2009.
- [23] R. Ramakrishna and A. Scaglione, "Grid-Graph Signal Processing (Grid-GSP): A Graph Signal Processing Framework for the Power Grid." *IEEE Transactions on*

Signal Processing, pp. 2725-2739, 2021

- [24] R. Ramakrishna, H.Wai, A.Scaglione, “A User Guide to Low-Pass Graph Signal Processing and its Applications”, IEEE Signal Processing, 2020.
- [25] L. Stankovi, D. Mandic, M. Dakovi, I. Kisil, E. Sejdi, A. G. Constantinides, “Understanding the Basis of Graph Signal Processing via an Intuitive Example-Driven Approach”, IEEE Signal Processing Magazine, pp. 133-145, 2019.
- [26] X. Dong, D. Thanou, L. Toni, M. Bronstein, and P. Frossard “Graph signal processing for machine learning: A review and new perspectives”, IEEE Signal Processing Magazine, pp. 117-127, 2020.
- [27] R. Li, X. Yuan, M. Radfar, P. Marendy, W. Ni, T. J O'Brien, P. M. Casillas-Espinosa, “Graph Signal Processing, Graph Neural Network and Graph Learning on Biological Data: A Systematic Review”, IEEE Reviews in Biomedical Engineering, 2021.

# Capítulo 2

## Estimación de la matriz Laplaciana en Modelos de Mezclas Gaussianas

### 2.1. Introducción

Un grafo puede ser representado en la forma  $G \{V, E, \mathbf{A}\}$ , donde  $V$  es un conjunto de  $N$  vértices o nodos,  $E$  es el conjunto de conexiones que unen los nodos y  $\mathbf{A}$  es la denominada matriz de adyacencia. El elemento genérico  $a_{nm}$  es el peso (asumiendo que es real y no negativo) correspondiente a la conexión que une el nodo  $m$  con el nodo  $n$ . Un peso igual a cero significa que no hay conexión del nodo  $m$  al nodo  $n$ . Considerando grafos no dirigidos, se tiene que  $a_{nm} = a_{mn}$ . El grafo puede ser caracterizado de una forma clara mediante su matriz de adyacencia, así como, de manera más interesante, por la matriz Laplaciana [1] definida como  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , donde  $\mathbf{D}$  es una matriz diagonal cuyos elementos vienen definidos como

$d_{mm} = \sum_{m=1}^N a_{mm}$ . La matriz  $\mathbf{L}$  es preferible por sus atractivas propiedades espectrales [2]. Por tanto, en procesamiento de señal sobre grafos, resulta fundamental el correcto cálculo de la matriz Laplaciana. En contextos específicos, es posible establecer los pesos considerando relaciones naturales entre nodos, como por ejemplo, a partir de aproximaciones espaciales, pero en un marco estadístico general, es más conveniente idear métodos para la estimación de la matriz Laplaciana a partir de conjuntos de datos de entrenamiento. De este modo, muchos de los trabajos recientes en procesamiento de señal sobre grafos [3] [4] [5] estiman la denominada matriz de precisión y la utilizan como matriz Laplaciana del grafo. Estos trabajos están relacionados con técnicas estadísticas para la estimación general de matrices de covarianza o precisión [6] [7] [8] [9] [10] [11]. El fundamento de esta idea es que los elementos fuera de la diagonal de la matriz de precisión (correctamente normalizados) coinciden con las correlaciones parciales entre cada par de elementos del vector de observación aleatorio  $\mathbf{x} = [x_1 \dots x_N]^T$ . La correlación parcial entre dos variables aleatorias  $x_n$  y  $x_m$   $n \neq m$  se define como la correlación entre los residuos obtenidos restando  $x_n$  y  $x_m$  de sus estimaciones lineales del mínimo error cuadrático medio (LLMSE, (del inglés *Linear Least Mean-Square Error*)) obtenidas del resto de variables. Por lo tanto, la correlación parcial es una medida real del grado de enlace o relación lineal entre los nodos  $n$  y  $m$ , no afectada por la presencia del resto de variables. Además, el estimador lineal que está implícito en la definición de la correlación parcial, es óptimo si la función de densidad de probabilidad multivariante (pdf) subyacente  $p(\mathbf{x})$  es Gaussiana. Sin embargo, si esta suposición no es correcta, la influencia del resto de variables en  $x_n$  y  $x_m$  no se eliminará del todo y la correlación parcial no capturará adecuadamente la dependencia real entre el nodo  $n$

y el nodo  $m$ . Hasta donde se sabe, faltan trabajos que traten el caso no gaussiano en el contexto de la estimación de la matriz Laplaciana. En esta tesis, se propone una definición más general de la correlación parcial para tener en cuenta el caso no gaussiano. Para ello, se propone reemplazar el estimador LLMSE de  $x_n$  y  $x_m$  por sus respectivas medias condicionadas  $E[x_n/\mathbf{x}_{-nm}]$  y  $E[x_m/\mathbf{x}_{-nm}]$ , es decir, sus respectivas estimaciones LMSE (no lineales en general). Como estas últimas dependen de la  $p(\mathbf{x})$  específica, en este capítulo se considerará un modelo no gaussiano para  $p(\mathbf{x})$  que pueda adaptarse a múltiples escenarios: un modelo de mezclas gaussianas (GMM) multivariante. De manera similar, en el próximo capítulo se considerará un modelo de Análisis de Componentes Independientes (ICA).

## 2.2. Coeficiente de correlación parcial generalizado GPCC

Un GMM viene definido como

$$p(\mathbf{x}) = \sum_{k=1}^K P_k \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_k|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b}_k)^T \mathbf{C}_k^{-1}(\mathbf{x}-\mathbf{b}_k)} . \quad (1)$$

Donde  $P_k$ ,  $\mathbf{C}_k$  y  $\mathbf{b}_k$  son, respectivamente, la probabilidad a priori, la matriz de covarianza y el valor de la media de cada componente  $k$  de la mezcla gaussiana. GMM es el modelo paramétrico no gaussiano más común por su buen balance entre versatilidad y simplicidad. Este modelo ha sido considerado en diferentes escenarios de la estadística y del procesamiento de señal (ver por ejemplo [12] [13] [14] [15] [16] [17] [18] [19], por mencionar algunas aplicaciones recientes). En esta tesis se propone un enfoque en el que como se va a mostrar, se pueden obtener mejores estimaciones de la matriz Laplaciana asumiendo un modelo (GMM) más

sofisticado que el modelo gaussiano multivariante implícito en el uso de cualquier estimación de la matriz de precisión.

Definiendo  $\mathbf{x}_{-nm}$  como el vector obtenido a partir del vector  $\mathbf{x}$  después de descartar los elementos  $x_n$  y  $x_m$ ,  $L(x_n/\mathbf{x}_{-nm})$  y  $L(x_m/\mathbf{x}_{-nm})$  las correspondientes estimaciones LLMSE, y  $r_n = x_n - L(x_n/\mathbf{x}_{-nm})$  y  $r_m = x_m - L(x_m/\mathbf{x}_{-nm})$  los residuos correspondientes. El coeficiente de correlación parcial o PCC (del inglés *Partial Correlation Coefficient*) viene dado por

$$\rho_{nm} = \frac{E[(r_n - E[r_n])(r_m - E[r_m])]}{\sqrt{E[(r_n - E[r_n])^2]} \sqrt{E[(r_m - E[r_m])^2]}} \quad (2)$$

Una propiedad interesante de  $\rho_{nm}$  es que puede ser estimada sin realizar explícitamente el cálculo de los residuos. Esto es así porque pueden ser calculados como la inversa de la matriz de covarianza, denominada matriz de precisión, tal como

$$\rho_{nm} = -\frac{q_{nm}}{\sqrt{q_{nn}q_{mm}}}, \quad (3)$$

donde  $q_{nm}$  es el elemento  $nm$  de la matriz  $\mathbf{Q}$ , siendo  $\mathbf{Q}^{-1} = \mathbf{C} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T]$ . Por tanto, se propone la correlación parcial generalizada sustituyendo en (2) los residuos lineales por los residuos obtenidos después de extraer de  $x_n$  y  $x_m$  sus respectivas medias condicionadas  $E[x_n/\mathbf{x}_{-nm}]$  y  $E[x_m/\mathbf{x}_{-nm}]$ , es decir, sus respectivas estimaciones LMSE. Se llamarán  $r_n^G = x_n - E[x_n/\mathbf{x}_{-nm}]$  y  $r_m^G = x_m - E[x_m/\mathbf{x}_{-nm}]$  los residuos generalizados, por lo que el coeficiente de correlación parcial generalizado o GPCC (del inglés *Generalized*



Partical Correlation Coefficient) es definido como

$$\rho_{nm}^G = \frac{E\left[\left(r_n^G - E[r_n^G]\right)\left(r_m^G - E[r_m^G]\right)\right]}{\sqrt{E\left[\left(r_n^G - E[r_n^G]\right)^2\right]}\sqrt{E\left[\left(r_m^G - E[r_m^G]\right)^2\right]}} = \frac{E\left[r_n^G r_m^G\right]}{\sqrt{E\left[\left(r_n^G\right)^2\right]}\sqrt{E\left[\left(r_m^G\right)^2\right]}}, \quad (4)$$

donde se tiene en cuenta que la media condicionada es un estimador insesgado, por lo que  $E[r_n^G] = E[r_m^G] = 0$ . Esta definición de GPCC está relacionada con otros conceptos ya propuestos para medir dependencias no lineales entre dos variables aleatorias. Uno de estos conceptos es la función de dependencia local  $H(x_n, x_m)$  [20], que mide la dependencia entre dos variables aleatorias condicionadas a algunos valores (locales) de ambas. En el contexto de señales aleatorias sobre grafos, esto inspira la función  $H_{nm}(\mathbf{x}_{-nm})$  para medir la dependencia entre  $x_n$  y  $x_m$  en los puntos locales  $\mathbf{x}_{-nm}$  en la forma

$$H_{nm}(\mathbf{x}_{-nm}) = \frac{E\left[\left(x_n - E[x_n/\mathbf{x}_{-nm}]\right)\left(x_m - E[x_m/\mathbf{x}_{-nm}]\right)\right]}{\sqrt{E\left[\left(x_n - E[x_n/\mathbf{x}_{-nm}]\right)^2\right]}\sqrt{E\left[\left(x_m - E[x_m/\mathbf{x}_{-nm}]\right)^2\right]}}, \quad (5)$$

donde la esperanza se calcula con respecto a la pdf conjunta  $p(x_n, x_m)$  para obtener la medida local  $H_{nm}(\mathbf{x}_{-nm})$ . Otro concepto relacionado es la correlación condicionada [21],

$$\rho_{nm}(\mathbf{x}_{-nm}) = \frac{E\left[\left(x_n - E[x_n/\mathbf{x}_{-nm}]/\mathbf{x}_{-nm}\right)\left(x_m - E[x_m/\mathbf{x}_{-nm}]/\mathbf{x}_{-nm}\right)\right]}{\sqrt{E\left[\left(x_n - E[x_n/\mathbf{x}_{-nm}]\right)^2/\mathbf{x}_{-nm}\right]}\sqrt{E\left[\left(x_m - E[x_m/\mathbf{x}_{-nm}]\right)^2/\mathbf{x}_{-nm}\right]}}. \quad (6)$$

La diferencia con (5) es que las esperanzas son calculadas con respecto a la pdf condicionada conjunta  $p(x_n, x_m/\mathbf{x}_{-nm})$  en vez de  $p(x_n, x_m)$ . Sin embargo, es interesante una asignación única para cada conexión del grafo. Por ello, es necesaria

una integración de  $H_{nm}(\mathbf{x}_{-nm})$  o  $\rho_{nm}(\mathbf{x}_{-nm})$  con respecto a  $\mathbf{x}_{-nm}$ . Esto se hace implícitamente en  $\rho_{nm}^G$ , porque las esperanzas en (4) son definidas con respecto a todas las variables aleatorias involucradas, es decir, con respecto a  $p(x_n, x_m, \mathbf{x}_{-nm}) = p(\mathbf{x})$ .

### 2.3. Calculando la GPCC de una GMM

En el *Apéndice A* de este capítulo, se deriva la media condicionada  $E[x_n / \mathbf{x}_{-nm}]$  necesaria para calcular los residuos generalizados  $r_n^G$ . A partir de (A11), se puede expresar el residuo generalizado en la forma

$$r_n^G = x_n - E[x_n / \mathbf{x}_{-nm}] = x_n - \mathbf{p}_{-nm}^T \hat{\mathbf{x}}_{n-nm}, \quad (7)$$

donde  $\mathbf{p}_{-nm}^T = [P(1/\mathbf{x}_{-nm}) \dots P(K/\mathbf{x}_{-nm})]^T$  y el elemento  $k$  del vector  $\hat{\mathbf{x}}_{n-nm}$  viene dado por  $\hat{\mathbf{x}}_{n-nm}(k) = \mathbf{c}_{kn-nm}^T \mathbf{C}_{k-nm}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k-nm}) + b_{kn}$   $k = 1 \dots K$ . Se define el vector  $\mathbf{e}_{n-nm} = \mathbf{x}_n - \hat{\mathbf{x}}_{n-nm}$  cuyos elementos son los residuos individuales correspondientes a cada componente de GMM (habiendo definido  $\mathbf{x}_n = [x_n \dots x_n]^T = x_n \cdot \mathbf{1}$ ). Se puede expresar:

$$r_n^G = x_n - \mathbf{p}_{-nm}^T (\mathbf{x}_n - \mathbf{e}_{n-nm}) = \mathbf{p}_{-nm}^T \mathbf{e}_{n-nm}, \quad (8)$$

ya que  $\mathbf{p}_{-nm}^T \mathbf{x}_n = x_n \sum_{k=1}^K \mathbf{p}_{-nm}(k) = x_n$ . Ahora se calcula la correlación correspondiente con los residuales generalizados

$$E[r_n^G r_m^G] = E[(\mathbf{p}_{-nm}^T \mathbf{e}_{n-nm})(\mathbf{p}_{-nm}^T \mathbf{e}_{m-nm})] = E[\mathbf{p}_{-nm}^T \mathbf{e}_{n-nm} \mathbf{e}_{m-nm}^T \mathbf{p}_{-nm}] \quad (9)$$

Por lo que se puede expresar

$$\rho_{nm}^G = \frac{E[r_n^G r_m^G]}{\sqrt{E[(r_n^G)^2]} \sqrt{E[(r_m^G)^2]}} = \frac{E[\mathbf{p}_{-nm}^T \mathbf{e}_{n-nm} \mathbf{e}_{m-nm}^T \mathbf{p}_{-nm}]}{\sqrt{E[\mathbf{p}_{-nm}^T \mathbf{e}_{n-nm} \mathbf{e}_{n-nm}^T \mathbf{p}_{-nm}]} \sqrt{E[\mathbf{p}_{-nm}^T \mathbf{e}_{m-nm} \mathbf{e}_{m-nm}^T \mathbf{p}_{-nm}]}} . \quad (10)$$

Hay que observar que  $\mathbf{p}_{-nm}$  puede calcularse de forma

$$\mathbf{p}_{-nm}(k) = P(k/\mathbf{x}_{-nm}) = \frac{p(\mathbf{x}_{-nm}/k)P_k}{p(\mathbf{x}_{-nm})} = \frac{p(\mathbf{x}_{-nm}/k)P_k}{\sum_{k'=1}^K P(\mathbf{x}_{-nm}/k')P_{k'}} , \quad (11)$$

donde  $p(\mathbf{x}_{-nm}/k)$  se puede obtener a partir de la ecuación (A7) . Los estadísticos en (10) son funciones de  $x_n, x_m, \mathbf{x}_{-nm}$  y de los parámetros de GMM, por lo que se podría calcular la media de las muestras o una integración numérica para obtener las esperanzas necesarias para la estimación de GPCCs. Sin embargo, la carga computacional será mucho mayor que la necesaria para estimar las PCCs mediante la estimación de la matriz de precisión. A continuación se propone una solución práctica después de asumir algunas simplificaciones.

Se puede expresar (9)

$$E[r_n^G r_m^G] = E[(\mathbf{p}_{-nm}^T \mathbf{e}_{n-nm})(\mathbf{p}_{-nm}^T \mathbf{e}_{m-nm})] = E\left[\sum_{k=1}^K \sum_{k'=1}^K P(k/\mathbf{x}_{-nm})P(k'/\mathbf{x}_{-nm})\mathbf{e}_{n-nm}(k)\mathbf{e}_{m-nm}(k')\right] . \quad (12)$$

Observando que  $P(k/\mathbf{x}_{-nm}), P(k'/\mathbf{x}_{-nm})$  son funciones sólo de  $\mathbf{x}_{-nm}$  y que  $\mathbf{e}_{n-nm}(k), \mathbf{e}_{m-nm}(k')$  son errores de predicción correspondientes a las predicciones lineales de las variables de predicción  $\mathbf{x}_{-nm}$  . Por lo tanto  $P(k/\mathbf{x}_{-nm}), P(k'/\mathbf{x}_{-nm})$  no están correlacionadas con  $\mathbf{e}_{n-nm}(k), \mathbf{e}_{m-nm}(k')$  y se puede escribir:

$$E[r_n^G r_m^G] = \sum_{k=1}^K \sum_{k'=1}^K E[P(k/\mathbf{x}_{-nm})P(k'/\mathbf{x}_{-nm})]E[\mathbf{e}_{n-nm}(k)\mathbf{e}_{m-nm}(k')] . \quad (13)$$

Pero

$$\begin{aligned}
 E[P(k/\mathbf{x}_{-nm})P(k'/\mathbf{x}_{-nm})] &= E\left[\frac{p(\mathbf{x}_{-nm}/k)P_k}{p(\mathbf{x}_{-nm})} \frac{p(\mathbf{x}_{-nm}/k')P_{k'}}{p(\mathbf{x}_{-nm})}\right] = \\
 &= \int_{-\infty}^{\infty} \frac{p(\mathbf{x}_{-nm}/k)P_k}{p(\mathbf{x}_{-nm})} \frac{p(\mathbf{x}_{-nm}/k')P_{k'}}{p(\mathbf{x}_{-nm})} p(\mathbf{x}_{-nm}) d\mathbf{x}_{-nm} \quad . \\
 &= P_k P_{k'} \int_{-\infty}^{\infty} \frac{p(\mathbf{x}_{-nm}/k)p(\mathbf{x}_{-nm}/k')}{p(\mathbf{x}_{-nm})} d\mathbf{x}_{-nm}
 \end{aligned} \tag{14}$$

La integral en (14) puede simplificarse si se asume que los componentes de GMM están “suficientemente separados” de modo que  $p(\mathbf{x}_{-nm}/k)p(\mathbf{x}_{-nm}/k') = 0 \quad k \neq k'$ . Este es el caso más interesante desde la perspectiva de mejora del rendimiento de PCC como medida de conectividad entre pares, ya que cuanto más separados estén los componentes, más se alejará de la gaussianidad. Esta separación puede definirse de diferentes maneras, utilizando diversas medidas de distancia entre dos pdfs normales (ver por ejemplo [22] y las referencias que aparecen). Sin embargo, en el contexto que se está tratando, “suficientemente separados” significa que  $p(\mathbf{x}_{-nm}/k)p(\mathbf{x}_{-nm}/k') = 0$  para todo  $\mathbf{x}_{-nm}$  y  $k \neq k'$ , por lo que en el *Apéndice B* de este capítulo, se propone un procedimiento práctico para medir la separación de los componentes de la mezcla en términos de los parámetros GMM definiendo un índice “*ad hoc*”  $0 \leq I_s \leq 1$  donde 1 significa una separación perfecta. Por otra parte, para  $k = k'$ , la suposición de “separación suficiente” conduce a  $\int_{-\infty}^{\infty} \frac{p(\mathbf{x}_{-nm}/k)p(\mathbf{x}_{-nm}/k)}{p(\mathbf{x}_{-nm})} d\mathbf{x}_{-nm} \simeq \int_{-\infty}^{\infty} \frac{p(\mathbf{x}_{-nm}/k)p(\mathbf{x}_{-nm}/k)}{P_k p(\mathbf{x}_{-nm}/k)} d\mathbf{x}_{-nm} = \frac{1}{P_k}$ , que puede aproximarse como

$$E[P(k/\mathbf{x}_{-nm})P(k'/\mathbf{x}_{-nm})] \simeq \begin{cases} P_k & k = k' \\ 0 & k \neq k' \end{cases} \quad . \tag{15}$$

Así considerando (15) en (13)

$$E[r_n^G r_m^G] \approx \sum_{k=1}^K P_k E[\mathbf{e}_{n-nm}(k) \mathbf{e}_{m-nm}(k)] = \tilde{E}[r_n^G r_m^G] \quad , \quad (16)$$

lo que significa que la correlación residual total se aproxima por la suma ponderada de la correlación residual individual debida a cada componente GMM, despreciando los términos cruzados entre diferentes componentes. Calculando ahora  $E[\mathbf{e}_{n-nm}(k) \mathbf{e}_{m-nm}(k)]$ , se observa que estos residuos corresponden a la suposición de que cada componente  $k$  de GMM está presente, por lo que

$$\begin{aligned} E[\mathbf{e}_{n-nm}(k) \mathbf{e}_{m-nm}(k)] &= E[(x_n - \hat{\mathbf{x}}_{n-nm}(k))(x_m - \hat{\mathbf{x}}_{m-nm}(k))] = \\ &= R_k(n, m) - E[x_n \hat{\mathbf{x}}_{m-nm}(k)] - E[x_m \hat{\mathbf{x}}_{n-nm}(k)] + E[\hat{\mathbf{x}}_{n-nm}(k) \hat{\mathbf{x}}_{m-nm}(k)] \quad . \end{aligned} \quad (17)$$

Donde  $R_k(n, m) = C_k(n, m) + b_{kn} b_{km}$  son los elementos de la matriz de correlación  $\mathbf{R}_k = \mathbf{C}_k + \mathbf{b}_k \mathbf{b}_k^T$  correspondientes al elemento  $k$  de GMM.

Pero

$$\begin{aligned} E[x_n \hat{\mathbf{x}}_{m-nm}(k)] &= E\left[x_n \left(\mathbf{c}_{km-nm}^T \mathbf{C}_{k-nm}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k-nm}) + b_{km}\right)\right] = \\ &= \mathbf{c}_{km-nm}^T \mathbf{C}_{k-nm}^{-1} (\mathbf{r}_{kn-nm} - \mathbf{b}_{k-nm} b_{kn}) + b_{kn} b_{km} = \mathbf{c}_{km-nm}^T \mathbf{C}_{k-nm}^{-1} \mathbf{c}_{kn-nm} + b_{kn} b_{km} \\ E[x_m \hat{\mathbf{x}}_{n-nm}(k)] &= \mathbf{c}_{kn-nm}^T \mathbf{C}_{k-nm}^{-1} \mathbf{c}_{km-nm} + b_{km} b_{kn} \quad , \end{aligned} \quad (18)$$

donde el vector  $\mathbf{r}_{kn-nm}$  está formado por los elementos de la  $n$ -ésima columna de la matriz  $\mathbf{R}_k$  eliminando las filas  $n$  y  $m$ . Y

$$\begin{aligned}
 E[\hat{\mathbf{x}}_{n_{-nm}}(k)\hat{\mathbf{x}}_{m_{-nm}}(k)] &= E\left[\left(\mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k_{-nm}}) + b_{kn}\right)\left(\mathbf{c}_{km_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k_{-nm}}) + b_{km}\right)\right] = \\
 &= \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} E\left[(\mathbf{x}_{-nm} - \mathbf{b}_{k_{-nm}})(\mathbf{x}_{-nm} - \mathbf{b}_{k_{-nm}})^T\right] \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}} \\
 &+ \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} (E[\mathbf{x}_{-nm}] - \mathbf{b}_{k_{-nm}}) b_{km} + \mathbf{c}_{km_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} (E[\mathbf{x}_{-nm}] - \mathbf{b}_{k_{-nm}}) b_{kn} + b_{kn} b_{km} = \\
 &= \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}} + b_{kn} b_{km}
 \end{aligned} \tag{19}$$

Así que finalmente

$$\begin{aligned}
 E[\mathbf{e}_{n_{-nm}}(k)\mathbf{e}_{m_{-nm}}(k)] &= C_k(n, m) + b_{kn} b_{km} - \mathbf{c}_{km_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{kn_{-nm}} - b_{kn} b_{km} - \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}} - b_{km} b_{kn} \\
 &+ \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}} + b_{kn} b_{km} = C_k(n, m) - \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}}
 \end{aligned} \tag{20}$$

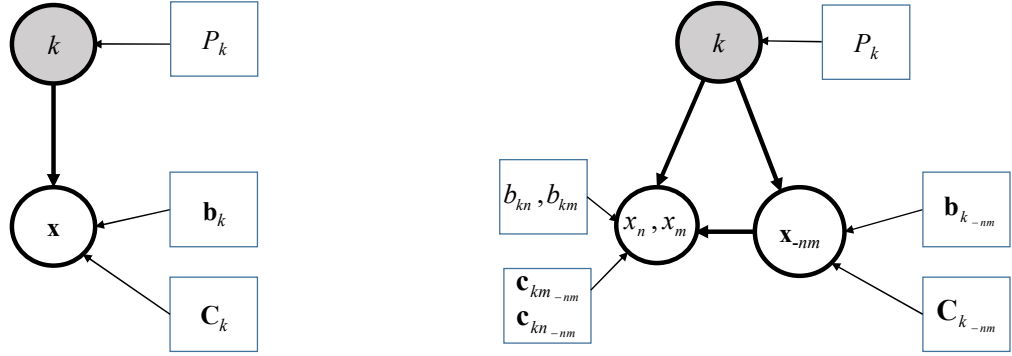
Como conclusión, dados los parámetros GMM, se puede calcular  $\tilde{E}[r_n^G r_m^G]$  a partir de las ecuaciones (16)-(20). Así que se puede definir un GPCC aproximado (aGPCC) como

$$\rho_{nm}^{aG} = \frac{\tilde{E}[r_n^G r_m^G]}{\sqrt{\tilde{E}[(r_n^G)^2]} \sqrt{\tilde{E}[(r_m^G)^2]}} = \frac{\sum_{k=1}^K P_k (C_k(n, m) - \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}})}{\sqrt{\sum_{k=1}^K P_k (C_k(n, n) - \mathbf{c}_{kn_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{kn_{-nm}})} \sqrt{\sum_{k=1}^K P_k (C_k(m, m) - \mathbf{c}_{km_{-nm}}^T \mathbf{C}_{k_{-nm}}^{-1} \mathbf{c}_{km_{-nm}})}} \tag{21}$$

La Figura 2.1 muestra el Grafo Acíclico Dirigido o DAG por sus siglas del inglés *Directed Acyclic Graph* de una Red Bayesiana (BN) correspondiente a un GMM. Este tipo de diagramas ayudan a comprender el rol y las dependencias de las diferentes variables y parámetros, y a definir algoritmos de aprendizaje (ver por ejemplo [23]).

a)  $p(\mathbf{x}, k) = p(\mathbf{x}/k)P_k$

b)  $p(\mathbf{x}, k) = p(x_1, x_2/\mathbf{x}_{-nm}, k) p(\mathbf{x}_{-nm}/k) P_k$



$$\rho_{nm}^G = f(P_k, b_{kn}, b_{km}, \mathbf{c}_{kn-nm}, \mathbf{c}_{km-nm}, \mathbf{b}_{k-nm}, \mathbf{C}_{k-nm}, \mathbf{x}_{-nm}) \quad \text{eq. (7)-(11), (A7)}$$

$$\rho_{nm}^{aG} = f(P_k, \mathbf{c}_{kn-nm}, \mathbf{c}_{km-nm}, \mathbf{C}_k, \mathbf{C}_{k-nm}) \quad \text{eq. (21)}$$

**Figura 2.1.** DAG de la BN asociada al GMM. Los círculos sombreados corresponden a las variables latentes, los círculos blancos a las variables observadas, los rectángulos corresponden a los parámetros que definen las probabilidades a priori o las pdfs condicionada. a) GMM compacto b) GMM dividido implícito en la estimación de las correlaciones parciales.

Así, en la Figura 2.1 se representa el modelo compacto que factoriza la pdf conjunta de la observación ( $\mathbf{x}$ ) y las ( $k$ ) variables latentes, como el producto de la probabilidad a priori por la pdf condicionada de las variables observadas  $p(\mathbf{x}, k) = p(\mathbf{x}/k)P_k$ . En la Figura 2.1a (en los rectángulos), también se incluyen los parámetros que deben ser estimados para la caracterización completa de la BN. En la Figura 2.1b se ha dividido el grafo compacto para hacer explícitas las dependencias involucradas en los cálculos de las correlaciones parciales. Aquí las variables  $x_n, x_m$  se han separado para que la pdf conjunta sea factorizada en la forma  $p(\mathbf{x}, k) = p(x_1, x_2/\mathbf{x}_{-nm}, k) p(\mathbf{x}_{-nm}/k) P_k$ . Los parámetros indicados en las cajas de

la Figura 2.1b han sido obtenidos a partir de los de la Figura 2.1a. Además, se ha incluido en la parte inferior de la Figura 2.1, los valores del DAG involucrados respectivamente en el cálculo de GPCC y aGPCC.

