

# MODELO DE MEDICIÓN PARA LA ESTIMACIÓN DE LA MATRIZ DE CORRELACIÓN DE DATOS ORDENADOS



**JUAN PABLO HOYOS SANCHEZ**

Tesis de Doctorado en Ciencias de la Electrónica

Director:

Ph.D. Pablo Emilio Jojoa Gomez

Universidad del Cauca  
Facultad de Ingeniería Electrónica y Telecomunicaciones  
Departamento de Electrónica y Telecomunicaciones  
Linea de Investigación en Procesamiento de Señales  
Popayán, Mayo de 2018



**JUAN PABLO HOYOS SANCHEZ**

**MODELO DE MEDICIÓN PARA LA  
ESTIMACIÓN DE LA MATRIZ DE  
CORRELACIÓN DE DATOS ORDENADOS**

Tesis presentada a la Facultad de Ingeniería Electrónica y Telecomunicaciones de la  
Universidad del Cauca para la obtención del título de:

**Doctor en:  
CIENCIAS DE LA ELECTRÓNICA**

Director:

Ph.D. Pablo Emilio Jojoa Gomez

Asesor Externo:

Ph.D. Rafael E. Carrillo

**Comisión Evaluadora**

Ph.D. Carlos Alberto Gaviria Lopez

Ph.D. Diego Alberto Bravo Montenegro

Ph.D. Evelio Astaiza Hoyos

Ph.D. Marcelo Fiori

Popayán  
2018



*A mi amada Natalia, y mis adorables hijos,  
Pablo Samuel y Luciana.*

*Más vale adquirir sabiduría que oro; más vale  
adquirir inteligencia que plata.*  
- Proverbios 16:16



## Agradecimientos

A Dios que es mi guía y mi soporte en los momentos más difíciles de mi vida.

A mi bella y amada familia por apoyarme incondicionalmente para alcanzar una nueva meta. A mi esposa Natalia Llanten por su amor y apoyo incondicional y mis amados hijos Pablo Samuel y Luciana por llenar de alegría mi vida.

A mis padres Paula Deicy Sanchez y Jorge Dario Hoyos por su amor, paciencia, tiempo y fortaleza para alcanzar mis sueños.

Al profesor Pablo Jojoa, por darme esta oportunidad, por guiarme, por levantarme y estar ahí siempre, espero la vida me de la oportunidad de retribuir toda su dedicación y apoyo.

Al profesor Sebastian Pazos de la Universidad Nacional de La Plata (Argentina) por brindarme su amistad, sus valiosos aportes y permitirme realizar mi estancia en tan importante universidad.

Al doctor Rafael Carrillo, por su conocimiento, aportes y su colaboración desinteresada en la realización de esta tesis.

Al profesor Gonzalo Arce de la Universidad de Delaware (USA) por sus aportes y permitirme realizar mi estancia en tan importante universidad.

A las personas del grupo GNTT que me colaboraron de alguna u otra forma en la realización de este trabajo.

Al comité del Doctorado por su gran apoyo en cada una de las actividades. A la Universidad del Cauca y a COLCIENCIAS, por el soporte económico brindado que hicieron posible el desarrollo de mis estudios.



## Resumen

Los métodos basados en estadísticos de orden o *order statistics*, son a menudo usados en finanzas, calidad de control, procesamiento de datos y señales, especialmente cuando las señales de interés están inmersas en ruido impulsivo. Estos permiten incluir información de tiempo a costa de incrementar la dimensión del problema. En problemas de grandes dimensiones, usualmente requerimos conocer solo estadísticos de segundo orden.

En esta tesis se diseñó un modelo de medición basado en mediciones cuadráticas de rango uno o *sketch* para estimar la matriz de correlación de datos que han sido ordenados. Se estudia la estructura de la matriz para un proceso aleatorio vectorial estacionario, y mostramos que su estructura depende de una submatriz diagonal y una submatriz con diagonal cero. Explotando esta estructura se diseñan dos algoritmos, el primero luego de la estimación realiza un promedio y el segundo algoritmo tiene en cuenta la información de la estructura para la definición de una nueva relajación convexa para recuperar la matriz. Se demuestra que el número de mediciones es proporcional al tamaño original del problema (sin ordenamiento), además se desarrolla un muestreo en bloques que optimiza el muestreo de la estructura de la matriz. Para ambos algoritmos se presentan simulaciones para mostrar el desempeño en la reconstrucción y la robustez de la estimación cuando el ruido uniforme está presente, superando el desempeño mostrado por el estimador referente.

Para los casos en que no es posible suponer un número infinito de realizaciones del proceso estacionario, se desarrolla una modificación al esquema de medición que permite una convergencia más rápida, con lo cual se consolida un modelo de medición para la estimación de la matriz de correlación.

Se desarrolló una extensión del segundo algoritmo basado en métodos de división proximal, en especial en los métodos *forward-backward* y primal-dual, que al solucionar iterativamente una secuencia de subproblemas logra una computación eficiente para el caso de datos de grandes dimensiones.

Por último, basados en el modelo de medición se diseñó un estimador para datos ordenados que presentan un comportamiento no-estacionario, mostrando una importante disminución en el consumo de recursos y una leve reducción en el número de mediciones respecto al estimador referente.

**Palabras clave:** Matriz de correlación, estadísticos de orden, mediciones cuadráticas, *sketch*, métodos de división proximal, *forward-backward*, primal-dual, ruido uniforme.



# Abstract

Methods based on ordered statistics are often used in finance, quality control, data and signals processing, especially when the signals of interest are immersed in impulsive noise. These allow to include information of time at the expense of increasing the dimension of the problem. In large problems, we usually need to know only second-order statistics.

In this thesis a measurement model based on quadratic measurements of rank one or textit sketch was designed to estimate the matrix of correlation of data that have been ordered. We study the structure of the matrix for a stationary vector random process, and show that its structure depends on a diagonal submatrix and a submatrix with zero diagonal. Exploiting this structure two algorithms are designed, the first after the estimation performs an average and the second algorithm takes into account the information of the structure for the definition of a new convex relaxation to recover the matrix. It is demonstrate that the number of measurements is proportional to the original size of the problem (without ordering), in addition a block sampling is developed to optimize the sampling of the structure of the matrix. For both algorithms we provide simulations to show the performance in the reconstruction and the robustness of the estimate when uniform noise is present, surpassing the performance shown by the referent estimator.

For cases in which it is not possible to assume an infinite number of realizations of the stationary process, a modification to the measurement scheme is developed that allows a faster convergence, which consolidates a measurement model for the estimation of the correlation matrix.

We developed an extension of the second algorithm based on proximal division methods, especially in the forward-backward and primal-dual methods, which when iteratively solving a sequence of sub-problems ensures an efficient computation for the case of large data.

Finally, based on the measurement model, an estimator was designed for ordered data that exhibit a non-stationary behavior, showing a significant decrease in the consumption of resources and a slight reduction in the number of measurements with respect to the referent estimator.

**Keywords:** Correlation matrix, ordered statistics, quadratic measurements, sketch, proximal splitting methods, forward-backward, primal-dual, uniform noise.



# Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos	3
1.1.1. Objetivo General	3
1.1.2. Objetivos específicos	3
1.2. Hipótesis	3
1.3. Principales Contribuciones	4
1.4. Organización y Materiales Publicados	4
<b>2. Aspectos Teóricos: Resultados Previos</b>	<b>6</b>
2.1. Notación	6
2.2. Introducción	6
2.3. Matrices de bajo rango	8
2.3.1. Aproximación de bajo rango	8
2.3.2. Recuperación bajo rango y minimización por traza	9
2.3.3. Recuperación de fase	10
2.3.4. Mediciones cuadráticas o sketch	11
2.4. Estimación de matrices dispersas o <i>sparse</i>	11
2.4.1. Umbral o <i>Thresholding</i>	12
2.4.2. Métodos aleatorios	12
2.5. Estimación de matrices Toeplitz	15
2.5.1. Muestreo de Covarianza Compresivo	16
2.5.2. Muestreo co-primo	17
2.5.3. Mediciones cuadráticas o sketch	18
2.6. Estimación de matrices Block Toeplitz	18
2.6.1. Expansiones en producto Kronecker	19
2.6.2. Muestreo de Covarianza Compresivo	19
<b>3. Estimación de la Matriz de Correlación</b>	<b>21</b>
3.1. Modelo de Medición Cuadrático	21
3.2. Principales Resultados y Algoritmos	25
3.2.1. Estructura de la Matriz de Correlación	25
3.2.2. Algoritmo 1: Estimación de la matriz de correlación	28
3.2.3. Experimentos numéricos - algoritmo 1	30
3.2.4. Algoritmo 2: Estimación de la matriz de correlación	32

3.2.5. Experimentos numéricos - algoritmo 2 . . . . .	35
3.3. Modelo de medición para la estimación de la matriz de correlación de datos ordenados . . . . .	38
3.3.1. Mediciones cuadráticas . . . . .	38
3.3.2. Modelo consolidado de medición . . . . .	40
<b>4. Estimación en grandes dimensiones</b>	<b>41</b>
4.1. Aspectos teóricos . . . . .	42
4.2. Métodos de división proximal . . . . .	43
4.2.1. División forward-backward . . . . .	44
4.3. Métodos Primal-Dual . . . . .	44
4.4. Resultados principales . . . . .	45
4.4.1. Algoritmo . . . . .	45
4.4.2. Complejidad computacional . . . . .	49
4.4.3. Experimentos numéricos . . . . .	49
4.5. Conclusiones . . . . .	52
<b>5. Estimación de la Matriz no-estacionaria</b>	<b>56</b>
5.1. Estructura de la matriz de correlación . . . . .	56
5.2. Estimador para el caso no-estacionario . . . . .	57
5.2.1. Resultados numéricos . . . . .	59
<b>6. Conclusiones y Trabajo Futuro</b>	<b>63</b>
6.1. Conclusiones . . . . .	63
6.2. Limitaciones . . . . .	64
6.3. Trabajo Futuro . . . . .	64
<b>A. Apéndice: Optimización convexa</b>	<b>66</b>
<b>B. Apéndice: Pruebas de proposiciones Capítulo 3</b>	<b>68</b>
B.1. Operadores lineales . . . . .	68
B.2. Operadores adjuntos . . . . .	69
<b>Bibliografía</b>	<b>71</b>

# Lista de Figuras

3.1. Flujo de datos . . . . .	22
3.2. Sketches . . . . .	23
3.3. Matriz de correlación de un proceso estacionario poisson para distintos números de realizaciones. . . . .	28
3.4. Convergencia de la matriz de correlación . . . . .	29
3.5. NMSE respecto al número de mediciones $m$ para las dimensiones $n = 5$ , $n=6$ y $n=7$ . . . . .	31
3.6. NMSE respecto al número de mediciones $m$ para una dimensión de $n=5$ y diferentes niveles de ruido. . . . .	31
3.7. NMSE respecto al número de mediciones $m$ para una dimensión de $n=6$ y diferentes niveles de ruido. . . . .	32
3.8. NMSE como función del número de mediciones $m$ con dimensión $n=7$ para la estimación general y la estimación propuesta mediante $\mathbf{D}$ y $\mathbf{B}$ . . . . .	36
3.9. NMSE como función del número de mediciones $m$ con dimensiones $n = 6,8,10,12$ , y $n=15$ para el esquema de estimación propuesto. . . . .	37
3.10. Comparación del NMSE entre el estimador general y el esquema propuesto para $n=6$ y diferentes niveles de ruido. . . . .	38
3.11. NMSE para el esquema propuesto para las dimensiones $n = 10$ and $n = 15$ con diferentes niveles de ruido. . . . .	39
3.12. NMSE para el muestreo de Chen et al. (2015) y el muestreo propuesto. . . . .	40
4.1. NMSE vs el número de iteraciones para las dimensiones $n = 10$ y $n = 17$ . . . . .	51
4.2. Comparación del NMSE del algoritmo 2, el algoritmo 4(primal-dual) y el estimador de Chen et al. . . . .	53
4.3. NMSE del estimador primal-dual para diferentes dimensiones. . . . .	54
4.4. NMSE para la dimensión $n=10$ con diferentes niveles de ruido. . . . .	55
4.5. NMSE para la dimensión $n=15$ con diferentes niveles de ruido. . . . .	55
5.1. Señales ECG de la base MIT-BIH . . . . .	60
5.2. Señales ECG de la base MIT-BIH en el rango 1975 a 2000 segundos . . . . .	61
5.3. Comparación del NMSE entre el estimador general y el esquema propuesto para $n = 3, 5, 7$ , caso no-estacionario. . . . .	61
5.4. Comparación del NMSE entre el estimador general y el esquema propuesto para $n = 5$ y diferentes niveles de ruido, caso no-estacionario. . . . .	62

5.5. Comparación del NMSE entre el estimador general y el esquema propuesto para $n = 7$ y diferentes niveles de ruido, caso no-estacionario. . . . .	62
---	----

# Lista de Tablas

5.1. Uso de la Memoria de Acceso Aleatorio (RAM) por parte del estimador estándar y el estimador propuesto. . . . .	60
---	----



## Lista de Abreviaturas

<b>ADMM</b>	Alternate Directions Method of Multipliers
<b>CoVALSA</b>	<i>CoVariance by Argumented Lagrangian Shrinkage Algorithm</i>
<b>ECG</b>	Electrocardiográficas
<b>i.i.d.</b>	Independientes e Idénticamente Distribuidas
<b>PCA</b>	Análisis de Componentes Principales
<b>RAM</b>	Memoria de Acceso Aleatorio
<b>STAP</b>	<i>Space-Time Adaptive Processing</i>
<b>SVD</b>	<i>Singular Value Decomposition</i>
<b>TIC</b>	Tecnologías de la Información y las Comunicaciones



# Capítulo 1

## Introducción

La teoría de señales ha sido un área de constante investigación desde sus primeros usos en la teoría de la información ([Shannon, 1948](#)). Sus avances han permitido el desarrollo de novedosos sistemas que han impactado notablemente la manera de vivir, llevando a la sociedad a la era de las Tecnologías de la Información y las Comunicaciones (TIC). El rápido crecimiento en la generación de datos e información por los diferentes sistemas de comunicaciones, y las actuales limitantes han ocasionado que los problemas de inferencia estadística y procesamiento de datos de altas dimensiones se hayan convertido en un reto de alta complejidad, ya que buscan cada vez métodos más eficientes de estimación de las estadísticas de segundo orden, manteniendo niveles óptimos de precisión ([Chen et al., 2015](#); [Leus y Tian, 2011](#)).

Además, si los datos han sido alterados por señales aleatorias indeseadas (ruido) que obedezcan a distribuciones de probabilidad Gaussianas, los modelos lineales son los más adecuados para su modelamiento y análisis ([Unser y Tafti, 2014](#)). Pero, cuando estas señales indeseadas presentan picos, es decir presentan comportamientos impulsivos, los modelos no-gaussianos son los más adecuados debido principalmente a que permiten tener una varianza infinita por lo que se hace necesario buscar métodos no lineales para su reducción ([Arce, 2004](#); [Unser y Tafti, 2014](#)). Entre estos métodos, uno muy utilizado son los filtros L1 ([Palmieri y Boncelet, 1989](#)) que ayudan a la eliminación de estos picos y permiten una óptima estimación basada en el ordenamiento temporal y de rango, también conocidos como estadísticos de orden o *order statistics*. El ordenamiento de rango preserva transiciones abruptas de la señal y remueve los componentes de ruido con cola pesada (impulsivos), mientras que el orden temporal de los datos se necesita mantener para realizar procesamiento lineal sobre los mismos, ya que permite efectuar modificaciones predecibles del contenido espectral de la secuencia observada ([Gandhi y Kassam, 1991](#)). En este tipo de estimadores como en aplicaciones en campos como finanzas, control de calidad, estudios de robustez y procesamiento de señales es esencial para su diseño la estimación de la matriz de correlación de los datos ordenados la cual presenta una forma Toeplitz en bloques dispersa (*sparse*) ([Palmieri y Boncelet, 1989](#); [Arce, 2004](#)). Pero, la principal desventaja al ordenar los datos en rango y tiempo radica en el incremento del tamaño de la matriz de correlación, ya que pasa de la dimensión original  $n \times n$  a tener una dimensión de  $n^2 \times n^2$ .

En radares aerotransportados una técnica muy utilizada para la eliminación de los ecos terrestres no deseados o *ground-clutter* es el procesamiento adaptativo tiempo-espacio o *Space-Time Adaptive Processing* (STAP), en donde se utilizan  $M$  antenas y  $N$  formas de ondas para la recolección de los datos (Abramovich et al., 2008b). El problema central al usar algoritmos 2D-STAP es la necesidad de estimar la matriz de covarianza usando generalmente un número limitado de muestras, ya que por diseño propio de los radares se presenta esta condición. Además, sin asumir de antemano una estructura de esta matriz la estimación no sería posible ya que el número de muestras necesario para la correcta estimación no estarían disponibles, por lo que en ciertos escenarios el *clutter* es estacionario y por tanto se tiene que la matriz de covarianza presenta una estructura de matriz Toeplitz en bloques (Guerci y Baranoski, 2006; Abramovich et al., 2008b).

Adicionalmente, la estructura de Toeplitz en bloques aparece en matemáticas, computación científica, ingeniería, como también en problemas de restauración de imágenes, series de tiempo, sistemas de almacenamiento magnético, procesos aleatorios estacionarios 2-D, ecuaciones diferencias e integrales, teoría de control, procesamiento de señales y radares (Krishna et al., 2013; Omar y Slock, 2008; Hong et al., 2015; Tsiligkaridis y Hero, 2013; Abramovich et al., 2008b; Oudin y Delmas, 2008; Gutiérrez-Gutiérrez y Crespo, 2012).

El incremento en el tamaño de la matriz de correlación cuando se hace uso de estadísticos de orden y las limitantes actuales en capacidad de procesamiento y memoria disponible, es lo que hace necesario desarrollar nuevas técnicas que puedan tratar este problema. Por ejemplo, técnicas como censado compresivo o *compressed sensing* (Bryan y Leise, 2013) y *sketches* (Chen et al., 2015; Bahmani y Romberg, 2015; Traganitis y Giannakis, 2017; Tang et al., 2017; Dasarathy et al., 2017) recurren al uso de optimizaciones y matrices aleatorias para realizar una recuperación de la matriz a partir de mediciones y procesamiento de los datos sin requerir del almacenamiento de todos los datos.

Además, las técnicas de *sketches* y mediciones compresivas requieren que la matriz exhiba alguna estructura para lograr una correcta recuperación. Cuando el vector aleatorio tiene una matriz de covarianza simultáneamente dispersa y de bajo rango, en Bahmani y Romberg (2015) se demuestra que si los *sketches* presentan una estructura especial, entonces se puede utilizar un algoritmo sencillo de dos estados que explota esta estructura para una estimación precisa. En Bioucas-Dias et al. (2014) se propone un método, denominado CoVALSA (*CoVariance by Argumented Lagrangian Shrinkage Algorithm*), que estima la matriz de covarianza mediante mediciones compresivas con información a priori sobre la estructura de la matriz incluyendo la estructura Toeplitz, dispersa y de bajo rango. En Dasarathy et al. (2015) se estudia la recuperación de la covarianza con una estructura llamada raleza distribuida mediante productos de tensores. En Tsiligkaridis y Hero (2013) se desarrolla un método de estimación de matrices de covarianza basado en representar la matriz Toeplitz a bloque en series de productos Kronecker para luego utilizar mínimos cuadrados penalizados en rango permutados. En Romero et al. (2016); Romero y Leus (2013a) se desarrolla un esquema de estimación de matrices, denominado Muestreo de Covarianza Compresivo, que busca recuperar la matriz de covarianza a partir de menos

muestras, obtenidas de acuerdo a un criterio llamado *reglas dispersas* partiendo de que la matriz de covarianza para estructuras dispersas, Toeplitz y de bajo rango pueda ser expresada como una combinación lineal de matrices de un conjunto base.

En [Chen et al. \(2015\)](#) se propone un modelo de medición cuadrático (rango uno) que permite una estimación precisa de la matriz de covarianza sin ruido, bajo almacenamiento de datos y robustez frente a ruido uniforme. En el modelo allí desarrollado se analizaron cuatro estructuras: bajo rango, bajo rango Toeplitz, dispersa y dispersa bajo rango. Para la estimación de la matriz de correlación de datos ordenados el modelo cuadrático solo explota la característica de *sparsity* de la matriz y la condición de ser semidefinida positiva. El presente trabajo extiende los métodos desarrollados en [Chen et al. \(2015\)](#) para estimar la matriz de correlación de datos ordenados y se analiza la estructura que presenta tal matriz.

Teniendo en cuenta lo anterior, la presente tesis busca darle respuesta al siguiente interrogante: ¿Cómo estimar la matriz de correlación con estructura Toeplitz en bloques dispersa utilizando un número de mediciones menor que la dimensión de los datos originales?

Para dar respuesta a este interrogante se debe desarrollar un modelo que capture la estructura particular de la matriz buscando obtener una reducción en la dimensión, y logre resolver eficiente el problema de estimación.

## 1.1. Objetivos

### 1.1.1. Objetivo General

Diseñar un modelo de medición para la estimación de la matriz de correlación de datos ordenados.

### 1.1.2. Objetivos específicos

- Determinar las características de los principales métodos de estimación de matrices con mediciones comprimidas.
- Diseñar un método para la recuperación de la matriz de correlación de datos ordenados.
- Evaluar por simulación el desempeño del modelo planteado.

## 1.2. Hipótesis

Como hipótesis principal se formuló: Un modelo de medición utilizando *sketches* permitirá la estimación estable de la matriz toeplitz en bloques con mediciones menores que la dimensión original de la matriz.

### 1.3. Principales Contribuciones

De forma resumida, los principales resultados alcanzados en la tesis fueron los siguientes:

- Se analizó la estructura de la matriz de correlación que presenta un proceso aleatorio estacionario cuando ha sido ordenado.
- Se diseñaron tres algoritmos para la estimación de la matriz de correlación utilizando un número de mediciones menor que la dimensión de los datos.
- Se diseñó un estimador para el caso de procesos no-estacionarios.
- Se consolidó un modelo de medición para la estimación de la matriz de correlación.
- Recurriendo a métodos de división proximal, se realizó una extensión del algoritmo 2 para la estimación eficiente computacionalmente de la matriz de correlación cuando el proceso aleatorio presenta una gran dimensión.

### 1.4. Organización y Materiales Publicados

A partir de los resultados de este trabajo de investigación, fueron realizadas algunas publicaciones, no obstante queda pendiente generar mas publicaciones donde sean presentados los resultados finales logrados (Hoyos et al., 2018). Esta tesis esta organizada de la siguiente manera:

En el **Capítulo 2** se discuten los principales métodos de estimación de matrices, especialmente para las estructuras mas relevantes: bajo rango, dispersas o *sparse*, Toeplitz y block Toeplitz. La sección 2.3 presenta la teoría de los métodos actuales para la estimación de matrices de bajo rango, entre los cuales se incluyen: Aproximación de bajo rango, minimización por traza, recuperación de fase y mediciones cuadráticas. La sección 2.4 presenta los métodos mas relevantes para la estimación de matrices dispersas o *sparse*, realizando una clasificación de acuerdo al tipo de técnica utilizada por los estimadores: umbral o aleatoria. La sección 2.5 presenta los métodos actuales para la estimación de matrices Toeplitz haciendo énfasis a métodos que utilizan mediciones compresivas. Por último, la sección 2.6 presenta los métodos mas relevantes para la estimación de matrices block Toeplitz. Estos resultados han sido presentados en Hoyos y Jojoa (2017a) y Hoyos y Jojoa (2017b)

El **Capítulo 3** presenta el problema de la estimación de la matriz de correlación de datos que han sido ordenados en tiempo como en rango. Se estudia la matriz de correlación y se presenta la estructura que se tiene cuando un proceso estacionario es ordenado. También se introducen dos nuevos algoritmos para la resolución de este problema, basados en relaciones convexas y la inclusión de la estructura descubierta de la matriz de correlación. El primer algoritmo consiste en reemplazar las submatrices por un promedio de las mismas lo que permite obtener una mejoría en la recuperación de la matriz.

El segundo algoritmo reformula la estructura del muestro, planteando un muestreo en bloques del mismo tamaño y número que los presentes en la matriz de correlación para poder incluir directamente la información de la estructura de la matriz de correlación en el programa convexo. Este nuevo problema de optimización convexo permite reducir notablemente el número de mediciones o muestras necesarias para alcanzar una estimación correcta de la matriz, pasando de depender de  $n^2(n^2 + 1)/2$  a solo depender de  $n(n + 1)/2$ . Basados en el nuevo esquema de medición y algoritmo 2 planteamos un nuevo modelo de medición para la estimación de la matriz de correlación de datos ordenados.

Por último, para ambos algoritmos se analiza el desempeño respecto al número de mediciones necesarias como su robustez frente al ruido uniforme. Estos resultados han sido publicados en [Hoyos et al. \(2017b\)](#), [Hoyos et al. \(2017a\)](#) y ([Hoyos y Jojoa, 2017c](#))

En el **Capítulo 4** se presenta un tercer algoritmo como extensión del algoritmo 2, cuando la dimensión de los datos es muy grande. Esto se hace necesario ya que la computación de los anteriores algoritmos es computacionalmente costosa y por tanto la computación de datos de grandes dimensiones se vuelve inviable. Recurriendo a métodos de división proximales o *proximal splitting methods*, en especial *forward-backward* y primal dual, se diseña el algoritmo, se analiza su tiempo de convergencia y su desempeño para datos con y sin ruido uniforme. Cabe aclarar que el ruido uniforme fue seleccionado en esta tesis para permitir una comparación con trabajos previos.

En el **Capítulo 5** basado en el modelo de medición desarrollado, se diseña un estimador para la recuperación de la matriz de correlación de datos no-estacionarios que han sido ordenados. Como caso de estudio se toman 7 señales Electrocardiográficas (ECG) a partir de las cuales utilizando el modelo de medición se recupera la matriz de correlación con y sin la presencia de ruido uniforme. El estimador diseñado reduce la complejidad computacional, mientras que el número de mediciones es ligeramente menor que el estimado general.

En el **Capítulo 6** se discuten los resultados de esta tesis, se resaltan los aportes originales de la misma y se presentan posibles trabajos futuros.

Finalmente, en los apéndices y anexos se brindan los aspectos matemáticos que soportan el desarrollo de la tesis.

# Capítulo 2

## Aspectos Teóricos: Resultados Previos

A continuación se presentan y analizan los resultados previos fundamentales que son la base de los modelos y algoritmos que se desarrollan en los capítulos 3 y 4 de esta tesis (Hoyos y Jojoa, 2017a,b). Para complementar este capítulo, en el apéndice A se dan algunas nociones de optimización.

### 2.1. Notación

Antes de proceder, introducimos varias notaciones usadas a lo largo de este capítulo. La notación utilizada para representar una matriz y un vector son respectivamente  $\mathbf{X}$  y  $\mathbf{x}$ . Dejamos que  $[n]$  represente el conjunto  $\{1, 2, \dots, n\}$ . La *Singular Value Decomposition* (SVD) de  $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$ , con  $*$  representando el operador hermitico (transpuesta conjugada). Denotamos por  $\|\mathbf{X}\|$ ,  $\|\mathbf{X}\|_F$  y  $\|\mathbf{X}\|_*$  la norma espectral (operador norma), norma de Frobenius y la norma nuclear de  $\mathbf{X}$ . La norma  $\ell_1$  elemento a elemento y el tamaño de soporte de  $\mathbf{X}$  se denota por  $\|\mathbf{X}\|_1$  y  $\|\mathbf{X}\|_0$  respectivamente. El producto interno matricial es definido como  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Tr}(\mathbf{X}\mathbf{Y}^*)$ , y el producto kronecker está denotado por  $\otimes$ . El operador adjunto viene denotado por  $\dagger$ .

### 2.2. Introducción

La estimación de matrices juega un papel preponderante en muchas aplicaciones, en especial la estimación de la matriz de covarianza es fundamental en aplicaciones como procesamiento de señales (Krim y Viberg, 1996), reconocimiento de patrones, geometría convexa computacional, estadística de altas dimensiones, ..., ya que captura la estructura de la covarianza de las variables aleatorias (Vershynin, 2012). Sea  $\mathbf{x}$  un vector aleatorio en  $\mathbf{R}^n$  distribuido acorde a alguna función de distribución  $\mu$ . La matriz de covarianza viene definida como

$$\mathbf{\Sigma} = \mathbb{E}\{(\mathbf{x} - \mathbb{E}\mathbf{x})(\mathbf{x} - \mathbb{E}\mathbf{x})^T\} = \mathbb{E}\{(\mathbf{x} - \mathbb{E}\mathbf{x}) \otimes (\mathbf{x} - \mathbb{E}\mathbf{x})\} \quad (2.1)$$

y tiene las propiedades de ser simétrica y semi-definida positiva. Los autovectores de la matriz de covarianza son llamados los componentes principales, he aquí la importancia en Análisis de Componentes Principales (PCA). La estimación de la matriz de covarianza se puede realizar a partir de muestras mediante

$$\Sigma_{\mathbf{k}} = \frac{1}{k} \sum_{i=1}^k (\mathbf{x} - \mathbb{E}\mathbf{x})(\mathbf{x} - \mathbb{E}\mathbf{x})^T, \quad (2.2)$$

y de acuerdo a la ley de grandes números se garantiza que el estimador se vuelva ajustado cuando el numero de muestras  $k$  va al infinito, es decir

$$\Sigma_{\mathbf{k}} \xrightarrow[k \rightarrow \infty]{} \Sigma. \quad (2.3)$$

La dificultad radica en cuan grande debe ser el numero de muestras para asegurar que la muestra de covarianza converja a la covarianza verdadera?, o equivalentemente cuantas muestras debo tener para lograr cierta precision ( $\epsilon = 0,01$ ) en el operador norma

$$\|\Sigma_{\mathbf{k}} - \Sigma\| \leq \epsilon \|\Sigma\|. \quad (2.4)$$

En (Tropp, 2015) se demuestra que si el número de muestras es mayor a  $c\epsilon^{-2}n \log n$  el estimador muestral provee una relativa precision en la estimación de la verdadera matriz de covarianza  $\Sigma$ . Pero con la gran cantidad de datos, generalmente las matrices son extremadamente grandes y con las facilidades en el censado y adquisición de datos, el tamaño tiende cada vez a aumentar. Pero por limitaciones de volumen y costos es común no poder muestrear completamente los datos de salida para obtener la matriz completa, o en campos como la biología o medicina el numero de muestras disponibles es reducido y generalmente menor que el tamaño de variables (Kwan, 2011). Por tanto se hace necesario desarrollar técnicas de estimación de matrices de covarianza capaces de lograr un bajo error sujetos a que el numero de muestras  $k$  sea moderado comparado a  $n$ .

El estimador bien condicionado desarrollado por (Ledoit y Wolf, 2003, 2004) conocido como estimación *shrinkage* que al igual que el estimador muestral tiene la ventaja de no necesitar conocer la estructura de la matriz de covarianza y logra un mejor desempeño que el estimador muestral (Ledoit y Wolf, 2012).

Otros métodos basados en teoría de matrices aleatorias, optimización convexa y muestreo compresivo han demostrado obtener una mejor estimación que los anteriores métodos, incluso cuando se tiene ruido (Walden y Schneider-Luftman, 2015; Bai y Shi, 2011). Además, sus resultados son mejores cuando se tiene en cuenta en la estimación la estructura que podría presentar la matriz de covarianza.

Por lo anterior, a continuación se presenta el estado actual en la estimación de matrices utilizando muestras comprimidas ( $k < n^2$ ), especificado principalmente las estructuras mas relevantes: bajo rango, dispersa o *sparse*, Toeplitz y block-Toeplitz.

## 2.3. Matrices de bajo rango

Las matrices de bajo rango son muy comunes debido a su aparición en diferentes campos: sistemas de identificación (Liu y Vandenberghe, 2009), restauración de películas (Yu et al., 2016), recuperación del movimiento en 3-D (Wang et al., 2016), ensamble de señales (Davies y Eldar, 2012), aprendizaje de maquina o machine learning (Guan et al., 2016; Argyriou et al., 2008).

**Definición:** (*Matriz de bajo rango*) una matriz  $\mathbf{X} \in \mathbf{R}^{m \times n}$  se le llamada de bajo rango si su rango  $r \ll \min(m, n)$ . Entonces la matriz puede ser expresada exactamente o aproximadamente mediante:

$$\mathbf{X} \approx \sum_{i=1}^r \alpha_i \mathbf{u}_i \mathbf{v}_i^T, \quad (2.5)$$

donde  $\alpha_i \geq 0, i = 1, \dots, r$  y  $\mathbf{u}_1, \dots, \mathbf{u}_r \in \mathbf{R}^m$  y  $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbf{R}^n$  son vectores ortonormales. Una interpretación es que los escalares son los valores singulares de  $\mathbf{X}$  y los vectores ortonormales son los correspondientes vectores singulares de la SVD.

Los diferentes métodos desarrollados hasta el momento se basan en variaciones en el diseño del operador  $\mathcal{A}$  que se encarga de la toma de mediciones o muestras, así encontramos (Davenport y Romberg, 2016):

- Operador aleatorio
- Operador que retorna un subconjunto de las entradas
- Operador de rango uno

A continuación se listan los estimadores más representativos en la recuperación de matrices bajo rango.

### 2.3.1. Aproximación de bajo rango

La idea es encontrar la mejor aproximación de bajo rango de una matriz dada  $\mathbf{X} \in \mathbf{R}^{m \times n}$ . Cuando se habla de la mejor aproximación significa que se desea determinar la matriz más cercana en norma de Frobenius, es decir

$$\underset{\mathbf{X}}{\text{minimize}} \|\mathbf{X}_l - \mathbf{X}\|_F^2 \text{ sujeto a } \text{rango}(\mathbf{X}_l) = r, \quad (2.6)$$

donde  $\|\cdot\|_F^2$  es el cuadrado de la norma de Frobenius, y  $r$  es el rango deseado de la aproximación. A pesar que el problema es no convexo, una forma para solucionarlo de manera fácil es utilizar la SVD

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (2.7)$$

donde  $\mathbf{U}, \mathbf{V}$  son matrices ortonormales y  $\mathbf{\Sigma}$  es una matriz diagonal que contiene los valores singulares ordenados. La solución se encuentra truncando la expansión, es decir aplicando

un umbral o *thresholding* a los valores singulares para mantener los mas grandes y a partir de ellos se reconstruya la matriz (Horn y Johnson, 1994)

$$\hat{\mathbf{X}} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \quad (2.8)$$

También se puede usar Lagrange en (A.1) para obtener

$$\underset{\mathbf{X}}{\text{minimize}} \|\mathbf{X}_l - \mathbf{X}\|_F^2 + \lambda \cdot \text{rango}(\mathbf{X}), \quad (2.9)$$

donde el conjunto de soluciones obtenidas para  $0 < \lambda < \infty$  es igual al conjunto de soluciones obtenidas en (A.1) para  $1 \leq r \leq k$  con  $k = \min(m, n)$ . Para un  $\lambda$  específico, (2.8) se puede solucionar calculando la SVD y sometiendo los valores singulares a un umbral duro. Cuando el umbral es igual a  $\lambda/2$  el resultado es igual a la solución de

$$\underset{\mathbf{X}}{\text{minimize}} \|\mathbf{X}_l - \mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_*, \quad (2.10)$$

donde  $\|\cdot\|_*$  es la norma nuclear o norma traza (suma de valores singulares) que es una relajación convexa del rango, muy utilizada en problemas de optimización (Boyd y Vandenberghe, 2004).

### 2.3.2. Recuperación bajo rango y minimización por traza

Generalmente, en cambio de observar directamente la matriz  $\mathbf{X}$ , lo que se hace es obtener mediciones de la forma  $\mathbf{y} = \mathcal{A}(\mathbf{X}) + \mathbf{n}$ , donde  $\mathbf{n}$  es un termino de error acotado y  $\mathcal{A} : \mathbf{R}^{m \times n} \rightarrow \mathbf{R}^k$  es un operador lineal de medición que actúa sobre la matriz  $\mathbf{X}$  tomando el producto interno estándar con matrices previamente establecidas  $\mathbf{A}_1, \dots, \mathbf{A}_k$ . Así, cada componente del vector  $\mathbf{y} \in \mathbf{R}^k$  viene dada por

$$y_i = \langle \mathbf{A}_i, \mathbf{X} \rangle + n_i. \quad (2.11)$$

El objetivo es recuperar una buena aproximación de  $\mathbf{X}$  con pocas medidas tanto como sean posibles. Una aproximación de bajo rango puede venir dada por

$$\underset{\mathbf{T}}{\text{minimize}} \|\mathbf{y} - \mathcal{A}(\mathbf{T})\|_2^2 \text{ sujeto a } \text{rango}(\mathbf{T}) = r \quad (2.12)$$

pero el problema de minimización de rango es en general NP-difícil (Boyd y Vandenberghe, 2004). Pero si en cambio se usa el programa de optimización convexa (2.10), se tiene

$$\underset{\mathbf{T}}{\text{minimize}} \|\mathbf{y} - \mathcal{A}(\mathbf{T})\|_2^2 + \lambda \|\mathbf{T}\|_* \quad (2.13)$$

el cual puede ser solucionado sistemáticamente mediante algoritmos proximal o FISTA (Davenport y Romberg, 2016). La soluciones a (2.13) para cualquier  $\beta > 0$  obedecerán la condición de punto fijo

$$\mathbf{T} = \text{prox}_\beta(\mathbf{T} - \beta \mathcal{A}^*(\mathcal{A}(\mathbf{T}) - \mathbf{y})), \quad (2.14)$$

donde el operador proximal viene dado por

$$\text{prox}(\mathbf{M}) = \arg \min_{\mathbf{T}} \|\mathbf{T} - \mathbf{M}\|_F^2 + \beta\lambda \|\mathbf{T}\|_* \quad (2.15)$$

y  $\lambda$  define el margen tolerable de error entre la solución y las mediciones  $\mathbf{y}$ . Si hay un nivel alto de fidelidad entonces se podría dejar  $\lambda = 0$  y la solución de (2.13) coincidirá con (Davenport y Romberg, 2016)

$$\underset{\mathbf{T}}{\text{minimize}} \|\mathbf{T}\|_* \text{ sujeto a } \mathcal{A}(\mathbf{T}) = \mathbf{y}, \quad (2.16)$$

presentando la ventaja de ser una optimización convexa, ya que la matriz objetivo  $\mathbf{X}$  pertenece a dos conjuntos convexos: el primero es la bola norma nuclear de radio  $\gamma_{\mathbf{X}} = \|\mathbf{X}\|_*$ , dada por

$$\mathbf{X} \in \mathcal{C} := \{\mathbf{T} : \|\mathbf{T}\|_* \leq \gamma_{\mathbf{X}}\}, \quad (2.17)$$

y el segundo es el espacio afín de todas las matrices  $m \times n$  que tienen las mismas mediciones

$$\mathbf{X} \in \mathcal{E} := \{\mathbf{T} : \mathcal{A}(\mathbf{T}) = \mathbf{y}\}. \quad (2.18)$$

entonces por definición  $\mathbf{X}$  será la única solución si y sólo si es la única matriz en la intersección  $\mathcal{C} \cap \mathcal{E}$ .