A continuación, con el fin de facilitar la comprensión y la implementación, se incluye la descripción de los algoritmos para la estimación de GPCCs y aGPCCs en pseudocódigo. En estos algoritmos se puede apreciar la gran simplificación de la estimación aGPCC, ya que no requiere la estimación de las muestras (el bucle  $l=1 \dots L$  deja de ser necesario).

<p><b>Algoritmo 2.1:</b> <i>Cálculo de la GPCC</i></p> <p>1: <b>Entrada:</b> Conjunto de entrenamiento <math>\{\mathbf{x}_{(l)}\} \quad l=1 \dots L</math></p> <p>2: <b>Estimar parámetros GMM</b> <math>\mathbf{b}_k, \mathbf{C}_k, P_k, k=1 \dots K</math> a partir del conjunto de entrenamiento (algoritmo EM)</p> <p>3: <b>for</b> <math>n=1 \dots N-1</math></p> <p>4: <b>for</b> <math>m=n+1 \dots N</math></p> <p>5: <b>for</b> <math>l=1 \dots L</math></p> <p>6: <b>Calcular</b> <math>\mathbf{p}_{(l) \rightarrow m}^r = \left[ P\left(1/\mathbf{x}_{(l) \rightarrow m}\right) \dots P\left(K/\mathbf{x}_{(l) \rightarrow m}\right) \right]^r</math>, ec. (11) y (A7)</p> <p>7: <b>Calcular</b> <math>\hat{\mathbf{X}}_{(l)n \rightarrow m}</math> y <math>\hat{\mathbf{X}}_{(l)m \rightarrow m}</math>, ec. (A11)</p> <p>8: <b>Calcular</b> <math>r_{(l)n}^G</math> y <math>r_{(l)m}^G</math>, ec. (7)</p> <p>9: <b>end for</b></p> <p>10: <b>Estimar</b> <math>\rho_{nm}^G</math> (estimaciones de muestras en ec. (10))</p> <p>11: <b>Calcular</b> <math>\rho_{mn}^G = \rho_{nm}^G</math></p> <p>12: <b>end for</b></p> <p>13: <b>end for</b></p> <p>14: <b>Salida:</b> <math>\rho_{nm}^G \quad n=1 \dots N \quad m=1 \dots M</math></p>	<p><b>Algoritmo 2.2:</b> <i>Cálculo de la aGPCC</i></p> <p>1: <b>Entrada:</b> Conjunto de entrenamiento <math>\{\mathbf{x}_{(l)}\} \quad l=1 \dots L</math></p> <p>2: <b>Estimar parámetros GMM</b> <math>\mathbf{b}_k, \mathbf{C}_k, P_k, k=1 \dots K</math> a partir del conjunto de entrenamiento (algoritmo EM)</p> <p>3: <b>for</b> <math>n=1 \dots N-1</math></p> <p>4: <b>for</b> <math>m=n+1 \dots N</math></p> <p>5: <b>Estimar</b> <math>\rho_{nm}^G</math>, ec. (21)</p> <p>6: <b>Calcular</b> <math>\rho_{mn}^G = \rho_{nm}^G</math></p> <p>7: <b>end for</b></p> <p>8: <b>end for</b></p> <p>9: <b>Salida:</b> <math>\rho_{nm}^G \quad n=1 \dots N \quad m=1 \dots M</math></p>
--	--

Finalmente, la Tabla 2.1 indica la complejidad computacional de los diferentes métodos en términos de dimensión de la observación  $N$ , el número de componentes gaussianos  $K$  y el tamaño del conjunto de datos de entrenamiento  $L$ .



Sólo se han tenido en cuenta los órdenes superiores, que son esencialmente los producidos por el cálculo de las inversas necesarias. Así pues, PCC tiene un coste computacional  $O(N^3)$  debido al cálculo de la inversa de la matriz de covarianza. Por otro lado, aGPCC tiene un coste  $O(N^5)$  dado que los cálculos deben realizarse para cada par  $nm$ . Para un  $N$  dado, el coste computacional de aGPCC depende linealmente de  $K$ , ya que los cálculos se deben realizar para cada componente GMM, mientras que  $L$  no tiene influencia. Finalmente GPCC tiene un coste computacional  $O(N^5)$  ya que también son necesarios los cálculos para cada par  $nm$ . Para un  $N$  dado, la complejidad computacional de aGPCC depende linealmente de  $K$ , pero también de  $L$ , dado que los cálculos se deben realizar para cada miembro del conjunto de entrenamiento. Esta dependencia de  $L$  es la principal causa del incremento del coste computacional de GPCC respecto a aGPCC.

	$O_N$	$O_K$	$O_L$
PCC	$O(N^3)$	—	—
aGPCC	$O(N^5)$	$O(K)$	—
GPCC	$O(N^5)$	$O(K)$	$O(L)$

**Tabla 2.1.** Complejidad computacional en términos de  $N$ ,  $K$  y  $L$ .

## 2.4. Experimentos

### 2.4.1. Simulaciones

Considerando que  $p(\mathbf{x})$  es una mezcla de dos componentes gaussianas con matrices de covarianza y probabilidades a priori idénticas, pero valores medios constantes y opuestos, es decir,

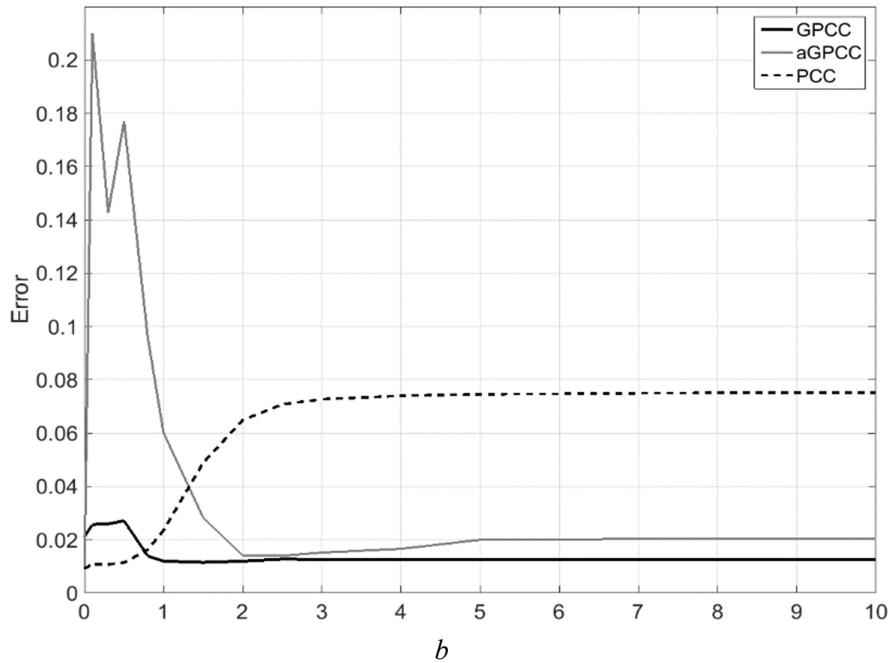
$$p(\mathbf{x}) = 0.5 \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \left[ e^{-\frac{1}{2}(\mathbf{x}-b\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x}-b\mathbf{1})} + e^{-\frac{1}{2}(\mathbf{x}+b\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x}+b\mathbf{1})} \right], \quad (22)$$

donde  $\mathbf{1} = [1 \dots 1]^T$ . Primero se calcula la matriz Laplaciana en la forma

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad w_{nm} = |\rho_{nm}^G| \quad d_{nn} = \sum_{m=1}^N |\rho_{nm}^G|, \quad (23)$$

donde  $\rho_{nm}^G$  se calcula con (10) considerando los valores reales de los parámetros de los dos componentes GMM. Las esperanzas en (10) se han calculado utilizando integración numérica. Con ello, la matriz obtenida se considera Laplaciana real. En la Figura 2.2 y Figura 2.3, se muestra el error normalizado (definido como el cociente de la norma de Frobenius de la diferencia entre la matriz Laplaciana real y la estimada, dividido por la norma de Frobenius de la Laplaciana real) correspondiente a reemplazar respectivamente en (23) la  $\rho_{nm}^G$  real por la  $\hat{\rho}_{nm}$  (PCC, ec. (3) con muestras estimadas de la matriz de precisión),  $\hat{\rho}_{nm}^{aG}$  (aGPCC, ec. (21), con los parámetros de GMM estimados mediante EM) y  $\hat{\rho}_{nm}^G$  (GPCC, ec. (10) con los parámetros GMM estimados mediante EM y las estimaciones muestrales de las esperanzas). Así pues, la Figura 2.2 muestra el error de estimación normalizado para varias separaciones ( $b$  en (22)) de las dos componentes GMM. La dimensión

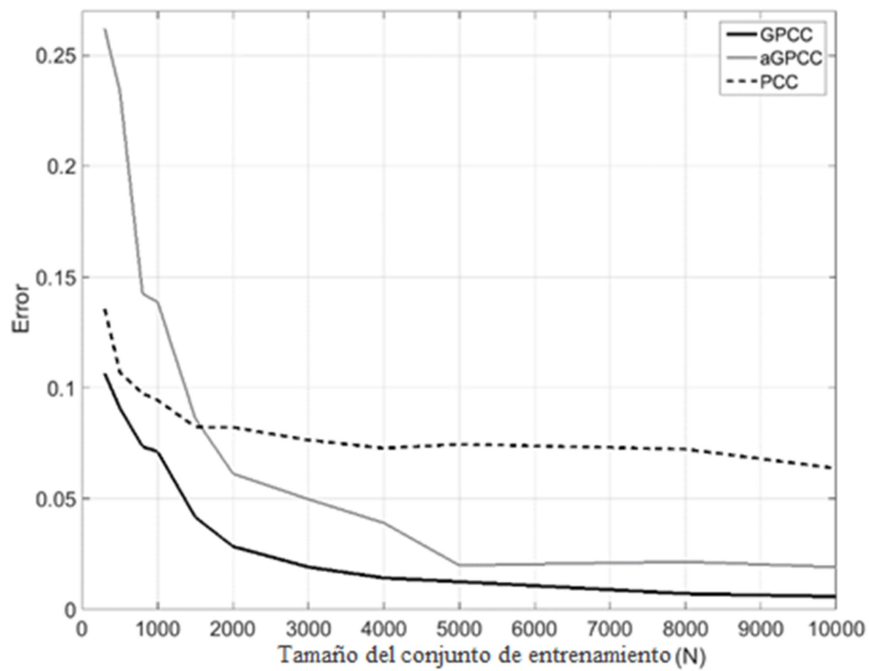
de observación fue  $N=20$ , con un tamaño del conjunto de entrenamiento de 5000, y la covarianza en (22)  $C(n,n)=1$   $C(n,m)=0.5$   $n \neq m$ .



**Figura 2.2.** Error en la estimación de la matriz Laplaciana para varias separaciones de los dos componentes de la GMM,  $N=20$ , 5000 de tamaño del conjunto de entrenamiento.

Se puede observar que para  $b=0$  todos los métodos tienen un error similar, en realidad este caso corresponde al caso gaussiano. Cuando  $b$  se incrementa GPCC es la mejor aproximación a la Laplaciana real. Además, como era de esperar, para un tamaño pequeño de  $b$ , PCC produce menos error que aGPCC, pero después de  $b=1,5$ , aGPCC supera a PCC y casi equivale a GPCC para  $b \geq 2$ . Esto es debido a la asunción de “buena separación” considerada en la derivación de aGPCC. En la Figura 2.3 se muestra la variación del error de estimación con el incremento del

tamaño del conjunto de entrenamiento. La separación entre componentes del GMM fue  $b=5$ , el tamaño de la observación y las matrices de covarianza fueron las mismas que en el caso anterior.



**Figura 2.3.** Error de estimación de la matriz Laplaciana según la variación del tamaño del conjunto de entrenamiento,  $N=20$ , 2 componentes GMM,  $b=5$ .

Como se puede observar, GPCC produce siempre la mejor aproximación a la Laplaciana verdadera, así como para tamaños de conjuntos de entrenamientos superiores a 1600 aGPCC supera a PCC y se mantiene razonablemente cerca de GPCC. Una conclusión general de ambas figuras es que a medida que el rendimiento de PCC se aleja de GPCC, aGPCC se aproxima a GPCC. Por lo tanto, hasta cierto punto, ambos métodos pueden considerarse métodos complementarios para la estimación eficiente de la correlación parcial. En la Figura 2.4 se muestra el error

normalizado al incrementar el número  $K$  de componentes GMM (desde 2 hasta 10). El tamaño del conjunto de entrenamiento fue de 5000 en todos los casos; por eso el error se incrementa con el número de parámetros a estimar. Al igual que en los experimentos previos se ha configurado  $N=20$  y  $C_k(n,n)=1$   $C_k(n,m)=0.5$   $n \neq m$  para todo  $k$ . Los vectores de medias correspondientes a GMM se han generado con valores seleccionados uniformemente entre  $\mathbf{b}_1 = -5(K-1)\mathbf{1}$  y  $\mathbf{b}_K = 5(K-1)\mathbf{1}$  de modo que la separación entre componentes es siempre múltiple de 5. Se observa de nuevo la mejora de comportamiento de aGPCC con respecto a PCC para cada valor de  $k$ .

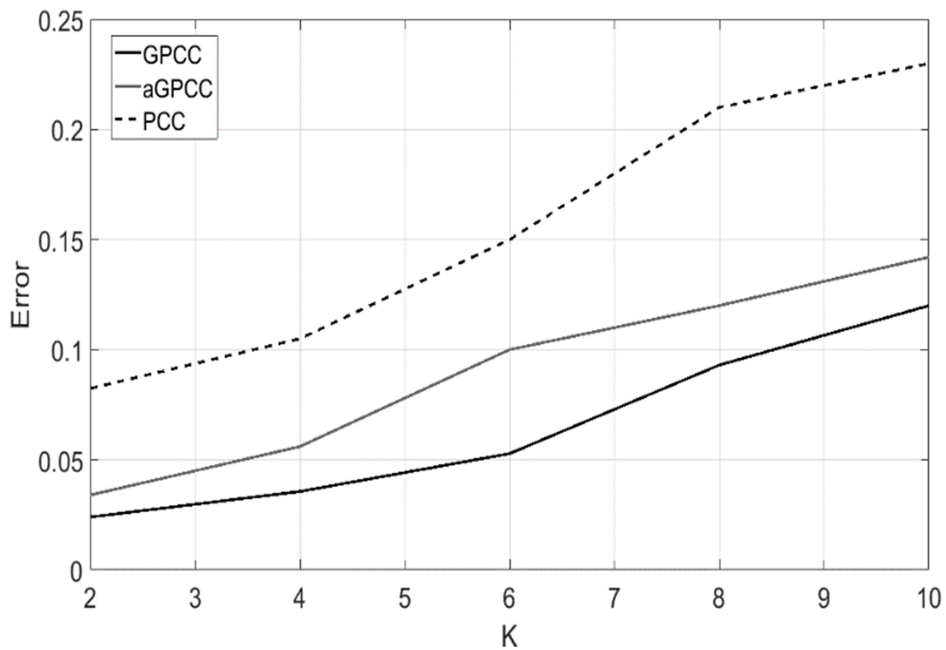
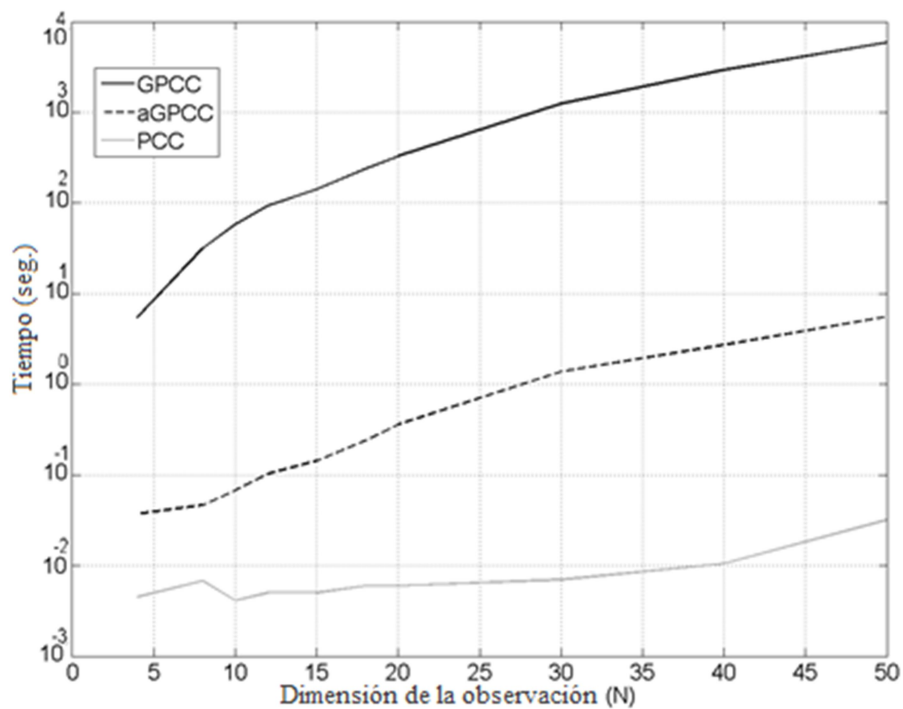


Figura 2.4. Error de estimación de la matriz Laplaciana para diferentes números de componentes  $K$  de GMM, tamaño del conjunto de entrenamiento de 5000,  $N=20$ .

Finalmente, la Figura 2.5 muestra el tiempo computacional necesario para estimar la matriz Laplaciana a través de los distintos métodos variando la dimensión  $N$  del vector de observación, con 2 componentes GMM,  $b=5$ , tamaño del conjunto de entrenamiento  $L=1000$  y las mismas matrices de covarianza utilizadas anteriormente. En esta figura se ilustra la reducción drástica (alrededor de 3 órdenes de magnitud) del método aGPCC respecto a GPCC. Por otra parte, aGPCC tiene entorno a 2 órdenes de magnitud sobre PCC, pero se puede considerar no relevante dado que el tiempo necesario es razonablemente pequeño:  $\approx 0.04$  segundos para  $N=4$  hasta  $\approx 5$  segundos para  $N=50$ .



**Figura 2.5.** Tiempo computacional necesario para los diferentes métodos de estimación de la matriz Laplaciana variando la dimensión  $N$  del vector de observación, 2 componentes GMM,  $b=5$ , tamaño del conjunto de entrenamiento de 1000.

### **2.4.2. Un experimento con datos reales**

Se han aplicado los diferentes métodos de estimación de la matriz Laplaciana a un experimento con datos reales. El marco del experimento ha sido la monitorización del sueño de pacientes con posibles desórdenes como apnea o epilepsia [24]. Estos desórdenes están caracterizados por la aparición regular de estados anormales del sueño, generalmente denominados como “arousal”. El paciente se ha monitorizado mediante 19 canales de EEG. La señal de cada canal se ha segmentado en intervalos de 1-3 segundos y cada característica es calculada promediando cada intervalo en ciclos de 25 segundos. Para cada ciclo se ha asociado un vector de observación  $\mathbf{x}$  con la característica extraída a partir de todos los canales (lo mismo para todos ellos), con  $N = 19$ . Establecer un modelo de grafos que relacione las 19 componentes de dicho vector resulta de interés para determinar una medida global de conectividad cerebral [22]. Por ello centraremos el experimento en la estimación de una matriz Laplaciana asociable al conjunto de vectores de características medidos. En total se han considerado 1000 ciclos, por lo que dispondremos de 1000 vectores de características para estimar dicha matriz.

En términos generales, los vectores de características obtenidos corresponden a dos clases diferentes: sueño normal y sueño anormal. En una primera aproximación, la pdf característica global puede considerarse un GMM de dos componentes, cuyos parámetros pueden estimarse mediante el algoritmo EM (*Expectation-Maximization*). Los resultados se muestran en la Tabla 2.2 para 6 pacientes diferentes con distintos niveles de desórdenes. Se han considerado por separado dos características diferentes. La primera es la “amplitud” que corresponde con la máxima amplitud en el intervalo correspondiente. La segunda es el índice “alfa-slow-index” (ASI) que es el ratio de potencia en la banda alfa (8.0–11 Hz) combinada con la potencia de las bandas delta (0.5–3.5 Hz) y theta (3.5–8.0 Hz). Para cada paciente y cada característica se ha calculado el error normalizado (definido previamente) de

la matriz Laplaciana estimada mediante aGPCC y PCC con respecto a las estimaciones obtenidas con GPCC. También se ha calculado el índice de separación  $I_s$  como se ha definido en el Apéndice B. En la Tabla 2.2 se puede ver que el método aGPCC es siempre mejor que el PCC cuando se utiliza la característica de amplitud. En el caso  $I_s=1$ , se mantiene la suposición de “separación suficiente”. Sin embargo,  $I_s$  disminuye al usar la característica ASI, esto produce que en la mayoría de pacientes, la estimación de la matriz Laplaciana sea mejor mediante PCC que mediante aGPCC.

Paciente	Amplitud			ASI		
	Error (aGPCC)	Error (PCC)	$I_s$	Error (aGPCC)	Error (PCC)	$I_s$
P1	<b>0,03</b>	0,19	1,00	<b>0,02</b>	0,04	0,67
P2	<b>0,24</b>	0,25	1,00	<b>0,01</b>	0,02	0,62
P3	<b>0,10</b>	0,11	1,00	0,11	<b>0,04</b>	0,71
P4	<b>0,14</b>	0,15	1,00	0,10	<b>0,04</b>	0,70
P5	<b>0,25</b>	0,26	1,00	0,12	<b>0,05</b>	0,71
P6	<b>0,17</b>	0,20	1,00	0,06	<b>0,04</b>	0,64

Tabla 2.2. Errores normalizados obtenidos en el experimento con datos reales

## 2.5. Conclusiones

En este capítulo se han propuesto nuevos métodos para estimar la matriz Laplaciana asociada a grafos no dirigidos. Al suponer que la pdf subyacente es una mezcla de gaussianas, los nuevos métodos capturan la conectividad del grafo de mane-



ra más completa que a través de la matriz de precisión. Los elementos normalizados de esta matriz son el PCC, una medida de dependencia lineal condicionada por pares. Se ha propuesto calcular el GPCC, capturando así las dependencias no lineales condicionadas implícitas en cualquier modelo no gaussiano. El GPCC puede ser estimado a partir de la estimación de los parámetros GMM y estimando las muestras o haciendo integración numérica. Para superar la alta carga computacional de este método se ha propuesto aGPCC, que es una solución aproximada bajo el supuesto de componentes de GMM “suficientemente separadas”.

Dado que el GMM multivariante es capaz de ajustarse razonablemente a cualquier pdf multivariante, los métodos propuestos pueden tener una aplicabilidad general en cualquier problema de procesado de señal sobre grafos donde la matriz Laplaciana pueda ser aprendida u obtenida mediante señales de entrenamiento.

### A. Apéndice: Derivación de la media condicionada

Empezando a partir de

$$E[x_n/\mathbf{x}_{-nm}] = \int x_n p(x_n/\mathbf{x}_{-nm}) dx_n = \frac{1}{p(\mathbf{x}_{-nm})} \int x_n p(\mathbf{x}_{-m}) dx_n \quad . \quad (A1)$$

Donde  $\mathbf{x}_{-m}$  es el vector formado por todos los elementos de  $\mathbf{x}$  excepto  $x_m$  y  $p(\mathbf{x}_{-m})$  es la marginal correspondiente. Esto último se puede calcular integrando  $p(\mathbf{x})$  en (1) sobre  $x_m$ , es decir

$$p(\mathbf{x}_{-m}) = \int_{x_m} p(\mathbf{x}) dx_m = \sum_{k=1}^K P_k \int_{x_m} \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_k|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b}_k)^T \mathbf{C}_k^{-1}(\mathbf{x}-\mathbf{b}_k)} dx_m \quad . \quad (A2)$$

Pero cualquier marginal de una gaussiana multivariante también es una gaussiana multivariante. El vector medio marginal se obtiene eliminando el  $m$ -ésimo elemento del vector de medias  $\mathbf{b}_k$ . De manera similar, la matriz de covarianza marginal se obtiene eliminando la  $m$ -ésima fila y la  $m$ -ésima columna de  $\mathbf{C}_k$ . Llamando a  $\mathbf{b}_{k-m}$  y  $\mathbf{C}_{k-m}$  respectivamente a la media marginal y a la covarianza. Entonces se puede escribir:

$$p(\mathbf{x}_{-m}) = \sum_{k=1}^K P_k \frac{1}{\sqrt{(2\pi)^{N-1} |\mathbf{C}_{k-m}|}} e^{-\frac{1}{2}(\mathbf{x}_{-m}-\mathbf{b}_{k-m})^T \mathbf{C}_{k-m}^{-1}(\mathbf{x}_{-m}-\mathbf{b}_{k-m})} \quad . \quad (A3)$$

Y sustituyendo en (7)

$$E[x_n/\mathbf{x}_{-nm}] = \frac{1}{p(\mathbf{x}_{-nm})} \sum_{k=1}^K P_k \int x_n \frac{1}{\sqrt{(2\pi)^{N-1} |\mathbf{C}_{k-m}|}} e^{-\frac{1}{2}(\mathbf{x}_{-m} - \mathbf{b}_{k-m})^T \mathbf{C}_{k-m}^{-1} (\mathbf{x}_{-m} - \mathbf{b}_{k-m})} dx_n \quad . \quad (\text{A4})$$

Ahora, definiendo la pdf gaussiana multivariante

$$p_k(\mathbf{x}_{-m}) = p_k(x_n, \mathbf{x}_{-nm}) = \frac{1}{\sqrt{(2\pi)^{N-1} |\mathbf{C}_{k-m}|}} e^{-\frac{1}{2}(\mathbf{x}_{-m} - \mathbf{b}_{k-m})^T \mathbf{C}_{k-m}^{-1} (\mathbf{x}_{-m} - \mathbf{b}_{k-m})} \quad . \quad (\text{A5})$$

En (10) se tiene que calcular

$$\int_{x_n} x_n p_k(x_n, \mathbf{x}_{-nm}) dx_n = p_k(\mathbf{x}_{-nm}) \int_{x_n} x_n p_k(x_n/\mathbf{x}_{-nm}) dx_n \quad , \quad (\text{A6})$$

pero

$$p_k(\mathbf{x}_{-nm}) = \frac{1}{\sqrt{(2\pi)^{N-2} |\mathbf{C}_{k-nm}|}} e^{-\frac{1}{2}(\mathbf{x}_{-nm} - \mathbf{b}_{k-nm})^T \mathbf{C}_{k-nm}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k-nm})} \quad . \quad (\text{A7})$$

Donde  $\mathbf{b}_{k-nm}$  se obtiene a partir de  $\mathbf{b}_k$  eliminando los elementos  $n$ -ésimo y  $m$ -ésimo, y  $\mathbf{C}_{k-nm}$  se obtiene eliminando la  $n$ -ésima y  $m$ -ésima filas y columnas de  $\mathbf{C}_k$ .