### 2.3.3. Recuperación de fase

Un problema que ha recibido gran atención ha sido la recuperación de la señal  $\mathbf{x} \in \mathbf{R}^n$  a partir de  $m$  mediciones cuadráticas (intensidad) de la forma  $\mathbf{x}_i = |\langle \mathbf{x}, \mathbf{a}_i \rangle|$ ,  $i = 1, \dots, m$ . Para recuperar la señal además de la magnitud se necesita determinar la fase, y este problema es NP-difícil (Sahinoglou y Cabrera, 1991). En un muy interesante trabajo conocido como *PhaseLift* desarrollado por Candès et al. (2013) demostraron que si los vectores Gaussianos  $\mathbf{a}_i$  son muestreados independiente y uniformemente sobre la esfera unitaria entonces la señal puede ser recuperada exactamente a partir de las mediciones de magnitud. El problema original puede ser reescrito como

$$|\langle \mathbf{x}, \mathbf{a}_i \rangle|^2 = \mathbf{x}^T \mathbf{a}_i \mathbf{a}_i^T \mathbf{x} = \langle \mathbf{X}, \mathbf{A}_i \rangle, \quad (2.19)$$

donde  $\mathbf{X} = \mathbf{x}\mathbf{x}^T$  y  $\mathbf{A}_i = \mathbf{a}_i \mathbf{a}_i^T$  son matrices simétricas de rango 1. Entonces el problema puede ser convertido a un problema de recuperación de matrices, donde la matriz  $\mathbf{X}$  sería la solución del problema de optimización

$$\underset{\mathbf{T}}{\text{minimize}} \|\mathbf{T}\|_* \text{ sujeto a } \langle \mathbf{T}, \mathbf{A}_i \rangle = y_i^2 \quad (2.20)$$

$$\mathbf{T} \succeq 0 \quad (2.21)$$

y la estimación es exitosa cuando  $m \geq n \log n$ .

### 2.3.4. Mediciones cuadráticas o sketch

En los anteriores estimadores los vectores de censado eran variables Gaussianas Independientes e Idénticamente Distribuidas (i.i.d.), en (Chen et al., 2015) se extendió el análisis a vectores de censado sub-Gaussianos, permitiendo utilizar una clase mas amplia de mecanismos de muestreo. Además, utilizando los vectores de medición o *sketch* se obtiene mediciones cuadráticas a partir de las cuales demostraron que es posible la estimación de una matriz de covarianza. Cada medición cuadrática viene dada por

$$y_i = \mathbf{a}_i^T \mathbf{C}_x \mathbf{a}_i + \eta_i, \quad (2.22)$$

y  $\mathbf{C}_x$  es la matriz de covarianza. Cuando las mediciones han sido contaminadas por ruido acotado  $\|\eta\| \leq \epsilon_1$ , y asumiendo una estructura de bajo rango en la matriz se puede plantear el siguiente problema de optimización

$$\hat{\mathbf{T}} = \underset{\mathbf{T}}{\operatorname{argmin}} \operatorname{rango}(\mathbf{T}) \text{ sujeto a } \mathbf{T} \succeq 0 \quad (2.23)$$

$$\|\mathbf{y} - \mathcal{A}(\mathbf{T})\| \leq \epsilon_1, \quad (2.24)$$

donde  $\|\cdot\|$  es a norma espectral y  $\epsilon_1$  es la cota superior del ruido. Pero como hemos visto el problema de la minimización del rango es un problema NP-difícil, y como para las matrices semidefinidas positivas la traza es un sustituto convexo de la función rango muy utilizada en problemas de recuperacion de fase (Candes et al., 2011; Candès et al., 2013), entonces una relajación convexa que permite volver el problema tratable viene dada por

$$\hat{\mathbf{T}} = \underset{\mathbf{T}}{\operatorname{argmin}} \operatorname{Tr}(\mathbf{T}) \text{ sujeto a } \mathbf{T} \succeq 0, \quad (2.25)$$

$$\|\mathbf{y} - \mathcal{A}(\mathbf{T})\| \leq \epsilon_1, \quad (2.26)$$

donde  $\operatorname{Tr}(\cdot)$  es la traza. Chen et al. (2015) demostró que en ausencia de ruido la recuperación es exacta y el numero de mediciones necesarias es del orden de  $nr$ . Cuando hay ruido demostraron que la estimación de la matriz es estable y el error esta acotado por

$$\|\mathbf{T} - \mathbf{C}_x\| = O\left(\frac{\epsilon}{m}\right) \quad (2.27)$$

Sus resultados coinciden con los encontrado con *PhaseLift* y al recurrir al uso de los sketch permite recuperar matrices de covarianza sin la necesidad de almacenar todas las realizaciones del vector de variables aleatoria.

## 2.4. Estimación de matrices dispersas o *sparse*

Las matrices dispersas o *sparse* se caracterizan por ser matrices con muchas de sus entradas iguales cero; es decir que contiene muy pocos elementos diferentes de cero por columna. La relación entre el número de entradas cero y el total de entradas es conocido como el nivel de dispersión o *sparsity*. La importancia de estas matrices radica en el impacto que tiene

en la reducción en la complejidad de los algoritmos, la computación y el almacenamiento de las matrices (Arora et al., 2006; Drineas y Zouzias, 2011), por lo cual se ha convertido en una área muy activa de investigación que ha dado origen a nuevos métodos de muestreo, estimación y aprendizaje (Chen et al., 2015; Dasarathy et al., 2015; Gilbert y Indyk, 2010). De forma general podemos clasificar los esquemas de estimación como

- Umbral
- Aleatorios

### 2.4.1. Umbral o *Thresholding*

Uno de los métodos mas fáciles para estimar matrices de covarianza dispersas ha sido el uso de un umbral, el cual establece elementos pequeños (menores al umbral) a cero (Bickel y Levina, 2008). Cada componente fuera de la diagonal principal de la matriz de covarianza  $\Sigma_{\mathbf{k}}$  es sometido a un umbral específico  $w_T$ , entonces la matriz dispersa viene dada por

$$\hat{\Sigma}_{\mathbf{k}} = (\hat{\sigma}_{ij})_{\mathbf{n} \times \mathbf{n}} = \begin{cases} \sigma_{ij} & \text{si } i = j \\ \sigma_{ij} \mathbf{1}(|\sigma_{ij}| > w_T) & \text{si } i \neq j \end{cases} \quad (2.28)$$

donde  $\mathbf{1}$  es la función indicador que realiza el proceso de umbral

$$\mathbf{1}(|x| > w_T) = \begin{cases} x & \text{si } |x| > w_T \\ 0 & \text{si } |x| \leq w_T \end{cases} \quad (2.29)$$

a este umbral se le conoce como umbral duro. Cuando los datos son Gaussianos o sub-Gaussianos, el umbral puede ser escogido como

$$w_T = c \sqrt{\frac{\log k}{n}}, \quad c > 0 \quad (2.30)$$

una ventaja del umbral es que evita estimar elementos pequeños y ayuda a reducir el nivel de ruido. Por otra parte Rothman et al. (2009) combino el método de umbral y shrinkage para crear reglas de umbral mas flexibles. Cai y Liu (2011) propusieron un umbral el cual es adaptativo a la variabilidad de las entradas y (Cai y Yuan, 2012) considero la estimación adaptativa de la matriz de covarianza utilizando un umbral en bloque. Pero, la gran desventaja de este tipo de estimadores es que no garantiza que la matriz de covarianza estimada vaya a ser definida positiva.

### 2.4.2. Métodos aleatorios

Dentro de las clasificaciones de métodos aleatorios se encuentran:

### 2.4.2.1. Producto de tensores

Cuando se desea recuperar la matriz  $\mathbf{T} \in \mathbf{R}^{n \times n}$  de la matriz de mediciones

$$\mathbf{Y} = \mathbf{A}\mathbf{T}\mathbf{B}^T \in \mathbf{R}^{m \times m} \quad (2.31)$$

donde  $\mathbf{A}$  y  $\mathbf{B}$  son *matrices Sketching*<sup>1</sup> de  $m \times n$  con  $m \ll n$ . Si la matriz tiene filas y columnas con pocas entradas diferentes de cero, a esta estructura se le conoce como dispersión distribuida. Dasarathy et al. (2015) mostraron que la estimación es posible resolviendo la optimización convexa

$$\underset{\mathbf{T}}{\text{minimize}} \|\mathbf{T}\|_1 \text{ sujeto a } \mathbf{A}\mathbf{T}\mathbf{B}^T = \mathbf{Y}, \quad (2.32)$$

donde  $\|\mathbf{T}\|_1$  es la norma de vectorizar a  $\mathbf{T}$  ( $\text{vect}(\mathbf{T})$ )<sup>2</sup>. El concepto de tensores aparece cuando reescribimos el producto matricial como

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

con  $\mathbf{x} = \text{vect}(\mathbf{X})$ ,  $\mathbf{y} = \text{vect}(\mathbf{Y})$  y  $\otimes$  el producto tensorial o kronecker. Este tipo de aproximaciones presenta dos grandes desventajas ya que al hacer uso del producto Kronecker se incrementa la dimension y se pierde algo de la estructura del problema, en especial es difícil hacer cumplir la propiedad de semi definida positiva en la forma vectorial. Entonces, la aproximación generalmente ignora este requerimiento y la matriz de covarianza estimada puede no ser una matriz de covarianza valida.

### 2.4.2.2. Minimización alternante

Para solucionar (2.32) Bioucas-Dias et al. (2014) propusieron explotar la información a priori sobre la matriz de covarianza para recuperar la matriz a partir de mediciones compresivas. Desarrollaron un algoritmo llamado *CoVariance by Argumented Lagrangian Shrinkage Algorithm* (CoVALSA), el cual produce formulaciones semi definidas para matrices dispersas, Toeplitz, bajo rango y bajo rango permutadas, las cuales son solucionadas eficientemente por el algoritmo SALSa (Afonso et al., 2011).

Sea  $\mathbf{x} \in \mathbf{R}^n$  un vector aleatorio de media cero con matriz de covarianza  $\mathbf{X}$ , la idea consiste en tomar mediciones compresivas dadas por

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^T, \quad (2.34)$$

donde  $\mathbf{A} \in \mathbf{R}^{m \times n}$  es la matriz de muestreo con  $m < n$ . Como el numero de mediciones es menor que la dimension del vector, entonces para poder recuperar la matriz de covarianza se debe tener en cuenta la estructura de la matriz, lo que les permitió plantear el siguiente problema de optimización

$$\underset{\mathbf{T}}{\min} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{T}\mathbf{A}^T\|_F^2 + \lambda\phi(\mathbf{T}) \text{ sujeto a } \mathbf{T} \succeq 0, \quad (2.35)$$

<sup>1</sup>Ensamble bipartita- $\delta$ : Consiste del conjunto de todas las matrices 0 – 1 que tienen al menos  $\delta$  unos por columna.

<sup>2</sup>La vectorización es una transformación lineal que toma las columnas de una matriz y las apila una tras otra para obtener un vector columna

donde  $\phi$  es una función cerrada, propia y convexa que captura la estructura deseada de la matriz (bajo rango, dispersa, Toeplitz).  $\lambda$  es un parámetro de regularización. Utilizando las ideas de los métodos de división (Combettes y Pesquet, 2011), el problema de optimización puede ser replanteado como

$$\min_{\mathbf{T}} g_1(\mathbf{T}) + g_2(\mathbf{V}) \text{ sujeto a } \mathbf{V} = \mathcal{R}(\mathbf{T}) \quad (2.36)$$

donde  $g_1(\mathbf{T}) = 0$ ,  $g_2(\mathbf{V}) = g_{21}(\mathbf{V}) + g_{22}(\mathbf{V}) + g_{23}(\mathbf{V})$  y  $\mathcal{R}$  es una matriz de regulación para cada tipo de estructura considerada. Para las funciones

$$\begin{aligned} g_{21}(\mathbf{V}_1) &= \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{V}_1\mathbf{A}^T\|_f^2 \\ g_{22}(\mathbf{V}_2) &= i_{S_+}(\mathbf{V}_2) \\ g_{s3}(\mathbf{V}_3) &= \lambda\phi(H(\mathbf{V}_3)) \end{aligned} \quad (2.37)$$

se obtienen sus operadores proximales correspondientes. Para resolver (2.36), se utiliza un método basado en métodos de multiplicadores de Lagrange llamado Alternate Directions Method of Multipliers (ADMM) (Boyd et al., 2011), el cual basado en los proximales permite resolver aproximadamente el problema minimizando sobre  $\mathbf{T}$  con  $\mathbf{V}$  fijo, y luego minimizando sobre  $\mathbf{V}$  con  $\mathbf{T}$  fija, de aquí el nombre minimización alternante. La minimización no es exacta, pero Bioucas-Dias et al. (2014) muestran que sobre ciertas condiciones que si el problema (2.36) tiene solución, entonces la minimización por ADMM converge a la solución.

La ventaja que esta aproximación presenta es que en el problema de optimización el espacio de matrices está limitado a semi-definidas positivas, lo cual nos asegura que la matriz recuperada es una matriz de covarianza válida, y aunque  $\mathbf{Y}$  es una matriz de menor dimensión se hace necesario un almacenamiento importante, por ejemplo muestran que para una matriz de dimensión  $n = 100$  se logra una buena recuperación si  $m = 50$ , que en últimas se pasa de tener una matriz de  $100 \times 100$  a una matriz de  $50 \times 50$ , aunque una buena relación estamos interesados en buscar algoritmos que permitan una mejor reducción, y además de ser posible  $\mathbf{Y}$  sea un vector y no una matriz.

### 2.4.2.3. Mediciones cuadráticas o *sketch*

El modelo de estimación desarrollado por (Chen et al., 2015) también abarca el caso cuando la matriz de covarianza a estimar es dispersa. Si  $\mathbf{X}$  es dispersa, es conocido que la norma  $\ell_0$  promueve la dispersión, entonces el problema de estimación puede ser planteado como

$$\hat{\mathbf{X}} = \underset{\mathbf{T}}{\operatorname{argmin}} \|\mathbf{T}\|_0 \text{ sujeto a } \mathbf{T} \succeq 0, \quad (2.38)$$

$$\|\mathbf{y} - \mathcal{A}(\mathbf{T})\|_1 \leq \epsilon, \quad (2.39)$$

donde  $\epsilon$  es una cota superior de  $\|\mathbf{n}\|_1$  y  $\mathcal{A}(\mathbf{T}) : \mathbf{T} \rightarrow \{\langle A_i, \mathbf{T} \rangle\}_{i=1}^m$ . Pero al utilizar la norma  $\ell_0$  el problema de minimización es NP-difícil (Foucart y Rauhut, 2013), es decir es

intratable. Entonces en (Chen et al., 2015) se planteo utilizar una relajación convexa dada por

$$\hat{X} = \underset{T}{\operatorname{argmin}} \|T\|_1 \text{ sujeto a } T \succeq 0, \quad (2.40)$$

$$\|y - \mathcal{A}(T)\|_1 \leq \epsilon, \quad (2.41)$$

donde la norma  $\ell_1$  es la relajacion conveza del tamaño de soporte y que ha demostrado ser exitosa en muchos algoritmos de *compressed sensing* promoviendo la dispersión (Foucart y Rauhut, 2013; Candès, 2008). Además, demostraron que cuando la matriz es exactamente  $p$ -dispersa y no hay presencia de ruido, la solución a (2.40) es exactamente igual la verdadera siempre que el numero de mediciones  $k > p \log(n^2/p)$ . E inclusive, para matrices aproximadamente dispersas la estimación es robusta indicando que las mediciones cuadráticas son ordenadas al menos tan buenas como las mediciones lineales.

La idea de sketch también ha sido utilizada en (Pedarsani et al., 2015) para la estimación de matrices dispersas mediante códigos de grafos dispersos, en (Bahmani y Romberg, 2015) para la estimación de matrices de bajo rango dispersas, o en el producto de tensores expuesto anteriormente. Otra forma de aproximarse a sketching es encontrar un pequeño subconjunto de columnas o filas de la matriz que aproximen la matriz entera, esto es conocido como Problema de Selección de Subconjunto Columna (Boutsidis et al., 2014, 2009; Drineas y Kannan, 2003). Un ultimo enfoque ha sido Direcciones Frecuentes, el cual se realiza de forma deterministica (Ghashami et al., 2016).

## 2.5. Estimación de matrices Toeplitz

Las matrices Toeplitz aparecen en soluciones de ecuaciones diferencias e integrales, funciones *splines*, en física, matemáticas, estadística y procesamiento de señales. En teoría del procesamiento de señales cuando se esta realizando estimación, detección, clasificación, regresión o en teoría de la información es común asumir que la media es constante y que la matriz de covarianza es Toeplitz, es decir  $T(k, j) = T(k - j)$  (Gray, 2005; Chen et al., 2015). En este caso se dice que el proceso aleatorio es estacionario débilmente o estacionario sentido amplio, o también se suele utilizar los términos estacionariedad de covarianza o estacionario de segundo orden. La matriz Toeplitz presenta la siguiente estructura

$$T = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & & \\ t_2 & t_1 & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} & & \cdots & & t_0 \end{bmatrix} \quad (2.42)$$

Dada la importancia de esta estructura se han desarrollado estimadores que logren una alta precision en su recuperación utilizando un numero reducido de muestras. En (Qiao

y Pal, 2015) propusieron un muestreo anidado generalizado (*Generalized Nested Sampling*) con el inconveniente que el intervalo de muestra debe cumplir el criterio de Nyquist imposibilitando desarrollar muestreo inferiores (muestras comprimidas), además asume un número infinito de datos muestra y cuando el número de muestras es finito no determina el desempeño de la reconstrucción. En (Qiao y Pal, 2016) determinan el desempeño del muestreador anidado generalizado para un número finito de muestras pero cuando la matriz es Toeplitz de bajo rango.

A continuación se describen algunos estimadores de la matriz de covarianza Toeplitz.

### 2.5.1. Muestreo de Covarianza Compresivo

Cuando la matriz de covarianza  $\Sigma_{\mathbf{x}}$  puede ser expresada como una combinación lineal de un conjunto de matrices Toeplitz Hermitianas  $\mathcal{S} = \{S_0, S_1, \dots, S_{I-1}\}$  (Romero et al., 2015), es decir

$$\Sigma_{\mathbf{x}} = \sum_{j=0}^{I-1} \alpha_j S_j, \quad (2.43)$$

donde  $I$  representa la dimensión del modelo. Al asumir que las matrices  $\mathcal{S}$  son linealmente independientes y los coeficientes son reales, el conjunto  $\mathcal{S}$  es una base. En (Romero et al., 2016; Romero y Leus, 2013a) se desarrolla un esquema de estimación de matrices llamado Muestro de Covarianza Compresivo, el cual busca recuperar la matriz de covarianza  $\Sigma_{\mathbf{x}}$  a partir de menos muestras, las cuales se obtienen de acuerdo a un criterio llamado *reglas dispersas*. Sea  $x(t)$  la señal original de tiempo continuo, el método tradicional (uniforme) de muestro consiste en

$$x[n] = x(nT_s), \quad n = 0, 1, \dots, N-1 \quad (2.44)$$

donde  $f_s = 1/T_s$  es la frecuencia de muestreo. Pero como se ha mencionado en muchas aplicaciones se tiene acceso a pocas muestras, por lo cual nuevos métodos como el muestreo *no uniforme* permiten recuperar la señal original. De forma general el muestreo no uniforme viene dado como

$$y[i] = x(a_i T_s), \quad \mathbf{A} = \{a_0, a_1, \dots, a_{A-1}\} \quad (2.45)$$

o de forma matricial

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.46)$$

donde  $\mathbf{A} \in \mathbb{C}^{A \times N}$ . Para el diseño de estos muestreadores se han tomado ideas de censado compresivo, que permiten obtener matrices de muestreo aleatorias con cotas probabilísticas sobre su admisibilidad. El primero es el muestreador disperso, el cual consiste matrices dispersas con unos en la posición  $(i, u_i)$  y ceros en las demás posiciones, se basa principalmente sobre matemáticas discretas y presenta un inconveniente ya que debe solucionar un problema combinatorial con solución de forma no cerrada. El segundo, recurre al uso de matrices densas basadas en argumentos probabilísticos. De manera interesante se tiene que cuando la covarianza tiene estructura Toeplitz se encuentra una forma cerrada de la

frecuencia optima del muestreador denso (Romero et al., 2015). Para la estimación de la matriz se sabe que

$$\Sigma_{\mathbf{y}} = \mathbf{A}\Sigma_{\mathbf{x}}\mathbf{A}^T \quad (2.47)$$

y recurriendo a la ec. (2.43), la anterior matriz puede ser representada como

$$\Sigma_{\mathbf{y}} = \sum_{j=0}^{I-1} \alpha_j \hat{\Sigma}_j, \quad (2.48)$$

donde  $\hat{\Sigma}_j = \mathbf{A}\mathbf{S}_j\mathbf{A}^H$ , lo cual muestra que las matrices de covarianza comparten los coeficientes  $\alpha_j$ , por tanto pueden ser estimador a partir de las mediciones  $\mathbf{y}$ . Esta relación de las covarianzas es muy útil para lograr la estimación, para lo cual hay varias formas de lograrlo.