La última integral en (12) coincide con la media condicionada de  $x_n$  con respecto a  $\mathbf{x}_{-nm}$  asumiendo el modelo gaussiano multivariante (11), es decir, asumiendo que  $\mathbf{x}$  ha sido generado por el componente  $k$ -ésimo de la mezcla. Por lo tanto, esta media condicionada es la estimación LLMSE  $x_n$  desde  $\mathbf{x}_{-nm}$ , que se puede obtener mediante las ecuaciones de Wiener-Hopf

$$\int_{x_n} x_n p_k(x_n/\mathbf{x}_{-nm}) dx_n = \mathbf{c}_{kn-nm}^T \mathbf{C}_{k-nm}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k-nm}) + b_{kn} \quad , \quad (\text{A8})$$

donde el vector  $\mathbf{c}_{kn-nm}$  está formado por los elementos de la  $n$ -ésima columna de la matriz  $\mathbf{C}_k$ , eliminando las filas  $n$  y  $m$ , y  $b_{kn}$  es el  $n$ -ésimo elemento del vector  $\mathbf{b}_k$ . Entonces se tiene

$$E[x_n/\mathbf{x}_{-nm}] = \frac{1}{p(\mathbf{x}_{-nm})} \sum_{k=1}^K P_k \cdot p_k(\mathbf{x}_{-nm}) \left[ \mathbf{c}_{kn-nm}^T \mathbf{C}_{k-nm}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k-nm}) + b_{kn} \right] \quad . \quad (\text{A9})$$

Pero, realmente, se puede expresar  $p_k(\mathbf{x}_{-nm}) = p(\mathbf{x}_{-nm}/k)$  como una pdf condicionada de clase, entonces

$$P_k \cdot p_k(\mathbf{x}_{-nm}) = P_k \cdot p(\mathbf{x}_{-nm}/k) = P(k/\mathbf{x}_{-nm}) p(\mathbf{x}_{-nm}) \quad . \quad (\text{A10})$$

Por lo tanto, finalmente

$$E[x_n/\mathbf{x}_{-nm}] = \sum_{k=1}^K P(k/\mathbf{x}_{-nm}) \left[ \mathbf{c}_{kn-nm}^T \mathbf{C}_{k-nm}^{-1} (\mathbf{x}_{-nm} - \mathbf{b}_{k-nm}) + b_{kn} \right] \quad . \quad (\text{A11})$$

Se puede observar que la media condicionada (no lineal) es una suma ponderada de  $K$  componentes. El  $k$ -ésimo componente es la estimación LLMSE de  $x_n$  desde  $\mathbf{x}_{-nm}$  asumiendo que el  $k$ -ésimo elemento de la mezcla está vigente. Por tanto, el peso correspondiente es la probabilidad de la presencia del  $k$ -ésimo elemento, condicionada a la observación  $\mathbf{x}_{-nm}$ . Esta ponderación hace que la media condicionada sea una función no lineal de  $\mathbf{x}_{-nm}$ .

## B. Apéndice: Una medida de separación entre componentes GMM

Para simplificar el procedimiento se considerará que una “buena separación” entre  $p(\mathbf{x}/k)$  y  $p(\mathbf{x}/k')$  que implica “buena separación” entre  $p(\mathbf{x}_{-nm}/k)$  y  $p(\mathbf{x}_{-nm}/k')$  para cada  $nm$ . Definiendo la función  $f(\mathbf{x}, k, k') = p(\mathbf{x}/k)p(\mathbf{x}/k')$   $k \neq k'$ . En el contexto de esta tesis, “buena separación” significa que  $\max f(\mathbf{x}, k, k') < \varepsilon$  para un pequeño número  $\varepsilon$ . Se define  $\mathbf{x}_{\max}$  como el valor que maximiza  $f(\mathbf{x}, k, k')$  para un par  $k, k'$  dado, es decir,

$$\begin{aligned} \mathbf{x}_{\max} &= \underset{\mathbf{x}}{\max} f(\mathbf{x}, k, k') = \underset{\mathbf{x}}{\max} \ln f(\mathbf{x}, k, k') = \underset{\mathbf{x}}{\max} \{ \ln p(\mathbf{x}/k) + \ln p(\mathbf{x}/k') \} = \\ &= \underset{\mathbf{x}}{\max} \left\{ -(\mathbf{x} - \mathbf{b}_k)^T \mathbf{C}_k^{-1} (\mathbf{x} - \mathbf{b}_k) - (\mathbf{x} - \mathbf{b}_{k'})^T \mathbf{C}_{k'}^{-1} (\mathbf{x} - \mathbf{b}_{k'}) \right\} . \end{aligned} \quad (\text{B1})$$

Derivando esta última expresión con respecto a  $\mathbf{x}$  e igualando a cero se obtiene

$$-2\mathbf{C}_k^{-1} (\mathbf{x} - \mathbf{b}_k) - 2\mathbf{C}_{k'}^{-1} (\mathbf{x} - \mathbf{b}_{k'}) = 0 \Rightarrow \mathbf{x}_{\max} = (\mathbf{C}_k^{-1} + \mathbf{C}_{k'}^{-1})^{-1} (\mathbf{C}_k^{-1} \mathbf{b}_k + \mathbf{C}_{k'}^{-1} \mathbf{b}_{k'}) . \quad (\text{B2})$$

Se tiene

$$\begin{aligned} \ln f(\mathbf{x}_{\max}, k, k') &= \\ &= -\ln(2\pi)^N \sqrt{|\mathbf{C}_k| |\mathbf{C}_{k'}|} - \frac{1}{2} (\mathbf{x}_{\max} - \mathbf{b}_k)^T \mathbf{C}_k^{-1} (\mathbf{x}_{\max} - \mathbf{b}_k) - \frac{1}{2} (\mathbf{x}_{\max} - \mathbf{b}_{k'})^T \mathbf{C}_{k'}^{-1} (\mathbf{x}_{\max} - \mathbf{b}_{k'}) \end{aligned} \quad (\text{B3})$$

Entonces  $\max \ln f(\mathbf{x}, k, k')$  puede obtenerse calculando (B3) para todos los pares  $k, k'$  y reteniendo la máxima. Nótese que si  $\mathbf{C}_k = \mathbf{C}_{k'}$  se tiene

$$\begin{aligned} \mathbf{x}_{\max} &= \frac{1}{2}(\mathbf{b}_k + \mathbf{b}_{k'}) \\ \ln f(\mathbf{x}_{\max}, k, k') &= -\ln(2\pi)^N |\mathbf{C}_k| - \frac{1}{2}(\mathbf{b}_{k'} - \mathbf{b}_k)^T \mathbf{C}_k^{-1} (\mathbf{b}_{k'} - \mathbf{b}_k) \end{aligned} \quad (B4)$$

la cual está estrechamente relacionada con la muy bien conocida distancia *Mahalanobis* [26] entre dos densidades normales multivariadas que tienen la misma matriz de covarianza. Finalmente llamando  $s = \max \ln f(\mathbf{x}, k, k')$  para que se pueda definir un índice de separabilidad normalizado “*ad hoc*”  $I_s$  como

$$I_s = \frac{e^{-s}}{1 + e^{-s}} \quad (B5)$$

De manera clara,  $0 \leq I_s \leq 1$  y, a medida que aumenta la separación,  $s \rightarrow -\infty$  y  $I_s \rightarrow 1$ .

## Referencias

- [1] R. Merris, "Laplacian matrices of a graph: a survey," *Linear Alg. Applicat.*, no. 197, pp. 143-176, 1994.
- [2] D. Spielman, "Spectral graph theory," in *Combinatorial Scientific Computing*, O. Schnek U. Naumann, Ed.: Chapman and Hall/CRC Press, 2012, ch. 16.
- [3] D. Thanou, P. Frossard, P. Vandergheynst X. Dong, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. on Signal Processing*, no. 64, pp. 6160-6173, 2016.
- [4] D. Florencio, P. A Chou Ch. Zhang, "Graph signal processing: a probabilistic framework," Microsoft, Microsoft Research Technical Report MSR-TR-2015-31, 2015.
- [5] A. Ortega E. Pávez, "Generalized precision matrix estimation for graph signal processing," in *Proc. of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, Sanghai (China), 2016, pp. 6350-6354.
- [6] T. Hastie, R. Tibshirani J. Friedman, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, pp. 432-441, 2008.
- [7] T. Hastie R. Mazumder, "The graphical lasso: New insights and alternatives," *Electronic Journal of Statistics*, vol. 6, pp. 2125-2149, 2012.
- [8] M.A. Sustik, I.S. Dhillon, P. Ravikumar C.J. Hsieh, "Sparse inverse covariance matrix estimation using quadratic approximation," *Advances in Neural Information Processing Systems*, vol. 24, pp. 2330-2338, 2011.
- [9] M. Xu, W. B. Wu X. Cheng, "Covariance and precision matrix estimation for high-dimensional time series," *The Annals of Statistics*, vol. 41, pp. 2994-3021, 2013.
- [10] C. Croux V. Öllerer, "Robust high-dimensional precision matrix estimation," in *Modern Multivariate and Robust Methods*, S.Taskinen K. Nodhausen, Ed. New York: Springer, 2015.
- [11] P. Wang, N. Zhou N, J. Zhu J. Peng J, "Partial correlation estimation by joint sparse regression model," *Journal of the American Statistical Association*, vol. 104, pp. 735-746, 2009.
- [12] A. Salazar, G. Safont, L. Vergara J. Igual, "Semi-Supervised Bayesian Classification of Materials with Impact-Echo Signals," *Sensors*, vol. 15, pp.

11528-11550, 2015.

- [13] W. Wang, S. I. Jeon, S. H. Son S. Aminikhanghahi, "A new fuzzy Gaussian mixture model (FGMM) based algorithm for mammography tumor image classification," *Multimedia Tools and Applications*, vol. 75, pp. 1-15, 2016.
- [14] M. Marczyk, "Improved Detection of 2D Gel Electrophoresis Spots by Using Gaussian Mixture Model," in *12-th International Symposium on Bioinformatics Research and Applications*, Minks (Belarus), 2016, pp. 284-294.
- [15] M. R. Taaffe J. Wang, "Multivariate Mixtures of Normal Distributions: Properties, Random Vector Generation, Fitting, and as Models of Market Daily Changes," *INFORMS Journal on Computing*, vol. 27, pp. 193-203, 2015.
- [16] P. Stockhammar M. U. Hassan, "Fitting probability distributions to economic growth: a maximum likelihood approach," *Journal of Applied Statistics*, vol. 43, pp. 1583-1603, 2016.
- [17] S. B. Popov A. P. Igoshev, "Gaussian mixture models and the population synthesis of radio pulsars," *Monthly Notices of the Royal Astronomical Society*, vol. 434, pp. 2229-2237, 2013.
- [18] D. Ciounzo, K. Kansanen, T. Ekman P. Salvo Rossi, "Performance Analysis of Energy Detection for MIMO Decision Fusion in Wireless Sensor Networks Over Arbitrary Fading Channels," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 7794-7806, 2016.
- [19] D. Ciounzo, K. Kansanen, T. Ekman P. Salvo Rossi, "On energy detection for MIMO decision fusion in wireless sensor networks over NLOS fading," *IEEE Communications Letters*, vol. 19, pp. 303-306, 2015.
- [20] S. Nadarajah S. Kotz, "Local dependence functions for elliptically symmetric distributions," *Sankhya: The Indian Journal of Statistics*, vol. 65, pp. 207-223, 2003.
- [21] R. Shibata, M. Sibuya K. Baba, "Partial correlation and conditional correlation as measures of conditional independence," *Australia & New Zealand of Statistics*, vol. 46, pp. 657-664, 2004.
- [22] A. Tomé, I.R. Keck, J. Górriz-Sáez, C. Puntonet E.W. Lang, "Brain connectivity analysis: a short survey," *Computational intelligence and neuroscience*, pp. 1-21, 2012.
- [23] J. Romagnoli D. Leung, "Dynamic probabilistic model-based expert system for fault diagnosis," *Journal of Computers & Chemical Engineering*, vol. 24, pp.



2473-2492, 2000.

- [24] L. Vergara, R. Miralles A. Salazar, "On including sequential dependence in ICA mixture models," *Signal Processing*, vol. 90, pp. 2314-2318, 2010.
- [25] L. Cheng, J. Sepulcre, G. E. Fakhri, Y. M. Lu, Q. Li C. Hu, "A graph theoretical regression model for brain connectivity learning of Alzheimer's disease," in *IEEE 10th International Symposium on Biomedical Imaging (ISBI2013)*, San Francisco (USA), 2013, pp. 616-619.
- [26] J. Koronacki, S. Lipovetsky E. Nowakowska, "Tractable measure of component overlap for gaussian mixture models," *ArXiv*, vol. 1, no. 1407.7172, 2014.

# Capítulo 3

## Estimación de la matriz Laplaciana en Modelos de Análisis de Componentes Independientes

### 3.1. Introducción

En este capítulo se sigue profundizando en el problema de estimación de la matriz Laplaciana de un grafo considerando el cálculo de la correlación parcial bajo un modelo no gaussiano, en particular un modelo de análisis de componentes independientes (ICA).

ICA [1] [2] [3] es una técnica consolidada que se puede encontrar en numerosas aplicaciones de tratamiento estadístico de la señal (p. ej., separación ciega de fuentes [3]) y reconocimiento de patrones [4]). Desde la perspectiva de esta tesis, ICA es un modelo que incorpora la no gaussianidad a través de variables independientes (fuentes) que se mezclan linealmente para crear las muestras de observación. Esto

permite modelar densidades no-gaussianas multivariantes con una gran versatilidad.

En la Sección 3.2 se define un nuevo coeficiente de correlación parcial: ICA-PCC. El concepto básico es nuevamente reemplazar la estimación lineal implícita del PCC convencional por una estimación no lineal (media condicionada) pero suponiendo ahora un modelo ICA subyacente. Además se presenta una fórmula general para calcular la matriz de covarianza residual a partir de la cual se calcularán los ICA-PCC. Una parte esencial de esta fórmula es la matriz diagonal que contiene los errores cuadráticos medios de la estimación de las fuentes del modelo ICA. Luego, en la Sección 3.3, se presenta un método práctico para estimar dicha matriz a partir de los parámetros del modelo ICA. Por último, la Sección 3.4 incluye algunas simulaciones para ilustrar la estimación mejorada de la correlación parcial por ICA-PCC en escenarios no gaussianos. También se presenta un ejemplo de datos reales con señales de electroencefalografía o EEG multicanal no gaussianas para cuantificar los cambios en la conectividad cerebral entre los estados normal y anormal de un paciente durante el sueño.

## **3.2. Correlaciones parciales en modelos ICA**

### **3.2.1. Planteamiento del problema**

Sea  $\mathbf{x} = [x_1 \dots x_N]^T$  el vector de observación teniendo la matriz de covarianza  $E[\mathbf{xx}^T] = \mathbf{C}_{xx}$ . Suponiendo que  $\mathbf{x}$  obedece a un modelo ICA, entonces

$$\mathbf{x} = \mathbf{U}\mathbf{s} \quad \mathbf{s} = \mathbf{W}\mathbf{x} \quad , \quad (1)$$

donde  $\mathbf{s} = [s_1 \dots s_N]^T$  es un vector de fuentes independientes y  $\mathbf{U}$  es una matriz de mezcla cuadrada e invertible ( $\mathbf{W} = \mathbf{U}^{-1}$  es la matriz de desmezcla). Las fuentes se consideran estandarizadas (media cero y varianza unitaria), de lo contrario pueden

tener diferentes densidades marginales no gaussianas que factorizan la función de densidad de probabilidad conjunta (pdf)  $p(\mathbf{s}) = p(s_1) \cdots p(s_N)$ . Nótese que

$$\begin{aligned} E[\mathbf{s}] &= \mathbf{0} & \mathbf{C}_{ss} &= E[\mathbf{ss}^T] = \mathbf{I} \\ E[\mathbf{x}] &= \mathbf{0} & \mathbf{C}_{xx} &= E[\mathbf{xx}^T] = E[\mathbf{Uss}^T \mathbf{U}^T] = \mathbf{U}\mathbf{U}^T \end{aligned} \quad (2)$$

Cada componente de  $\mathbf{x}$  se asigna a cada nodo del grafo  $G\{V, E, \mathbf{A}\}$ , donde  $V$  es el conjunto de  $N$  nodos,  $E$  es el conjunto de conexiones que unen los nodos y  $\mathbf{A}$  es la matriz de adyacencia, tal y como ya se definieron en el Capítulo anterior. El problema es aprender u obtener  $\mathbf{A}$  a partir de un conjunto de vectores de observación disponible. Como ya se ha mencionado en el Capítulo 2, los coeficientes de PCC son candidatos razonables ya que pueden medir la correlación entre dos nodos eliminando el efecto del resto de nodos. Además, hay que recordar que los coeficientes PCC se pueden calcular a partir de la matriz de precisión  $\mathbf{Q}_{xx} = \mathbf{C}_{xx}^{-1}$  en la forma

$$\rho_{nm}^{PCC} = -\frac{q_{nm}}{\sqrt{q_{nn}q_{mm}}}, \quad (3)$$

donde  $q_{nm}$  es el elemento  $nm$  de la matriz  $\mathbf{Q}$  y  $\rho_{nm}^{PCC}$  es el PCC de los nodos  $n$  y  $m$ . La ecuación (3) podría usarse para cualquier densidad de probabilidad conjunta  $p(\mathbf{x})$  subyacente, sin embargo, es óptima sólo para el caso gaussiano. Esto se debe a que la definición formal de  $\rho_{nm}^{PCC}$  viene dada por

$$\rho_{nm}^{PCC} = \frac{E[(x_n - L[x_n/\mathbf{x}_{-nm}])(x_m - L[x_m/\mathbf{x}_{-nm}])]}{\sqrt{E[(x_n - L[x_n/\mathbf{x}_{-nm}])^2]} \sqrt{E[(x_m - L[x_m/\mathbf{x}_{-nm}])^2]}} \quad (4)$$

Donde  $\mathbf{x}_{-nm}$  es el vector formado por todas las muestras de  $\mathbf{x}$  excepto  $x_n$  y  $x_m$ , y  $L[x_n/\mathbf{x}_{-nm}]$ ,  $L[x_m/\mathbf{x}_{-nm}]$  son respectivamente las estimaciones LLMSE de  $x_n$  y  $x_m$  desde  $\mathbf{x}_{-nm}$ . Sin embargo, la eliminación óptima del efecto de  $\mathbf{x}_{-nm}$  implica el

uso de las medias condicionadas  $E[x_n/\mathbf{x}_{-nm}]$  y  $E[x_m/\mathbf{x}_{-nm}]$ , que coinciden respectivamente con  $L[x_n/\mathbf{x}_{-nm}]$  y  $L[x_m/\mathbf{x}_{-nm}]$  sólo cuando  $p(\mathbf{x})$  es gaussiana multivariante. Por lo tanto en el caso no gaussiano, el PCC convencional no captura con precisión la correlación parcial y, por lo tanto, la conectividad del grafo.

En este capítulo se va a calcular la media condicionada  $E[x_n/\mathbf{x}_{-nm}]$ , correspondiente al modelo ICA (1), de modo que la correlación parcial puede estimarse mejor mediante

$$\rho_{nm}^{ICA-PCC} = \frac{E[(x_n - E[x_n/\mathbf{x}_{-nm}])(x_m - E[x_m/\mathbf{x}_{-nm}])]}{\sqrt{E[(x_n - E[x_n/\mathbf{x}_{-nm}])^2]} \sqrt{E[(x_m - E[x_m/\mathbf{x}_{-nm}])^2]}}. \quad (5)$$

A este nuevo coeficiente de correlación parcial se le llamará ICA-PCC.

### 3.2.2. Una fórmula general para la covarianza residual

Definiendo el vector  $\mathbf{x}_{nm} = [x_n, x_m]^T$ . Se puede expresar  $\mathbf{x}$  en la forma

$$\mathbf{x} = \mathbf{T}_{-nm} \mathbf{x}_{-nm} + \mathbf{T}_{nm} \mathbf{x}_{nm}, \quad (6)$$

donde  $\mathbf{T}_{-nm}$  es una matriz de dimensión  $(N \times (N-2))$  obtenida a partir de una matriz identidad  $(N \times N)$  eliminando las columnas  $n$ -ésima y  $m$ -ésima. De modo similar,  $\mathbf{T}_{nm}$  es una matriz de dimensión  $(N \times 2)$  obtenida a partir de una matriz identidad  $(N \times N)$  eliminando todas las columnas excepto la  $n$ -ésima y  $m$ -ésima.

Definiendo también el vector residual  $\mathbf{e}_{nm} = [e_n, e_m]^T$   $e_n = x_n - E[x_n/\mathbf{x}_{-nm}]$   $e_m = x_m - E[x_m/\mathbf{x}_{-nm}]$ . Nótese que la media condicionada es una estimador insesgado, por lo que los residuos son de media cero y la matriz de covarianza residual será

$$\mathbf{C}_{e_{nm}e_{nm}} = E[\mathbf{e}_{nm}\mathbf{e}_{nm}^T] = \begin{bmatrix} E[e_n^2] & E[e_n e_m] \\ E[e_m e_n] & E[e_m^2] \end{bmatrix} \quad (7)$$

Se quiere calcular la matriz de covarianza residual para que se pueda aplicar (5). Asumiendo un modelo ICA. Primero se observa que  $\mathbf{e}_{nm} = \mathbf{x}_{nm} - E[\mathbf{x}_{nm}/\mathbf{x}_{-nm}]$ , pero considerando (1) y (6), se puede escribir

$$E[\mathbf{s}/\mathbf{x}_{-nm}] = \mathbf{W}(\mathbf{T}_{-nm}\mathbf{x}_{-nm} + \mathbf{T}_{nm}E[\mathbf{x}_{nm}/\mathbf{x}_{-nm}]) \quad (8)$$

Pudiéndose resolver para  $E[\mathbf{x}_{nm}/\mathbf{x}_{-nm}]$

$$E[\mathbf{x}_{nm}/\mathbf{x}_{-nm}] = (\mathbf{W}\mathbf{T}_{nm})^+ (E[\mathbf{s}/\mathbf{x}_{-nm}] - \mathbf{W}\mathbf{T}_{-nm}\mathbf{x}_{-nm}). \quad (9)$$

Donde  $(\bullet)^+$  significa la pseudoinversa de Moore-Penrose (izquierda). Así en (9) se está expresando la media condicionada de  $\mathbf{x}_{nm}$  en términos de la media condicionada de las fuentes y los parámetros del modelo ICA. Esto permite derivar la siguiente fórmula general, que, a pesar de su simplicidad, requiere una derivación bastante tediosa que se ha relegado al Apéndice A de este capítulo

$$\mathbf{C}_{e_{nm}e_{nm}} = (\mathbf{W}\mathbf{T}_{nm})^+ \mathbf{M}_{nm} ((\mathbf{W}\mathbf{T}_{nm})^+)^T, \quad (10)$$

donde  $\mathbf{M}_{nm}$  es una matriz diagonal ( $N \times N$ ) teniendo en su diagonal principal los MSEs (*Mean Squared Error*) de la estimación óptima de las fuentes a partir de  $\mathbf{x}_{-nm}$ , es decir,

$$\begin{aligned} \mathbf{M}_{nm}(i,i) &= mse_{nmi} = E[(s_i - E[s_i/\mathbf{x}_{-nm}])^2] \\ 0 &\leq mse_{nmi} \leq 1 \end{aligned} \quad (11)$$

Nótese que (10) es una combinación de las contribuciones de cada fuente a la matriz de covarianza residual. Esto se puede ver mejor expresando (10) en la forma alternativa

$$\mathbf{C}_{e_{nm}e_{nm}} = \sum_{i=1}^N mse_{nmi} \cdot \mathbf{u}_{nmi}^+ \mathbf{u}_{nmi}^{+T} \quad , \quad (12)$$

donde  $\mathbf{u}_{nmi}^+$  es la columna  $i$ -ésima de  $(\mathbf{W}\mathbf{T}_{nm})^+$ . Obsérvese que  $\mathbf{W}\mathbf{T}_{nm}$  es una matriz  $(N \times 2)$  formada por las columnas  $n$ -ésimas y  $m$ -ésimas de  $\mathbf{W}$ , es decir, formada por los coeficientes que definen las contribuciones de  $x_n$  y  $x_m$  a  $\mathbf{S}$ . Entonces  $(\mathbf{W}\mathbf{T}_{nm})^+$  es una matriz  $(2 \times N)$ ,  $\mathbf{u}_{nmi}^+$  es un vector  $(2 \times 1)$  y  $\mathbf{u}_{nmi}^+ \mathbf{u}_{nmi}^{+T}$  es una matriz  $(2 \times 2)$  que puede ser interpretada como la contribución de la fuente  $s_i$  a  $\mathbf{C}_{e_{nm}e_{nm}}$ . Esta contribución se pondera mediante  $mse_{nmi}$ . Por lo tanto,  $mse_{nmi} = 0$  indica que la fuente  $s_i$  ha sido perfectamente estimada por  $\mathbf{x}_{-nm}$ , y por ello  $s_i$  no contribuye a la correlación parcial entre  $x_n$  y  $x_m$ . En el otro extremo,  $mse_{nmi} = 1$  indica que  $s_i$  es independiente de  $\mathbf{x}_{-nm}$ , por lo que tiene máxima contribución a la correlación parcial entre  $x_n$  y  $x_m$ .

### 3.3. Calculando ICA-PCC

Se quiere estimar  $\rho_{nm}^{ICA-PCC}$  mediante

$$\hat{\rho}_{nm}^{ICA-PCC} = \frac{\hat{E}[e_n e_m]}{\sqrt{\hat{E}[(e_n)^2]} \sqrt{\hat{E}[(e_m)^2]}} \quad . \quad (13)$$

Entonces, de acuerdo con (7), se tiene que estimar  $C_{e_{nm}e_{nm}}$ . Considerando (10), se necesita realizar la estimación de  $\mathbf{W}$  y  $\mathbf{M}_{nm}$ . Ya que la estimación de  $\mathbf{W}$  y los parámetros del modelo ICA, puede obtenerse utilizando diferentes algoritmos [1] [2] [3] [4] a continuación, se estudiará la estimación de  $\mathbf{M}_{nm}$ , es decir, la estimación  $mse_{nmi} = E\left[(s_i - E[s_i/\mathbf{x}_{-nm}])^2\right] \quad i = 1 \dots N$ . Para calcular  $E[s_i/\mathbf{x}_{-nm}]$  se considerará una forma particular de una estructura Wiener que fue propuesta en [5], esto es:

$$E[s_i/\mathbf{x}_{-nm}] \simeq E[s_i/\hat{s}_i^l] \quad , \quad (14)$$

donde  $\hat{s}_i^l$  es el estimador LLMSE de  $s_i$  a partir de  $\mathbf{x}_{-nm}$  (se eliminará la dependencia en  $nm$  para facilitar la notación) y la media condicionada unidimensional puede aproximarse mediante [5] [6]

$$E[s_i/\hat{s}_i^l] = \sum_{k=1}^{\infty} \frac{1}{k!} E[s_i \cdot (\hat{s}_{in}^l)^k] H_k(\hat{s}_{in}^l) \quad , \quad (15)$$

donde  $H_k(x)$  es el  $k$ -ésimo polinomio de Hermite y  $\hat{s}_{in}^l = \frac{\hat{s}_i^l}{(\text{var}[\hat{s}_i^l])^{\frac{1}{2}}}$  es una variable aleatoria gaussiana estandarizada.

Aproximando (15) por los dos primeros términos. Teniendo en cuenta que  $H_1(x) = x$   $H_2(x) = x^2 - 1$ , se puede escribir

$$E[s_i/\hat{s}_i^l] = E[s_i \cdot \hat{s}_{in}^l] \hat{s}_{in}^l + E[s_i \cdot (\hat{s}_{in}^l)^2] \left( \frac{(\hat{s}_{in}^l)^2 - 1}{2} \right). \quad (16)$$

Pero



$$\begin{aligned}
 E[s_i \cdot \hat{s}_{in}^l] \hat{s}_{in}^l &= E[s_i \cdot \hat{s}_i^l] \frac{\hat{s}_i^l}{\text{var}(\hat{s}_i^l)} = E[\hat{s}_i^l \cdot \hat{s}_i^l] \frac{\hat{s}_i^l}{\text{var}(\hat{s}_i^l)} \\
 &= \text{var}(\hat{s}_i^l) \frac{\hat{s}_i^l}{\text{var}(\hat{s}_i^l)} = \hat{s}_i^l \quad ,
 \end{aligned} \tag{17}$$

donde se ha considerado que  $E[s_i \cdot \hat{s}_i^l] = E[\hat{s}_i^l \cdot \hat{s}_i^l]$  (debido a la ortogonalidad entre el error de estimación y la estimación lineal), y que  $\hat{s}_i^l$  es insesgado, por lo que  $E[\hat{s}_i^l] = E[s_i] = 0$ . Entonces se puede expresar la media condicionada en (16) como la combinación del término lineal  $\hat{s}_i^l$  más un término no lineal  $s_i^{nl} = E[s_i \cdot (\hat{s}_{in}^l)^2] \left( \frac{(\hat{s}_{in}^l)^2 - 1}{2} \right)$ . Se puede ahora expresar la estimación de  $\mathbf{S}$  a partir

$\mathbf{x}_{-nm}$  i de manera más compacta en la forma

$$E[\mathbf{s} / \mathbf{x}_{-nm}] = \hat{\mathbf{s}} = \hat{\mathbf{s}}^l + \hat{\mathbf{s}}^{nl} \quad . \tag{18}$$

Se puede escribir

$$\begin{aligned}
 \mathbf{M}_{nm} &= \text{diag} \left( E \left[ (\mathbf{s} - \hat{\mathbf{s}}) (\mathbf{s} - \hat{\mathbf{s}})^T \right] \right) \\
 &= \text{diag} \left( E \left[ ((\mathbf{s} - \hat{\mathbf{s}}^l) - \hat{\mathbf{s}}^{nl}) ((\mathbf{s} - \hat{\mathbf{s}}^l) - \hat{\mathbf{s}}^{nl})^T \right] \right) \quad . \\
 &= \mathbf{M}_{nm}^l + \text{diag} \left( E \left[ \hat{\mathbf{s}}^{nl} (\hat{\mathbf{s}}^{nl})^T \right] \right) - 2 \text{diag} \left( E \left[ (\mathbf{s} - \hat{\mathbf{s}}^l) (\hat{\mathbf{s}}^{nl})^T \right] \right)
 \end{aligned} \tag{19}$$

Donde  $\mathbf{M}^l$  es una matriz diagonal cuyos elementos son las MSEs correspondientes a la estimación lineal de  $s_i$  a partir de  $\mathbf{x}_{-nm}$ , que es

$$\begin{aligned}
 \mathbf{M}_{nm}^l &= \text{diag} \left( E \left[ (\mathbf{s} - \hat{\mathbf{s}}^l) (\mathbf{s} - \hat{\mathbf{s}}^l)^T \right] \right) \\
 &= \text{diag} \left( E \left[ \mathbf{s} \mathbf{s}^T \right] \right) - \text{diag} \left( E \left[ \hat{\mathbf{s}}^l (\hat{\mathbf{s}}^l)^T \right] \right) \quad .
 \end{aligned} \tag{20}$$

Se ha considerado en (20) ortogonalidad entre el vector de error y el vector estimado. Pero  $\hat{\mathbf{s}}^l$  es la estima LLMSE, por lo que se puede obtener mediante las ecuaciones de Wiener-Hopft

$$\begin{aligned}\hat{\mathbf{s}}^l &= \mathbf{C}_{sx-nm} \mathbf{C}_{x-nm}^{-1} \mathbf{x}_{-nm} \\ \mathbf{C}_{sx-nm} &= E[\mathbf{s} \mathbf{x}_{-nm}^T] \quad \mathbf{C}_{x-nm} = E[\mathbf{x}_{-nm} \mathbf{x}_{-nm}^T]\end{aligned}\quad (21)$$

Por lo tanto se puede escribir

$$\begin{aligned}\mathbf{M}_{nm}^l &= \mathbf{I} - \text{diag}\left(E\left[\mathbf{C}_{sx-nm} \mathbf{C}_{x-nm}^{-1} \mathbf{x}_{-nm} \mathbf{x}_{-nm}^T \mathbf{C}_{x-nm}^{-1} \mathbf{C}_{sx-nm}^T\right]\right) = \\ &= \mathbf{I} - \text{diag}\left(\mathbf{C}_{sx-nm} \mathbf{C}_{x-nm}^{-1} \mathbf{C}_{sx-nm}^T\right)\end{aligned}\quad (22)$$

Teniendo en cuenta que  $\mathbf{C}_{sx-nm} = \mathbf{W} \mathbf{C}_{xx} \mathbf{T}_{-nm}$ ,  $\mathbf{C}_{x-nm} = \mathbf{T}_{-nm}^T \mathbf{C}_{xx} \mathbf{T}_{-nm}$  y  $\mathbf{C}_{xx} = \mathbf{W}^{-1} (\mathbf{W}^{-1})^T$ , finalmente se puede expresar  $\mathbf{M}_{nm}^l$  en términos de los parámetros del modelo ICA

$$\mathbf{M}_{nm}^l = \mathbf{I} - \text{diag}\left(\left(\mathbf{W}^{-1}\right)^T \mathbf{T}_{-nm} \left(\mathbf{T}_{-nm}^T \mathbf{W}^{-1} \left(\mathbf{W}^{-1}\right)^T \mathbf{T}_{-nm}\right)^{-1} \mathbf{T}_{-nm}^T \mathbf{W}^{-1}\right).\quad (23)$$

Considerando ahora los otros dos términos en (19). Primero, nótese que  $\hat{s}_i^{nl}$  puede ser interpretado como una estimación lineal de  $s_i$  a partir de  $(\hat{s}_{in}^l)^2$ , ya que asumiendo que  $\hat{s}_{in}^l$  es una variable aleatoria gaussiana estandarizada, entonces  $(\hat{s}_{in}^l)^2$  es  $\chi^2$  teniendo la media igual a 1 y la varianza igual a 2. Por lo tanto, podemos aplicar de nuevo la ortogonalidad:

$$\begin{aligned}\text{diag}\left(E\left[\left(\mathbf{s} - \hat{\mathbf{s}}^{nl}\right) \left(\hat{\mathbf{s}}^{nl}\right)^T\right]\right) &= 0 \Rightarrow \\ \Rightarrow \text{diag}\left(E\left[\mathbf{s} \left(\hat{\mathbf{s}}^{nl}\right)^T\right]\right) &= \text{diag}\left(E\left[\hat{\mathbf{s}}^{nl} \left(\hat{\mathbf{s}}^{nl}\right)^T\right]\right)\end{aligned}\quad (24)$$

En consecuencia, se tiene que

$$\begin{aligned} & \text{diag}\left(E\left[\hat{\mathbf{s}}^{nl}(\hat{\mathbf{s}}^{nl})^T\right]\right) - 2\text{diag}\left(E\left[(\mathbf{s} - \hat{\mathbf{s}}^l)(\hat{\mathbf{s}}^{nl})^T\right]\right) = \\ & = -\text{diag}\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{nl})^T\right]\right) + 2\text{diag}\left(E\left[\hat{\mathbf{s}}^l(\hat{\mathbf{s}}^{nl})^T\right]\right) \end{aligned} \quad (25)$$

El segundo término en (25) es cero, ya que

$$\begin{aligned} \left[\text{diag}\left(E\left[\hat{\mathbf{s}}^l(\hat{\mathbf{s}}^{nl})^T\right]\right)\right]_{ii} &= E\left[\hat{s}_i^l E\left[s_i \cdot (\hat{s}_m^l)^2\right] \left(\frac{(\hat{s}_m^l)^2 - 1}{2}\right)\right] = \\ &= \frac{1}{2} E\left[s_i \cdot (\hat{s}_m^l)^2\right] \left(E\left[\hat{s}_i^l (\hat{s}_m^l)^2\right] - E\left[\hat{s}_i^l\right]\right), \end{aligned} \quad (26)$$

donde  $E\left[\hat{s}_i^l\right] = 0$  y  $E\left[\hat{s}_i^l (\hat{s}_m^l)^2\right] = (\text{var}\left[\hat{s}_i^l\right])^{\frac{1}{2}} E\left[(\hat{s}_m^l)^3\right] = 0$ , porque se ha asumido que  $\hat{s}_{in}^l$  es gaussiano y por tanto sus momentos impares son cero. En cuanto al primer término de (25)

$$\begin{aligned} & \left[\text{diag}\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{nl})^T\right]\right)\right]_{ii} = \\ &= E\left[s_i E\left[s_i \cdot (\hat{s}_m^l)^2\right] \left(\frac{(\hat{s}_m^l)^2 - 1}{2}\right)\right] \\ &= \frac{1}{2} E\left[s_i \cdot (\hat{s}_m^l)^2\right] \left(E\left[s_i (\hat{s}_m^l)^2\right] - E\left[s_i\right]\right) \\ &= \frac{1}{2 \text{var}^2\left[\hat{s}_i^l\right]} E^2\left[s_i \cdot (\hat{s}_i^l)^2\right] \end{aligned} \quad (27)$$

Definiendo el vector  $\hat{\mathbf{s}}^{l(2)} = \left[(\hat{s}_1^l)^2 \dots (\hat{s}_N^l)^2\right]^T$  y teniendo en cuenta que  $\text{var}\left[\hat{s}_i^l\right] = \left[\mathbf{C}_{sx-nm} \quad \mathbf{C}_{x-nm}^{-1} \quad \mathbf{C}_{sx-nm}^T\right]_{ii}$  se puede escribir

$$\begin{aligned} & \text{diag}\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{nl})^T\right]\right) \\ &= \frac{1}{2} \text{diag}^2\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{l(2)})^T\right]\right) \text{diag}^{-2}\left(\mathbf{C}_{sx-nm} \quad \mathbf{C}_{x-nm}^{-1} \quad \mathbf{C}_{sx-nm}^T\right), \end{aligned} \quad (28)$$

y considerando (21) y (23), (28) puede expresarse en términos de parámetros del modelo ICA

$$\begin{aligned}
 & \text{diag}\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{nl})^T\right]\right) = \frac{1}{2} \\
 & \cdot \text{diag}^2\left(\mathbf{W}E\left[\mathbf{x}\left[\left((\mathbf{W}^{-1})^T \mathbf{T}_{-nm} \left(\mathbf{T}_{-nm}^T \mathbf{W}^{-1} (\mathbf{W}^{-1})^T \mathbf{T}_{-nm}\right)^{-1} \mathbf{x}_{-nm}\right)^{(2)}\right]^T\right]\right)\right) \cdot \\
 & \cdot \text{diag}^{-2}\left((\mathbf{W}^{-1})^T \mathbf{T}_{-nm} \left(\mathbf{T}_{-nm}^T \mathbf{W}^{-1} (\mathbf{W}^{-1})^T \mathbf{T}_{-nm}\right)^{-1} \mathbf{T}_{-nm}^T \mathbf{W}^{-1}\right)
 \end{aligned} \tag{29}$$

En conclusión, se puede expresar la matriz  $\mathbf{M}_{nm}$  como

$$\mathbf{M}_{nm} = \mathbf{M}_{nm}^l - \text{diag}\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{nl})^T\right]\right) \quad , \tag{30}$$

donde  $\mathbf{M}_{nm}^l$  y  $\text{diag}\left(E\left[\mathbf{s}(\hat{\mathbf{s}}^{nl})^T\right]\right)$  se pueden obtener respectivamente mediante (23)

y (29) utilizando estimaciones  $\hat{\mathbf{W}}$  de los parámetros del modelo y una media muestral para evaluar la esperanza necesaria en (29). El Algoritmo 3.1 presentado a continuación describe el proceso de estimación.

**Algoritmo 3.1:** Cálculo ICA-PCC

---

- 1: **Entrada:** Conjunto de datos de entrenamiento  $\mathbf{x}^{(l)}$   $l = 1 \dots L$
  - 2: **Calcular**  $\hat{\mathbf{W}}$  del conjunto de datos de aprendizaje (mediante cualquier algoritmo ICA)
  - 3: **for**  $n = 1, 2 \dots N$
  - 4: **for**  $m = n \dots N$
  - 5: Calcular  $\hat{\mathbf{M}}_{nm}$  (ec. (30), (23), (29))
  - 6: Calcular  $\hat{\mathbf{C}}_{e_{nm}e_{nm}} = -(\hat{\mathbf{W}}\mathbf{T}_{nm})^+ \hat{\mathbf{M}}_{nm} \left( (\hat{\mathbf{W}}\mathbf{T}_{nm})^+ \right)^T$
  - 7: Calcular  $\hat{\rho}_{nm}^{ICA-PCC}$  (ec. (13))
  - 8: Calcular  $\hat{\rho}_{mn}^{ICA-PCC} = \hat{\rho}_{nm}^{ICA-PCC}$
  - 9: **end for**
  - 10: **end for**
  - 11: **Salida**  $\hat{\rho}_{mn}^{ICA-PCC}$   $n = 1 \dots N$   $m = 1 \dots M$
- 

La ecuación (30) proporciona una interesante descomposición de  $mse_{nmi}$ . Si se llama  $mse_{nmi}^l$  a las entradas de la diagonal de la matriz  $\mathbf{M}_{nm}^l$ , entonces  $mse_{nmi}$  puede expresarse como  $mse_{nmi}^l$  menos un término no negativo (ver ecuación (29)), de modo que  $mse_{nmi} \leq mse_{nmi}^l$ . La condición  $mse_{nmi} = mse_{nmi}^l \Leftrightarrow \mathbf{M}_{nm} = \mathbf{M}_{nm}^l$  se cumple para el caso gaussiano, debido a que  $E\left[s_i \cdot (\hat{s}_{in}^l)^2\right]$  (ver ecuación (27)) se vuelve cero (es un momento impar de orden superior de una variable gaussiana multivariada). En tal caso,  $E[\mathbf{s}/\mathbf{x}_{-nm}] = \hat{\mathbf{s}}^l$  se convierte en una función lineal de  $\mathbf{x}_{-nm}$  y lo mismo sucede con  $E[\mathbf{x}/\mathbf{x}_{-nm}]$  en (9). Por tanto, el segundo término en (30) es responsable de la reducción mejorada de la influencia de  $\mathbf{x}_{-nm}$  en la estimación de la correlación parcial entre  $x_n$  y  $x_m$ , en el caso no gaussiano. Además, se deberían esperar resultados similares de ICA-PCC y PCC para el caso gaussiano.

### 3.4. Experimentos

#### 3.4.1. Experimentos con datos sintéticos

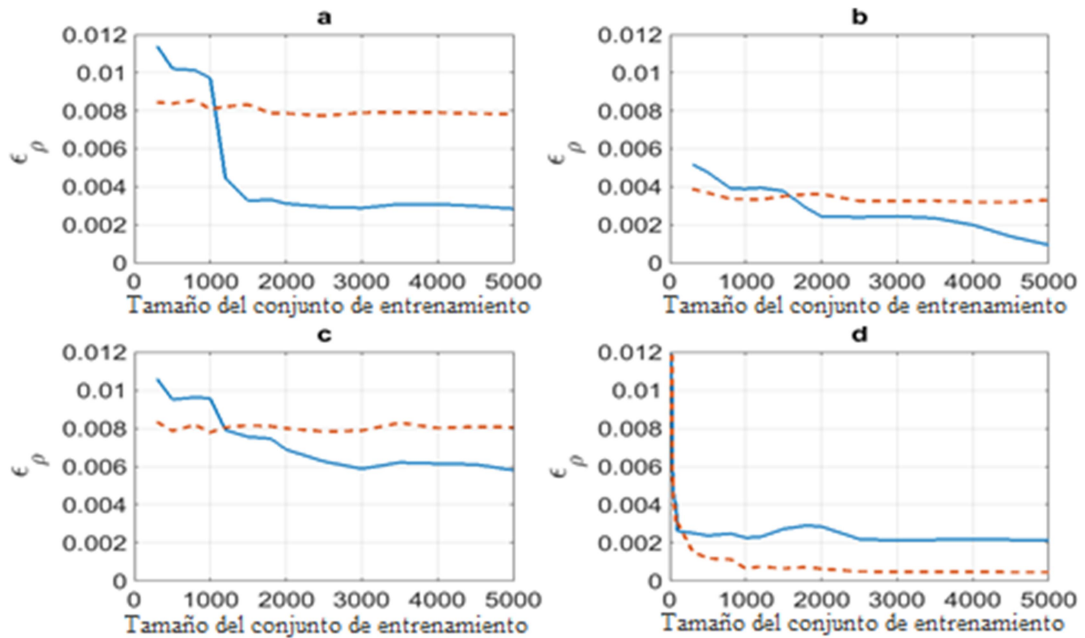
En este experimento se quiere evaluar la influencia del tamaño del conjunto de entrenamiento en la estimación de  $\rho_{nm}^{ICA-PCC}$  así como comparar la calidad de la estimación con la obtenida mediante la matriz de precisión. Para ello, se han generado datos sintéticos correspondientes a tres modelos ICA distintos. En el primero, las fuentes  $s_i$  son variables aleatorias independientes e idénticamente distribuidas (i.i.d.) que tienen una pdf uniforme de media cero con varianza unitaria. Esto corresponde a un ejemplo de distribución sub-gaussiana, ya que el exceso de curtosis es negativo  $\kappa - \kappa_G = -1.2$ , donde  $\kappa$  es la curtosis y  $\kappa_G = 3$  es la curtosis de una pdf gaussiana. En el segundo modelo, las fuentes  $s_i$  son variables aleatorias i.i.d. que tienen una pdf Laplaciana de media cero con varianza unitaria. Esto corresponde a un ejemplo de distribución súper-gaussiana, ya que el exceso de curtosis es positivo  $\kappa - \kappa_G = 3$ . En el tercer modelo algunas fuentes son uniformes y el resto son Laplacianas. Finalmente también se ha considerado el caso gaussiano generando fuentes con una pdf gaussiana estándar. La Figura 3.1 muestra los errores correspondientes a la estimación de  $\rho_{nm}^{ICA-PCC}$  para los cuatro modelos. Cada curva es una ponderación de 10 curvas correspondientes a 10 ejecuciones distintas. En cada ejecución, se ha seleccionado aleatoriamente una matriz ICA  $\mathbf{U} = \mathbf{W}^{-1}$ : cada entrada se ha obtenido muestreando una pdf gaussiana estándar. También se han generado un diferente número de vectores de entrenamiento  $\mathbf{x}$  a partir del vector de fuentes  $\mathbf{s}$  teniendo componentes independientes muestreados de las pdf marginales mencionadas: sub-gaussiana (Figura 3.1a), súper-gaussiana (Figura 3.1b), sub/súper-gaussiana mezcladas (Figura 3.1c) y gaussiana (Figura 3.1d). El error se ha calculado como

$$\epsilon_{\rho}^{ICA-PCC} = \frac{1}{N^2 - N} \sum_{n=1}^N \sum_{m \neq n} (|\rho_{nm}^{ICA-PCC}| - |\hat{\rho}_{mn}^{ICA-PCC}|)^2, \quad (31)$$

y promediando el resultado de 10 ejecuciones para cada tamaño de conjunto de entrenamiento. Obsérvese que  $0 \leq \epsilon_{\rho}^{ICA-PCC} \leq 1$ , ya que  $\epsilon_{\rho_{\min}}^{ICA-PCC} = 0$ , cuando  $|\rho_{nm}^{ICA-PCC}| = |\hat{\rho}_{nm}^{ICA-PCC}| \quad \forall n \quad \forall m \neq n$  y  $\epsilon_{\rho_{\max}}^{ICA-PCC} = 1$ , cuando  $|\rho_{nm}^{ICA-PCC}| - |\hat{\rho}_{nm}^{ICA-PCC}| = \pm 1 \quad \forall n \quad \forall m \neq n$ . In (31),  $\rho_{nm}^{ICA-PCC}$  se ha obtenido mediante el Algoritmo 3.1 utilizando la matriz  $\mathbf{W}$  real y un número extremadamente grande de instancias para la media muestral para calcular la esperanza en (29). Por otra parte,  $\hat{\rho}_{mn}^{ICA-PCC}$  se ha calculado mediante el Algoritmo 3.1 utilizando estimaciones de  $\mathbf{W}$  obtenidas con el conjunto de entrenamiento finito. Se ha utilizado el algoritmo *Extended Infomax* descrito en [7]. Para evaluar la esperanza en (29) se ha utilizado el mismo conjunto de entrenamiento finito. En todos los casos se ha considerado  $N=20$ . A modo de comparación, también se ha calculado el error

$$\epsilon_{\rho}^{PCC} = \frac{1}{N^2 - N} \sum_{n=1}^N \sum_{m \neq n} (|\rho_{nm}^{ICA-PCC}| - |\hat{\rho}_{mn}^{PCC}|)^2, \quad (32)$$

correspondiente a los PCCs obtenidos mediante estimaciones empíricas  $\hat{\mathbf{Q}}$  de la matriz de precisión como se indica en (3):  $\hat{\rho}_{nm}^{PCC} = -\hat{q}_{nm} / \sqrt{\hat{q}_{nm} \hat{q}_{mm}}$



**Figura 3.1.**  $\epsilon_p^{ICA-PCC}$  (continua) y  $\epsilon_p^{PCC}$  (discontinua) a) Caso sub-gaussiano b) Caso súper-gaussiano c) Caso sub/súper-gaussiano mezclado d) Caso gaussiano.

Nótese que también es  $0 \leq \epsilon_p^{PCC} \leq 1$ . Se pueden sacar varias conclusiones de la Figura 3.1. Primero, se puede observar que en los casos no gaussianos a) b) y c), el PCC no consigue disminuir el error al aumentar el tamaño del conjunto de entrenamiento. Esto demuestra el desajuste del modelo debido a la gaussianidad implícita de PCC. En estos tres casos, ICA-PCC mejora a PCC después de un número suficiente de muestras de entrenamiento y mantiene un error decreciente al aumentar el tamaño del conjunto de entrenamiento. Las curvas de convergencia son claramente diferentes para los casos sub-gaussianos y súper-gaussianos, y, como era de esperar, el caso mixto sub/súper-gaussiano presente un comportamiento intermedio. Finalmente, para el caso gaussiano, PCC produce un error muy pequeño, que disminuye al aumentar el tamaño del conjunto de entrenamiento. En este caso,



ICA-PCC es peor que PCC, aunque el error es razonablemente pequeño. Cabe recordar que, en el caso gaussiano, se esperaban resultados similares en ambos métodos, sin embargo, el camino seguido para la estimación ha sido distinto: en PCC la matriz de precisión se ha estimado directamente, mientras que en ICA-PCC las matrices  $\mathbf{W}$  y  $\mathbf{M}_{nm}$  han sido estimadas mediante el Algoritmo 3.1. Esta puede ser la explicación de la diferencia observada entre las dos curvas de error en la Figura 1d).

### **3.4.2. Una aplicación con datos reales**

Como se verá a continuación, el problema médico subyacente, así como el tipo y la configuración de las medidas realizadas, son básicamente coincidentes con las del experimento con datos reales presentado en el capítulo anterior. Sin embargo el objetivo último es diferente. Así, en el capítulo anterior se pretendía mejorar la estimación de una matriz Laplaciana global para cada paciente que eventualmente pudiera estar relacionada con algún parámetro de conectividad global del cerebro. Por ello se modelaba todo el conjunto de medidas (sueño normal y despertares) a partir de una mezcla de dos Gaussianas, calculándose la matriz Laplaciana mediante los diferentes métodos considerados. No se profundizó en la estimación de la conectividad global al no disponer de un *groundtruth* que permitiera valorar la calidad de dicha estima. Los métodos propuestos en este capítulo fueron también aplicados a dicho objetivo, no consiguiéndose mejorar los resultados del modelo de mezclas Gaussianas. Ello es indicativo de que el modelo ICA aplicado al conjunto global de datos no mejora los anteriores modelos. Interpretamos que cuando hay dos componentes distintas claramente definidas (nótese que cada componente corresponde a una clase distinta), el modelo GMM es capaz de capturar suficientemente la distribución global de todo el conjunto de datos ya que está concebido como un modelo de mezclas, cosa que no ocurre con el modelo ICA. Por ello, el

objetivo del experimento aquí presentado se orientó hacia un objetivo distinto. En este caso se estimaron Laplacianas separadamente para cada clase y a partir de ellas se obtuvieron medidas de conectividad. El objetivo era demostrar que era posible determinar variaciones en la conectividad cerebral cuando el paciente pasaba del modo de sueño normal al modo despertar y viceversa. Por lo tanto, en este enfoque es necesario ajustar un modelo separadamente para cada clase. Al no existir unas componentes claramente definidas en cada clase por separado, el modelo de mezclas GMM no mostró tan buen comportamiento, de hecho no mejoraba el modelo Gaussiano. Sin embargo el modelo ICA, sí permitió, como veremos ajustarse mejor a la distribución de los datos separadamente para cada clase. Mediante dicho modelo mostraremos que fue posible determinar variaciones en la conectividad entre clases que no eran perceptibles mediante el modelo Gaussiano. En consecuencia, se ha aplicado el método propuesto para cuantificar la importancia de los cambios en la conectividad cerebral durante el sueño de pacientes con trastornos como apnea o epilepsia [7]. Estos desórdenes están caracterizados por la aparición regular de despertares, que son etapas anormales de sueño degradado. La frecuencia de aparición de estos despertares en un periodo de tiempo dado está relacionada con la gravedad de la patología. Pero la intensidad de los despertares también puede ser relevante para un diagnóstico adecuado. Asumiendo que los despertares están asociados a cambios en la conectividad cerebral [8], una medida relacionada con el cambio de magnitud, puede ser útil para cuantificar la importancia de la patología. Para ello, el paciente fue monitorizado durante el sueño a través de 19 canales de registro EEG. La señal de cada canal fue segmentada en intervalos de 2 segundos y se obtuvo una característica calculada en cada intervalo y promediándolas en ciclos de 26 segundos. Cada ciclo se ha etiquetado de forma manual o automática [7] en dos posibles estados: sueño normal (estado 0) o “*arousal*” (estado 1). Por tanto, asociado a cada ciclo, un vector de observación  $\mathbf{x}$  se ha creado

con una característica extraída desde todos los canales (la misma característica para todos ellos), por lo que  $N = 19$ . En este experimento, hay disponibles un total de 2000 ciclos para cada estado. Dados estos conjuntos de datos, se ha calculado una medida promedio relacionada con la conectividad cerebral para cuantificar la importancia de los cambios cerebrales entre los dos estados definidos.

Hay muchas definiciones posibles de conectividad general, aquí se ha considerado la conectividad algebraica [9], que se puede calcular como el segundo autovalor más pequeño  $\lambda_2$  de la matriz Laplaciana del grafo [10]  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , siendo  $\mathbf{D}$  una matriz diagonal con las entradas  $d_{nn} = \sum_{m \neq n} a_{nm}$  y  $\mathbf{A}$  la matriz de adyacencia con entradas  $a_{nm} \geq 0$ . La matriz Laplaciana es semidefinida positiva con el autovalor más pequeño  $\lambda_1$  igual a cero, por lo que  $\lambda_2 \geq 0$ . Además, se demuestra en [9] que  $\lambda_2 = N$  para un grafo completo (un grafo con  $a_{nm} = 1 \quad \forall n \neq m$ ). También se demuestra en [12] que

$$\lambda_2 \leq \frac{N}{N-1} \min[d_{nn}] \quad . \quad (33)$$

Por lo tanto, suponiendo que  $0 \leq a_{nm} \leq 1$  (como será este caso), la mayor cota superior para  $\lambda_2$  en (33) corresponde al grafo completo ( $a_{nm} = 1 \quad \forall n \neq m \Rightarrow d_{nn} = N - 1 \quad \forall n$ ), por tanto  $0 \leq \lambda_2 \leq N$ . En consecuencia, se propone una versión normalizada de  $\lambda_2$  para medir la conectividad

$$\zeta = \frac{\lambda_2}{N} \quad 0 \leq \zeta \leq 1 \quad . \quad (34)$$

El límite inferior  $\zeta = 0$  corresponde a un grafo discontinuo, ya que implica un orden de multiplicidad del autovalor más pequeño mayor que 1. El límite superior corresponde a un grafo completo ya que es el que tendría máxima conectividad

bajo la restricción  $0 \leq a_{nm} \leq 1$ . Se han obtenido estimaciones de conectividad para cada estado (0 o 1) y para cada método (ICA-PCC o PCC):  $\hat{\zeta}_0^{ICA-PCC}, \hat{\zeta}_1^{ICA-PCC}, \hat{\zeta}_0^{PCC}, \hat{\zeta}_1^{PCC}$ . Esto se ha hecho a partir de (34) con  $N=19$ , después de haber calculado el segundo autovalor más pequeño de la matriz Laplaciana, considerando que las entradas de la matriz de adyacencia asociada son las respectivas magnitudes de las estimaciones de correlación parcial obtenidas del conjunto de entrenamiento 0 o 1:

$$\begin{aligned} a_{nm0}^{ICA-PCC} &= |\hat{\rho}_{nm0}^{ICA-PCC}|, & a_{nm1}^{ICA-PCC} &= |\hat{\rho}_{nm1}^{ICA-PCC}| \\ a_{nm0}^{PCC} &= |\hat{\rho}_{nm0}^{PCC}|, & a_{nm1}^{PCC} &= |\hat{\rho}_{nm1}^{PCC}| \end{aligned} \quad (35)$$

Se han considerado por separado dos características diferentes. La primera es la “amplitud” (*Amp*) considerada como el valor máximo de amplitud correspondiente a 2 segundos de intervalo. La segunda es la “alfa-slow-index” (*Asi*) que es el ratio entre la potencia de la banda alfa (8.0–11 Hz) combinado con la potencia en las bandas delta (0.5–3.5 Hz) y theta (3.5–8.0 Hz). La Tabla 3.1 y Tabla 3.2 muestran los resultados para la características *Amp* y *Asi* respectivamente para 6 pacientes distintos. Junto a la conectividad normalizada se ha incluido la variación de conectividad entre los estados definida como  $\Delta^{ICA-PCC} = |\hat{\zeta}_1^{ICA-PCC} - \hat{\zeta}_0^{ICA-PCC}|$  y  $\Delta^{PCC} = |\hat{\zeta}_1^{PCC} - \hat{\zeta}_0^{PCC}|$ . Además se ha incluido la estimación de la curtosis para cada paciente y estado. Esta estimación se ha obtenido como la media de las 19 curtosis empíricas calculadas por separado para cada componente del vector  $\mathbf{x}$ , es decir, la curtosis empírica de las distribuciones marginales de  $\mathbf{x}$ . Hay que tener en cuenta que la curtosis estimada está claramente por encima de la referencia gaussiana  $\kappa_G = 3$ , por lo que la suposición de gaussianidad no se cumple en este caso.