**Máxima verosimilitud:** Si se conoce la distribución de probabilidad de las mediciones, se puede utilizar el estimador muestral

$$\hat{\Sigma}_{\mathbf{y}} = \frac{1}{k} \sum_{i=1}^k \mathbf{y}_i \mathbf{y}_i^T \quad (2.49)$$

y la minimización del log de la función de verosimilitud da el siguiente problema no convexo

$$\underset{\alpha_i}{\text{minimize}} \log |\Sigma_{\mathbf{y}}| + \text{Tr} \Sigma_{\mathbf{y}}^{-1} \hat{\Sigma}_{\mathbf{y}} \quad (2.50)$$

**Mínimos cuadrados:** Dados los altos costos computacionales de la aproximación usando máxima verosimilitud, una alternativa es utilizar la vectorización de  $\Sigma_{\mathbf{x}}$ ,  $\Sigma_{\mathbf{y}}$  para obtener  $\sigma_x$  y  $\sigma_y$ , las cuales se pueden relacionar mediante  $\sigma_y = (\mathbf{A}^* \otimes \mathbf{A})\sigma_x$  que se deriva de vectorizar (2.47). Al vectorizar (2.43) se obtiene  $\sigma_x = \sum_{i=0}^{I-1} \alpha_i \sigma_i$  o si se define una matriz  $\mathbf{T}$  donde sus columnas son los vectores  $\sigma_i$  se obtiene la representación matricial  $\sigma_x = \mathbf{T}\alpha$ . En definitiva, se tiene que  $\sigma_y = (\mathbf{A}^* \otimes \mathbf{A})\mathbf{T}\alpha$ . En (Romero et al., 2016) se muestra que la solución por mínimos cuadrados produce la estimación de forma cerrada

$$\Sigma_{\mathbf{x}} = \text{vec}^{-1} \{ \mathbf{T} [(\mathbf{A}^* \otimes \mathbf{A})\mathbf{T}]^\dagger \sigma_{\mathbf{y}} \}, \quad (2.51)$$

donde  $\text{vec}^{-1}$  convierte el vector en una matriz cuadrada. Este tipo de aproximación tiene la desventaja de no asegurar que la matriz estimada sea semidefinida positiva.

### 2.5.2. Muestreo co-primo

En (Qin et al., 2017) se desarrolla un esquema de muestreo que permite reducir la tasa de muestreo al representar la señal de alta dimension por una de baja dimension, y a partir de esta nueva señal reconstruir la matriz de covarianza. El esquema co-primo utiliza 2 muestreadores uniformes con intervalos de muestreo  $M$  y  $N$ . Los valores de muestreo son enteros y se encogen tal que sean números co-primos. Así, en (Qin et al., 2017) desarrollan una matriz de muestro basada en el esquema co-primo y a partir del vector de muestras obtenido se utiliza el estimador muestral para recuperar la matriz de covarianza.

### 2.5.3. Mediciones cuadráticas o sketch

(Chen et al., 2015) propone una recuperación cuando la matriz es simultáneamente de bajo rango y Toeplitz, la cual puede representar la matriz de covarianza de una proceso aleatoria estacionario en sentido amplio. Para lograrlo recurre al minimizador norma nuclear, dado origen al siguiente estimador

$$\hat{\mathbf{X}} = \underset{\mathbf{T}}{\operatorname{argmin}} \operatorname{Tr}(\mathbf{T}) \text{ sujeto a } \mathbf{T} \succeq 0, \quad (2.52)$$

$$\|\mathbf{y} - \mathcal{A}(\mathbf{T})\|_2 \leq \epsilon, \text{ } \mathbf{T} \text{ es Toeplitz} \quad (2.53)$$

La relajación semidefinida exacta con alta probabilidad sobre mediciones libres de ruido, ya que la intersección entre el cono Toeplitz PSD (*Positive Semidefinite*) y el hiperplano aleatorio pasando a través de  $\mathbf{X}$  a menudo contiene solo un punto. Cuando las mediciones han sido contaminadas por ruido, Chen et al. (2015) establece que la estimación es estable si el numero de mediciones satisface  $k > cr \log n$ .

En (Chi, 2016) se utilizan los sketch para estimar matrices que satisfacen la expansión en producto Kronecker (Tsiligkaridis y Hero, 2013).

## 2.6. Estimación de matrices Block Toeplitz

A pesar de no ser muy conocidas las matrices block Toeplitz, se presentan en áreas del procesamiento de señales por ejemplo MIMO (Tsiligkaridis y Hero, 2013) y radares como por ejemplo STAP (Abramovich et al., 2008a,b; Guerci y Baranoski, 2006), censado de espectro de banda ancha (crucial en radios cognitivos (Sun et al., 2013)), datos ordenados (Arce, 2004), también en matemáticas, restauración de imágenes, procesos aleatorios estacionarios 2-D, ecuaciones diferencias (Krishna et al., 2013; Omar y Slock, 2008; Hong et al., 2015; Oudin y Delmas, 2008). La estructura de una matriz Toeplitz en bloques viene dada por

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_0 & \mathbf{T}_{-1} & \mathbf{T}_{-2} & \cdots & \mathbf{T}_{-(n-1)} \\ \mathbf{T}_1 & \mathbf{T}_0 & \mathbf{T}_{-1} & & \\ \mathbf{T}_2 & \mathbf{T}_1 & \mathbf{T}_0 & & \\ \vdots & & & \ddots & \\ \mathbf{T}_{n-1} & & & \cdots & \mathbf{T}_0 \end{bmatrix} \quad (2.54)$$

Dado que se encuentran muchos estimadores lo cuales explotan la estructura Toeplitz en bloques pero no el número necesario de muestras o no establecen las cotas mínimas en la estimación, entonces priorizamos los estimadores que tratan el problema de estimar la matriz con pocas muestras y reducir la computación sin sacrificar la capacidad de recuperación de la información de la covarianza. A continuación se listan algunos estimadores para este tipo de matrices.

### 2.6.1. Expansiones en producto Kronecker

En (Tsiligkaridis y Hero, 2013) se desarrolla un método de estimación de matrices de covarianza el cual se basa en representar la matriz block Toeplitz en series de productos Kronecker para luego utilizar mínimos cuadrado penalizados en rango permutados,

$$\Sigma = \sum_{i=0}^r A_i \otimes B_i, \quad (2.55)$$

donde las matrices  $A_i$  y  $B_i$  son matrices linealmente independientes. Las muestras  $\{z_i, i = 1, \dots, n\}$  se suponen ser Gaussianas *i.i.d.*. Para estimar la matriz de covarianza Tsiligkaridis y Hero (2013) propone la relajación convexa

$$P \in \arg \min \|\hat{P} - P\|_F^2 + \lambda \|P\|_*, \quad (2.56)$$

donde  $\hat{P} = \mathcal{R}(\hat{\Sigma})$  es la matriz de covarianza muestral permutada. Dado que la relajación es estrictamente convexa existe una única solución la cual puede ser encontrada eficientemente utilizando métodos numéricos, además determinaron que la solución tiene forma cerrada. También demuestran que ciertas matrices block Toeplitz son aproximables por baja separación de rango y establecen la mínima separación de rango que asegura un nivel dado de sesgo, y su estimador lo utilizan para realizar la predicción de la velocidad del viento.

### 2.6.2. Muestreo de Covarianza Compresivo

El censado del espectro es requerido en radios cognitivos ya que es necesario monitorear la potencia de diferentes transmisores operando sobre bandas de frecuencias anchas, para lo cual determinar las estadísticas de segundo orden es más importante que la señal en sí mismas. Por eso en (Romero y Leus, 2013b) se desarrollan tres algoritmos de estimación explotando la redundancia en la estructura Toeplitz para el modelo de estimación de Muestreo de Covarianza Compresivo desarrollado anteriormente. La redundancia de la matriz permite que sea necesario calcular solo los bloques hasta un cierto valor  $M$  fuera de la diagonal principal, es decir

$$\hat{T} = \begin{bmatrix} T_0 & T_{-1} & T_{-2} & \cdots & T_{-M} \\ T_1 & T_0 & T_{-1} & & \\ T_2 & T_1 & T_0 & & \vdots \\ \vdots & & & \ddots & \\ T_L & & & \cdots & T_0 \end{bmatrix} \quad (2.57)$$

lo cual permite obtener una reducción en la dimensión de la matriz. Para estimar la matriz en el dominio del tiempo utilizan el algoritmo *LIKES* (*Likelihood-based estimation of sparse parameters*), adaptándolo para el caso de estudio y resolviendo el problema

$$\underset{\theta_{\geq 0}}{\text{minimize}} \mathbf{y}^* T^{-1} \mathbf{y} + \sum_{i=0}^{I-1} \text{Tr}(T^{-1} T_i) \theta_i \quad (2.58)$$

donde la matriz de interés es  $\mathbf{T} = \sum_i \theta_i \mathbf{T}_i$ . . Luego diseñan el algoritmo *SIIA* (*Simplified inverse iteration algorithm*) para resolver el problema no lineal

$$\text{Tr}((\mathbf{T}^{-1} - \mathbf{T}^{-1} \mathbf{A} \mathbf{T}^{-1}) \mathbf{T}_i) = 0 \quad (2.59)$$

y para obtener la aproximación por mínimos cuadrados utilizan la colección de algoritmos *SPICE*. Los algoritmos *SPICES* formulan el problema de estimación como un programa de la forma

$$\text{minimize } \text{Tr}(\mathbf{A}^* \mathbf{T}^{-1} \mathbf{A}) + \mathbf{z}^T \boldsymbol{\theta} \quad (2.60)$$

entonces el problema es resuelto como un programa cónico de segundo orden o como punto fijo.

# Capítulo 3

## Estimación de la Matriz de Correlación a Partir de Mediciones Cuadráticas

En este capítulo investigamos la recuperación de la matriz de correlación de datos que han sido ordenados en tiempo y en rango, a partir de mediciones cuadráticas aleatorias, las cuales brindan un método eficiente de adquisición y procesamiento de los datos, además de requerir un almacenamiento reducido.

Presentamos la estructura de la matriz de correlación que tienen los procesos aleatorios estacionarios cuando son ordenados en tiempo y rango. Introducimos dos nuevos algoritmos que al incluir la información de la estructura de la matriz de correlación superan el desempeño del estimador general (Hoyos et al., 2017b,a), y concluimos presentando un nuevo modelo de medición para la estimación de la matriz de correlación de datos ordenados.

### 3.1. Modelo de Medición Cuadrático

Cuando se desea explorar las características de un flujo de datos, es muy importante obtener la matriz de correlación<sup>1</sup> ya que contiene información sustancial de los datos que permitirán su reconstrucción. Cuando la matriz exhibe alguna estructura, es posible estimarla desde un esquema de mediciones (*sketches measurements*) (Chen et al., 2015; Bahmani y Romberg, 2015). Chen et al. (2015) propone un modelo de medición cuadrático (rango uno) que permite una estimación precisa de la matriz de covarianza sin ruido y además es robusta frente a ruido Gaussiano. En el modelo que desarrollaron se tuvieron en cuenta cuatro estructuras: bajo rango, bajo rango Toeplitz, dispersa, dispersa y rango uno. Para lograr la estimación definieron unos vectores llamados *sketches*, que permiten una estimación no sesgada al recurrir a paradigmas de relajación convexa, logrando darle solución al problema NP de recuperar la matriz de covarianza de dimensión  $n \times n$  a partir

---

<sup>1</sup>La matriz de correlación  $C_{\text{corr}}$  y la matriz de covarianza  $C_{\text{cov}}$  viene relacionadas por  $C_{\text{corr}} = (\text{diag}(C_{\text{cov}}))^{-1/2} C_{\text{cov}} (\text{diag}(C_{\text{cov}}))^{-1/2}$ .

de  $m$  mediciones menores que  $n(n+1)/2$ . La presente tesis busca retomar ideas de [Chen et al. \(2015\)](#) para desarrollar un modelo de estimación de la matriz de correlación de datos que han sido ordenados en tiempo y rango.

Consideremos un proceso vectorial aleatorio  $\mathbf{x}$  de media cero, donde cada realización es denotada por  $\mathbf{x}_t \in \mathbf{R}^n$  con una matriz de covarianza  $\mathbf{C}_x$ , como se aprecia en la Figura 3.1.

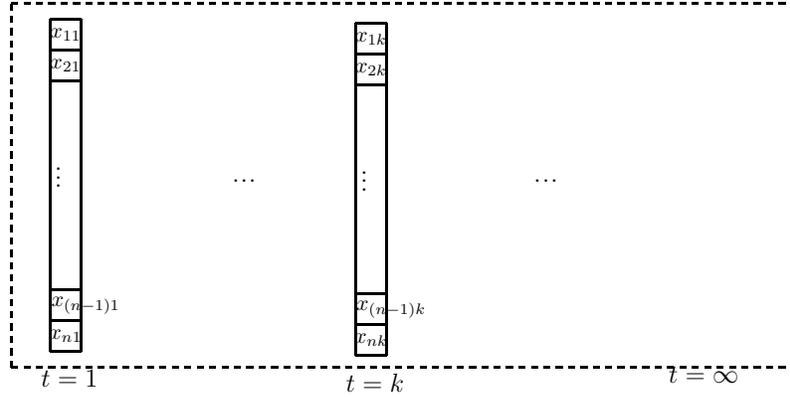


Figura 3.1: Flujo de datos

Para recuperar la matriz de covarianza usualmente es necesarios almacenar cada realización  $\mathbf{x}_t$ , pero utilizando vectores de muestreo (sketching)  $\{\mathbf{a}_i\}_{i=1}^m$ , solo necesitamos almacenar una medida escalar para cada realización. Cada uno de estos vectores  $\mathbf{a}_i = [a_1, \dots, a_n]$ , tiene componentes i.i.d sub-Gaussianos, satisfaciendo una distribución de probabilidad con las siguientes propiedades,

$$\mathbb{E}[a_i] = 0 \quad \mathbb{E}[a_i^2] = 1 \quad \mathbb{E}[a_i^4] > 1 \quad (3.1)$$

donde  $\mathbb{E}$  es el valor esperado. Una vez se han definido los  $m$  vectores sketches, un sketches  $\mathbf{a}_i$ , con  $i \in \{1, \dots, m\}$ , es escogido aleatoriamente para cada realización  $\mathbf{x}_t$ , y se realiza el producto interno cuadrático obteniéndose el sketch cuadrático  $(\mathbf{a}_i^T \mathbf{x}_j)^2 \in \mathbb{R}$ , como se muestra en la Figura 3.2.

Para cada realización los sketches cuadráticos utilizando el mismo vector sketching  $\mathbf{a}_i$  son sumados y normalizados

$$y_i = \frac{1}{|K_i|} \sum_{j \in K_i} (\mathbf{a}_i^T \mathbf{x}_j)^2 \quad (3.2)$$

donde el conjunto  $K_i$  contiene todos los instantes de tiempo en los cuales los vectores sketching  $\mathbf{a}_i$  fueron usados. Cuando el número de realizaciones se incrementa, esta expresión

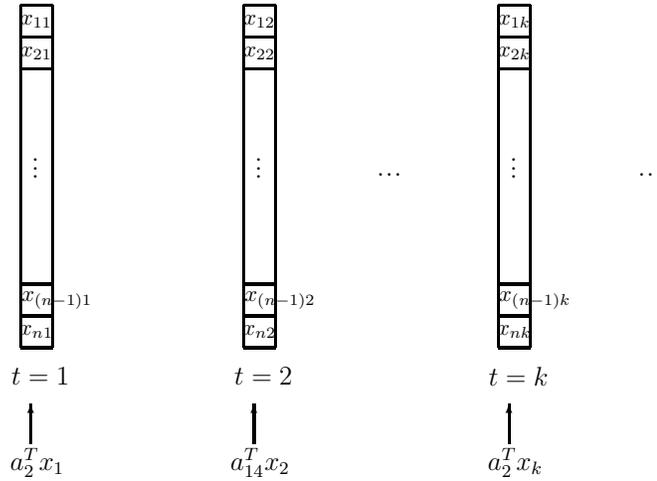


Figura 3.2: Sketches

converge rápidamente a

$$\begin{aligned}
 y_i &= \mathbb{E} \left[ \left( \mathbf{a}_i^T \mathbf{x}_t \right)^2 \right] + \eta_i \\
 y_i &= \mathbb{E} \left[ \left( \mathbf{a}_i^T \mathbf{x}_t \right) \left( \mathbf{a}_i^T \mathbf{x}_t \right)^T \right] + \eta_i \\
 y_i &= \mathbf{a}_i^T \mathbf{C}_x \mathbf{a}_i + \eta_i
 \end{aligned} \tag{3.3}$$

donde  $\eta_i$  representa el error presente en la estimación. Si se define la matriz  $A_i = \mathbf{a}_i \mathbf{a}_i^T$ , entonces  $y_i = \langle A_i, C \rangle + \eta_i$ , y se puede definir el operador  $\mathcal{A}(C) : \mathcal{C} \rightarrow \{ \langle A_i, C \rangle \}_{i=1}^m$  entonces el *vector de mediciones* puede ser escrito como

$$\mathbf{y} = \mathcal{A}(C) + \eta. \tag{3.4}$$

Lo que se desea es a partir del vector de mediciones  $\mathbf{y}$  recuperar la matriz de covarianza  $\mathbf{C}_x$  con mediciones  $m < n(n+1)/2$ , el cual es un problema mal planteado si no se explota la estructura subyacente dentro de la matriz. [Chen et al. \(2015\)](#) hacen la estimación para cuatro tipos de estructuras de la matriz de covarianza, pero para la presente tesis solo es de interés la estructura dispersa o *sparse*.

Cuando la matriz de covarianza presenta una estructura dispersa, se podría utilizar

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{T}} \|\mathbf{T}\|_0 \quad \text{sujeto a} \quad \mathbf{T} \succeq 0, \|\mathbf{y} - \tilde{\mathbf{y}}\|_1 \leq \epsilon_1, \tag{3.5}$$

donde  $\epsilon_1$  es el error presente en la estimación y  $\tilde{\mathbf{y}} = \mathcal{A}(\mathbf{T})$ . Ya que la norma  $\ell_0$  resulta en un problema mal planteado, podemos definir una relajación convexa que usa la norma  $\ell_1$  como una aproximación a la norma  $\ell_0$ ,

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{T}} \|\mathbf{T}\|_1 \quad \text{sujeto a} \quad \mathbf{T} \succeq 0, \|\mathbf{y} - \mathcal{A}(\mathbf{T})\|_1 \leq \epsilon_1, \tag{3.6}$$

donde la estabilidad y la precisión de la estimación son garantizadas por el siguiente teorema

**Teorema 1.** (*Chen et al., 2015*) *Considere el muestreo sub-Gaussiano (3.1) y asuma que  $\|\eta\| \leq 1$ . Entonces con probabilidad excediendo a  $1 - a_0 \exp(-b_0 m)$ , la solución a (3.26) satisface*

$$\|\hat{\mathbf{C}} - \mathbf{C}\|_F \leq a_1 \frac{\|\mathbf{C} - \mathbf{C}_\Omega\|_1}{\sqrt{k}} + a_2 \frac{\epsilon_1}{m}, \quad (3.7)$$

*simultáneamente para todas las  $\mathbf{C} \in \mathbb{R}^{n \times n}$ , tal que  $m > b_1 k \log(n^2/k)$ .  $\mathbf{C}_\Omega$  denota la mejor aproximación  $k$ -dispersa, y  $a_0, a_1, a_2, b_0$  y  $b_1$  son constantes universales positivas.*

Para el establecimiento de este teorema, se planteo para los sketches un nuevo concepto llamado *propiedad de isometría restrictiva de norma mezclada- RIP( $\ell_2/\ell_1$ )*. Pero, dado que el operador  $\mathcal{A}$  no puede satisfacer  $\text{RIP}(\ell_2/\ell_1)$ , porque cada matriz de mediciones no tiene media cero, entonces se desarrolla una matriz de medición auxiliar

$$\mathbf{B}_i = \mathbf{A}_{2i-1} - \mathbf{A}_{2i}, \quad (3.8)$$

así  $\mathcal{B}_i(\mathbf{C}) := \langle \mathbf{B}_i, \mathbf{C} \rangle$ , por tanto se puede definir el operador  $\mathcal{B}(\mathbf{C}) : \mathbf{C} \rightarrow \{\mathcal{B}_i(\mathbf{C})\}_{i=1}^m$ . Dado que el operador  $\mathcal{B}$  exhibe la  $\text{RIP-}\ell_2/\ell_1$ , en el siguiente teorema se determina los límites para el caso de matrices dispersas.