Suj.	$\kappa_0$	$\kappa_1$	$\hat{\zeta}_0^{ICA-PCC}$	$\hat{\zeta}_1^{ICA-PCC}$	$\Delta^{ICA-PCC}$	$\hat{\zeta}_0^{PCC}$	$\hat{\zeta}_1^{PCC}$	$\Delta^{PCC}$
S1	6,46	4,58	0.30	0.33	<b>0.03</b>	0.03	0.03	0.00
S2	8,05	5,29	0.74	0.39	<b>0.35</b>	0.04	0.04	0.00
S3	9,84	6,76	0.57	0.28	<b>0.29</b>	0.03	0.02	0.01
S4	9,04	8,87	0.39	0.66	<b>0.27</b>	0.04	0.02	0.02
S5	9,61	15,13	0.31	0.44	<b>0.13</b>	0.02	0.03	0.01
S6	9,14	13,82	0.24	0.36	<b>0.12</b>	0.02	0.02	0.00

**Tabla 3.1.** Resultados correspondientes a la característica *Amp*.

Suj.	$\kappa_0$	$\kappa_1$	$\hat{\zeta}_0^{ICA-PCC}$	$\hat{\zeta}_1^{ICA-PCC}$	$\Delta^{ICA-PCC}$	$\hat{\zeta}_0^{PCC}$	$\hat{\zeta}_1^{PCC}$	$\Delta^{PCC}$
S1	16,32	22,51	0.31	0.51	<b>0.20</b>	0.02	0.02	0.00
S2	10,52	9,09	0.34	0.60	<b>0.26</b>	0.02	0.02	0.00
S3	9,91	7,05	0.68	0.48	<b>0.20</b>	0.02	0.03	0.01
S4	8,39	11,69	0.37	0.74	<b>0.37</b>	0.03	0.02	0.01
S5	7,72	13,15	0.22	0.71	<b>0.49</b>	0.02	0.03	0.01
S6	11,86	9,24	0.43	0.56	<b>0.13</b>	0.02	0.03	0.01

**Tabla 3.2.** Resultados correspondientes a la característica *Asi*.

Como se puede ver en la Tabla 3.1 y Tabla 3.2, el método PCC obtiene valores de conectividad muy pequeños para todos los sujetos y estados, por lo que no es sensible a posibles cambios de estado. Sin embargo, ICA-PCC proporciona valores de conectividad altos y cambios entre estados significativos. La Figura 3.2 muestra las matrices de adyacencia estimadas correspondientes a los diferentes sujetos, métodos y estados. En esta figura se puede observar que las magnitudes de PCC

son en general más bajas que las magnitudes ICA-PCC, por lo que PPC tiene mayores dificultades para revelar las interrelaciones entre los diferentes canales de EEG debido a la actividad cerebral. Esto puede explicarse en términos de los residuos  $e_n$  y  $e_m$ . Nótese de (12) que

$$E[e_n^2] = \sum_{i=1}^N mse_{nmi} (v_{1nmi}^+)^2 \quad E[e_m^2] = \sum_{i=1}^N mse_{nmi} (v_{2nmi}^+)^2, \quad (36)$$

donde  $\{v_{1nmi}^+\}$  y  $\{v_{2nmi}^+\}$  son respectivamente los elementos de los vectores  $\mathbf{v}_{1nm}^+$  y  $\mathbf{v}_{2nm}^+$ , y estos últimos son respectivamente la primera y la segunda fila de  $(\mathbf{W}\mathbf{T}_{nm})^+$ . Se ha mostrado en la Sección 3 que PCC debería ser similar a ICA-PCC para  $mse_{nmi} = mse_{nmi}^l$ , pero para observaciones no gaussianas  $mse_{nmi} < mse_{nmi}^l$ , por lo que se deduce de (36) que

$$\begin{aligned} E[e_n^2] &\leq E^l[e_n^2] = \sum_{i=1}^N mse_{nmi}^l (v_{1nmi}^+)^2 \\ E[e_m^2] &\leq E^l[e_m^2] = \sum_{i=1}^N mse_{nmi}^l (v_{2nmi}^+)^2 \end{aligned}, \quad (37)$$

donde se cumple la igualdad en el caso gaussiano. Por tanto, PCC proporciona residuos sobreestimados donde la correlación parcial real entre  $x_n$  y  $x_m$  puede eventualmente quedar ocultada. Este efecto de enmascaramiento debería aumentar con la no gaussianidad de las observaciones. En este experimento, las características son altamente no gaussianas tal y como demuestran los valores de curtosis de las tablas 1 y 2. Por ello, cuando se utiliza PCC, los “verdaderos” residuos parecen estar sobreestimados por residuos bastante no correlacionados, lo que proporciona una estimación demasiado baja de la interrelación real entre los diferentes canales

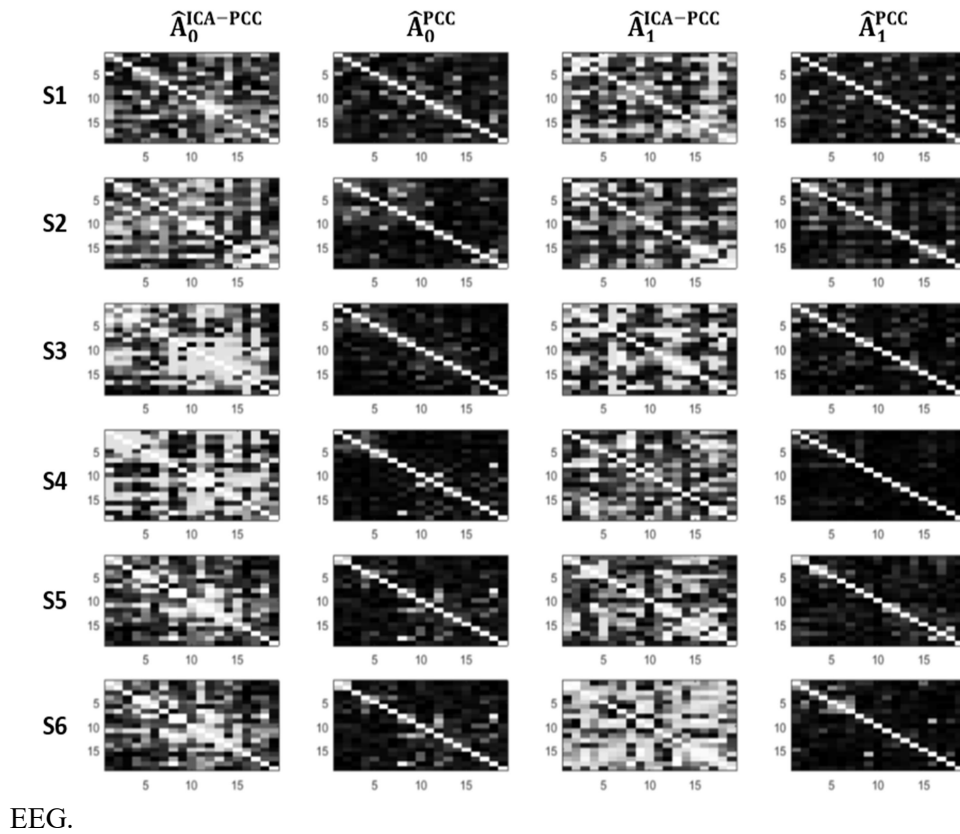


Figura 3.2. Matrices de adyacencia correspondientes a la característica *Amp*.

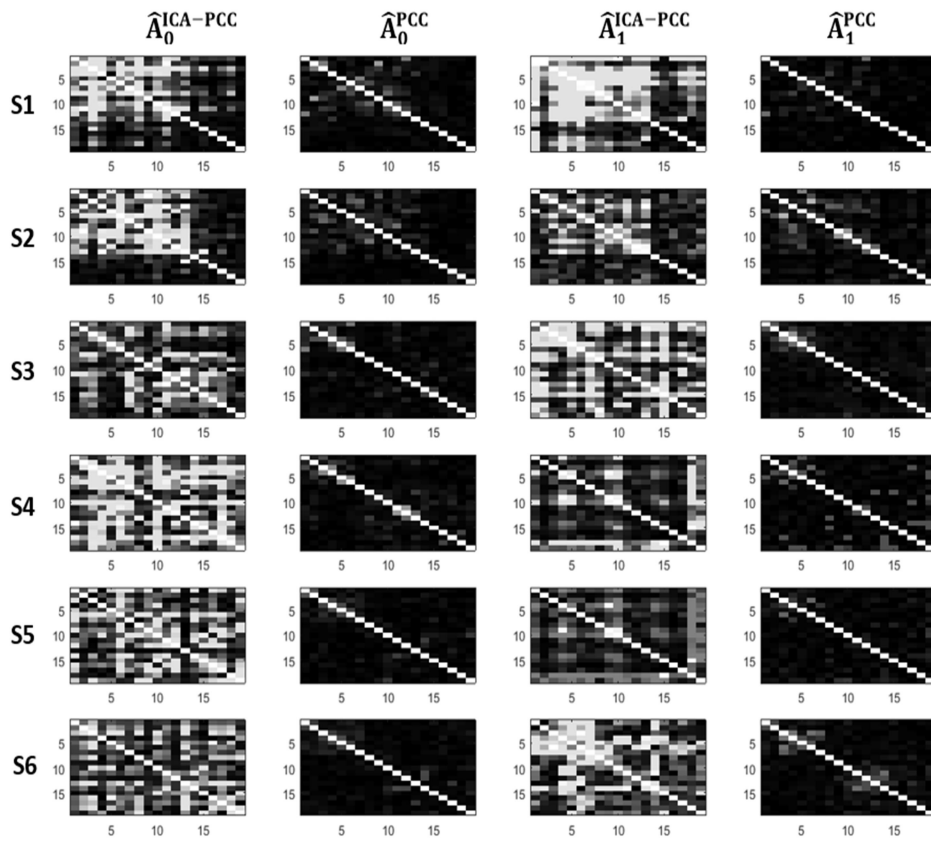


Figura 3.2. Matrices de adyacencia correspondientes a la característica  $As_i$ .



### 3.5. Conclusiones

Las correlaciones parciales pueden utilizarse para definir las conexiones de un grafo no dirigido para el posterior procesado de la señal sobre grafos. Convencionalmente, las correlaciones parciales se obtienen de la matriz de precisión, pero esta forma es óptima sólo bajo el supuesto de gaussianidad. Por lo tanto, se ha propuesto un nuevo método para calcular la correlación parcial, asumiendo un modelo no gaussiano (ICA). Este método resulta ser un modelo versátil que se adapta a una diversidad de pdfs no gaussianas.

El método propuesto requiere del cálculo de los parámetros del modelo ICA, que pueden ser obtenidos mediante cualquier algoritmo existente para este propósito. Además, también precisa del cálculo de los errores cuadráticos medios correspondientes a la estimación óptima de las fuentes. Por tanto, se ha propuesto una aproximación de segundo orden de la media condicionada. Se podrían intentar órdenes superiores con la implicación de una mayor complejidad.

Se ha verificado, mediante simulaciones y experimentos con datos reales, que el nuevo método captura mejor la conectividad por pares y general del grafo que mediante la matriz de precisión en escenarios no gaussianos.

Este trabajo se puede extender en líneas futuras. Sería deseable algún tipo de regularización con tal de enfatizar la información relevante proporcionada por la conectividad del grafo y/o establecer relaciones más naturales entre los nodos conectados. Por ejemplo [11] un grafo tiene más interés si manifiesta conexiones significativas entre un número reducido de nodos (condición denominada *sparsity* que podemos traducir como “escasez”). Teniendo en cuenta la ecuación (10), la *sparsity* podría imponerse seleccionando sólo aquellas fuentes que contribuyan significativamente a la correlación parcial entre  $x_n$  y  $x_m$ , es decir, mediante un umbral bajo o alto en  $mse_{nm}$ . Por otro lado, la regularización podría intentarse de manera simi-

lar al enfoque propuesto en [12] para el caso gaussiano. Para ello, se podría considerar que la matriz  $U$  puede descomponerse en una matriz de correlación multiplicada por una matriz de rotación (unitaria) [13]. Entender la relación de esta rotación con la conectividad del grafo puede permitir definir funciones de coste que incluyan algunos términos relacionados con la denominada *smoothness* (“suavidad”). Otras restricciones estructurales [14] también podrían ser compatibles con el modelo no gaussiano.

### A. Apéndice: Derivación de la fórmula general

Debido a la ortogonalidad entre el error y la media condicionada, se puede escribir

$$\mathbf{C}_{e_{nm}e_{nm}} = E[\mathbf{x}_{nm}\mathbf{x}_{nm}^T] - E[E[\mathbf{x}_{nm}/\mathbf{x}_{-nm}]E^T[\mathbf{x}_{nm}/\mathbf{x}_{-nm}]]. \quad (\text{A1})$$

Definiendo

$$\mathbf{f}(\mathbf{x}_{-nm}) = E[\mathbf{s}/\mathbf{x}_{-nm}] \equiv \mathbf{f} \quad \mathbf{q}(\mathbf{x}_{-nm}) = -\mathbf{W}\mathbf{T}_{-nm}\mathbf{x}_{-nm} = \mathbf{q}, \quad (\text{A2})$$

Obteniendo de (9), se puede escribir  $E[\mathbf{x}_{nm}/\mathbf{x}_{-nm}] = (\mathbf{W}\mathbf{T}_{nm})^+(\mathbf{f} + \mathbf{q})$ , por lo que

$$\begin{aligned} \mathbf{C}_{e_{nm}e_{nm}} &= \\ &= E[\mathbf{x}_{nm}\mathbf{x}_{nm}^T] - E[(\mathbf{W}\mathbf{T}_{nm})^+(\mathbf{f} + \mathbf{q})(\mathbf{f}^T + \mathbf{q}^T)((\mathbf{W}\mathbf{T}_{nm})^+)^T] \\ &= \underbrace{E[\mathbf{x}_{nm}\mathbf{x}_{nm}^T]}_1 + \underbrace{(\mathbf{W}\mathbf{T}_{nm})^+ E[\mathbf{q}\mathbf{q}^T] ((\mathbf{W}\mathbf{T}_{nm})^+)^T}_2 \\ &\quad + \underbrace{(\mathbf{W}\mathbf{T}_{nm})^+ E[\mathbf{f}\mathbf{f}^T] ((\mathbf{W}\mathbf{T}_{nm})^+)^T}_3 \\ &\quad + \underbrace{(\mathbf{W}\mathbf{T}_{nm})^+ E[\mathbf{f}\mathbf{q}^T] ((\mathbf{W}\mathbf{T}_{nm})^+)^T + (\mathbf{W}\mathbf{T}_{nm})^+ E^T[\mathbf{q}\mathbf{f}^T] ((\mathbf{W}\mathbf{T}_{nm})^+)^T}_4 \end{aligned} \quad (\text{A3})$$

Se ha de calcular los 4 términos de (A3).

-Término 1

$$E[\mathbf{x}_{nm}\mathbf{x}_{nm}^T] = \mathbf{C}_{x_{nm}x_{nm}} = \mathbf{T}_{nm}^T \mathbf{C}_{xx} \mathbf{T}_{nm} \quad (\text{A4})$$

-Término 2

$$\begin{aligned}
 E[\mathbf{q}\mathbf{q}^T] &= E[\mathbf{W}\mathbf{T}_{-nm}\mathbf{x}_{-nm}\mathbf{x}_{-nm}^T\mathbf{T}_{-nm}^T\mathbf{W}^T] \quad ( ) \\
 &= \mathbf{W}\mathbf{T}_{-nm}\mathbf{C}_{\mathbf{x}_{-nm}\mathbf{x}_{-nm}}\mathbf{T}_{-nm}^T\mathbf{W}^T \\
 &= \mathbf{W}\mathbf{T}_{-nm}\mathbf{T}_{-nm}^T\mathbf{C}_{xx}\mathbf{T}_{-nm}\mathbf{T}_{-nm}^T\mathbf{W}^T \\
 &= \mathbf{W}(\mathbf{I}-\mathbf{T}_{nm}\mathbf{T}_{nm}^T)\mathbf{C}_{xx}(\mathbf{I}-\mathbf{T}_{nm}\mathbf{T}_{nm}^T)\mathbf{W}^T \\
 &= \mathbf{W}\mathbf{C}_{xx}\mathbf{W}^T + \mathbf{W}\mathbf{T}_{nm}\mathbf{T}_{nm}^T\mathbf{C}_{xx}\mathbf{T}_{nm}\mathbf{T}_{nm}^T\mathbf{W}^T \\
 &\quad - \mathbf{W}\mathbf{T}_{nm}\mathbf{T}_{nm}^T\mathbf{C}_{xx}\mathbf{W}^T - \mathbf{W}\mathbf{C}_{xx}\mathbf{T}_{nm}\mathbf{T}_{nm}^T\mathbf{W}^T \\
 &(\mathbf{W}\mathbf{T}_{nm})^+ E[\mathbf{q}\mathbf{q}^T] (\mathbf{W}\mathbf{T}_{nm})^+{}^T \\
 &= (\mathbf{W}\mathbf{T}_{nm})^+ ((\mathbf{W}\mathbf{T}_{nm})^+)^T + \mathbf{T}_{nm}^T\mathbf{C}_{xx}\mathbf{T}_{nm} \\
 &\quad - \mathbf{T}_{nm}^T\mathbf{W}^{-1}((\mathbf{W}\mathbf{T}_{nm})^+)^T - (\mathbf{W}\mathbf{T}_{nm})^+(\mathbf{W}^{-1})^T\mathbf{T}_{nm}
 \end{aligned} \tag{A5}$$

-Término 3

$$\begin{aligned}
 &i \neq j \\
 E[\mathbf{f}\mathbf{f}^T](i, j) &= E[E[s_i/\mathbf{x}_{-nm}]E[s_j/\mathbf{x}_{-nm}]] \\
 &= E[E[s_i/\mathbf{x}_{-nm}]]E[E[s_j/\mathbf{x}_{-nm}]] = E[s_i]E[s_j] = 0
 \end{aligned}$$

$$\begin{aligned}
 &i = j \\
 E[\mathbf{f}\mathbf{f}^T](i, i) &= E[E^2[s_i/\mathbf{x}_{-nm}]] \\
 &= \text{var}[E[s_i/\mathbf{x}_{-nm}]] + E^2[E[s_i/\mathbf{x}_{-nm}]] \\
 &= \text{var}[E[s_i/\mathbf{x}_{-nm}]] + E^2[s_i] = \text{var}[E[s_i/\mathbf{x}_{-nm}]]
 \end{aligned} \tag{A6}$$

Expresando  $\text{var}[E[s_i/\mathbf{x}_{-nm}]]$  en términos de  $mse_i$  tal y como viene definido en (A7)

$$\begin{aligned}
 mse_{nmi} &= E\left[\left(s_i - E[s_i/\mathbf{x}_{-nm}]\right)^2\right] \\
 &= E[s_i^2] + E\left[E^2[s_i/\mathbf{x}_{-nm}]\right] - 2E[s_i E[s_i/\mathbf{x}_{-nm}]] \\
 E\left[\left(s_i - E[s_i/\mathbf{x}_{-nm}]\right)E[s_i/\mathbf{x}_{-nm}]\right] &= 0 \\
 \Rightarrow E[s_i E[s_i/\mathbf{x}_{-nm}]] &= E\left[E[s_i/\mathbf{x}_{-nm}]E[s_i/\mathbf{x}_{-nm}]\right] \\
 \Rightarrow mse_{nmi} &= E[s_i^2] - E\left[E_i^2[s_i/\mathbf{x}_{-nm}]\right] = 1 - \text{var}\left[E[s_i/\mathbf{x}_{-nm}]\right]
 \end{aligned} \tag{A7}$$

donde se ha tenido en cuenta que la media condicionada es un estimador insesgado. Por lo que se tiene  $E[\mathbf{ff}^T](i,i) = \text{var}\left[E[s_i/\mathbf{x}_{-nm}]\right] = 1 - mse_{nmi}$  y definiendo  $\mathbf{M}_{nm}$  como una matriz diagonal ( $N \times N$ ) en la que los valores  $mse_{nmi}$   $i=1\dots N$  están en su diagonal principal, se puede escribir

$$(\mathbf{W}\mathbf{I}_{nm})^+ E[\mathbf{ff}^T] ((\mathbf{W}\mathbf{I}_{nm})^+)^T = (\mathbf{W}\mathbf{I}_{nm})^+ (\mathbf{I} - \mathbf{M}_{nm}) ((\mathbf{W}\mathbf{I}_{nm})^+)^T. \tag{A8}$$

$$\begin{aligned}
 E[\mathbf{f}\mathbf{q}^T] &= -E[\mathbf{f}\mathbf{x}_{-nm}^T \mathbf{T}_{-nm}^T \mathbf{W}^T] \\
 E[E[s_i/\mathbf{x}_{-nm}]x_j] & \\
 &\stackrel{x_j \in \{\mathbf{x}_{-nm}\}}{=} \int_{\mathbf{x}_{-nm}} x_j E[s_i/\mathbf{x}_{-nm}] p(\mathbf{x}_{-nm}) d\mathbf{x}_{-nm} \\
 &= \int_{\mathbf{x}_{-nm}} x_j \left[ \int_{s_i} s_i p(s_i/\mathbf{x}_{-nm}) ds_i \right] p(\mathbf{x}_{-nm}) d\mathbf{x}_{-nm} \\
 &= \int_{s_i} s_i \left[ \int_{\mathbf{x}_{-nm}} x_j p(s_i/\mathbf{x}_{-nm}) p(\mathbf{x}_{-nm}) d\mathbf{x}_{-nm} \right] ds_i \\
 &= \int_{s_i} s_i \left[ \int_{\mathbf{x}_{-nm}} x_j p(\mathbf{x}_{-nm}/s_i) p(s_i) d\mathbf{x}_{-nm} \right] ds_i \\
 &= \int_{s_i} s_i p(s_i) \left[ \int_{\mathbf{x}_{-nm}} x_j p(\mathbf{x}_{-nm}/s_i) d\mathbf{x}_{-nm} \right] ds_i \\
 &= \int_{s_j} s_j p(s_j) \left[ \int_{x_j} x_j p(x_j/s_j) dx_j \right] ds_j \\
 &= \int_{s_j} \int_{x_j} s_j x_j p(s_j, x_j) ds_j dx_j = E[s_j x_j]
 \end{aligned}$$

$$\begin{aligned}
 E[\mathbf{f}\mathbf{q}^T] &= -E[\mathbf{f}\mathbf{x}_{-nm}^T \mathbf{T}_{-nm}^T] \mathbf{W}^T = E[\mathbf{s}\mathbf{x}_{-nm}^T \mathbf{T}_{-nm}^T] \mathbf{W}^T \\
 &= -E[\mathbf{W}\mathbf{x}\mathbf{x}_{-nm}^T \mathbf{T}_{-nm}^T] \mathbf{W}^T = -\mathbf{W}\mathbf{C}_{\mathbf{x}\mathbf{x}_{-nm}} \mathbf{T}_{-nm}^T \mathbf{W}^T = \\
 &= -\mathbf{W}\mathbf{C}_{\mathbf{x}\mathbf{x}} \mathbf{T}_{-nm} \mathbf{T}_{-nm}^T \mathbf{W}^T = -(\mathbf{W}^{-1})^T (\mathbf{I} - \mathbf{T}_{nm} \mathbf{T}_{nm}^T) \mathbf{W}^T \\
 &= (\mathbf{W}^{-1})^T \mathbf{T}_{nm} \mathbf{T}_{nm}^T \mathbf{W}^T - \mathbf{I} \\
 (\mathbf{W}\mathbf{T}_{nm})^+ E[\mathbf{f}\mathbf{q}^T] ((\mathbf{W}\mathbf{T}_{nm})^+)^T & \quad \quad \quad (A9) \\
 &= (\mathbf{W}\mathbf{T}_{nm})^+ (\mathbf{W}^{-1})^T \mathbf{T}_{nm} - (\mathbf{W}\mathbf{T}_{nm})^+ ((\mathbf{W}\mathbf{T}_{nm})^+)^T \\
 (\mathbf{W}\mathbf{T}_{nm})^+ E^T[\mathbf{f}\mathbf{q}^T] ((\mathbf{W}\mathbf{T}_{nm})^+)^T & \\
 &= \mathbf{T}_{nm}^T \mathbf{W}^{-1} ((\mathbf{W}\mathbf{T}_{nm})^+)^T - (\mathbf{W}\mathbf{T}_{nm})^+ ((\mathbf{W}\mathbf{T}_{nm})^+)^T
 \end{aligned}$$

Finalmente, considerando (A3), (A4), (A5), (A6), (A8) se puede escribir:

$$\begin{aligned}
 \mathbf{C}_{\epsilon_{nm}\epsilon_{nm}} &= \mathbf{T}_{nm}^T \mathbf{C}_{xx} \mathbf{T}_{nm} \\
 &- (\mathbf{W}\mathbf{T}_{nm})^+ \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T - \mathbf{T}_{nm}^T \mathbf{C}_{xx} \mathbf{T}_{nm} \\
 &+ \mathbf{T}_{nm}^T \mathbf{W}^{-1} \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T + (\mathbf{W}\mathbf{T}_{nm})^+ (\mathbf{W}^{-1})^T \mathbf{T}_{nm} \\
 &- (\mathbf{W}\mathbf{T}_{nm})^+ (\mathbf{I} - \mathbf{M}_{nm}) \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T \quad . \quad (A10) \\
 &- (\mathbf{W}\mathbf{T}_{nm})^+ (\mathbf{W}^{-1})^T \mathbf{T}_{nm} + (\mathbf{W}\mathbf{T}_{nm})^+ \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T \\
 &- \mathbf{T}_{nm}^T \mathbf{W}^{-1} \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T + (\mathbf{W}\mathbf{T}_{nm})^+ \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T \\
 &= (\mathbf{W}\mathbf{T}_{nm})^+ (\mathbf{M}_{nm}) \left( (\mathbf{W}\mathbf{T}_{nm})^+ \right)^T
 \end{aligned}$$

## Referencias

- [1] T.W. Lee, *Independent Component Analysis: Theory and Applications*, 1st ed. Norwell: Kluwer, 1998.
- [2] A. Hyvriinen, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411-430, 2000.
- [3] C. Jutten P. Common, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. USA: Academic Press, 2010.
- [4] L. Vergara, A. Serrano, J. Igual A. Salazar, "A General Procedure for Learning Mixtures of Independent Component Analyzers," *Pattern Recognition*, vol. 43, pp. 69-85, 2010.
- [5] P. Bernabeu L. Vergara, "Simple approach to nonlinear prediction," *Electron. Lett.*, vol. 37, pp. 928-936, 2001.
- [6] E. Celebi, "General formula for conditional mean using higher-order statistics," *Electron. Lett.*, vol. 33, pp. 2097-2099, 1997.
- [7] L. Vergara, R. Miralles A. Salazar, "On including sequential dependence in ICA mixture models," *Signal Processing*, vol. 90, pp. 2314-2318, 2010.
- [8] A. Tomé, I.R. Keck, J. Górriz-Sáez, C. Puntonet E.W. Lang, "Brain connectivity analysis: a short survey," *Computational intelligence and Neuroscience*, vol. 2012, pp. 1-21, 2012.
- [9] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, pp. 298-305, 1973.
- [10] R. Merris, "Laplacian matrices of a graph: a survey," *Linear Alg. Applicat.*, vol. 197, p. 143-176, 1994.
- [11] B. Lake and J. Tenenbaum, "Discovering structure by learning sparse graph," in *Proc. 32nd Annu. Meet. Cogn. Sci. Soc.*, 2010, p. 778-783.
- [12] D. Thanou, P. Frossard and P. Vandergheynst X. Dong, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. on Signal Processing*, vol. 64, p. 6160-1673, 2016.
- [13] L. Vergara, J. Gosálbez J. Moragues, "Generalized matched subspace filter for nonindependent noise based on ICA," *IEEE Trans. on Signal Processing*, vol. 59,



pp. 3430-3434, 2011.

- [14] E. Pavez, A. Ortega H. E. Egilmez, "Graph learning from data under Laplacian and structural constraints," *IEEE J. Sel. Topics Signal Processing*, vol. 11, no. 6, p. 825–841, 2017.
- [15] M. Girolami, T.J. Sejnowski. T.W. Lee, "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources," *Neural Computation*, vol. 11, pp. 409-433, 1999.

# Capítulo 4

## Un nuevo algoritmo de subrogación para la Transformada de Fourier Compleja sobre Grafos (CGFT)

### 4.1. Introducción

#### 4.1.1. Planteamiento del problema y trabajos relacionados

A partir de una señal disponible, se pueden generar señales sintéticas que sean subrogadas de la misma. Esencialmente, el concepto de señal subrogada significa que la amplitud<sup>1</sup> del espectro (Fourier) de la señal original se retiene mientras la fase se aleatoriza. Los métodos de subrogación han sido considerados en diferentes aplicaciones [1]-[7], donde las señales sintéticas se han utilizado para obtener una estimación no paramétrica de la distribución de alguna estadística considerada en test de hipótesis. A pesar de su enfoque atractivo, los métodos de subrogación existentes están demasiado restringidos debido al uso de la Transformada de Fourier (FT): tanto

---

<sup>1</sup> Se usará el término amplitud para definir la magnitud o el módulo de un número complejo o de un número real.

la señal original como sus subrogadas comparten la misma función de autocorrelación. Esto implica que se asuma que las interrelaciones por pares son las mismas para muestras igualmente separadas. Si bien esto puede ser apropiado para algunos casos, se podría obtener una mayor flexibilidad si se pudieran imponer relaciones específicas por pares entre las muestras subrogadas. Así, por ejemplo, se podrían mejorar las propiedades a corto plazo priorizando las interrelaciones entre muestras vecinas. También se podrían introducir restricciones de interrelación de dominio de aplicación dadas por un experto que debe estar de acuerdo con las subrogaciones. De hecho, se mostrará que los métodos FT son casos particulares del nuevo método propuesto.

En este capítulo se propone un nuevo método de subrogación que explota los conceptos de señal sobre grafos [8]-[10]. Un trabajo anterior relacionado es [11], donde los autores proponen una aleatorización de los signos en la transformada de Fourier sobre grafos (GFT). La GFT representa la señal sobre grafos en un dominio expandido por los autovectores de la matriz<sup>2</sup> Laplaciana [8], [12]. El algoritmo se aplica para generar subrogadas de señales EEG definidas en el dominio espacial, que conservan la *smoothness* de las señales originales en el grafo definido. Los autores demuestran el interés de este método en la detección de los efectos de tareas conductuales en señales de electroencefalografía. Los buenos resultados de este novedoso enfoque sugieren seguir trabajando y profundizando en él.

#### ***4.1.2. Nuevas contribuciones y organización del capítulo***

En este capítulo se va a profundizar en diferentes aspectos del concepto. En primer lugar se considerará el caso complejo. Nótese que la GFT de una señal sobre grafos real es también real, por lo que la aleatorización de la fase en realidad está limitada a

---

<sup>2</sup> Otras GFT son posibles como la propuesta en [13], que considera los autovectores generalizados de la matriz de adyacencia. Esto puede conducir a nuevas extensiones del trabajo presentado en esta tesis.

una aleatorización de signo tal y como hacen los autores en [11]. Sin embargo, trabajar con señales complejas permite una aleatorización de fase sin restricciones. Hay que tener en cuenta que una señal compleja debe expresarse como una combinación lineal de autovectores complejos. Mientras que la mayor parte del trabajo previo sobre GSP asume matrices Laplacianas reales, en esta tesis se ha definido una matriz Laplaciana Hermítica compleja (Sección 4.2.1). Derivando propiedades de la Laplaciana Hermítica, en particular una generalización de la *smoothness* [8] [14]-[16] relativa a los pesos complejos conjugados. Así pues, se ha generalizado el GFT para el caso complejo (Sección 4.2.2) definiendo una GFT Compleja (CGFT).

Además, con el objetivo de entender el algoritmo de subrogación propuesto, se ha realizado un análisis de las propiedades de las señales originales que se conservan en las señales subrogadas. Esto se presenta en la Sección 4.3 a través del concepto de invariantes de la Amplitud del Espectro sobre Grafos (GSA). En particular, se demuestra que la *smoothness*, la matriz de precisión [17], [18] y la matriz de covarianza de las señales sobre grafos estacionarias en sentido amplio (GWSS, *Graph Wide-Sense Stationary*), tal como se han definido recientemente en [19], [20] son GSA invariantes.

La Sección 4.2 presenta el concepto de matriz Laplaciana Hermítica, que permite llegar a una definición de la CGFT. La Sección 4.3 está dedicada a los invariantes de la GSA, es decir a propiedades de la señal sobre grafos original que se preservarán en la subrogada. La Sección 4.4 presenta los nuevos algoritmos de subrogación basados en el uso de la CGFT. Finalmente, la Sección 4.5 presenta experimentos para evaluar las mejoras de los nuevos métodos de subrogación en comparación con los basados en FT. El objetivo de los experimentos es el entrenamiento de un clasificador con escasez de datos. Esto se puede considerar como una generalización del problema clásico de test de hipótesis en el que habitualmente se consideran las subrogaciones. Así se presenta un experimento de simulación para la clasificación de vectores alea-

torios complejos pertenecientes a dos GMM multivariantes diferentes. Luego, se presentan experimentos de reconocimiento de gestos realizados con las manos. Este es un problema de clasificación difícil debido a la gran cantidad de clases, y a la gran cantidad y variedad de características.

## **4.2. La matriz Laplaciana Hermítica**

### **4.2.1. Definición y propiedades**

No hay muchos precedentes del uso de matrices Laplacianas complejas. En [21] se propone una matriz Laplaciana Hermítica para incluir algunas posibles orientaciones de las conexiones, pero esto conduce a una forma muy limitada de matrices Laplacianas Hermíticas: sus elementos fuera de la diagonal son 1 para las conexiones no orientadas y  $\pm j$  para las conexiones orientadas. En [22] se propone una matriz Laplaciana Hermítica más general añadiendo un término de fase a los elementos fuera de la diagonal de la matriz Laplaciana real. Esta extensión implica que no se garantiza que la parte real de la matriz de adyacencia correspondiente tenga entradas positivas. En este trabajo se propone una matriz Laplaciana Hermítica diferente añadiendo una parte imaginaria a los elementos de la parte (positiva) fuera de la diagonal de la matriz de adyacencia, mientras los elementos  $n$ -ésimos de la matriz diagonal  $\mathbf{D}$  se mantiene como la suma de las partes reales de las conexiones incidentes al nodo  $n$ . Así, considerando el grafo ponderado  $G\{V, E, \mathbf{A}\}$ , el elemento genérico  $\mathbf{A}(n, m) = a_{nm}$  es el peso correspondiente a la conexión que une el nodo  $m$  al nodo  $n$ .  $\mathbf{A}$  se asume que es una matriz simétrica y real con elementos no negativos  $a_{nm} \geq 0 \quad \forall n, m$ . Si bien esto parece razonable para señales sobre un grafo real, puede ser demasiado limitado para señales sobre grafos complejos. Así, se considerará que  $\mathbf{A}$  es en general una matriz Hermítica  $\mathbf{A} = \mathbf{A}^H$ ,  $a_{nm} = a_{nm}^*(n, m) \quad \forall n, m$ . También se

considerará que la parte real es semipositiva  $\text{Re}[a_{nm}] \geq 0 \quad \forall n, m$ . Esto permite incluir el caso real como un caso especial. Nótese que en el caso real todas las conexiones son no dirigidas porque  $a_{nm} = a_{mn} \quad \forall n, m$ . Sin embargo, en el caso complejo, la parte imaginaria de los pesos complejos cambian el signo dependiendo de la dirección  $\text{Im}[a_{nm}] = -\text{Im}[A(a_{nm})]$ , por lo que estas conexiones son dirigidas. Esta situación a veces se conoce como grafos mixtos [21].

Se puede definir una matriz Laplaciana Hermítica en la forma  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  donde  $\mathbf{D}$  es una matriz diagonal con el elemento genérico  $\mathbf{D}(n, n) = \sum_{m=1}^N \text{Re}[a_{nm}]$ . Esto es una generalización obvia del caso real donde  $\mathbf{D}(n, n) = \sum_{m=1}^N a_{nm}$  y  $\mathbf{D}$  se denomina matriz de grado. La definición anterior de la matriz Hermítica Laplaciana es apropiada para extender el concepto de *smoothness* a señales sobre grafos complejas. Definiendo la señal sobre grafos  $\mathbf{s} = [s_1, \dots, s_N]^T \in \mathbb{C}^N$ , donde el valor  $s_n$  está asociado al  $n$ -ésimo nodo del grafo, para  $n = 1 \dots N$ . Nótese que  $\mathbf{s}$  puede corresponder a cualquier dominio arbitrario.

La *smoothness* del grafo puede ser definida como una generalización simple del caso real

$$S(\mathbf{s}) = \mathbf{s}^H \mathbf{L} \mathbf{s} \quad . \quad (1)$$

Se muestra en el Apéndice A de este capítulo que

$$S(\mathbf{s}) = \mathbf{s}^H \mathbf{L} \mathbf{s} = \sum_{n=1}^{N-1} \sum_{m>n}^N a_{nm}^R |s_n - s_m|^2 + \sum_{n=1}^{N-1} \sum_{m>n}^N 2a_{nm}^I \text{Im}[s_n^* s_m], \quad (2)$$

Nótese que si la matriz Laplaciana y/o la señal sobre el grafo es real, el segundo término en (2) desaparece, y la *smoothness* se convierte en la medida ponderada clásica de las diferencias (cuadradas) de señal entre cada dos nodos. Pero en el caso más

general donde tanto la matriz Laplaciana como la señal sobre grafos pueden ser complejas, aparece un nuevo término:

$$\sum_{n=1}^{N-1} \sum_{m>n}^N 2a'_{nm} \operatorname{Im}[s_n^* s_m] = \sum_{n=1}^{N-1} \sum_{m>n}^N 2a'_{nm} |s_n| |s_m| \sin(\angle s_m - \angle s_n), \quad (3)$$

donde  $s_n = |s_n| \exp(j\angle s_n)$ . Por lo tanto, al utilizar la matriz Laplaciana Hermítica así definida, se está incorporando a la *smoothness* información sobre la diferencia de fase del valor asignado al nodo  $m$  con respecto al nodo  $n$ . Ilustrando esto con un ejemplo simple: un grafo de línea, es decir, un grafo en el que el nodo  $n$  está conectado sólo con el nodo  $n+1$  y el nodo  $n-1$ , excepto el nodo 1 que está conectado sólo con el nodo 2, y el nodo  $N$  que está conectado sólo con el nodo  $N-1$ . Suponiendo también que los pesos son constantes, y definiendo  $\mathbf{s}_r = [s_N, \dots, s_1]^T$ , una versión invertida de  $\mathbf{s}$ . Es sencillo mostrar que el primer término de (1) es el mismo para la *smoothness* de  $\mathbf{s}$  y  $\mathbf{s}_r$ , mientras que el segundo término cambia de signo. Así, la *smoothness* depende de la variabilidad de la amplitud de la señal sobre grafos (como sucede en el caso real) y del sentido de rotación al moverse en el plano complejo de un nodo al otro. Este segundo término puede ser positivo o negativo, por lo que la *smoothness* podría eventualmente ser negativa para algunas señales sobre grafos. Esto implicaría que la matriz Laplaciana Hermítica correspondiente no podría ser semidefinida positiva como lo es la matriz Laplaciana real. Esto puede demostrarse utilizando el criterio de Sylvester [23]: una condición necesaria y suficiente para que una matriz Hermítica sea definida positiva es que todos sus menores principales superiores (determinantes de las submatrices principales) sean positivos. Si llamamos  $\mathbf{L}_2$  a la submatriz superior (2x2) de  $\mathbf{L}$ .

$$\mathbf{L}_2 = \begin{pmatrix} a_{12}^R + \sum_{m>2} a_{1m}^R & -a_{12}^R - ja_{12}^I \\ -a_{21}^R - ja_{21}^I & a_{21}^R + \sum_{m>2} a_{2m}^R \end{pmatrix} = \begin{pmatrix} a_{12}^R + \sum_{m>2} a_{1m}^R & -a_{12}^R - ja_{12}^I \\ -a_{12}^R + ja_{12}^I & a_{12}^R + \sum_{m>2} a_{2m}^R \end{pmatrix} \quad (4)$$

$$\det \mathbf{L}_2 = \left[ a_{12}^R \sum_{m>2} (a_{1m}^R + a_{2m}^R) + \sum_{m>2} a_{1m}^R \sum_{m>2} a_{2m}^R \right] - (a_{12}^I)^2$$

Se puede ver que dependiendo de la selección de los pesos correspondientes a los nodos 1 y 2, el menor principal superior de orden 2 podría ser no positivo. El resultado (4) podría aplicarse a cualquier par de nodos, ya que reordenar los nodos conduce a una nueva matriz Laplaciana Hermítica que es una permutación similar a la original [24], y la condición semidefinida positiva o no positiva sigue siendo la misma.

En consecuencia, dado que las contribuciones del segundo término en (2) pueden ser positivas o negativas, se podría plantear cuál es el significado de aumentar o disminuir la *smoothness* en el caso complejo. En el caso real, esto tiene una respuesta intuitiva, ya que la *smoothness* es una medida de la variabilidad de la señal a medida que nos movemos por el grafo. Sin embargo, en el caso complejo la medida de variabilidad del primer término, puede eventualmente ser compensada por las contribuciones negativas del segundo término. Por lo tanto, para dar un significado (incluso intuitivo) a (2) se deben considerar topologías particulares (como se hizo en el ejemplo anterior de un grafo de línea) o definir algunas restricciones a las señales sobre grafos cuya *smoothness* se está comparando. Así, por ejemplo, (2) es una medida de la variabilidad de la señal sobre el grafo, para el subconjunto de señales que satisfacen  $\text{sign}[a_{mm}^I] = \text{sign}[\sin(\angle s_m - \angle s_n)]$  de modo que todas las contribuciones del segundo término en (2) son positivas. De esta manera se garantiza que tanto las distancias como las diferencias de fase contribuyen positivamente a la *smoothness*, y que la suavidad cero se obtiene sólo cuando  $s_n = s_m \quad \forall nm$ , como en el caso real.



#### 4.2.2. La transformada de Fourier Compleja sobre grafos (CGFT)

Cualquier matriz Hermítica puede ser diagonalizada por una matriz unitaria, por lo que

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \sum_{n=1}^N \lambda_n \mathbf{u}_n \mathbf{u}_n^H, \quad (5)$$

donde  $\lambda_n$   $n=1\dots N$  son los autovalores de  $\mathbf{L}$  ( $\mathbf{\Lambda}$  una matriz diagonal,  $\mathbf{\Lambda}(n,n) = \lambda_n$ ) y  $\mathbf{u}_n$   $n=1\dots N$  (columnas de  $\mathbf{U}$ ) son los correspondientes autovectores,  $\mathbf{U}^H = \mathbf{U}^{-1}$  por lo que  $\mathbf{U}$  es unitaria. Los autovalores son reales debido a la propiedad Hermítica. La GFT se define para el caso real como la proyección de la señal sobre grafos sobre el espacio vectorial expandido por una base formada por los autovectores de la matriz Laplaciana real. La noción de frecuencia se asimila al autovalor asociado a cada autovector de la misma manera que la exponencial compleja se define para una frecuencia específica cuando se usa la FT. Se puede extender este concepto al caso complejo simplemente considerando los autovectores de la matriz Laplaciana Hermítica, definiendo así la CGFT directa e inversa como

$$\mathbf{r} = CGFT(\mathbf{s}) = \mathbf{U}^H \mathbf{s}, \quad \mathbf{s} = CGFT^{-1}(\mathbf{r}) = \mathbf{U} \mathbf{r}. \quad (6)$$

En el caso real la matriz Laplaciana es semidefinida positiva, por lo que  $\lambda_n \geq 0 \quad \forall n$ , de hecho, el autovalor mínimo es 0 [25]. Sin embargo, se ha visto que en el caso complejo la matriz Hermítica  $\mathbf{L}$  podría no ser semidefinida positiva, por lo que al menos un autovalor será negativo. Obsérvese que los diferentes modos (autovectores) de una matriz Hermítica se pueden ordenar de acuerdo con la smoothness creciente porque  $\mathbf{u}_n^H \mathbf{L} \mathbf{u}_n = \lambda_n$  y que para cualquier vector  $\mathbf{s}$   $\lambda_{\min} \leq \frac{\mathbf{s}^H \mathbf{L} \mathbf{s}}{\mathbf{s}^H \mathbf{s}} \leq \lambda_{\max}$ . Por lo tanto, aún se puede asimilar el concepto de frecuencia al autovalor, es decir,  $CGFT^{-1}$  en (6) expresa la señal sobre grafos como una combinación lineal de modos que tie-

nen una *smoothness* creciente (de  $\lambda_{\min}$  a  $\lambda_{\max}$ ). Sin embargo, hay que tener en cuenta que, como se explica al final de la Sección 4.2.1 es difícil obtener una interpretación general del significado de *smoothness*. Así, la aparición de autovalores negativos sugiere que el sentido de rotación al moverse en el plano complejo de un nodo a otro puede ser relevante para una caracterización completa de una señal sobre grafos por la CGFT. Esto ha sido verificado en el ejemplo particular de grafo lineal de la sección anterior. A continuación se muestra otro ejemplo numérico simple. Considerando un grafo de 2 nodos, donde  $w_{12} = 1 + j = w_{21}^*$ . Si se calculan los dos autovalores y los dos autovectores de la correspondiente Laplaciana Hermítica 2x2 se obtiene  $\lambda_1 = -0.41$  y  $\lambda_2 = 2.41$ ,  $\mathbf{u}_1 = [0.5 + 0.5j, 0.707]^T$  y  $\mathbf{u}_2 = [-0.5 - 0.5j, 0.707]^T$ . Se observa que aparece un autovalor negativo y que en el primer autovector hay una rotación en el sentido de las agujas del reloj al moverse en el plano complejo desde su primer elemento hasta el segundo. Sin embargo, en el segundo autovector, la rotación es en sentido contrario a las agujas del reloj. De manera similar, se ha mostrado al final de la Sección 4.2.1 que la nueva *smoothness* definida puede mantener el significado clásico de medida de variabilidad al imponer algunas restricciones a las señales sobre grafos que se compararán. Esta nueva medida incorpora distancias así como diferencias de fase, lo cual es consistente con el dominio complejo. Finalmente, obsérvese que la FT puede interpretarse como un caso particular de una CGFT (como lo es la GFT) donde la matriz Laplaciana Hermítica asociada es una matriz circulante (ver más abajo, el final de la Sección 4.2).

### **4.3. Invariantes de la Amplitud del Espectro sobre grafos**

El elemento clave de los algoritmos de subrogación es mantener la GSA de la señal sobre grafos original mientras las fases del espectro sobre grafos son aleatorizadas. A continuación se determinan los invariantes de GSA antes de presentar los algoritmos

detallados en la siguiente sección. Esto ayuda a motivar los métodos de grafos y a obtener información sobre las implicaciones de las señales sobre grafos subrogadas. Los invariantes de GSA son aquellas funciones o estadísticas de la señal sobre grafos original que no cambia si la GSA no cambia. La relación entre dos señales sobre grafos  $\mathbf{s}$  y  $\hat{\mathbf{s}}$  que tienen la misma GSA se puede expresar en la forma

$$\hat{\mathbf{s}} = (\mathbf{U}^H)^{-1} \Phi \mathbf{U}^H \mathbf{s} = \mathbf{U} \Phi \mathbf{U}^H \mathbf{s} \quad (7)$$

Donde  $\Phi$  es una matriz diagonal con entradas fasoriales  $\Phi(n,n) = \exp(j\phi_n)$   $n = 1 \dots N$ . Obsérvese que las amplitudes espectrales del vector transformado se conservan, pero se añade una fase  $\phi_n$  a la fase n-ésima del espectro del grafo, antes de la transformación inversa. Definiendo la matriz unitaria  $\mathbf{P}^H = \mathbf{U} \Phi \mathbf{U}^H = \mathbf{P}^{-1}$ , desde (7) se puede escribir

$$\hat{\mathbf{s}} \hat{\mathbf{s}}^H = \mathbf{U} \Phi \mathbf{U}^H \mathbf{s} \mathbf{s}^H \mathbf{U} \Phi \mathbf{U}^H = \mathbf{P}^{-1} \mathbf{s} \mathbf{s}^H \mathbf{P} \quad (8)$$

Entonces las matrices de correlación de la señal sobre grafos original y sus subrogaciones son unitariamente similares [26]. El criterio de Specht [27] establece una condición necesaria y suficiente para la similitud unitaria de dos matrices en términos de invariancia de algunas trazas. La aplicación de este criterio a las matrices de correlación muestral conduce a

$$\text{trace}\left(\left(\hat{\mathbf{s}} \hat{\mathbf{s}}^H\right)^k\right) = \text{trace}\left(\left(\mathbf{s} \mathbf{s}^H\right)^k\right), \quad (9)$$

para cada entero  $k$ . La ecuación (9) define funciones invariantes que involucran polinomios de los elementos  $\mathbf{s}$  y  $\hat{\mathbf{s}}$ . En particular, para  $k=1$  esto implica que la energía de la señal sobre grafos es un invariante GSA

$$\sum_{n=1}^N |\hat{s}_n|^2 = \sum_{n=1}^N |s_n|^2, \quad (10)$$

y para  $k=2$  se puede ver que la norma de Frobenius de la matriz de correlación muestral es también invariante GSA

$$\sum_{n=1}^N \sum_{m=1}^N |\hat{s}_n \hat{s}_m^*|^2 = \sum_{n=1}^N \sum_{m=1}^N |s_n s_m^*|^2 \quad (11)$$

Nótese que, hasta ahora, el valor del invariante GSA es el mismo para todas las posibles matrices Laplacianas Hermíticas. Considerando ahora invariantes que dependen de la matriz Laplaciana Hermítica. Primero, obsérvese que  $\mathbf{s}$  y  $\hat{\mathbf{s}}$  son igualmente *smooth* en el grafo  $G \{V, E, \mathbf{A}\}$

$$\begin{aligned} S(\hat{\mathbf{s}}) &= \hat{\mathbf{s}}^H \mathbf{L} \hat{\mathbf{s}} = \mathbf{s}^H \mathbf{U} \Phi^H \mathbf{U}^H \mathbf{L} \mathbf{U} \Phi \mathbf{U}^H \mathbf{s} \\ &= \mathbf{s}^H \mathbf{U} \Phi^H \mathbf{U}^H \mathbf{U} \Lambda \mathbf{U}^H \mathbf{U} \Phi \mathbf{U}^H \mathbf{s} \\ &= \mathbf{s}^H \mathbf{L} \mathbf{s} = S(\mathbf{s}) \quad . \end{aligned} \quad (12)$$

Entonces *la smoothness* en el grafo definido por el Laplaciano Hermítico es invariante GSA. Adoptando ahora una perspectiva estocástica, suponiendo que  $\mathbf{s}$  es un vector aleatorio de media cero que tiene una matriz de precisión conocida  $\mathbf{Q}_{ss} = \mathbf{C}_{ss}^{-1}$ , donde  $\mathbf{C}_{ss} = E[\mathbf{s}\mathbf{s}^H]$  es la matriz de covarianza. Tomando esperanzas e invirtiendo en (8) se concluye que las matrices de precisión también son unitariamente similares

$$\mathbf{Q}_{\hat{s}\hat{s}} = \mathbf{P}^{-1} \mathbf{Q}_{ss} \mathbf{P} \quad (13)$$

Los elementos de  $\mathbf{Q}_{ss}$  normalizados adecuadamente son las correlaciones parciales [28]-[31] definiendo dependencias condicionadas lineales entre cada dos elementos de  $\mathbf{s}$ . Esto motiva el uso extendido de  $\mathbf{Q}_{ss}$  como matriz Laplaciana. Suponiendo entonces que se selecciona  $\mathbf{L} = \mathbf{Q}_{ss}$ , considerando (5), y la definición de  $\mathbf{P}$

$$\mathbf{Q}_{\hat{s}\hat{s}} = \mathbf{U} \Phi \mathbf{U}^H \mathbf{Q}_{ss} \mathbf{U} \Phi^H \mathbf{U}^H = \mathbf{U} \Phi \mathbf{U}^H \mathbf{U} \Lambda \mathbf{U}^H \mathbf{U} \Phi^H \mathbf{U}^H = \mathbf{Q}_{ss} \quad (14)$$

Por lo tanto, la matriz de precisión es invariante a la GSA cuando se usa como Laplaciana Hermítica.

Finalmente, algunos trabajos recientes [19] [20] han propuesto extensiones equivalentes del concepto de estacionaridad para señales sobre grafos al considerar definiciones de traslación de señales sobre grafos. En [19][20] los autores demuestran que una señal sobre grafos estocástica es GWSS si la matriz de covarianza es diagonalizable conjuntamente con la matriz Laplaciana. Entonces suponiendo que las interrelaciones por pares se definen de tal manera que  $s$  es una realización de una señal sobre grafos aleatoria (media cero) que es GWSS con respecto a la Laplaciana utilizada para generar las subrogaciones, es decir, que la señal sobre grafos es invariante a algún operador de traslación, entonces  $C_{ss} = \mathbf{U}\mathbf{\Gamma}_{ss}\mathbf{U}^H$ , donde  $\mathbf{\Gamma}_{ss}$  es una matriz diagonal, entonces

$$\mathbf{C}_{\tilde{s}\tilde{s}} = \mathbf{U}\mathbf{\Phi}\mathbf{U}^H\mathbf{U}\mathbf{\Gamma}_{ss}\mathbf{U}^H\mathbf{U}\mathbf{\Phi}^H\mathbf{U}^H = \mathbf{C}_{ss} . \quad (15)$$

Por lo tanto, la matriz de covarianza es un invariante GSA para las señales GWSS, lo que es consistente con la bien conocida invariancia de la función de autocovarianza con respecto a la densidad espectral de potencia de las señales WSS convencionales. Nótese que la matriz de covarianza preservada no es necesariamente Toeplitz, como sí ocurre al preservar la función de auto correlación con métodos de subrogación clásicos.

La Tabla 4.1 resume los invariantes GSA presentados en esta sección.

<i>Invariante</i>	<i>Ecuación</i>
Energía de la señal	$\sum_{n=1}^N  s_n ^2 = \sum_{n=1}^N  \hat{s}_n ^2$
Norma de Frobenious de la matriz de correlación de la muestra	$\sum_{n=1}^N \sum_{m=1}^N  s_n s_m^* ^2 = \sum_{n=1}^N \sum_{m=1}^N  \hat{s}_n \hat{s}_m^* ^2$
Rastros de las potencias de la matriz de correlación de la muestra	$\text{trace}((\mathbf{s}\mathbf{s}^H)^k) = \text{trace}((\hat{\mathbf{s}}\hat{\mathbf{s}}^H)^k)$
<i>Smoothness</i>	$\mathbf{s}^H \mathbf{L} \mathbf{s} = \hat{\mathbf{s}}^H \mathbf{L} \hat{\mathbf{s}}$
Matriz de precisión	If $\mathbf{L} = \mathbf{Q}_{ss}$ then $\mathbf{Q}_{\hat{s}\hat{s}} = \mathbf{Q}_{ss}$
Graph Wide-Sense Stationarity	If $\mathbf{C}_{ss} = \mathbf{U}\mathbf{\Gamma}_{ss}\mathbf{U}^H$ then $\mathbf{C}_{\hat{s}\hat{s}} = \mathbf{C}_{ss}$

**Tabla 4.1.** Invariantes GSA.

## 4.4. Algoritmos de subrogación

### 4.4.1. Algoritmo iterativo de subrogación de señales sobre grafos

Una de las primeras propuestas de un algoritmo de subrogación fue el Amplitude Adjusted Fourier Transform (AAFT), [32]. En éste se calcula la FT de la señal original (real), luego se aleatoriza la fase del espectro de las muestras obtenidas y se realiza un corrección de la transformada inversa de Fourier para cumplir con la distribución empírica de las muestras originales. Las fases aleatorias se obtienen muestreando una distribución uniforme entre  $-\pi$  y  $\pi$ , restringida para mantener la propiedad de antisimetría de las fases espectrales de las señales reales. Las correcciones consisten en reemplazar cada muestra de la subrogada por el valor obtenido de la muestra original que tiene el mismo orden de rango, recuperando así la distribución empírica de las señales originales. De modo que, la subrogada es una versión barajada de la señal original.