**Teorema 2.** (*Chen et al., 2015*) *RIP( $\ell_2/\ell_1$ ): Considere el muestreo sub-Gaussiano (3.1). Entonces con probabilidad excediendo a  $1 - a_3 \exp(-c_3 m)$ ,  $\mathcal{B}$  satisface la RIP ( $\ell_2/\ell_1$ ) para todas las matrices  $\mathbf{X}$  de dispersión (sparsity) al menos  $k$ , y obedece*

$$1 - \gamma_K^{li} \geq \frac{c_1}{2}, \quad 1 + \gamma_K^{ls} \leq 2c_2, \quad (3.9)$$

*tal que  $m > c_4 k \log(n^2/k)$ , y  $c_1, c_2, c_3$  y  $c_4$  son constantes universales.*

Esta estimación de la matriz de covarianza es óptima cuando los datos han sido contaminados con ruido gaussiano, pero en el caso de un ruido impulsivo se hace necesario utilizar procesamiento no lineal, uno de estos es *order statistics* y en especial los filtros LL. Pero, para su diseño se hace necesario la estimación de la matriz de correlación, la cual por el ordenamiento de los datos presenta una estructura toeplitz en bloques.

Diferentes autores (*Arce, 2004; Palmieri y Boncelet, 1989; Duffin y Schaeffer, 1952; Unser y Tafti, 2014*) han mostrado que si los datos son ordenados se puede combatir efectos impulsivos adversos que obedezcan modelos no gaussianos, esto es llamado estadísticos de orden o *order statistics*. Los datos para cada vector  $\mathbf{x}(t)$  pueden ser ordenados en tiempo o en rango. El vector ordenado temporalmente es definido como

$$\mathbf{x}_\ell = [x_1, x_2, \dots, x_n]^T, \quad (3.10)$$

y el vector ordenado en rango como

$$\mathbf{x}_L = [x_{(1)}, x_{(2)}, \dots, x_{(n)}]^T, \quad (3.11)$$

donde  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Entonces, para el flujo de datos  $\{\mathbf{x}_t\}_{t=1}^{\infty}$  donde cada  $x(t) \in \mathbf{R}^n$ , para cada instante de tiempo  $t$  se obtiene un nuevo vector ordenado  $\mathbf{z}_t \in \mathbf{R}^{n^2}$ , construido combinando el ordenamiento temporal y de rango (Arce, 2004)

$$\mathbf{z} = [z_{1(1)}, \dots, z_{1(n)} | \dots, z_{i(j)}, \dots | z_{n(1)}, \dots, z_{n(n)}]^T, \quad (3.12)$$

con

$$z_{i(j)} = \begin{cases} x_i & \text{si } x_i \leftrightarrow x_{(j)} \\ 0 & \text{(en otro caso)} \end{cases} \quad (3.13)$$

donde  $x_i \leftrightarrow x_{(j)}$  indica que el componente  $x_i$  tiene el rango  $j$ , entonces cada vector ordenado  $z_{i(1)}, z_{i(2)}, \dots, z_{i(n)}$  tiene solo una entrada diferente de cero e igual a  $x_i$ . Por ejemplo, para el vector  $\mathbf{x} = [4 \ 1 \ 2/3]^T$  con  $n = 3$ , los rango correspondientes son  $r_4 = 3, r_1 = 2$  y  $r_{2/3} = 1$ , entonces el vector ordenado es

$$\mathbf{z} = [0 \ 0 \ 4 \ | \ 0 \ 1 \ 0 \ | \ 2/3 \ 0 \ 0]^T \quad (3.14)$$

donde la dimensión del vector  $\mathbf{z}$  se ha incrementado a  $n^2 = 9$ .

## 3.2. Principales Resultados y Algoritmos

### 3.2.1. Estructura de la Matriz de Correlación

De acuerdo al estado actual del arte, la matriz de correlación de los datos que han sido ordenados presenta la siguiente estructura (Arce, 2004)

$$\mathbf{C}_z = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1n} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \cdots & \mathbf{C}_{nn} \end{bmatrix} \quad (3.15)$$

donde las submatrices o bloques de la matriz  $\mathbf{C}_z$  vienen dadas por  $\mathbf{C}_{ij} = \mathbb{E}\{\mathbf{z}_i \mathbf{z}_j^T\}$ , y el vector  $\mathbf{z}_i$  representa el subvector resultante de el orden de rango del  $i$ -ésimo elemento de  $\mathbf{x}$ , es decir  $\mathbf{z}_i = [z_{i(1)}, z_{i(2)}, \dots, z_{i(n)}]$ . Dado que el vector  $\mathbf{z}_i$  es disperso porque solo un elemento es diferente de cero, entonces muchas combinaciones en las submatrices no pueden ocurrir, por tanto las submatrices también son dispersas. Además, para la diagonal principal,  $i = j$  las submatrices son diagonales.

Cuando se tienen observaciones *estacionarias*<sup>2</sup>  $\mathbb{E}\{z_{i(a)} z_{j(b)}^T\} = \mathbb{E}\{z_{i(c)} z_{j(d)}^T\}$  para cualquier  $a, b, c, d \in \mathbb{Z}^+$  la matriz de correlación es block-toeplitz.

---

<sup>2</sup>Un proceso estacionario es aquel que sus parámetros estadísticos son independientes del tiempo

Adicionalmente, la matriz de covarianza de los datos originales  $\mathbf{C}_x = \mathbb{E}\{\mathbf{x}_t \mathbf{x}_t^T\}$  puede ser recuperada de la matriz ordenada, usando

$$\mathbf{C}_x = [\mathbf{I}_n \otimes \mathbf{e}_n^T] \mathbf{C}_z \{[\mathbf{I}_n \otimes \mathbf{e}_n]\}, \quad (3.16)$$

donde  $\mathbf{I}_n$  es la matriz identidad,  $\mathbf{e}_n$  es un vector de unos y  $\otimes$  es el producto de Kronecker. Es importante resaltar que order statistics permite obtener una mejor recuperación de la matriz original ya que contiene más información.

Para el caso de una matriz  $3 \times 3$  la estructura viene dada como

$$\left( \begin{array}{ccc|ccc|ccc} z_{1(1)}z_{1(1)} & z_{1(1)}z_{1(2)} & z_{1(1)}z_{1(3)} & z_{1(1)}z_{2(1)} & z_{1(1)}z_{2(2)} & z_{1(1)}z_{2(3)} & z_{1(1)}z_{3(1)} & z_{1(1)}z_{3(2)} & z_{1(1)}z_{3(3)} \\ z_{1(2)}z_{1(1)} & z_{1(2)}z_{1(2)} & z_{1(2)}z_{1(3)} & z_{1(2)}z_{2(1)} & z_{1(2)}z_{2(2)} & z_{1(2)}z_{2(3)} & z_{1(2)}z_{3(1)} & z_{1(2)}z_{3(2)} & z_{1(2)}z_{3(3)} \\ z_{1(3)}z_{1(1)} & z_{1(3)}z_{1(2)} & z_{1(3)}z_{1(3)} & z_{1(3)}z_{2(1)} & z_{1(3)}z_{2(2)} & z_{1(3)}z_{2(3)} & z_{1(3)}z_{3(1)} & z_{1(3)}z_{3(2)} & z_{1(3)}z_{3(3)} \\ \hline z_{2(1)}z_{1(1)} & z_{2(1)}z_{1(2)} & z_{2(1)}z_{1(3)} & z_{2(1)}z_{2(1)} & z_{2(1)}z_{2(2)} & z_{2(1)}z_{2(3)} & z_{2(1)}z_{3(1)} & z_{2(1)}z_{3(2)} & z_{2(1)}z_{3(3)} \\ z_{2(2)}z_{1(1)} & z_{2(2)}z_{1(2)} & z_{2(2)}z_{1(3)} & z_{2(2)}z_{2(1)} & z_{2(2)}z_{2(2)} & z_{2(2)}z_{2(3)} & z_{2(2)}z_{3(1)} & z_{2(2)}z_{3(2)} & z_{2(2)}z_{3(3)} \\ z_{2(3)}z_{1(1)} & z_{2(3)}z_{1(2)} & z_{2(3)}z_{1(3)} & z_{2(3)}z_{2(1)} & z_{2(3)}z_{2(2)} & z_{2(3)}z_{2(3)} & z_{2(3)}z_{3(1)} & z_{2(3)}z_{3(2)} & z_{2(3)}z_{3(3)} \\ \hline z_{3(1)}z_{1(1)} & z_{3(1)}z_{1(2)} & z_{3(1)}z_{1(3)} & z_{3(1)}z_{2(1)} & z_{3(1)}z_{2(2)} & z_{3(1)}z_{2(3)} & z_{3(1)}z_{3(1)} & z_{3(1)}z_{3(2)} & z_{3(1)}z_{3(3)} \\ z_{3(2)}z_{1(1)} & z_{3(2)}z_{1(2)} & z_{3(2)}z_{1(3)} & z_{3(2)}z_{2(1)} & z_{3(2)}z_{2(2)} & z_{3(2)}z_{2(3)} & z_{3(2)}z_{3(1)} & z_{3(2)}z_{3(2)} & z_{3(2)}z_{3(3)} \\ z_{3(3)}z_{1(1)} & z_{3(3)}z_{1(2)} & z_{3(3)}z_{1(3)} & z_{3(3)}z_{2(1)} & z_{3(3)}z_{2(2)} & z_{3(3)}z_{2(3)} & z_{3(3)}z_{3(1)} & z_{3(3)}z_{3(2)} & z_{3(3)}z_{3(3)} \end{array} \right)$$

Dado que el vector ordenado es disperso, muchas de sus componentes son cero, por ejemplo  $\mathbb{E}\{z_{1(1)}z_{1(2)}\} = 0$ , ya que la variable aleatoria de menor rango (1) solo tiene un elemento distinto de cero, y como  $\mathbb{E}\{z_{1(2)}z_{1(1)}\} = 0$ , se satisface la condición de simetría de la matriz de correlación. Entonces las únicas componentes distintas de cero serán las de la diagonal principal, por tanto las submatrices de la diagonal  $\mathbf{C}_{ii} = \mathbb{E}\{z_{i(\cdot)}z_{i(\cdot)}^T\}$  son a su vez matrices diagonales

$$\mathbf{C}_{ii} = \begin{bmatrix} \mathbb{E}\{z_{i(1)}^2\} & 0 & \cdots & 0 \\ 0 & \mathbb{E}\{z_{i(2)}^2\} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbb{E}\{z_{i(n)}^2\} \end{bmatrix} \quad (3.17)$$

donde  $\mathbb{E}\{z_{i(1)}^2\}$  representa la autocorrelación de la variable aleatoria de menor rango (1), y dado que las variables  $x(n)$  poseen la misma distribución, se tiene

$$\mathbb{E}\{z_{i(1)}^2\} = \mathbb{E}\{z_{j(1)}^2\} = \alpha_{i(1)} \quad (3.18)$$

entonces todas las variables aleatorias que comparten el mismo rango tiene la misma varianza, por tanto

$$\mathbf{C}_{ii} = \mathbf{C}_{jj} = \mathbf{D} \quad \forall i, j = 1, \dots, n \quad (3.19)$$

Como la matriz de correlación  $C_z$  es simétrica, las submatrices fuera de la diagonal cumplen que  $C_{ij} = C_{ji}$ . Así,  $\mathbb{E}\{z_{i(m)}z_{j(p)}\} = \mathbb{E}\{z_{i(p)}z_{j(m)}\}$  y su estructura resulta

$$C_{ij} = \begin{bmatrix} 0 & \mathbb{E}\{z_{i(1)}z_{j(2)}\} & \cdots & \mathbb{E}\{z_{i(1)}z_{j(n)}\} \\ \mathbb{E}\{z_{i(1)}z_{j(2)}\} & 0 & \cdots & \mathbb{E}\{z_{i(2)}z_{j(n)}\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\{z_{i(1)}z_{j(n)}\} & \mathbb{E}\{z_{i(2)}z_{j(n)}\} & \cdots & 0 \end{bmatrix} \quad (3.20)$$

donde la diagonal principal es nula ya que dos muestras  $x_i, x_j$  no pueden tener el mismo rango. Adicionalmente, al tener variables i.i.d se tiene que  $\mathbb{E}\{z_{i(m)}z_{j(p)}\} = \mathbb{E}\{z_{k(m)}z_{l(p)}\}, \forall i, j, k, l$ , resultando en la relación

$$C_{ij} = B \quad \forall i \neq j \quad (3.21)$$

y por tanto la matriz de correlación de los datos ordenados presenta la siguiente estructura (Hoyos et al., 2017b)

$$C_z = \begin{bmatrix} D & B & \cdots & B \\ B & D & \cdots & B \\ \vdots & \vdots & \ddots & \vdots \\ B & B & \cdots & D \end{bmatrix}. \quad (3.22)$$

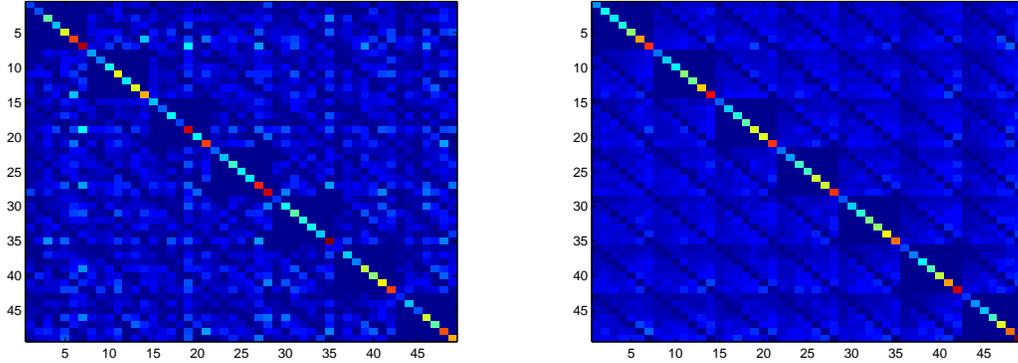
la cual solo tiene  $n(n+1)/2$  grados de libertad contrario a los  $n^2 \times n^2$  que presenta la matriz de correlación en la dimensión ambiente de la matriz, reduciendo considerablemente la cantidad de muestras necesarias para su reconstrucción.

En (3.2.1), se presenta la estructura obtenida para un proceso estacionario generado por variables aleatorias Poisson de dimensión  $n = 3$  y  $\lambda = 15$  el cual ha sido ordenado de acuerdo a 3.13, observándose como la estructura coincide con la obtenida teóricamente.

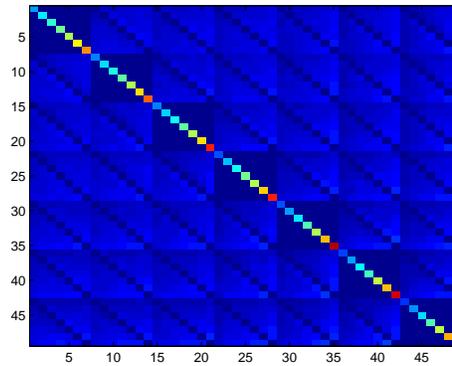
$$\left( \begin{array}{ccc|ccc|ccc} 48,6477 & 0 & 0 & 0 & 29,8630 & 36,3983 & 0 & 29,8630 & 36,3983 \\ 0 & 76,0415 & 0 & 29,8630 & 0 & 46,1557 & 29,8630 & 0 & 46,1557 \\ 0 & 0 & 115,1812 & 36,3983 & 46,1557 & 0 & 36,3983 & 46,1557 & 0 \\ 0 & 29,8630 & 36,3983 & 48,6477 & 0 & 0 & 0 & 29,8630 & 36,3983 \\ 29,8630 & 0 & 46,1557 & 0 & 76,0415 & 0 & 29,8630 & 0 & 46,1557 \\ 36,3983 & 46,1557 & 0 & 0 & 0 & 115,1812 & 36,3983 & 46,1557 & 0 \\ 0 & 29,8630 & 36,3983 & 0 & 29,8630 & 36,3983 & 48,6477 & 0 & 0 \\ 29,8630 & 0 & 46,1557 & 29,8630 & 0 & 46,1557 & 0 & 76,0415 & 0 \\ 36,3983 & 46,1557 & 0 & 36,3983 & 46,1557 & 0 & 0 & 0 & 115,1812 \end{array} \right)$$

Para mirar la convergencia, se generó un proceso aleatorio Poisson  $\mathbf{x}$  de dimensión 7 el cual es ordenado obteniendo el proceso ordenado  $\mathbf{z}$ , y dado el incremento al ser ordenado,

se pasa de  $7 \times 7$  a  $49 \times 49$ , en la Figura 3.3 para su mejor observación se presenta la estructura desplegada como una imagen.



(a) matriz de correlación para 100 realizaciones. (b) matriz de correlación para 1000 realizaciones.



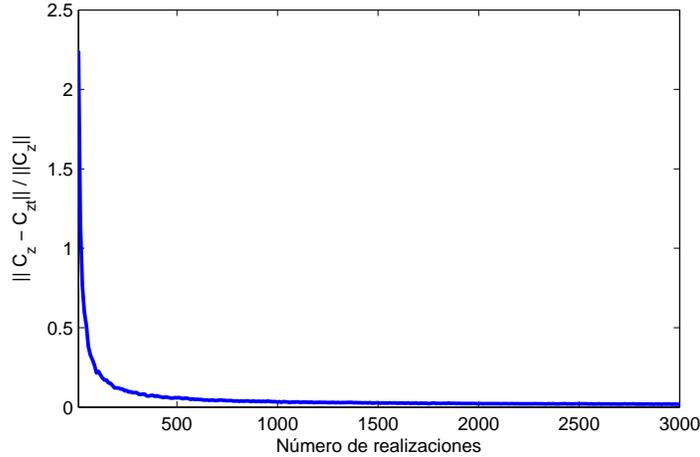
(c) matriz de correlación para 10000 realizaciones.

**Figura 3.3:** Matriz de correlación de un proceso estacionario poisson para distintos números de realizaciones.

Se observa como a medida que se incrementa el número de iteraciones la estructura de la matriz de correlación del proceso ordenado converge a la matriz teórica (3.22). Por último, en la Figura 3.4 se observa como el error mínimo cuadrado normalizado  $\|C_z - \hat{C}_z\|/\|C_z\|$  entre la matriz real  $\hat{C}_z$  y la matriz teórica  $C_z$  disminuye a medida que el número de realizaciones del proceso aleatorio aumenta, convergiendo en el infinito.

### 3.2.2. Algoritmo 1: Estimación de la matriz de correlación

Partiendo de la redundancia presente en la estructura de la matriz de correlación (3.22), a partir de la matriz estimada  $\hat{C}$  se diseñó una mejora basada en tomar el promedio de las



**Figura 3.4:** Convergencia de la matriz de correlación

matrices, es decir, la matriz recuperada por (3.26) tendría una estructura

$$\hat{C} = \begin{bmatrix} D_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & D_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & D_{nn} \end{bmatrix} \quad (3.23)$$

donde las submatrices  $B_{ij}$  son muy similares pero no iguales. Algo similar resulta para las submatrices  $D_{ii}$ . Dado que se conoce de antemano la estructura que debe tener la matriz de correlación (3.22), las submatrices  $B_{ij}$  se pueden reemplazar por

$$\hat{B} = \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{j \neq i} B_{ij} \quad \longleftrightarrow \quad \hat{B} = P_1(\hat{C}) \quad (3.24)$$

con  $P_1$  representando el proceso de promedio de las submatrices  $B_{ij}$ . Las matrices  $D_{ii}$  pueden ser reemplazadas por

$$\hat{D} = \frac{1}{n} \sum_{i=1}^n D_{ii} \quad \longleftrightarrow \quad \hat{D} = P_2(\hat{C}). \quad (3.25)$$

con  $P_2$  representando el proceso de promedio de las submatrices  $D_{ii}$ . Así, el algoritmo consiste en su primer parte de realizar la estimación mediante el programa convexo de Chen et al.

$$\hat{T} = \arg \min_{T} \|T\|_1 \quad \text{dado que} \quad T \succeq 0, \|\mathbf{y} - \tilde{\mathbf{y}}\|_1 \leq \epsilon_1, \quad (3.26)$$

donde la estabilidad y la confianza en la estimación vienen dadas por el teorema de Chen et al. (2015). Luego de la estimación, el Algoritmo 1 toma la matriz  $\hat{T}$  y obtiene las matrices  $D$  y  $B$  de acuerdo a (3.24, 3.25), para conformar la nueva matriz  $C$ . Lo anterior se resume a continuación (Hoyos et al., 2017b).

---

**Algorithm 1** Programa convexo con promedios
 

---

```

1: dato  $\mathbf{y}, \mathbf{a}, \epsilon$ 
2: minimizar
3:    $\|\mathbf{T}\|_1$  dado que  $\mathbf{T} \succeq 0, \|\mathbf{y} - \tilde{\mathbf{y}}\|_1 \leq \epsilon_1$ 
4: end
5:  $\hat{\mathbf{B}} = P_1(\mathbf{T}), \hat{\mathbf{D}} = P_2(\mathbf{T})$ 
6:  $\mathbf{C} = F(\hat{\mathbf{B}}, \hat{\mathbf{D}})$ 

```

---

donde  $F$  representa el proceso de conformación de la matriz  $\mathbf{C}$  a partir de las matrices  $\hat{\mathbf{B}}$  y  $\hat{\mathbf{D}}$ .

### 3.2.3. Experimentos numéricos - algoritmo 1

Para evaluar el desempeño del algoritmo propuesto se generaron procesos estocásticos vectoriales con dimensiones sin ordenar  $n = 5, 6, 7$  y distribución Poisson de parámetro  $\lambda = 15$ . Se ordenaron de acuerdo a (3.13), obteniéndose una matriz de correlación  $\mathbf{C}_z$  de  $625 = 25 \times 25$ ,  $1296 = 36 \times 36$  y  $2401 = 49 \times 49$  respectivamente, las cuales se utilizan para obtener las mediciones (3.3). Para brindar una estabilidad estadísticas, se realizaron 10 experimentos independientes y cada matriz de censado  $\mathbf{A}_i$  es generada con entradas Gaussianas i.i.d. Por último, para resolver la relajación convexa (3.26) se utilizó el software de optimización CVX Grant y Boyd (2014).

En la Figura 3.5 se muestra el Error Cuadrático Medio Normalizado (NMSE) definido como  $\|\hat{\mathbf{C}} - \mathbf{C}\|_F^2 / \|\mathbf{C}\|_F^2$  respecto al número de mediciones  $m$  para la estimación general (sin tener en cuenta la estructura) y teniendo en cuenta la estructura indicada en (3.22). Se observa que los sketches permiten la recuperación de la matriz con un reducido número de muestras, alrededor de 300 contra 625 para  $n = 5$ , 600 contra 1296 para  $n = 6$  y 1200 contra 2401 para  $n = 7$ . Esto se traduce en que no se requiere del almacenamiento de las realizaciones del proceso ya que sólo es necesario almacenar las mediciones cuadráticas.

Adicionalmente, se muestra la reducción en el NMSE cuando se tiene en cuenta la estructura, la cual varía proporcionalmente en función de la dimensión de la matriz, por tanto a grandes dimensiones se obtendrá una mejor estimación al utilizar la estructura de la matriz de correlación con un menor número de mediciones. Por ejemplo, para  $n = 6$  se obtiene el mismo NMSE de  $10^{-1}$  con cerca de 150 mediciones menos y para  $n = 7$  se obtiene con cerca de 300 mediciones menos el mismo NMSE de  $10^{-1}$ . En resumen se observa que para NMSE altos se tiene una gran reducción en el número de mediciones mientras que para valores bajos de NMSE la reducción es baja.

Es de resaltar que, dada la estructura particular de la matriz, no es necesario recuperar

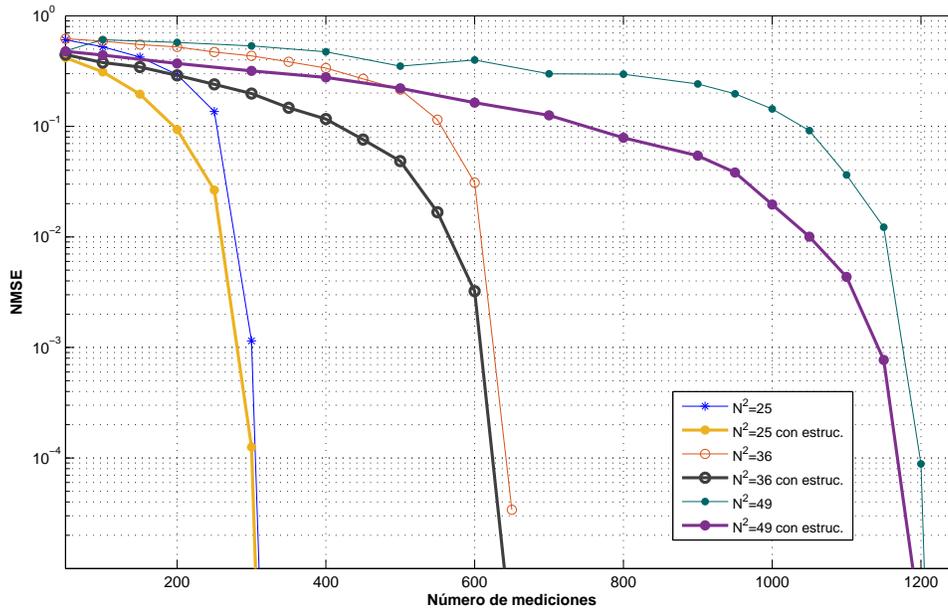


Figura 3.5: NMSE respecto al número de mediciones  $m$  para las dimensiones  $n=5$ ,  $n=6$  y  $n=7$ .

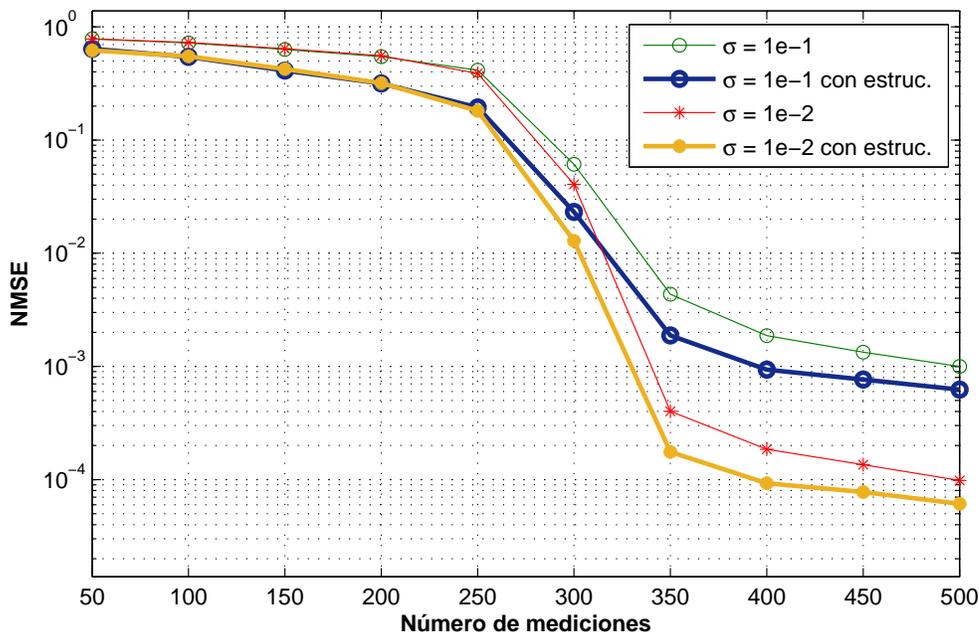
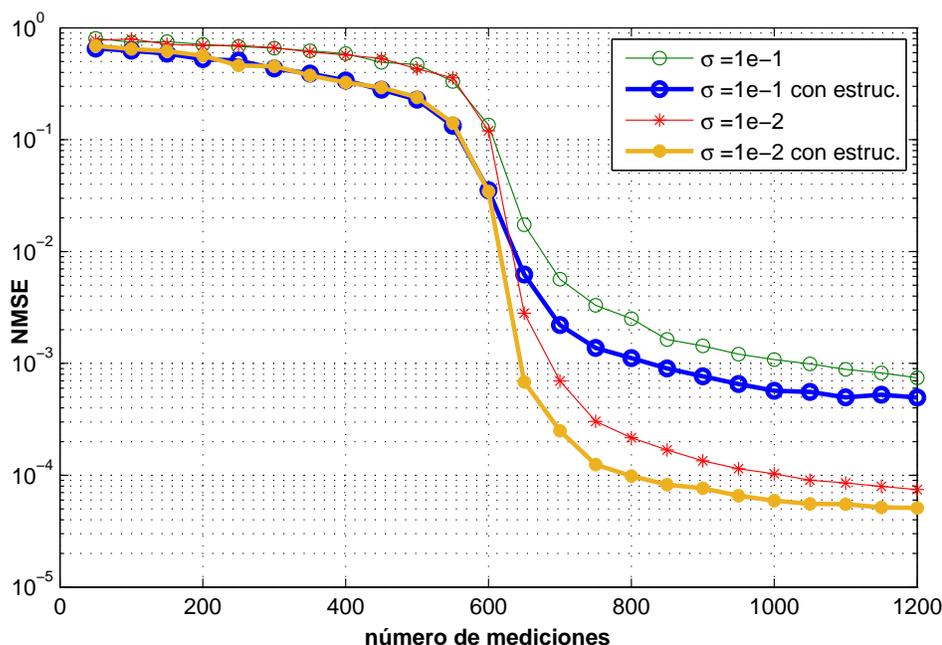


Figura 3.6: NMSE respecto al número de mediciones  $m$  para una dimensión de  $n=5$  y diferentes niveles de ruido.



**Figura 3.7:** NMSE respecto al número de mediciones  $m$  para una dimensión de  $n = 6$  y diferentes niveles de ruido.

toda la matriz sino que se podría limitar a estimar las matrices  $A$  y  $B$  que son de dimensión  $n \times n$ , con la ventaja de reducción en la complejidad y un menor NMSE respecto a la matriz  $C_z$  de dimensión  $n^2 \times n^2$ .

Para las dimensiones  $n = 5, 6$  se adicionó a las mediciones ruido uniforme en  $[-\sigma, \sigma]$  con  $\sigma = 10^{-2}$  y  $10^{-1}$ . En la Figura 3.6 y Figura 3.7 se muestra para la estimación de las matrices de correlación, el NMSE frente a  $m$ , para los diferentes niveles de ruido y con  $\epsilon = \sigma m$  en (3.26). El NMSE se desarrolló tanto para el caso general y cuando se tiene en cuenta la estructura de la matriz de correlación. Para un NMSE de  $10^{-4}$  se observa una reducción de 100 y 200 mediciones para  $n = 5, 6$  respectivamente, por tanto, para valores bajos y altos de NMSE se observa una reducción en el número de mediciones necesarias al tener en cuenta la estructura de la matriz de correlación en la estimación cuando han sido alteradas por ruido uniforme.

### 3.2.4. Algoritmo 2: Estimación de la matriz de correlación

Partiendo de la estructura de la matriz de correlación encontrada (3.22), esta puede ser reescrita como

$$\mathbf{C}_z = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{0} & \cdots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{0} \end{bmatrix}. \quad (3.27)$$

recordando el producto kronecker

$$\mathbf{F} \otimes \mathbf{G} = \begin{bmatrix} f_{11}\mathbf{G} & \cdots & f_{1n}\mathbf{G} \\ \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ f_{n1}\mathbf{G} & \cdots & f_{nn}\mathbf{G} \end{bmatrix} \quad (3.28)$$

entonces, se tiene que

$$\begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D} \end{bmatrix} = \mathbf{I}_n \otimes \mathbf{D} \quad (3.29)$$

y

$$\begin{bmatrix} \mathbf{0} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{0} & \cdots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{0} \end{bmatrix} = \mathbf{1}_n \otimes \mathbf{B} - \mathbf{I}_n \otimes \mathbf{B} \quad (3.30)$$

donde  $\mathbf{1}_n$  es la matriz  $n \times n$  de unos. Así, la matriz de correlación puede ser reescrita como (Hoyos et al., 2017a)

$$\mathbf{C}_z = \mathbf{I}_n \otimes \mathbf{D} + (\mathbf{1}_n \otimes \mathbf{B} - \mathbf{I}_n \otimes \mathbf{B}). \quad (3.31)$$

Explotando la redundancia presente en la estructura de la matriz de correlación (3.22), los vectores sketching puede ser escritos como

$$\mathbf{a}_i = [\mathbf{a}_{i1}^T, \mathbf{a}_{i2}^T, \dots, \mathbf{a}_{in}^T]^T. \quad (3.32)$$

Es importante notar que al tener en cuenta la estructura de la matriz se esta definiendo un nuevo esquema de muestro o medición, el cual llamaremos *medición en bloques*. Para una medición particular  $i$ , tenemos

$$\mathbf{y}_i = \mathbf{a}_i^T (\mathbf{I}_n \otimes \mathbf{D}) \mathbf{a}_i + \mathbf{a}_i^T ((\mathbf{1}_n \otimes \mathbf{B} - \mathbf{I}_n \otimes \mathbf{B})) \mathbf{a}_i, \quad (3.33)$$

y usando (3.32) obtenemos

$$\begin{aligned} \mathbf{y}_i = & [\mathbf{a}_{i1}^T \mathbf{D}, \mathbf{a}_{i2}^T \mathbf{D}, \dots, \mathbf{a}_{in}^T \mathbf{D}] \mathbf{a}_i + \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{B} \cdot \mathbf{e}_n \mathbf{a}_i \\ & - [\mathbf{a}_{i1}^T \mathbf{B}, \mathbf{a}_{i2}^T \mathbf{B}, \dots, \mathbf{a}_{in}^T \mathbf{B}] \mathbf{a}_i, \end{aligned}$$

$$y_i = \sum_{k=1}^n \mathbf{a}_{ik}^T \mathbf{D} \mathbf{a}_{ik} + \sum_{j=1}^n \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{B} \mathbf{a}_{ij} - \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{B} \mathbf{a}_{ik}, \quad (3.34)$$

o

$$y_i = \sum_{k=1}^n \mathbf{a}_{ik}^T \mathbf{D} \mathbf{a}_{ik} + \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik} \mathbf{B} \mathbf{a}_{ij}. \quad (3.35)$$

Entonces, el vector de muestras puede ser expresado como

$$\mathbf{y} = \mathcal{A}(\mathbf{C}_z) = \mathcal{A}_1(\mathbf{D}) + \mathcal{A}_2(\mathbf{B}), \quad (3.36)$$

donde

$$\mathcal{A}_1(\mathbf{D}) : \mathbb{D} \rightarrow \left\{ \sum_{k=1}^n \mathbf{a}_{ik}^T \mathbf{D} \mathbf{a}_{ik} \right\}_{i=1}^m, \quad (3.37)$$

define el operador que muestrea las matrices diagonales  $\mathbb{D}^n$  de  $n \times n$ , es decir  $\mathcal{A}_1 : \mathbb{D}^n \rightarrow \mathbf{R}^m$ , y

$$\mathcal{A}_2(\mathbf{B}) : \mathbf{B} \rightarrow \left\{ \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik}^T \mathbf{B} \mathbf{a}_{ij} \right\}_{i=1}^m. \quad (3.38)$$

el operador que muestra las matrices con diagonal igual a 0, es decir  $\mathcal{A}_2 : \mathbb{B}^n \rightarrow \mathbf{R}^m$ . Los operadores adjuntos necesarios para resolver el problema de optimización son definidos como (ver B.2)

$$\mathcal{A}_1^\dagger(\mathbf{y}) = \mathcal{T} \left( \sum_{i=1}^m y_i \left( \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{a}_{ik}^T \right) \right), \quad (3.39)$$

donde  $\mathcal{T}$  representa el operador de diagonalización, y

$$\mathcal{A}_2^\dagger(\mathbf{y}) = \mathcal{N} \left( \sum_{i=1}^m y_i \left( \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik} \mathbf{a}_{ij}^T \right) \right), \quad (3.40)$$

donde  $\mathcal{N}$  representa el operador que remueve la diagonal principal de la matriz. Con estas descomposición la matriz de correlación  $\mathbf{C}_z$  puede ser recuperada del vector de mediciones  $\mathbf{y}$  usando el siguiente problema convexo (Hoyos et al., 2017a):

$$\begin{aligned} & \min_{\mathbf{D} \in \mathbb{D}^n, \mathbf{B} \in \mathbb{B}^n} \|\mathbf{D}\|_1 + \|\mathbf{B}\|_1 & (3.41) \\ & \text{subject to } \|\mathbf{y} - \mathcal{A}_1(\mathbf{D}) - \mathcal{A}_2(\mathbf{B})\|_1 \leq \epsilon \\ & \text{diag}(\mathbf{D}) \succeq 0 \\ & \text{diag}(\mathbf{B}) = 0 \end{aligned}$$

donde  $\mathbb{D}^n$  denota el conjunto de  $n \times n$  matrices diagonales,  $\mathbb{S}^n$  el conjunto de  $n \times n$  matrices simétricas y  $\text{diag}(\cdot)$  la diagonal principal. Una vez las matrices  $\mathbf{D}$  y  $\mathbf{B}$  son obtenidas, la matriz de correlación  $\mathbf{C}_z$  puede ser construida utilizando (3.31). Esta optimización resulta en matrices  $\mathbf{D}$  y  $\mathbf{B}$  de  $n \times n$ , contrario a la solución al resolver directamente (3.26), que obtiene una matriz  $\mathbf{C}_z$  de  $n^2 \times n^2$ .

Para obtener la garantías teóricas en cuanto a estabilidad y estimación, se puede observar que

$$\|\mathbf{D} + \mathbf{B}\|_1 = \|\mathbf{D}\|_1 + \|\mathbf{B}\|_1 \quad (3.42)$$

debido a que los soportes de los conjuntos donde las matrices pertenecen son disyuntos,  $\mathbb{D}^n \cap \mathbb{B}^n = \emptyset$ . Por tanto, la función objetivo puede reescribirse como

$$\min_{\mathbf{D} \in \mathbb{D}, \mathbf{B} \in \mathbb{B}^n} \|\mathbf{D} + \mathbf{B}\|_1$$

y se puede partir de la matriz densa  $\mathbf{C} = \mathbf{D} + \mathbf{B} \in \mathbb{S}^n$ . Por otra parte, se tiene que a partir de los operador se puede definir un nuevo operador  $\mathcal{L}$ , definido como

$$\mathcal{L} = \mathcal{A}_1 + \mathcal{A}_2 \quad \longleftrightarrow \quad \mathcal{L} : \mathbb{S}^n \rightarrow \mathbf{R}^m \quad (3.43)$$

el cual es lineal (ver B.1). En resumen, el problema convexo puede ser rescrito como

$$\begin{aligned} & \min_{\mathbf{C} \in \mathbb{S}^n} \|\mathbf{C}\|_1 \\ & \text{subject to } \|\mathbf{y} - \mathcal{L}(\mathbf{C})\|_1 \leq \epsilon \\ & \text{diag}(\mathbf{C}) \succeq 0 \end{aligned} \quad (3.44)$$

donde al igual que [Chen et al. \(2015\)](#) se está minimizando la norma  $\ell_1$ , sujeto a la condición de fidelidad de los vectores de mediciones, con la restricción que la diagonal debe ser mayor o igual a cero, que en otras palabras es la condición de las matrices semidefinidas positivas utilizada en ([Chen et al., 2015](#)), por lo cual las condiciones teóricas de estabilidad y estimación desarrolladas por Chen se mantienen para nuestro problema de optimización.