Se ha propuesto una versión mejorada que ajusta iterativamente la amplitud del espectro y corrige la distribución empírica de las muestras: AAFT iterativo o Iterative

AAFT (IAAFT) [1] [33]. Se han propuesto mejoras menores o variaciones de la AAFT en [34]-[36]. Más recientemente, se han considerado extensiones al caso multivariante donde el problema adicional es mantener la correlación entre los diferentes canales subrogados [37]-[40].

Hasta ahora, estos algoritmos de subrogación se han aplicado sólo a señales reales en el dominio del tiempo. En esta sección se presenta un algoritmo iterativo para subrogar señales complejas sobre grafos por medio de la CGFT (ICGFT). Todo empieza con la señal sobre grafos original  $\mathbf{s} = [s_1, \dots, s_N]^T \in \mathbb{C}^N$ . Se calcula una subrogación inicial no corregida  $\hat{\mathbf{s}}^{(0)}$  al aleatorizar las fases de la CGFT of  $\mathbf{s}$  (ec. (16a)) mientras se mantiene la GSA. Las fases aleatorias se obtienen mediante el muestreo de una distribución uniforme entre  $-\pi$  y  $\pi$  (no se impone ninguna restricción de antisimetría ya que se consideran señales complejas). Luego se corrige  $\hat{\mathbf{s}}^{(0)}$ , para que coincida con la distribución empírica de la señal sobre grafos original. Más específicamente, cada muestra de la subrogación obtenida se reemplaza por la muestra de la señal sobre grafos original que tiene el mismo orden de rango de amplitud. Con ello se obtiene la subrogación inicial corregida  $\mathbf{s}^{(0)}$  (ec. (16b))

$$\hat{\mathbf{s}}^{(0)} = \mathbf{U}\Phi\mathbf{U}^H \mathbf{s} \quad , \quad (16a)$$

$$\mathbf{s}^{(0)} = C\left(\hat{\mathbf{s}}^{(0)}\right) \quad , \quad (16b)$$

Luego se procede iterativamente, como en IAAFT, calculando la CGFT de la subrogación corregida, recuperando la GSA original y generando una nueva subrogación no corregida  $\hat{\mathbf{s}}^{(i)}$  (17a). Esta luego se corrige nuevamente para que coincida con la distribución empírica de la señal sobre grafos (17b), y así sucesivamente:

$$\hat{\mathbf{s}}^{(i)} = \mathbf{U}\mathbf{A}^{(i-1)}\mathbf{U}^H \mathbf{s}^{(i-1)} \quad , \quad (17a)$$

$$\mathbf{s}^{(i)} = C\left(\hat{\mathbf{s}}^{(i)}\right) \quad i = 1 \dots I \quad . \quad (17b)$$

Definiendo  $\mathbf{r}^{(i-1)} = \mathbf{U}^H \mathbf{s}^{(i-1)}$  entonces  $\mathbf{A}^{(i-1)}$  una matriz diagonal con elementos  $\mathbf{A}^{(i-1)}(n,n) = |r_n| / |r_n^{(i-1)}|$ . El algoritmo itera hasta la convergencia, es decir, hasta que no se requieran más correcciones. El Algoritmo 4.1 a continuación es una descripción de pseudocódigo de ICGFT.

**Algoritmo 4.1:** Subrogación compleja de señales sobre grafos mediante CGFT (ICGFT).

---

1: **Entrada:** Señal sobre grafos  $\mathbf{s}$ , Laplaciana Hermítica  $\mathbf{L}$ , número de subrogaciones  $K$ , número de iteraciones  $I$

2: **Calcular matriz de autovectores  $\mathbf{U}$**

3: **for**  $k = 1, 2 \dots K$  **do:**

4: Generar aleatorización de fases en el espectro del grafo  $\Phi_k$

5: Calcular subrogación inicial, ec. (16a, b)  $\mathbf{s}_k^{(0)} = C(\mathbf{U}\Phi_k\mathbf{U}^H\mathbf{s})$

6: **for**  $i = 1, 2 \dots I$  **do:**

7: Calcular subrogación, ec. (17a, b)

$$\mathbf{s}^{(i)} = C(\mathbf{U}\mathbf{A}^{(i-1)}\mathbf{U}^H\mathbf{s}^{(i-1)})$$

8: Si  $\mathbf{s}^{(i)} \approx \mathbf{s}^{(i-1)}$  ir a 9

9: **end for**

10:  $\mathbf{s}_k = \mathbf{s}_k^{(I)}$

11: **end for**

12: **Output:**  $\mathbf{s}_1 = \mathbf{s}_1^{(I)} \dots \mathbf{s}_K = \mathbf{s}_K^{(I)}$

---

#### 4.4.2. Seleccionando la matriz Laplaciana Hermítica

Un aspecto clave del algoritmo ICGFT es seleccionar una matriz Laplaciana (o de adyacencia) adecuada. Esto es un problema clásico en procesamiento de señal sobre grafos. Una opción podría ser estimar  $\mathbf{Q}_{ss}$  a partir del conjunto de señales sobre grafos originales disponibles (ver por ejemplo [17], [18], [28]-[31]), y usar esta como matriz Hermítica. En tal caso, se ha visto en la Sección 4.3 que de este modo la matriz de



precisión se conserva en las subrogaciones. Otras alternativas hacen una estimación directa de la matriz Laplaciana minimizando funciones de coste que incluyen algunas restricciones de regularización [15], [16]. Sin embargo, en un escenario de escasez, sería difícil obtener estimaciones estables de la matriz Laplaciana Hermítica mediante todos estos métodos.

Una opción diferente es definir conexiones a partir de una conectividad deseada general o específica que debe conservarse en las subrogaciones. Por ejemplo, una opción general podría ser conectar cada nodo sólo con sus vecinos (esto se hace en [11] en el dominio del espacio), de modo que las propiedades locales de la señal sobre grafos puedan conservarse en las subrogaciones. Se pueden establecer conexiones más específicas según el contexto y el dominio de la aplicación.

Queda una pregunta sobre el signo de  $a_{nm}^l$ . Cabe recordar que  $a_{nm}^R = a_{mn}^R \geq 0$ , pero  $a_{nm}^l = -a_{mn}^l$ , por ejemplo, se pueden conectar dos muestras porque son vecinas, pero esto se puede hacer con un valor positivo o negativo de la parte imaginaria  $a_{nm}^l$ . Como se indica en (2) una selección diferente de signo  $a_{nm}^l$  conduce a una *smoothness* diferente que, eventualmente, podría ser negativa. Aquí se propone seleccionar los signos de tal manera que se maximice  $S(\mathbf{s})$ , es decir, ajustando  $\text{sign}[a_{nm}^l] = \text{sign}[\sin(\angle s_m - \angle s_n)]$ . En el Apéndice B de este capítulo se muestra que el ancho de banda del espectro  $B = \lambda_{\max} - \lambda_{\min}$  de la matriz Laplaciana Hermítica tiene un límite inferior por

$$B \geq \max\left(l_{\max}, \frac{S(\mathbf{s})}{\mathbf{s}^H \mathbf{s}}\right) - l_{\min}, \quad (18)$$

donde  $l_{\max}$  y  $l_{\min}$  son respectivamente el máximo y mínimo de los elementos de la diagonal principal de la matriz Laplaciana Hermítica (que depende únicamente de  $a_{nm}^R$ ). Por lo que, dados  $a_{nm}^R, a_{nm}^l$  y  $\mathbf{s}$ , el límite inferior se maximiza cuando  $S(\mathbf{s})$  es

máxima. Como la aleatorización de la fase en el dominio transformado es un paso clave de la subrogación, es apropiado tener disponible el máximo ancho de banda posible.

Finalmente, nótese que la FT (más precisamente su implementación numérica a través de la DFT) puede interpretarse como una forma particular de selección de matriz Laplaciana Hermítica. Esto se debe a que como es bien sabido, los vectores base de la DFT son autovectores de matrices circulantes [41]. Por tanto, implícitamente, el uso de la DFT impone restricciones en los elementos de la matriz de adyacencia. En particular, la conectividad por pares es la misma para nodos ubicados a la misma “distancia” en la numeración de nodos, lo que implica que se asume estacionaridad (del concepto convencional de traslación). Además, la simetría cíclica requerida de los coeficientes impone restricciones adicionales que pueden no tener sentido en un contexto general. También, obsérvese que los autovectores de una matriz Laplaciana Hermítica circulante están restringidos al no depender del valor específico de los pesos. Por lo tanto, se puede obtener mucha más flexibilidad definiendo una conectividad específica de acuerdo con las propiedades deseadas de la señal sobre grafos original que deben conservarse en las subrogaciones.

#### **4.5. Experimentos**

Muchas aplicaciones de procesamiento de señales se caracterizan por una carencia de señales durante la fase de desarrollo. Esto a su vez produce escasez de instancias (vectores de características extraídas de las señales preprocesadas) para el entrenamiento automático de clasificadores. Por ejemplo, en las pruebas automáticas no destructivas de materiales, se debe entrenar a un clasificador para cada clase de defecto. Esto requiere una colección de una gran cantidad de señales que incluyan todos los posibles tipos de defectos, lo que puede ser poco práctico o incluso imposible

[42] [43]. Otro ejemplo totalmente diferente es la detección de fraudes con tarjetas de crédito [44], donde el número de operaciones legítimas es un orden de magnitud mayor que las operaciones fraudulentas. Esto también sucede en problemas de varias clases donde algunas clases pueden tener un conjunto de entrenamiento reducido [45]. Una solución para paliar la escasez es generar réplicas sintéticas usando métodos tradicionales (ver por ejemplo [46] y sus referencias). Pero los métodos tradicionales están orientados al modelo, por ejemplo, el modelo se ajusta a algunas propiedades de correlación y densidad de probabilidad de primer orden. Los métodos orientados a modelo precisan la definición de un modelo apropiado y la estimación de los parámetros del modelo, pero en escenarios de escasez ambas cosas son bastante limitadas. Por el contrario, la subrogación puede considerarse un método de síntesis orientado a la señal; en realidad, ya se ha propuesto como un método general para sintetizar datos multivariados [38]. Por lo tanto, en esta sección experimental se ha considerado la aplicación de subrogación para aumentar un conjunto de datos para el aprendizaje del clasificador. Además, obsérvese que, hasta cierto punto, esta es una generalización natural del uso de subrogaciones en las pruebas de hipótesis. En este último caso, la hipótesis nula es que la instancia bajo la prueba satisface un modelo dado (por ejemplo, un modelo lineal seguido de una no linealidad escalar). Pero, en última instancia, se asumen algunas propiedades empíricas de la instancia original (autocorrelación, distribución de amplitud...) representativas del modelo y, en consecuencia, se imponen en las subrogaciones. El mismo principio se puede aplicar para generar réplicas de instancias para clases deficientes en el entrenamiento del clasificador: las réplicas (subrogadas) deben mantener algunas propiedades relevantes de las instancias originales. En particular, considerando que las instancias son señales sobre grafos, las subrogaciones preservan los invariantes GSA deducidas en la Sección 3, que se relaciona con preservar las interrelaciones por pares definidas por la matriz Laplaciana Hermítica correspondiente.

#### 4.5.1. Clasificación de instancias aleatorias

En este experimento, se ilustra la aplicabilidad potencial de los métodos de subrogación para mejorar el entrenamiento de un discriminante lineal de dos clases en caso de que sólo se disponga de un número reducido de instancias originales para el entrenamiento. Para ello, se han generado vectores aleatorios sintéticos correspondientes a dos GMM complejos multivariantes diferentes, con los siguientes parámetros:

- Dimensión del vector:  $N=10$
- Número de clases: 2
- Número de componentes de GMM: 2 para ambas clases
- Parámetros GMM:
  - class1:  $mean1=0.5(1+j)\mathbf{1}$   $mean2=(1+j)\mathbf{1}$ ,  $\mathbf{1}=[1\dots1]^T$
  - class2:  $mean1=0.5(1-j)\mathbf{1}$   $mean2=(1-j)\mathbf{1}$
  - la matriz de covarianza se ha generado aleatoriamente y ha sido la misma para ambas clases

Los vectores aleatorios generados de este modo se consideran instancias originales sobre un dominio arbitrario para ser clasificados por el discriminante lineal. Cada componente del vector aleatorio se ha asignado al nodo de un grafo, formando así la señal sobre grafos. Para calcular la matriz Laplaciana Hermítica necesaria en el Algoritmo 2.1, se ha utilizado el método propuesto en el Capítulo 2. Básicamente, la matriz de adyacencia se determina a partir del módulo de los coeficientes de correlación parcial generalizados correspondientes a un GMM complejo. Nótese que estos coeficientes son diferentes a los obtenidos de la matriz de precisión, que corresponden a un modelo gaussiano multivariante [17].

Se define un Conjunto de Entrenamiento Ampliado o *Enlarged Training Set* (ETS): un conjunto formado por sólo un número reducido de instancias originales ampliadas por un número variable de subrogaciones obtenidas de ellos. También se ha definido un Conjunto de Entrenamiento Original u *Original Training Set* (OTS) formado sólo

por las instancias originales. Se han considerado tres tipos diferentes de conjuntos de entrenamientos:

- OTS, conjunto de entrenamiento original formado por las instancias originales.
- ETS-ICGFT, conjunto de entrenamiento ampliado formado por un número reducido de instancias originales ampliado con un número variable de subrogaciones obtenidas mediante el método ICGFT (Algoritmo 4.1).
- ETS-IAAFT, conjunto de entrenamiento ampliado formado por un número reducido de instancias originales (parte real) ampliado por un número variable de subrogados obtenidos de ellos por el método IAAFT.
- ETS- $n$ IGFT, conjunto de entrenamiento ampliado formado por un número reducido de las instancias originales (parte real) ampliado por un número variable de subrogados obtenidos de ellos por el método propuesto en [11], que utiliza la GFT (real). Este método no itera ya que no se realizan correcciones en el dominio del grafo original, por lo que se ha llamado GFT no iterativo ( $n$ IGFT). Para calcular la matriz Laplaciana (real), se ha utilizado el método propuesto en el Capítulo 2.

En la Figura 4.1 se presentan las tasas de error promedio del clasificador de dos clases en función del tamaño del conjunto de entrenamiento. Cada curva corresponde respectivamente a cada uno de los cuatro tipos de conjuntos de entrenamiento. El ETS estuvo formado en todos los casos por 5 instancias originales ampliadas por un número variable de subrogaciones (0 a 20), por lo que el tamaño del conjunto de entrenamiento varía de 5 a 25. En este experimento se consideraron 10 particiones de validación cruzada. La tasa de error se estimó en cada partición utilizando 50 instancias originales de prueba (diferentes de las incluidas en los conjuntos de entrenamiento para evitar el sobreajuste), y luego se promedió sobre las 10 particiones.

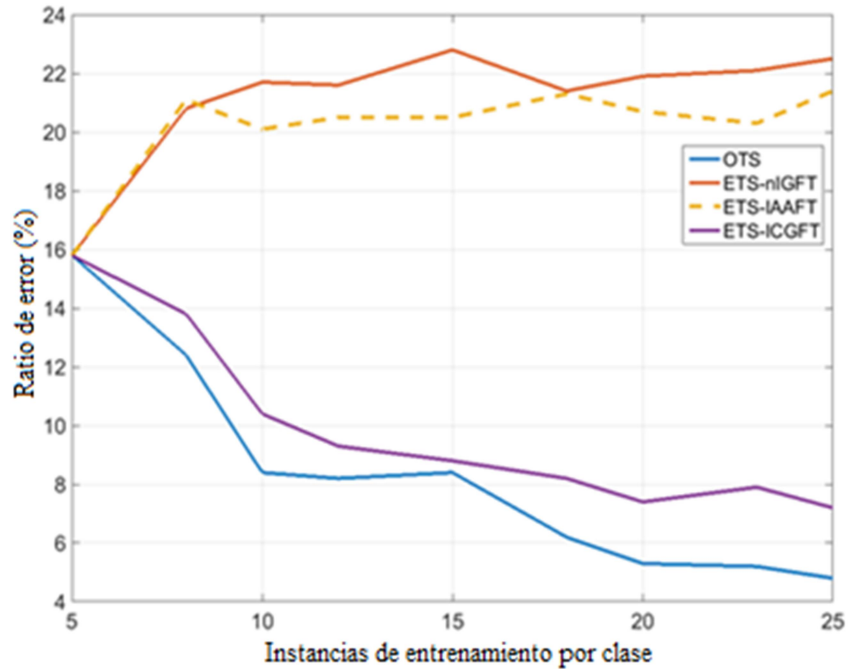


Figura 4.1. Tasa de error de clasificación para un número creciente de instancias de entrenamiento.

Como era de esperar, el entrenamiento con OTS obtiene la tasa de error mínima para cada tamaño de conjunto de entrenamiento. Sin embargo, se observa que la tasa de error correspondiente a un OTS con sólo 5 instancias originales se reduce monótonamente al entrenar con ETS-ICGFT para un número creciente de subrogaciones. Además, se acerca razonablemente a la tasa de error del entrenamiento con un OTS del mismo tamaño. Esto ilustra que, en caso de que sólo esté disponible un número reducido de instancias originales, se puede ampliar el conjunto de entrenamiento con subrogados para mejorar la calidad del entrenamiento del clasificador. Respecto al entrenamiento con ETS-IAAFT o ETS-nIGFT, se observa que no se puede disminuir la tasa de error. Esto se debe principalmente a la limitación de trabajar con señales reales, pero también a la falta de flexibilidad para imponer interrelaciones por pares

entre las muestras de señales sobre grafos (IAAFT) o a la aleatorización de fase restringida y falta de iteraciones (nIGFT). Finalmente, la Tabla 4.2 muestra las tasas de error promediadas sobre el número de instancias de entrenamiento.

	OTS	ETS-ICGFT	ETS-IAAFT	ETS-nIGFT
<i>Ratio de error (%)</i>	8.30	9.86	20.19	21.18

**Tabla 4.2.** Tasa de error promediada sobre el número de instancias de entrenamiento.

#### **4.5.2. Reconocimiento automático de gestos con las manos**

En esta sección se considera el uso de los métodos de subrogación en un estudio de caso de datos reales: el reconocimiento automático de gestos con las manos. Este ha sido seleccionado debido a la complejidad del problema: gran cantidad de clases, gran cantidad y número variable de características. Por lo tanto, es de esperar que los resultados se extrapolen a casos más simples de clasificación automática. El reconocimiento de gestos con las manos es muy útil en la vida cotidiana para la interacción hombre-máquina en aplicaciones como la navegación en entornos virtuales o el reconocimiento de posturas de las manos [47]. Originalmente, las imágenes de la cámara rastreaban las manos para determinar su pose y su trayectoria en 2D, pero la introducción de sensores de profundidad como *Kinect* o *Leap Motion* permitió obtener información en 3D para reconocer los gestos [48]. Estas aplicaciones suelen utilizar clasificadores basados en modelos ocultos de Markov (*Hidden Markov Models*, HMM) [49] o máquinas de vectores de soporte o (*Support Vector Machines*, SVM) [50]. Se ha utilizado un conjunto de datos público (disponible en <https://github.com/Sasanita/spanish-sign-language-db.git>) compuesto por 91 aproximaciones de los gestos de la lengua de signos española adquiridas por un sensor *Leap*

*Motion.* Este conjunto de datos contiene 40 instancias por signo. Cada instancia es una matriz que tiene 42 columnas. Cada columna es una señal correspondiente a la evolución temporal de una variable observada (característica). El total de 42 variables observadas describen la trayectoria de los dedos y las manos y su inclinación según los ejes X, Y y Z de un sistema tridimensional. La frecuencia de muestreo es de 30 Hz para todos los casos, sin embargo, el tiempo de observación puede variar entre diferentes casos, ya que la señal manual se puede producir a diferentes velocidades. Así, la dimensión de la columna de las matrices de instancia varía de 7 (0,23 s) a 101 (3,36 s) con un valor promedio de 26 (0,86 s). En todos los casos, el clasificador se basó en un HMM con 7 estados. La probabilidad condicionada de las observaciones a los estados ocultos se consideró una mezcla gaussiana de 2 componentes. Se utilizó el algoritmo de Viterbi para estimar la secuencia de estados más probable. De hecho, se utilizó el kit de herramientas *Hidden Markov Model Toolkit* (los detalles se pueden encontrar en [51] y sus referencias). En todos los casos la prueba del clasificador una vez entrenado, se realizó con instancias originales no incluidas en los conjuntos de entrenamiento para evitar sobreajustes.

Los diferentes métodos de subrogación se han utilizado para generar columnas subrogadas de las instancias originales. Varios detalles son relevantes acerca de la implementación de ambos métodos. En IAAFT, la transformada discreta de Fourier (DFT) se calculó con el mismo tamaño que la columna transformada para que la dimensión del dominio original se pueda conservar una vez que se calcule la DFT inversa. Por otro lado, ICGFT está concebido para trabajar con señales sobre grafos complejas, por lo que primero se ha calculado la señal analítica correspondiente a cada una de las 42 columnas de la matriz<sup>3</sup> de instancia. Con respecto a la matriz

---

<sup>3</sup> Esta es una posibilidad a la que siempre se puede recurrir para aplicar ICGFT a instancias reales, beneficiándose así de la aleatorización de fases sin restricciones de CGFT en contraste con la aleatorización de signos limitada de GFT.



Laplaciana necesaria en el Algoritmo 4.1, se calculó a partir de una matriz de adyacencia definida como

$$a_{nm} = \begin{cases} 1 \pm j & |n-m|=1 \\ 0 & otherwise \end{cases} = a_{nm}^* \quad , \quad (19)$$

donde se han seleccionado los signos de la parte imaginaria como se indica en la Sección 4.4.2 (la parte real de (19) se conserva para implementar nIGFT). Esto se hace en un esfuerzo por replicar el comportamiento local de las señales sobre grafos originales (los vecinos están conectados), pero son posibles muchas otras opciones. En particular, se podría intentar replicar algunas propiedades especiales de las instancias originales que se sabe que son relevantes para el objetivo final de la clasificación. Esto normalmente debe hacerse con la colaboración de un especialista en el dominio de la aplicación. Otra pregunta importante se refiere a la etapa de corrección en el Algoritmo 4.1. En el reconocimiento de gestos con las manos, el clasificador debe capturar la dinámica de las señales. Esto lo hace el HMM determinando primero los estados sucesivos a través de los cuales pasan las señales. Recuperar la distribución empírica de las señales gráficas originales mediante el procedimiento de corrección del Algoritmo 4.1 puede afectar seriamente la dinámica de la señal, por lo que se ha considerado en este caso un tipo diferente de corrección, a saber, recuperar la frecuencia instantánea original, que es una medida aceptada de la dinámica de señales complejas [52]. De esta manera, se aprovecha trabajar el dominio complejo y la evolución de la "forma de onda" de las señales sobre grafos original y las subrogaciones serán mucho más comparable. Esta corrección alternativa se puede expresar modificando la ecuación (17b) en la forma

$$\mathbf{s}^{(i)} = \mathbf{C}^{(i)} \hat{\mathbf{s}}^{(i)} \quad i=1..I \quad , \quad (20)$$

donde  $\mathbf{C}^{(i)}$  es una matriz diagonal con elementos  $\mathbf{C}^{(i)}(n,n) = \exp\left(j\left(\angle s_n - \angle \hat{s}_n^{(i)}\right)\right)$ .

Se han concebido dos experimentos para ilustrar el interés de utilizar subrogaciones en dos escenarios representativos. El primero será el de la escasez general: se supondrá que sólo hay disponible un número reducido de instancias originales en cada clase, de manera que el conjunto de instancias se ampliará con subrogadas para mejorar la calidad del entrenamiento. Esto es similar al experimento anterior con datos sintéticos. En el segundo escenario, se considerará una situación de desequilibrio en la que hay una cantidad significativamente menor de instancias disponibles en algunas clases respecto a otras.

*Experimento 1. Ampliación del conjunto de entrenamiento original: escasez general*

En este experimento, se considera que sólo un número reducido de instancias está disponible para entrenamiento en cada clase, específicamente 10. Este conjunto reducido se amplía agregando instancias subrogadas para formar un ETS. La Figura 4.2 muestra la tasa de error promedio lograda por los diferentes métodos de subrogación en función del tamaño del ETS, de 10 (sólo las 10 instancias de entrenamiento originales por clase) a 30 (10 instancias de entrenamiento originales más 20 subrogadas por clase). La tasa de error promedio se evaluó sobre 10 instancias originales diferentes a las disponibles del entrenamiento. Se promediaron las tasas de error de 4 particiones. En cada partición las subrogadas se generaron a partir de un conjunto reducido de 10 instancias originales para reproducir el caso práctico de escasez. También se ha incluido en la Figura 4.2, a modo de comparación, las curvas correspondientes a un OTS que tiene el mismo tamaño del ETS correspondiente.

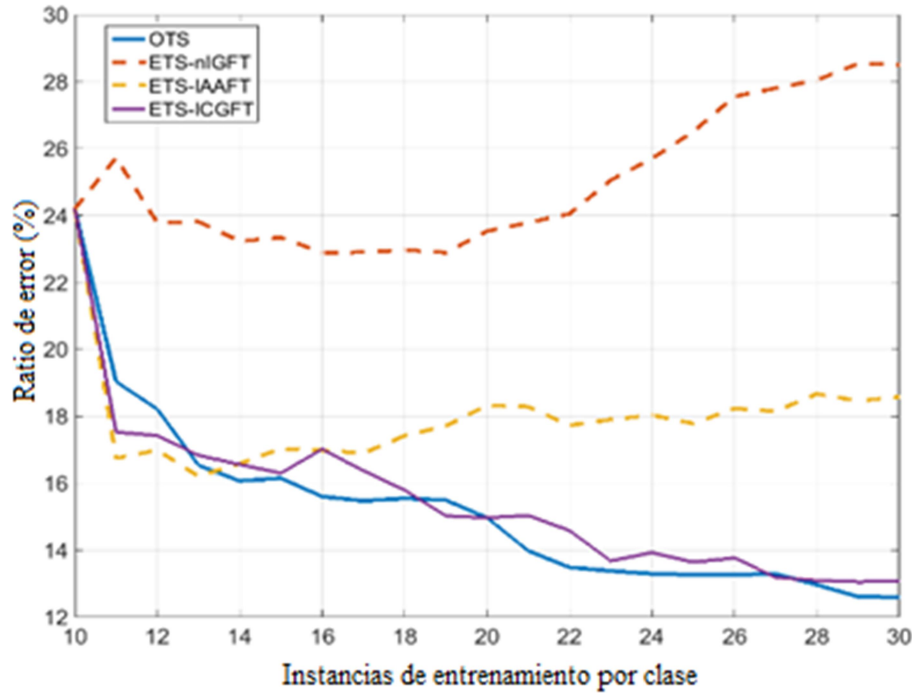


Figura 4.2. Tasa de error de clasificación para un número creciente de instancias de entrenamiento.

Se puede observar en la Figura 4.2 que el entrenamiento con ETS-ICGFT arroja resultados similares al entrenamiento con OTS, por lo que en caso de escasez se pueden agregar subrogaciones manteniendo la calidad del entrenamiento. Esto no es lo mismo cuando se entrena con ETS-IAAFT porque, después de una disminución inicial, la tasa de error aumenta con la cantidad de subrogaciones agregadas. Algo similar sucede al entrenar con ETS-nIGFT, pero con tasas de error mucho mayores. Esto sugiere que ICGFT puede implementar subrogaciones más realistas y un mejor sobremuestreo del espacio de instancias. La Tabla 4.3 compara las tasas de error correspondientes al uso de un ETS del tamaño máximo (10 originales + 20 subrogadas) para cada clase, obtenidas de diferentes métodos de subrogación. También se incluyen las tasas de error correspondientes a OTS de tamaño 10 y OTS de tamaño 30.