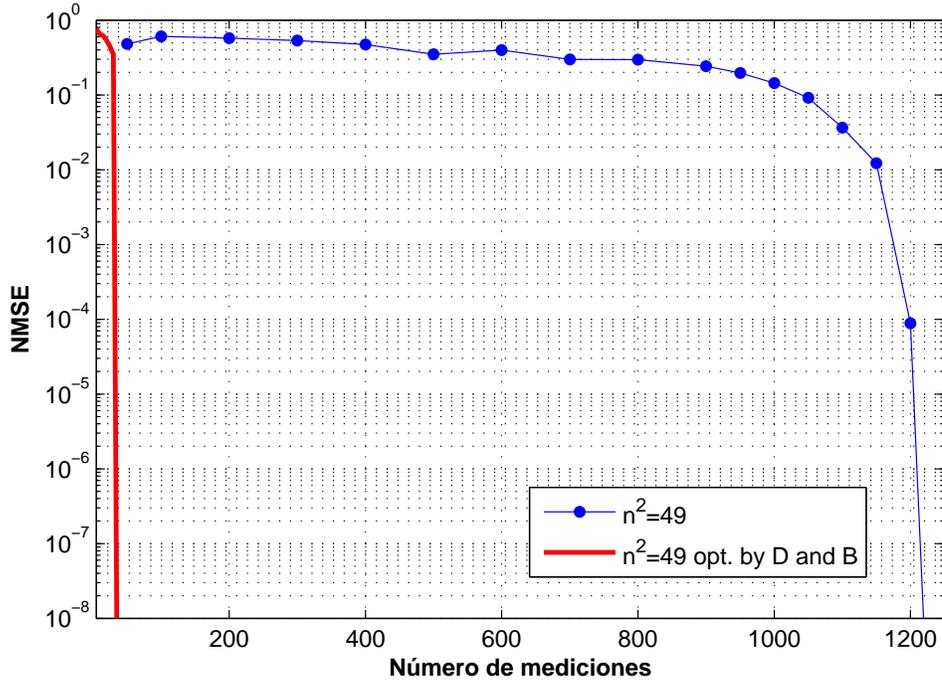
### 3.2.5. Experimentos numéricos - algoritmo 2

Para estudiar el desempeño del algoritmo se diseñaron tres escenarios con el fin de poder observar y comparar su desempeño.

#### Escenario 1

El escenario 1 compara el estimador propuesto con el estimador general de ([Chen et al., 2015](#)). Generamos un proceso aleatorio vectorial con dimensión  $n=7$  y distribución Poisson con parametro  $\lambda = 15$ . A continuación, es ordenado este vector de acuerdo a (3.12) y computamos la matriz de correlación  $\mathbf{C}_z$  de  $49 \times 49$ , la cual utilizamos para obtener las mediciones (3.3). Analizamos 10 experimentos independientes donde las entradas de la matriz de censado  $\mathbf{A}_i$  son i.i.d. generadas desde una distribución Gaussiana  $(0, 1)$ . En orden para encontrar la solución al problema convexo (3.26) utilizamos el paquete software de optimización CVX [Grant y Boyd \(2014\)](#).

La Figura 3.8 muestra el Error Cuadrático Medio Normalizado (NMSE)  $\|\hat{\mathbf{C}} - \mathbf{C}\|_F / \|\mathbf{C}\|_F$  como función del número de mediciones  $m$  para la estimación general (3.26) y para la relajación convexa propuesta (3.41) que tiene en cuenta la estructura en la minimización. En

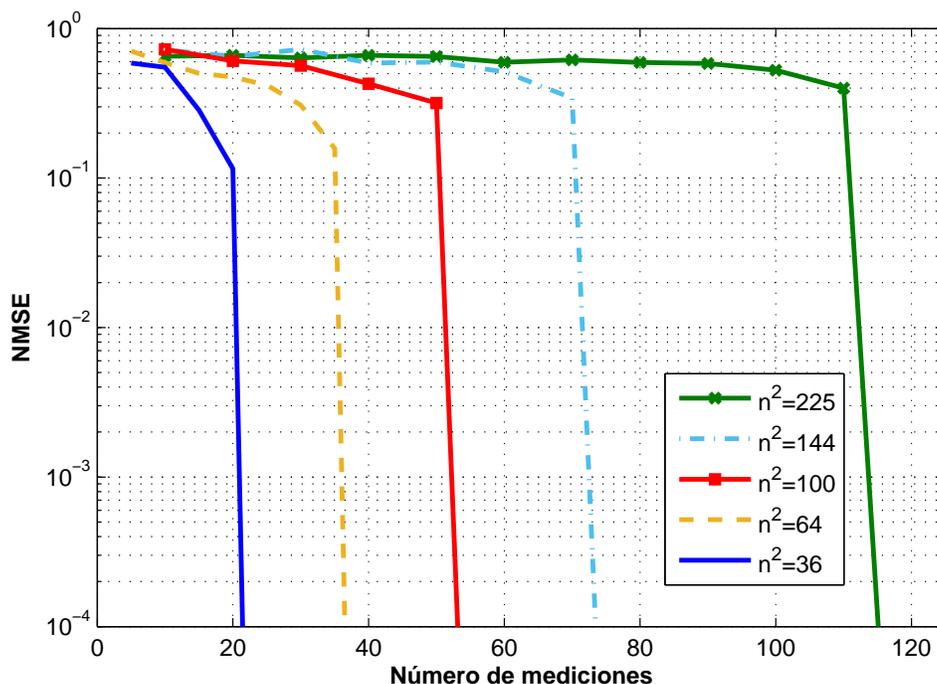


**Figura 3.8:** NMSE como función del número de mediciones  $m$  con dimensión  $n=7$  para la estimación general y la estimación propuesta mediante D y B.

ambos casos podemos notar que los sketches permiten la recuperación de la matriz, pero cuando la estructura es considerada, se obtiene una significativa reducción en el número de mediciones requeridas. Esto es debido al hecho que la estimación es desarrollada directamente sobre las matrices D y B de dimensiones  $n \times n$ , por tanto el número de mediciones  $m$  necesarias pasa de depender de  $n^2$  a  $n$ . Por ejemplo, para la matriz de correlación  $49 \times 49$  necesitaríamos alrededor de  $1200 \sim n^2(n^2 + 1)/2$  mediciones, pero usando la relajación propuesta (3.41) solo necesitamos alrededor de  $30 \sim n(n + 1)/2$  mediciones. Esto implica que no necesitamos almacenar todas las realizaciones del proceso, solo las mediciones cuadráticas para cada vector sketching (muestra), las cuales con nuestra propuesta pasaron a depender de la dimensión del problema antes de ordenarse (original)  $n$ .

## Escenario 2

En este escenario estudiamos el comportamiento del estimador propuesto para diferentes dimensiones. Generamos un proceso aleatorio vectorial con dimensiones  $n = 6, 10, 15$  y distribución Poisson con parámetro  $\lambda = 15$ , los cuales son ordenados de acuerdo a (3.12), obteniendo las matrices de correlación  $C_z$  de  $49 \times 49$ ,  $100 \times 100$  y  $225 \times 225$  respectivamente, las cuales son usadas para obtener las mediciones (3.3). Como en el escenario anterior, analizamos 20 experimentos independientes donde las entradas de la matriz de censado  $A_i$  son i.i.d. y generadas acorde a la distribución Gaussiana.



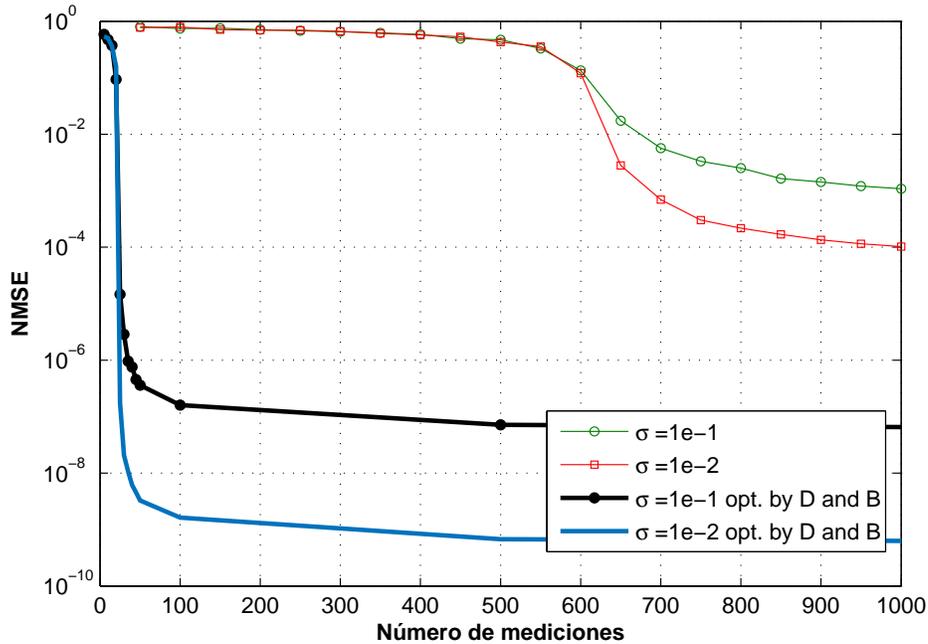
**Figura 3.9:** NMSE como función del número de mediciones  $m$  con dimensiones  $n = 6, 8, 10, 12$ , y  $n = 15$  para el esquema de estimación propuesto.

La Figura 3.9 muestra el número reducido de muestras necesarias para la reconstrucción de las matrices de covarianza de grandes dimensiones. Por ejemplo, la matriz  $225 \times 225$  con 50,625 entradas requiere almacenar solo  $n(n+1)/2 = 120$  mediciones cuando se usa el esquema propuesto para recuperarlas. Adicionalmente, podemos notar que la relación entre las mediciones requeridas y  $n(n+1)/2$  es constante para todas las dimensiones.

### Escenario 3

El escenario 3 estudia la robustez del estimador propuesto frente al ruido. Para las dimensiones  $n = 6, 10$  y  $15$  contaminamos el vector de mediciones con ruido uniforme entre  $[-\sigma, \sigma]$ . La Figura 3.10 y la Figura 3.11 muestra el NMSE de la estimación de la matriz de correlación como una función del número de mediciones  $m$ , para  $\sigma = 10^{-2}$  y  $10^{-1}$  y  $\epsilon = \sigma m$  dado en (3.26). Debido a las limitaciones computacionales, solo pudimos comparar nuestro esquema con el de Chen's para el caso de  $n = 6$ . Puede notarse que la relajación propuesta requiere significativamente menos mediciones y presenta un comportamiento robusto frente al ruido.

La Figura 3.11 muestra el desempeño del esquema del estimador propuesto cuando el ruido está presente para las dimensiones  $n = 10$  y  $15$ . Como se esperaba, el número de mediciones



**Figura 3.10:** Comparación del NMSE entre el estimador general y el esquema propuesto para  $n=6$  y diferentes niveles de ruido.

requeridas para una estimación precisa es más alto que para el caso sin ruido, pero todavía se encuentra en un margen razonable.

### 3.3. Modelo de medición para la estimación de la matriz de correlación de datos ordenados

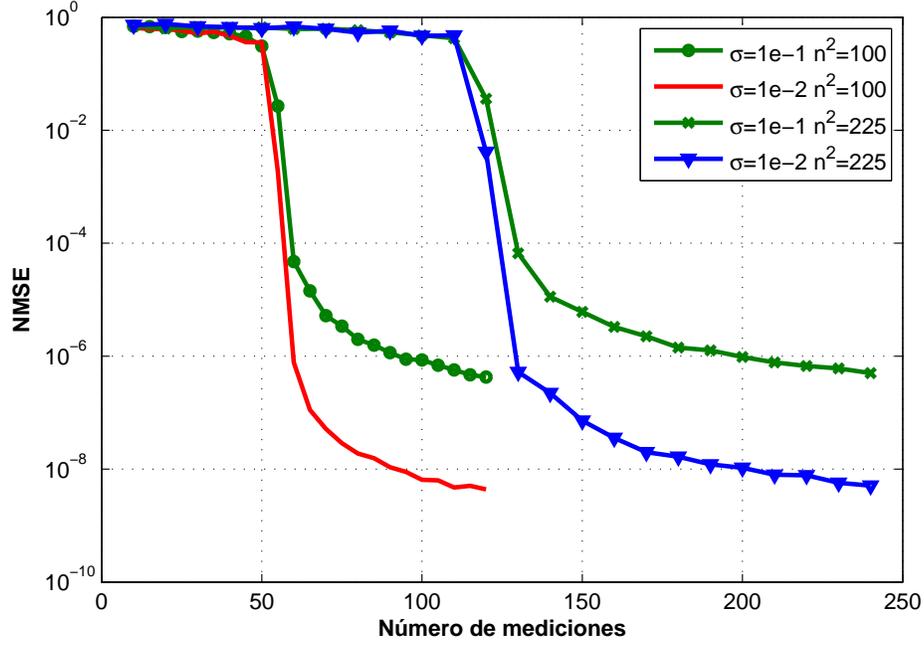
#### 3.3.1. Mediciones cuadráticas

En el esquema de muestreo planteado por [Chen et al. \(2015\)](#) se supone que número de realizaciones es infinito, pero sabemos que en la práctica es difícil asegurar esta condición, por eso para superar esta limitante hemos realizado una pequeña modificación en el muestreo, ahora proponemos que para cada realización del proceso estocástico se obtengan  $m$  sketches cuadráticos y no uno, es decir para cada realización  $t$  obtenemos un vector

$$y_t = \left\{ \left( \mathbf{a}_i^T \mathbf{x}_t \right)^2 \right\}_{i=1}^m \quad (3.45)$$

luego sumamos todos los vectores  $\mathbf{y}_t$  resultantes y normalizamos para obtener el vector de mediciones

### 3.3. MODELO DE MEDICIÓN PARA LA ESTIMACIÓN DE LA MATRIZ DE CORRELACIÓN DE D.



**Figura 3.11:** NMSE para el esquema propuesto para las dimensiones  $n = 10$  and  $n = 15$  con diferentes niveles de ruido.

$$y = \frac{1}{I} \sum_{j \in I} \mathbf{y}_t \quad (3.46)$$

donde  $I$  representa el número de realizaciones del proceso estacionario. De manera equivalente, para cada sketch se tiene que

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t (\mathbf{a}_i^T \mathbf{x}_k)^2 &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t (\mathbf{a}_i^T \mathbf{x}_k) (\mathbf{a}_i^T \mathbf{x}_k)^T \\ &= \mathbf{a}_i^T \left\{ \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t (\mathbf{x}_k \mathbf{x}_k^T) \right\} \mathbf{a}_i \\ &= \mathbf{a}_i^T \mathbf{C} \mathbf{a}_i \end{aligned} \quad (3.47)$$

y se observa que esta forma de obtener la matriz de correlación disminuye el error en la estimación, además que no es necesaria hacer ninguna suposición de convergencia del flujo de datos  $\mathbf{x}_t$ . Al final tenemos que cada medición es igual a

$$y_i = \mathbf{a}_i^T \mathbf{C} \mathbf{a}_i \quad i = 1, \dots, m \quad (3.48)$$

#### Experimento numérico

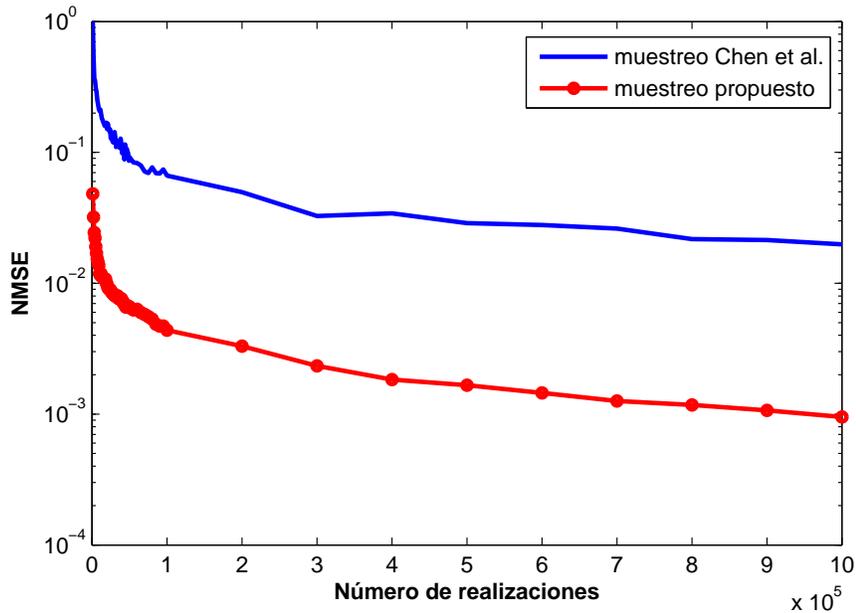


Figura 3.12: NMSE para el muestreo de [Chen et al. \(2015\)](#) y el muestreo propuesto.

Se generó un proceso vectorial aleatorio Gaussiano  $(0, 1)$  para diferentes número de realizaciones, se obtuvieron los vectores de mediciones del esquema de Chen (3.3) y del esquema propuesto (3.46), y se tomó como referencia el vector de mediciones obtenido directamente de la matriz de covarianza. Los resultados obtenidos se ilustran en la Figura 3.12 de la cual se puede apreciar como el NMSE entre el vector de mediciones decrece más rápido para nuestro esquema propuesto a medida que el número de realizaciones del proceso aleatorio aumenta.

### 3.3.2. Modelo consolidado de medición

Definido ya el esquema de muestreo, en resumen nuestro modelo de medición vendría dado por los siguientes componentes

1. Ordenamiento temporal y de rango del proceso estocástico de acuerdo a (3.13)
2. Obtención del vector de mediciones de acuerdo a (3.46), resaltando que en caso de utilizar el algoritmo 2 se realizaría un muestreo en bloques.
3. Por último, para la estimación de la matriz de correlación se escoge entre el algoritmo 1 (1) o el algoritmo 2 (3.41).

# Capítulo 4

## Estimación de la Matriz de Correlación Escalable para Grandes Dimensiones

En el inicio, los métodos de procesamiento de señales al estar fundamentados en análisis funcional y álgebra lineal eran netamente lineales, pero con el desarrollo de métodos no lineales en las matemáticas y con computadores cada vez más rápidos, lentamente las técnicas no lineales en el procesamiento de señales han empezado prevalecer ([Combettes y Pesquet, 2011](#); [Cevher et al., 2014](#); [Komodakis y Pesquet, 2015](#)). Particularmente, optimización es un paradigma muy popular ya que proporciona algoritmos eficientes para computar soluciones confiables en muchas áreas de matemáticas aplicadas e ingeniería, procesamiento de señales, finanzas, visión por computadora, aprendizaje de máquina, problemas inversos, redes,... ([Boyd y Vandenberghe, 2004](#)).

Gran parte de la importancia del uso de optimización radica en que en muchas aplicaciones es difícil encontrar una solución cerrada además de la presencia de incertidumbre. La incertidumbre puede ser generada por ruido o imperfección en los dispositivos de muestreo lo cual genera que soluciones exactas sean difíciles de obtener, pero por el contrario soluciones inexactas pero óptimas pueden ser fácilmente computadas.

Por otra parte hoy día con las facilidades en la adquisición de datos se presentan problemas de muy alta dimensionalidad, lo cual hace indispensable el uso de técnicas de optimización que sean capaces de explotar la estructura del problema buscando solucionarlo eficientemente. Entre estas técnicas encontramos los métodos de división proximal o *proximal splitting methods* y métodos primal-dual que permiten diseñar algoritmos escalables.

El foco principal de este capítulo es proponer un algoritmo escalable para resolver tareas de optimización convexa de datos ordenados de conjunto de datos de grandes dimensiones.

A continuación, se darán algunos aspectos teóricos necesarios para el desarrollo del capítulo.

## 4.1. Aspectos teóricos

El dominio de una función  $f : \mathbf{R}^n \rightarrow \mathbf{R} \cup \infty$  es el subconjunto de  $\mathbf{R}^n$  donde la función toma valores finitos, es decir

$$\text{dom } f = \{\mathbf{x} \in \mathbf{R}^n \mid f(\mathbf{x}) < +\infty\} \quad (4.1)$$

Una función se dice ser propia si su dominio es no vacío, es decir  $\text{dom } f \neq \emptyset$ .