	OTS10	ETS- <i>n</i> IGFT	ETS-IAAFT	ETS-ICGFT	OTS30
<i>Ratio de error (%)</i>	24.1	28.5	18.58	13.07	12.6

**Tabla 4.3.** Tasas de error para diferentes conjuntos de entrenamiento: originales y ampliados.

Experimento 2. Ampliación del conjunto de entrenamiento original: clases desequilibradas

En este experimento se asume que el número de instancias de entrenamiento originales es diferente para cada clase. Así, se considera que para la mitad de las clases se dispone de 30 instancias originales de entrenamiento, mientras que para la otra mitad sólo se dispone de un número  $N \leq 30$  (igual para todas). La Figura 4.3 muestra los resultados correspondientes a este experimento: tasa de error media (4 particiones) en función del número de desequilibrio. Este último se define como la diferencia  $N-30$  donde  $N$  varía de 30 (sin desequilibrio) a 10 (desequilibrio más grande). En la Figura 4.3 se ha representado la curva correspondiente al entrenamiento con sólo el conjunto de entrenamiento desequilibrado original (OITS, *Original Imbalanced Training Set*), como era de esperar la tasa de error (evaluada nuevamente en un conjunto de 10 instancias originales diferentes a las disponibles del entrenamiento) aumenta significativamente con el número de desequilibrio. Asimismo, se muestran en la Figura 4.3 las tasas de error obtenidas al ampliar los conjuntos de entrenamiento de clases deficientes con instancias subrogadas, de modo que todas las clases tengan 30 instancias para el entrenamiento. Se ha denominado el conjunto de entrenamiento así obtenido como conjunto de entrenamiento desequilibrio ampliado (EITS, *Extended Imbalanced Training Set*), para distinguirlo del ETS del experimento 1, pero obsérvese que el EITS en realidad está equilibrado debido a la adición de instancias subrogadas.

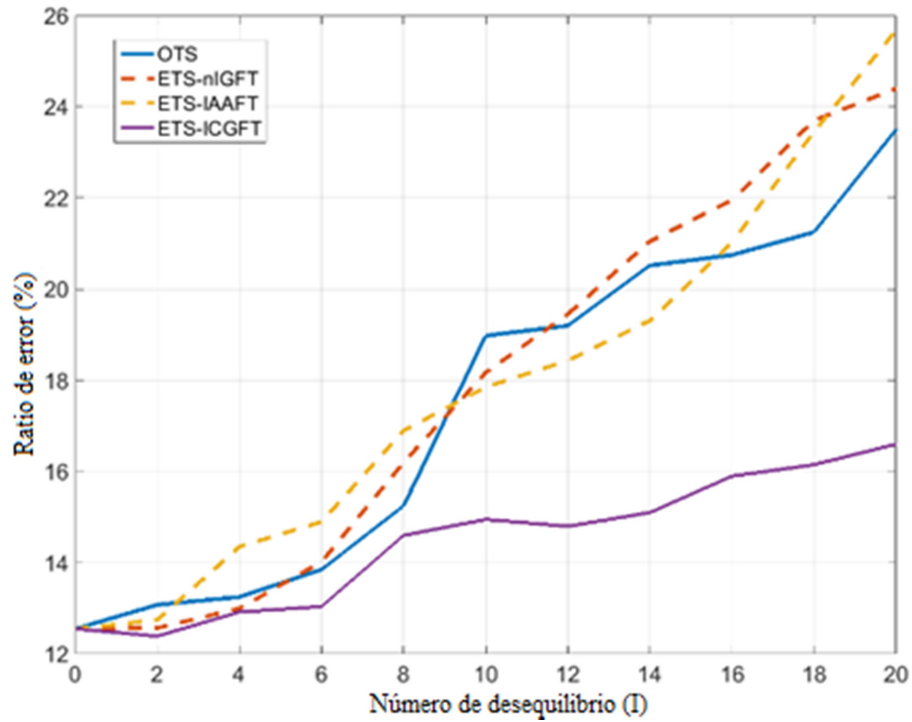


Figura 4.3. Tasa de error de clasificación para el aumento del número de desequilibrio.

Se puede ver en la Figura 4.3 que sólo EITS-ICGFT es capaz de reducir significativamente el error correspondiente al entrenamiento con OITS, aliviando así el problema del entrenamiento desequilibrado. La Tabla 4.4 muestra la tasa de error promediada sobre el dominio del número de desequilibrio para la OITS y para la EITS obtenida de los diferentes métodos de subrogación.

	OTS10	ETS- nIGFT	ETS-IAAFT	ETS-ICGFT
<i>Ratio de error (%)</i>	17.47	17.91	17.92	14.82

**Tabla 4.4.** Tasa de error promediada sobre el número de desequilibrio para los conjuntos de entrenamiento originales y ampliados.

#### 4.6. Conclusiones

Se ha presentado un nuevo algoritmo para generar señales subrogadas. Al considerar que las instancias de cualquier dominio arbitrario se definen en los vértices de un grafo, los nuevos algoritmos pueden obtener mucha más flexibilidad en comparación con los métodos tradicionales de subrogación basados en FT. Esta flexibilidad emana de la definición de la topología del grafo por la matriz Laplaciana. Se ha considerado el caso general de señales sobre grafos complejas, por lo que se ha definido una matriz Laplaciana Hermítica y una CGFT asociada. En consecuencia, se han realizado algunos análisis teóricos e interpretaciones con respecto a estas nuevas definiciones. Esto puede tener interés general en GSP. Además, mediante el concepto de invariantes GSA, se han deducido propiedades significativas de las señales sobre grafos originales que se conservan en las subrogadas.

Los experimentos sobre vectores aleatorios sintéticos y sobre reconocimiento automático de gestos manuales han confirmado la eficacia de ampliar el conjunto de entrenamiento original con señales subrogadas y que los mejores resultados corresponden al uso de ICGFT.

### Apéndice A: Derivación de la ecuación (2)

Separamos la matriz de adyacencia Hermítica en sus partes real e imaginaria:  $\mathbf{A} = \mathbf{A}_R + j\mathbf{A}_I$ . Es claro que  $\mathbf{A}_R$  es simétrica con elementos reales no negativos y  $\mathbf{A}_I$  es antisimétrica con elementos reales, más específicamente

$$\begin{aligned}\mathbf{A}_R(n, m) &= \text{Re}[a_{nm}] = a_{nm}^R = a_{mn}^R \geq 0, \\ \mathbf{A}_I(n, m) &= \text{Im}[a_{nm}] = a_{nm}^I = -a_{mn}^I\end{aligned}\quad (A1)$$

Calculando la forma cuadrática de la matriz Laplaciana Hermítica:

$$\begin{aligned}S(\mathbf{s}) &= \mathbf{s}^H \mathbf{L} \mathbf{s} = \mathbf{s}^H (\mathbf{D} - \mathbf{A}) \mathbf{s} \\ &= \mathbf{s}^H (\mathbf{D} - \mathbf{A}_R) \mathbf{s} - j \mathbf{s}^H \mathbf{A}_I \mathbf{s}\end{aligned}\quad (A2)$$

Pero

$$\begin{aligned}& \mathbf{s}^H (\mathbf{D} - \mathbf{A}_R) \mathbf{s} \\ &= \sum_{n=1}^N \sum_{m=1}^N a_{nm}^R (|s_n|^2 - s_n^* s_m) \\ &= \sum_{n=1}^N \sum_{m=1}^N a_{nm}^R |s_n - s_m|^2 - \sum_{n=1}^N \sum_{m=1}^N a_{nm}^R (|s_m|^2 - s_n^* s_m) \\ &= \sum_{n=1}^N \sum_{m=1}^N a_{nm}^R |s_n - s_m|^2 - \sum_{n=1}^N \sum_{m=1}^N a_{mn}^R (|s_m|^2 - s_m^* s_n) \\ &= \sum_{n=1}^N \sum_{m=1}^N a_{nm}^R |s_n - s_m|^2 - (\mathbf{s}^H (\mathbf{D} - \mathbf{A}_R) \mathbf{s}).\end{aligned}\quad (A3)$$

Donde se ha hecho uso de la simetría de  $\mathbf{A}_R$ , i.e.,  $a_{nm}^R = a_{mn}^R$ . A partir de (A3) se concluye que

$$\begin{aligned}& \mathbf{s}^H (\mathbf{D} - \mathbf{A}_R) \mathbf{s} \\ &= \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N a_{nm}^R |s_n - s_m|^2 \\ &= \sum_{n=1}^{N-1} \sum_{m>n}^N a_{nm}^R |s_n - s_m|^2.\end{aligned}\quad (A4)$$

Por otra parte

$$\begin{aligned}
 \mathbf{s}^H \mathbf{A}_I \mathbf{s} &= \sum_{n=1}^N \sum_{m=1}^N a_{nm}^I s_n^* s_m \\
 &= \sum_{n=1}^{N-1} \sum_{m>n}^N (a_{nm}^I s_n^* s_m + a_{mn}^I s_m^* s_n) \\
 &= \sum_{n=1}^{N-1} \sum_{m>n}^N a_{nm}^I (s_n^* s_m - s_m^* s_n) \\
 &= \sum_{n=1}^{N-1} \sum_{m>n}^N a_{nm}^I j2 \operatorname{Im}[s_n^* s_m] .
 \end{aligned} \tag{A5}$$

En la primera igualdad de (A5) se ha tenido en cuenta que  $a_{nn}^I = 0 \quad \forall n$ , y en la segunda igualdad que  $a_{nm}^I = -a_{mn}^I \quad \forall n \neq m$ . Finalmente se ha obtenido:

$$\begin{aligned}
 S(\mathbf{s}) &= \mathbf{s}^H \mathbf{L} \mathbf{s} = \\
 &= \sum_{n=1}^{N-1} \sum_{m>n}^N a_{nm}^R |s_n - s_m|^2 + \sum_{n=1}^{N-1} \sum_{m>n}^N 2a_{nm}^I \operatorname{Im}[s_n^* s_m] .
 \end{aligned} \tag{A6}$$

## Apéndice B: Derivación de la ecuación (19)

El teorema de Schur-Horn [53] aplicado a las Laplacianas Hermíticas establece que

$$\sum_{n=1}^M (\lambda_n - l_n) \geq 0 \quad M \leq N \tag{B1}$$

Donde  $\lambda_n$  y  $l_n \quad n=1\dots N$  son, respectivamente, los autovalores y los elementos de la diagonal principal de la Laplaciana Hermítica ordenados en orden no creciente. Considerando en (B1)  $M = 1$ , se obtiene

$$\lambda_1 \geq l_1 \Leftrightarrow \lambda_{\max} \geq l_{\max} \tag{B2}$$



Por otra parte, para  $M=N$  (B1) se cumple con la igualdad, entonces se puede escribir:

$$\begin{aligned}\lambda_N + \sum_{n=1}^{N-1} \lambda_n &= l_N + \sum_{n=1}^{N-1} l_n \Rightarrow \sum_{n=1}^{N-1} (\lambda_n - l_n) = l_N - \lambda_N \geq 0 \\ &\Rightarrow \lambda_N \leq l_N \Leftrightarrow \lambda_{\min} \leq l_{\min}\end{aligned}\tag{B3}$$

Se sabe que para cualquier vector  $\mathbf{s}$

$$\lambda_{\max} \geq \frac{S(\mathbf{s})}{\mathbf{s}^H \mathbf{s}}\tag{B4}$$

Por lo tanto, dado un vector  $\mathbf{s}$  se puede escribir un límite inferior para el ancho de banda del espectro  $B = \lambda_{\max} - \lambda_{\min}$

$$B \geq \max\left(l_{\max}, \frac{S(\mathbf{s})}{\mathbf{s}^H \mathbf{s}}\right) - l_{\min}\tag{B5}$$

## Referencias

- [1] T. Schreiber and A. Schmitz, “Surrogate time series,” *Physica D*, vol. 142, pp. 346–382, 2000.
- [2] T. Maiwald, E. Mammen, S. Nandi and J. Timmer, “Surrogate data – a qualitative and quantitative analysis,” *Mathematical Methods in Signal Processing and Digital Image Analysis*, Eds. R. Dahlhaus, J. Kurths, P. Maass and J. Timmer (Springer, NY), pp. 41–74, 2008.
- [3] R. Miralles, L. Vergara, A. Salazar and J. Igual , “Blind Detection of Nonlinearities in Ultrasonic Grain Noise,” *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 3, pp. 637-647, 2008.
- [4] D.P. Mandic, M. Chen, T. Gautama, M.M. Van Hulle and A. Constantinides: “On the characterization of the deterministic/stochastic and linear/nonlinear nature of time series,” *Proc. of the Royal Society A*, vol. 464, pp. 1141–1160, 2008.
- [5] R. Araújo, M. Small, and R. Fernandes, “Testing for Linear and Nonlinear Gaussian Processes in Nonstationary Time Series,” *Int. J. Bifurcation and Chaos*, vol. 25, pp. 1550013-1–1550013-19, 2015.
- [6] B. Jelfs, P. Vayanos, S. Javidi, V.S.L. Goh and D.P. Mandic, “Collaborative adaptive filters for online knowledge extraction and information fusion,” in *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, Eds. D. Mandic, M Golz, A. Kuh, D. Obradovic and T. Tanaka, Eds. (Springer, NY), pp. 3–21, 2007.
- [7] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, “Testing stationarity with surrogates: A time-frequency approach,” *IEEE Trans. on Signal Processing*, vol. 58, pp. 3459–3470, 2010.

- [8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst: “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE Signal Processing Magazine*, vol. 30, pp. 83–98, 2013.
- [9] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs,” *IEEE Trans. on Signal Processing*, vol. 61, pp. 1644–1656, 2013.
- [10] A. Sandryhaila and J. M. F. Moura, “Big Data Analysis with signal processing on graphs,” *IEEE Signal Processing Magazine*, vol. 31, pp. 80-90, 2014.
- [11] E. Pirondini, A. Vybornova, M. Coscia, and D. Van De Ville, “A Spectral Method for Generating Surrogate Graph Signals,” *IEEE Signal Processing Letters*, vol. 13, pp. 1275-1278, 2016.
- [12] D. I. Shuman, B. Ricaud and P. Vandergheynst, “Vertex-frequency analysis on graphs,” *Applied and Computational Harmonic Analysis*, vol. 40, pp. 260–291, 2016.
- [13] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs: Frequency analysis,” *IEEE Trans. on Signal Processing*, vol. 62, pp. 3042–3054, 2014.
- [14] D. Spielman, “Spectral graph theory,” *Combinatorial Scientific Computing*, Eds. U. Naumann and O. Schnek (Chapman and Hall/CRC Press), Ch 16, pp 1-23, 2012.
- [15] C. Hu, L. Cheng, J. Sepulcre, G. E. Fakhri, Y. M. Lu, and Q. Li, “A graph theoretical regression model for brain connectivity learning of Alzheimer’s disease,” 2013 *IEEE 10th International Symposium on Biomedical Imaging*, 2013, pp. 616-619.
- [16] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, “Learning Laplacian matrix in smooth graph signal representations,” *IEEE Trans. on Signal Processing*, vol. 64, pp. 6160–1673, 2016.

- [17] C. Zhang, D. Florencio, and P. A. Chou, “Graph signal processing— A probabilistic framework,” Microsoft Res., Redmond, WA, USA, Tech.Rep. MSR-TR-2015–31, 2015.
- [18] E.Pávez and A. Ortega, “Generalized precision matrix estimation for graph signal processing,” in Proc. of 2016 IEEE Int. Conf, on Acoust., Speech Signal Process. 2016, pp. 6350-6354.
- [19] B. Girault, “Stationary graph signals using an isometric graph translation,” in Proc. 23rd Eur. Signal Process. Conf., 2015, pp. 1516–1520.
- [20] N. Perraudin and P. Vandergheynst, “Stationary Signal processing on graphs”, IEEE Trans. on Signal Processing, vol. 65, pp. 3462–3477, 2017
- [21] G. Yu and H. Qu, “Hermitian Laplacian Matrix and positive of mixed graphs,” Applied Mathematics and Computation, vol. 269, pp 70-76, 2015.
- [22] R.C. Wilson and E.R. Hancock, “Spectral Analysis of Complex Laplacian Matrices,” Structural, Syntactic and Statistical Pattern Recognition LNCS, Eds. A. Fred, T.M. Caelli, R.P.W. Duin, A. Campilho and D. de Ridder (Springer, Heidelberg), vol. 3138, pp. 57–65, 2004.
- [23] G.T. Gilbert, “Positive definite matrices and Sylvester's criterion,” The American Mathematical Monthly, vol. 98, pp. 44–46, 1991.
- [24] R. Merris, “Laplacian matrices of a graph: a survey,” Linear Algebra with Applications, vol. 197, pp. 143–176, 1994.
- [25] X. D. Zhang, “The Laplacian Eigenvalues of Graphs: A Survey,” Linear Algebra Research Advances, Ed G. D. Ling (Nova Science Publishers Inc.), pp. 201–228, 2007.
- [26] H. Shapiro, “A survey of canonical forms and invariants for unitary similarity,” Linear Algebra Appl. vol.147, pp.101–167, 1991.

[27] V. Futorny, R.A. Horn and V.V. Sergeichuk, “Spetch’s criterion for systems of linear mapping,” *Linear Algebra Appl.* vol.519, pp.278–295, 2017.

[28] R. Mazumder and T. Hastie, “The graphical lasso: New insights and alternatives,” *Electronic journal of statistics*, vol. 6, pp. 2125-2149, 2012.

[29] K. Baba, R. Shibata and M. Sibuya: “Partial correlation and conditional correlation as measures of conditional independence,” *Australian & New Zealand Journal of Statistics*, vol 46, pp. 657-664. 2004.

[30] X. Chen, M. Xu and W. B. Wu: “Covariance and precision matrix estimation for high-dimensional time series,” *The Annals of Statistics*, vol. 41, pp. 2994-3021, 2013.

[31] V. Öllerer and C. Croux: “Robust high-dimensional precision matrix estimation,” *Modern Multivariate and Robust Methods*, Eds K. Nordhausen and S.Taskinen. Springer NY: Springer, pp 329–354, 2015.

[32] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J.D. Farmer, “Testing for nonlinearity in time series: the method of surrogate data,” *Physica D*, vol. 58, pp 77-94, 1992.

[33] T. Schreiber and A. Schmitz, “Improved surrogate data for nonlinearity tests,” *Physical Review Letters*, vol. 77, pp. 635-638, 1996.

[34] E. Mammen, S. Nandi, T. Maiwald and J. Timmer, “Effect of Jump Discontinuity for Phase-Randomized Surrogate Data Testing,” *Int. J. Bifurcation and Chaos*, vol. 19, pp. 403-408, 2009.

[35] J. H. Lucio, R. Valdes and L.R. Rodriguez, “Improvements to surrogate data methods for nonstationary time series,” *Physical Review E*, vol. 85, pp. 056202-1–056202-19, 2012.

[36] T. Schreiber, “Constrained randomization of time series data,” *Phys. Rev. Lett.*, vol. 80, pp. 2105-2108, 1998.

[37] D. Prichard and J. Theiler, “Generating surrogate data for time series with several simultaneously measured variables,” *Physical Review Letters*, vol. 73, pp. 951–954, 1994.

[38] P. Borgnat, P. Abry and P. Flandrin, “Using surrogates and optimal transport for synthesis of stationary multivariate series with prescribed covariance function and non-Gaussian joint distribution,” *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*, Kyoto (Japan), pp. 3729-3732, 2012.

[39] A. Salazar, G. Safont and L. Vergara, “Surrogate Techniques for Testing Fraud Detection Algorithms in Credit Card Operations,” *2014 IEEE International Carnahan Conference on Security Technology (ICCSR 2014)*, Rome (Italy), 2014.

[40] Y. Hirata, H. Suzuki, and K. Aihara, “Wind Modelling and its Possible Application to Control of Wind Farms,” *Signal Processing Techniques for Knowledge Extration and Information Fusion*. Eds. D. Mandic, M.Golz, A. Kuh, D. Obradovic and T. Tanaka (Springer, New York, NY), pp. 23–35, 2008.

[41] R. M. Gray, *Toeplitz and circulant matrices: a review*, Information System Laboratory, Stanford University, 1971.

[42] T. W. Liao, “Classification of weld flaws with imbalanced class data,” *Expert Systems with Applications*, vol. 35, pp. 1041–1052, 2008.

[43] S.J. Song and Y.G. Shin. “Eddy current flaw characterization in tubes by neural networks and finite element modeling,” *NDT&E International*, vol. 33, pp. 233–243, 2000.

[44] S. Bhattacharyya, S. Jha, K. Tharakunnel and J. C. Westland, “Data mining for credit card fraud: A comparative study,” *Decision Support Systems*, vol. 50, pp. 602-613, 2011.

[45] M. N. A. Kumar and H. S. Sheshadri, “On the classification of imbalanced datasets,” *Int. J. on Computer Applications*, vol. 44, pp. 1–7, 2012.

[46] J. Bardet, G. Lang, G. Oppenheim, A. Philippe and M. Taqqu, “Generators of Long-Range Dependent Processes: A Survey,” Long-range dependence: theory and applications. Eds. P. Doukhan, G. Oppenheim and M. Taqqu, (Birkhauser, Boston, MA), pp. 579–623, 2003

[47] S. Mitra and T. Acharya, “Gesture Recognition: A Survey,” IEEE Trans. on System, Man, and Cybernetics, vol. 37, pp. 311-324, 2007.

[48] G. Marin, F. Dominio and P. Zanuttigh, “Hand Gesture Recognition with Leap Motion and Kinect Devices,” Proceedings IEEE International Conference on Image Processing (ICIP2014), París (France), pp. 1565-1569, 2014.

[49] M.A. Moni and A.B.M. Shawkat Ali, “HMM based hand gesture recognition: A review on techniques and approaches”, IEEE International Conference on Computer Science and Information Technology (ICCSIT 2009), Beijing (China), pp. 433-437, 2009.

[50] N.H. Dardas and N.D. Georganas, “Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques”, IEEE Trans. on Instrumentation and Measurement, vol. 60, pp 3592-3606, 2011.

[51] Z. Parcheta and C.D. Martínez-Hinarejos, “Sign language gesture recognition using HMM,” accepted at 8th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2017), Faro (Portugal), 2017.

[52] B. Boashash, “Estimating and interpreting the instantaneous frequency of a signal – part 1: Fundamentals,” Proc. of IEEE, vol. 80, pp. 520–538, 1992.

[53] A. Horn, “Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix,” Amer. J. Math., vol. 76, pp. 620-630, 1954.

# Capítulo 5

## Conclusiones generales y líneas futuras

### 5.1. Conclusiones generales

Esta tesis realiza contribuciones novedosas en el ámbito emergente del procesado de señal sobre grafos. Así, los Capítulos 2 y 3 abordan la problemática de estimar una matriz de conexiones de los nodos del grafo (matriz Laplaciana y/o matriz de adyacencia) asumiendo modelos no Gaussianos multivariantes para los vectores de observaciones. En el Capítulo 2 se considera un modelo de mezclas Gaussianas y en el Capítulo 3 un modelo de análisis de componentes independientes. Por otra parte, en el Capítulo 4 se propone un nuevo algoritmo para la síntesis de señales subrogadas, que explota el posible conocimiento del modelo de grafo asociado. Este último puede obtenerse a partir de los métodos propuestos en los dos capítulos anteriores o definirse a partir de consideraciones del contexto de la aplicación.

Al final de cada capítulo se han incluido sus conclusiones particulares, algunas conclusiones generales son posibles, tales como las indicadas a continuación:



- Estimar las matrices de conectividad de un grafo a partir de coeficientes de correlación parcial, que tienen en cuenta la posible no-Gaussianidad de los datos, es una forma natural de extender los procedimientos basados en la matriz de precisión.
- Los resultados obtenidos en los experimentos simulados y reales confirman que se mejora la estima de la conectividad a partir de los métodos propuestos, cuando asumir el modelo multivariante Gaussiano no es adecuado. La cuantificación de la mejora depende de cada contexto concreto, pero es en todo caso significativa.
- La aportación teórica relativa a una nueva definición de una matriz Laplaciana compleja (Hermítica), así como la definición de la transformada de Fourier sobre grafos compleja que lleva asociada, tiene un interés general en el procesado de señales sobre grafos, limitado hasta ahora a matrices Laplacianas reales.
- La inclusión de un posible modelo de grafo en la síntesis de una señal mediante la generación de réplicas subrogadas ha demostrado su interés en los experimentos realizados, proporcionando resultados claramente superiores con respecto a métodos clásicos de subrogación, que no consideran modelos de grafos.

## **5.2. Líneas futuras**

Como en todo trabajo de investigación, los logros alcanzados animan a profundizar en ciertos aspectos relevantes. Indicamos a continuación un conjunto de temas que podrían constituir líneas de investigación sobre las que continuar los trabajos presentados en esta tesis.

### 5.2.1. “Sparsity”

Está bien asumido en teoría de grafos que, con carácter general, un grafo tiene más interés si manifiesta conexiones significativas entre un número reducido de nodos (condición denominada “sparsity” que podemos traducir como “escasez”). Esta situación es habitual en numerosas aplicaciones por lo que resulta de gran interés imponerla como parte del propio diseño del grafo. Varias opciones son posibles en los métodos de estimación de la matriz Laplaciana propuestos en esta tesis. La opción más simple es una umbralización de los coeficientes de correlación parcial, facilitada por el hecho de que dichos coeficientes se estiman normalizados entre 0 y 1. Otra opción (que puede ser complementaria con la anterior) es considerar que los modelos no-Gaussianos asumidos implican una mezcla de contribuciones. En el caso de los modelos GMM se mezclan densidades de probabilidad Gaussianas, en el caso de los modelos ICA se mezclan fuentes independientes no-Gaussianas. Si podemos separar la contribución de cada uno de los elementos de la mezcla al coeficiente de correlación parcial, podríamos desestimar las contribuciones que no fueran significativas. Por ejemplo en la ecuación (21) del Capítulo 2, considerar sólo los términos que superen cierto umbral en el sumatorio. Asimismo en las conclusiones del Capítulo 3 ya se sugirió una umbralización de las contribuciones separadas de cada fuente al error cuadrático medio de la estima no-lineal.

### 5.2.2. “Smoothness”

La condición de “smoothness” o “suavidad” en la matriz Laplaciana es también conveniente para obtener modelos de grafos de señales de interés práctico. Esencialmente implica que en general las señales modeladas mediante un mismo grafo deben tener una “smoothness” similar medida a través de dicho grafo (véase la definición de “smoothness” en la ecuación (1) del Capítulo 4). Esta medida puede interpretarse como el nivel de variabilidad de la señal si nos movemos por los diferentes nodos del

grafo. La mayor parte de los métodos que controlan la “smoothness” lo hacen a través de funciones de coste que establecen un compromiso entre la estimación muestral de la matriz de precisión (inversa de la estimación muestral de la matriz de correlación) y la maximización de la “smoothness”. Los métodos propuestos en esta tesis, en última instancia proporcionan estimas de una matriz de precisión mejorada, podríamos por tanto plantar reajustar las funciones de coste incorporando la matriz mejorada.

### **5.2.3. Aplicación en detección/clasificación**

Tal y como se ha visto en los ejemplos de los Capítulos 2 y 3, el interés último del modelo de grafo era proporcionar una caracterización alternativa de la que pudieran deducirse otras características (por ejemplo la conectividad del grafo) para resolver problemas de detección/clasificación automáticas. Desde esa perspectiva, puede resultar conveniente profundizar en la investigación sobre la relación entre la calidad de la estimación de la Laplaciana y la calidad de la estimación de las características que se extraigan de ella para detección/clasificación automáticas. Así por ejemplo, no tiene sentido realizar un gran esfuerzo en conseguir una mejora de la calidad de la estima de la Laplaciana, que luego no supone una mejora significativa en la calidad de la estima de la característica extraída de ella, y viceversa. Establecer, sea por análisis y/o empíricamente, una cuantificación de este aspecto puede ser una línea futura de trabajo.

### **5.2.4. Definición de la matriz Laplaciana en los métodos de subrogación**

Tal y como vimos en el Capítulo 4, el éxito de la subrogación de señales basada en grafos depende de una buena definición del modelo de grafo (la matriz Laplaciana). En el ejemplo de aplicación considerado se priorizó la preservación de propiedades “locales”, conectando una muestra con sus vecinas. Esta es sólo una opción entre muchas posibles, pueden enfatizarse propiedades específicas de la señal, o utilizarse

modelos probabilísticos a partir de los cuales se construye el modelo de grafos como hicimos en los Capítulos 3 y 4. Comparar en profundidad diferentes opciones y entender en qué forma se preservan las propiedades de la señal original en las subrogadas es una evidente línea futura de investigación.

### **5.3. Publicaciones generadas a partir de los resultados de la tesis**

J. Belda, L.Vergara, A. Salazar, G. Safont: Estimating the Laplacian matrix of Gaussian mixtures for signal processing on graphs, *Signal Processing*, vol. 148, pp 241-249, Julio 2018. IF: 4.662, Q1.

J. Belda, L.Vergara, G. Safont, A. Salazar: Computing the partial correlation of ICA models for Non-Gaussian Graph Signal Processing, *Entropy*, vol 21, 22. (Special Issue in Information Theory Applications in Signal Processing), pp 1-16, 2019. IF: 3.23, Q2.

J. Belda, L.Vergara, G. Safont, A. Salazar, Z. Parcheta: A new surrogating method by the Complex Graph Fourier Transform, *Entropy*, vol 21,759, pp 1-18, 2019. IF: 3.23, Q2.