Una función  $f : \mathbf{R}^n \rightarrow \mathbf{R} \cup \infty$  es semicontinua inferiormente si su epigrafo

$$\text{epi} f = \{(\mathbf{x}, t) \in \text{dom } f \times \mathbf{R} \mid f(\mathbf{x}) \leq t\} \quad (4.2)$$

es un conjunto cerrado. Lo anterior también puede verse como las funciones que satisfacen que  $\liminf_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \geq f(\mathbf{x}_0)$ . Una función es convexa si para todo  $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$  y  $\alpha, \beta \in \mathbf{R}$  con  $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$  se tiene que

$$f_i(\alpha \mathbf{x} + \beta \mathbf{y}) \leq \alpha f_i(\mathbf{x}) + \beta f_i(\mathbf{y}). \quad (4.3)$$

Para el interés de nuestra tesis, vamos a considerar las funciones que son propias, convexas, semicontinuas inferiormente de  $\mathbf{R}^n$  a  $\mathbf{R} \cup \infty$ , denotadas por  $\Gamma_0(\mathbf{R}^n)$ .

La función indicador de un conjunto  $C \in \mathbf{R}^n$  pertenece a  $\Gamma_0(\mathbf{R}^n)$  si y solo si  $C$  es un conjunto convexo cerrado no vacío, y viene definida como

$$i_C(\mathbf{x}) = \begin{cases} 0 & \text{si } \mathbf{x} \in C \\ +\infty & \text{en otro caso} \end{cases} \quad (4.4)$$

La idea central detrás de la dualidad en análisis convexo es ver un conjunto convexo cerrado como la intersección de todos los semiespacios conteniéndolo. Entonces, cuando aplicamos esta idea a los epigrafos nos permite ver una función convexa cerrada como un supremo de funciones afines <sup>1</sup>. Estas ideas dan origen a la importante noción de *función conjugada*. La conjugada de una función  $f : \mathbf{R}^n \rightarrow \mathbf{R} \cup \infty$  viene dada por

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbf{R}^n} \mathbf{x}^T \mathbf{y} - f(\mathbf{x}) \quad (4.5)$$

Una visión geométrica de la función conjugada se da cuando buscamos representar  $f$ , para este caso sólo necesitamos el mejor minorante afín lineal con pendiente  $\mathbf{y}$ , entonces

$$\begin{aligned} f(\mathbf{x}) \geq \mathbf{x}^T \mathbf{y} - \alpha &\Leftrightarrow \alpha \geq \mathbf{x}^T \mathbf{y} - f(\mathbf{x}) \\ &\Leftrightarrow \alpha \geq \sup_{\mathbf{x} \in \mathbf{R}^n} \mathbf{x}^T \mathbf{y} - f(\mathbf{x}). \end{aligned} \quad (4.6)$$

---

<sup>1</sup> Una función es afín si es la suma de una funciones lineal y una constante,  $f(x) = Ax + b$ .

entonces la mejor selección para la constante  $\alpha$  es la función conjugada. Cuando  $f \in \Gamma_0(\mathbf{R}^n)$ , entonces  $f^* \in \Gamma_0(\mathbf{R}^n)$  y también la biconjugada de  $f$  es igual a  $f$ , es decir

$$\begin{aligned} f(\mathbf{x}) &= \sup_{\mathbf{y} \in \mathbf{R}^n} \mathbf{x}^T \mathbf{y} - f^*(\mathbf{y}) \\ &= f^{**}(\mathbf{x}) \end{aligned} \quad (4.7)$$

esta fórmula nos dice como recuperar a  $f$  de su representación dual  $f^*$ .

**Conjugada de una norma:** sea  $f(\mathbf{x}) = \|\mathbf{x}\|$ , donde  $\|\cdot\|$  es cualquier norma. Entonces  $f^*$  es la función indicador de la norma de la bola unitaria dual.

**Subdiferencial Moreau:** El subdiferencial de una función  $f : \mathbf{R}^n \rightarrow \mathbf{R} \cup \infty$  en  $\mathbf{x} \in \mathbf{R}^n$  es definido como

$$\partial f(\mathbf{x}) = \{\mathbf{u} \in \mathbf{R}^n \mid f(\mathbf{x}) + \mathbf{u}^T(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{R}^n\} \quad (4.8)$$

y cualquier  $\mathbf{u}$  en  $\partial f(\mathbf{x})$  es llamado un *subgradiente* de  $f$  en  $\mathbf{x}$ .

**Operador proximal:** El operador de proximidad de una función  $f \in \Gamma_0(\mathbf{R}^n)$  es la solución única denotada por  $\text{prox}_f$  al problema de optimización

$$\underset{\mathbf{y}}{\text{minimize}} \quad f(\mathbf{y}) + \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2. \quad (4.9)$$

El operador de proximidad es una generalización de los operadores de proyección. Cuando  $f = i_C$ , el operador de proximidad se reduce a la proyección  $P_C$  sobre este conjunto, es decir  $P_C = \underset{\mathbf{y}}{\text{minimize}} \|\mathbf{y} - \mathbf{x}\|$ . Una propiedad interesante de  $\text{prox}_f$  es que es firmemente no-expansivo, es decir

$$\|\text{prox}_f \mathbf{x} - \text{prox}_f \mathbf{y}\|^2 + \|(\mathbf{x} - \text{prox}_f \mathbf{x}) - (\mathbf{y} - \text{prox}_f \mathbf{y})\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2, \quad (4.10)$$

esta propiedad nos permite ver la posibilidad de desarrollar algoritmos basados en estos operadores.

**Descomposición Moreau:** Sea  $\lambda > 0$ , entonces

$$\mathbf{x} = \text{prox}_{\lambda f} \mathbf{x} + \lambda \text{prox}_{\lambda^{-1} f^*} \lambda^{-1} \mathbf{x} \quad (4.11)$$

la descomposición muestra que el operador proximal de  $f^*$  puede ser fácilmente obtenido a partir de la función original  $f$ .

## 4.2. Métodos de división proximal

Muchos problemas de optimización incluyen funciones que no son suaves, por tanto no son diferenciables, quedando fuera de los problemas de optimización convencionales. Para

resolverlo, a cada función no suave se le puede asociar un operador de proximidad. Para funciones convexas, propias y semicontinuas, los métodos de división proximal resuelven problemas de la forma

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) + g_1(\mathbf{x}) + \cdots + g_m(\mathbf{x}) \quad (4.12)$$

donde ninguna suposición en las funciones acerca de la suavidad es requerida ya que cada función que no es diferenciable se relaciona en la minimización a través de su operador proximal (4.9). Haciendo uso de la función indicador es posible reescribir problemas de optimización con restricciones a problemas de la forma (4.12) (Onose et al., 2016). Las técnicas de división proximal son muy interesantes ya que permiten diseñar algoritmos con estructuras escalables, algunos métodos proximales son forward-backward, Douglas-Rachford, multiplicadores de dirección alternante o multiplicadores de dirección-simultanea (Combettes y Pesquet, 2011).

Para el interés de la tesis sólo detallaremos el siguiente método.

### 4.2.1. División forward-backward

Se desea solucionar el problema

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{x}) \quad (4.13)$$

donde  $f \in \Gamma_0(\mathbf{R}^n)$ , y  $g : \mathbf{R}^n \rightarrow \mathbf{R}$  es una función convexa y diferenciable con gradiente continuo Lipschitz  $\nabla_g$ , es decir

$$\|\nabla_g(\mathbf{x}) - \nabla_g(\mathbf{y})\| \leq \gamma \|\mathbf{x} - \mathbf{y}\| \quad (4.14)$$

donde  $\gamma \in (0, +\infty)$  (Cevher et al., 2014). Este problema admite al menos una solución, la cual es caracterizada por la ecuación de punto fijo

$$\mathbf{x} = \text{prox}_{\lambda f}(\mathbf{x} - \lambda \nabla g(\mathbf{x})) \quad (4.15)$$

que formula la posibilidad de una iteración de la forma

$$x_{n+1} = \text{prox}_{\lambda_n f}(x_n - \lambda_n \nabla g(x_n)) \quad (4.16)$$

donde  $\lambda_n$  define el tamaño el paso. Este algoritmo de división forward-backward consiste de un paso de gradiente hacia adelante (explícito) usando la función  $g$ , y un paso hacia atrás (implícito) usando la función  $f$  (Combettes y Pesquet, 2011).

## 4.3. Métodos Primal-Dual

Los métodos primal-dual que como su nombre lo indica buscan solucionar el problema primario (primal) como también la formulación dual del problema, son capaces de alcanzar

una completa división (4.12), lo que significa que cada uno de los operadores relacionados al problema son usados por separado, es decir, la solución al problema es iterativamente computada a través de la solución de una secuencia de subproblemas sencillos.

Los esquemas de división primal-dual logran manejar términos diferenciables como no-diferenciables, a través de operadores gradiente (paso explícito), operadores proximal (paso implícito) y operadores lineales. Explotando las características del problema, se busca lograr que los pasos explícitos como implícitos sean fáciles de implementar, para obtener algoritmos eficientes. Ya que los operadores pueden ser utilizados por separado no es necesaria la inversión de operadores, que es una tarea demandante computacionalmente y de gran impacto en problemas de grandes dimensiones (Komodakis y Pesquet, 2015), brindando una ventaja sobre otros esquemas de división.

Muchos problemas pueden ser expresados de la forma

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{L}\mathbf{x}) \quad (4.17)$$

donde  $f$  y  $g$  son propias y semicontinuas y  $\mathbf{L}$  un operador lineal. Este problema de minimización es usualmente conocido como *el problema primal* al cual se le asocia el *problema dual*

$$\underset{v}{\text{minimize}} \quad f^*(-\mathbf{L}^\dagger v) + g^*(v) \quad (4.18)$$

Este método alterna iterativamente entre resolver el problema primal y el dual para converger hacia un punto Karush-Kuhn-Tucker (KKT) (Onose et al., 2016; Parikh y Boyd, 2014). Para más de dos funciones, una generalización del problema primal-dual estaría dada por

$$\min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^k g_i(\mathbf{L}_i \mathbf{x}) \quad \Leftrightarrow \quad \min_{v_1, \dots, v_m} f^*\left(-\sum_{i=1}^k \mathbf{L}_i^\dagger v_i\right) + \sum_{i=1}^k g_i^*(v_i)$$

## 4.4. Resultados principales

### 4.4.1. Algoritmo

Partiendo del problema de optimización del algoritmo 2,

$$\begin{aligned} & \min_{\mathbf{D} \in \mathbb{D}^n, \mathbf{B} \in \mathbb{S}^n} \quad \|\mathbf{D}\|_1 + \|\mathbf{B}\|_1 & (4.19) \\ & \text{subject to} \quad \|\mathbf{y} - \mathcal{A}_1(\mathbf{D}) - \mathcal{A}_2(\mathbf{B})\|_1 \leq \epsilon \\ & \quad \text{diag}(\mathbf{D}) \succeq 0 \\ & \quad \text{diag}(\mathbf{B}) = 0 \end{aligned}$$

para reescribir este problema de la forma (4.12), primero debemos observar que se están minimizando dos variables  $\mathbf{D}$  y  $\mathbf{B}$  lo cual para el caso de primal dual generaría más iteraciones, por tanto para buscar un algoritmo más eficiente computacionalmente primero reescribamos el problema como

$$\begin{aligned} & \min_{\mathbf{C} \in \mathbb{S}^n} \|\mathbf{C}\|_1 \\ & \text{subject to } \|\mathbf{y} - \mathcal{L}(\mathbf{C})\|_1 \leq \epsilon \\ & \text{diag}(\mathbf{C}) \succeq 0 \end{aligned}$$

lo cual es posible dado que los soportes son disjuntos y  $\|\mathbf{D}\|_1 + \|\mathbf{B}\|_1 = \|\mathbf{D} + \mathbf{B}\|_1 = \|\mathbf{C}\|_1$ . Haciendo uso de la función indicador podemos definir las siguientes funciones

$$\begin{aligned} f &= i_S, \quad S = \mathbb{S}^n \\ g_1 &= i_D, \quad D = \{\mathbf{X} \in \mathbf{R}^{n \times n} \mid \text{diag}(\mathbf{X}) \succeq 0\} \\ g_2 &= \|\cdot\|_1, \\ g_3(\hat{\mathbf{y}}) &= i_C(\hat{\mathbf{y}}) \quad C = \{\hat{\mathbf{y}} \in \mathbb{C}^m : \|\mathbf{y} - \hat{\mathbf{y}}\|_2 \leq \epsilon\} \end{aligned}$$

por tanto se puede plantear el problema equivalente

$$\min_{\mathbf{X}} f(\mathbf{X}) + g_1(\mathbf{X}) + g_2(\mathbf{X}) + g_3(\mathcal{L}(\mathbf{X})) \quad (4.20)$$

al cual utilizando la generalización primal-dual

$$\min_x f(x) + \sum_{i=1}^k g_i(\mathcal{L}_i x) \quad \Leftrightarrow \quad \min_{v_1, \dots, v_m} f^*\left(-\sum_{i=1}^k \mathcal{L}_i^\dagger v_i\right) + \sum_{i=1}^k g_i^*(v_i)$$

se obtiene el dual de (4.20)

$$\min_{\mathbf{V}_1, \mathbf{V}_2, v_3} f^*(-\mathbf{V}_1 - \mathbf{V}_2 - \mathcal{L}^\dagger(v_3)) + \sum_{i=1}^3 g_i^*(v_i)$$

y podemos encontrar un punto Kuhn-Tucker. Para determinar el operador adjunto  $\mathcal{L}^\dagger$  primero veamos que el operador se puede expresar como

$$\mathcal{L}(\mathbf{X}) = \mathcal{A}_1(\mathbf{D}) + \mathcal{A}_2(\mathbf{B})$$

donde  $\mathcal{A}_1$  y  $\mathcal{A}_2$  son los operadores definidos en (3.37) y (3.38). Entonces

$$\mathcal{L}^\dagger(\mathbf{y}) = \mathcal{A}_1^\dagger(\mathbf{y}) + \mathcal{A}_2^\dagger(\mathbf{y}).$$

Partiendo de (Pesquet y Repetti, 2014; Komodakis y Pesquet, 2015) proponemos un algoritmo basado en primal-dual usando iteraciones forward-backward como un solucionador al problema de minimización (4.20).

---

**Algorithm 3** PD con iteraciones forward-backward - 4 funciones
 

---

- 1: **dato**  $\mathbf{X}^{(0)}, \hat{\mathbf{X}}^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_2^{(0)}, v_3^{(0)}, \mu, \tau, \sigma_1, \sigma_2, \sigma_3$
  - 2: **repetir para**  $t = 1, \dots$
  - 3:   **hacer en paralelo**
  - 4:      $\mathbf{V}_1^{(t)} = \mathbf{V}_1^{(t-1)} + \hat{\mathbf{X}}^{(t-1)} - P_D \left( \mathbf{V}_1^{(t-1)} + \hat{\mathbf{X}}^{(t-1)} \right)$
  - 5:      $\mathbf{V}_2^{(t)} = \mathbf{V}_2^{(t-1)} + \hat{\mathbf{X}}^{(t-1)} - \text{prox}_{g_2} \left( \mathbf{V}_2^{(t-1)} + \hat{\mathbf{X}}^{(t-1)} \right)$
  - 6:      $v_3^{(t)} = v_3^{(t-1)} + \mathcal{L}(\hat{\mathbf{X}}^{(t-1)}) - P_C \left( v_3^{(t-1)} + \mathcal{L}(\hat{\mathbf{X}}^{(t-1)}) \right)$
  - 7:   **end**
  - 8:    $\mathbf{X}^{(t)} = P_{\mathbb{S}^n} \left( \mathbf{X}^{(t-1)} - \tau \left( \sigma_1 \mathbf{V}_1^{(t)} + \sigma_2 \mathbf{V}_2^{(t)} + \sigma_3 \mathcal{L}^\dagger(v_3^{(t)}) \right) \right)$
  - 9:    $\hat{\mathbf{X}}^{(t)} = 2\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}$
  - 10: **hasta que converja**
- 

Este algoritmo es capaz de alcanzar completa división de las funciones involucradas en el problema de minimización y además presenta una estructura paralelizable. Por ejemplo, los pasos 4 – 6 podrían ejecutarse en tres nodos contrario a un gran nodo central, de tal manera que la carga por cada nodo es baja y facilitaría el manejo de conjuntos de datos de grandes dimensiones. Dado que el costo de envío de información para los nodos es bajo, una arquitectura distribuida podría ser una buena opción.

A pesar que el algoritmo es completamente funcional, puede ser optimizado respecto a la complejidad computacional si replanteamos la función  $f$  para que incluya la condición de diagonal positiva, dando origen a solo tres funciones

$$\begin{aligned}
 f &= i_S, \quad S = \{\mathbf{X} \in \mathbb{S}^n \mid \text{diag}(\mathbf{X}) \succeq 0\} \\
 g_1 &= \|\cdot\|_1, \\
 g_2(\hat{y}) &= i_C(\hat{y}) \quad C = \{\hat{y} \in \mathbb{C}^m : \|y - \hat{y}\|_2 \leq \epsilon\}
 \end{aligned}$$

entonces el problema primal se puede replantear

$$\min_{\mathbf{X}} f(\mathbf{X}) + g_1(\mathbf{X}) + g_2(\mathcal{L}(\mathbf{X}))$$

y se obtiene el dual

$$\min_{\mathbf{V}_1, v_2} f^*(-\mathbf{V}_1 - \mathcal{L}^\dagger(v_2)) + \sum_{i=1}^2 g_i^*(v_i)$$

de los cuales podemos encontrar un punto Kuhn-Tucker. Para este problema dual sólo estamos requiriendo de dos variables duales  $\mathbf{V}_1$  y  $v_2$ , contrario a las tres variables duales del algoritmo 3, además hemos reducido el número de iteraciones necesarias para la convergencia del algoritmo. Todos los detalles son presentados en el algoritmo 4. Para

incrementar la velocidad en la convergencia del algoritmo incluimos un parámetro libre  $\kappa > 0$  el cual puede ser utilizado en ambos algoritmos.

---

**Algorithm 4** PD con iteraciones forward-backward - 3 funciones
 

---

```

1: dado  $\mathbf{X}^{(0)}, \hat{\mathbf{X}}^{(0)}, \mathbf{V}_1^{(0)}, v_2^{(0)}, \mu, \tau, \sigma_1, \sigma_2, \gamma$ 
2: repetir para  $t = 1, \dots$ 
3:   hacer en paralelo
4:      $\mathbf{V}_1^{(t)} = \mathbf{V}_1^{(t-1)} + \hat{\mathbf{X}}^{(t-1)} - \kappa \mathbf{T}_{\gamma/\sigma_1} \left( \mathbf{V}_1^{(t-1)} + \hat{\mathbf{X}}^{(t-1)} \right)$ 
5:      $v_2^{(t)} = v_2^{(t-1)} + \mathcal{L}(\hat{\mathbf{X}}^{(t-1)}) - P_C \left( v_2^{(t-1)} + \mathcal{L}(\hat{\mathbf{X}}^{(t-1)}) \right)$ 
6:   end
7:    $\bar{\mathbf{X}}^{(t)} = P_{\mathbb{S}^n} \left( \mathbf{X}^{(t-1)} - \tau \left( \sigma_1 \mathbf{V}_1^{(t)} + \sigma_2 \mathcal{L}^\dagger(v_2^{(t)}) \right) \right)$ 
8:    $\mathbf{X}^{(t)} = P_D(\bar{\mathbf{X}}^{(t)})$ 
9:    $\hat{\mathbf{X}}^{(t)} = 2\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}$ 
10: hasta que converja

```

---

La descomposición Moreau (4.11) es utilizada para expresar los pasos del algoritmo respecto a los operadores proximal de las diferentes funciones y no a los conjugados.

Todas las actualizaciones son desarrolladas sobre las variables duales de manera paralela. La actualización de la variable primal  $\mathbf{X}^{(t)}$  requiere de las variables duales  $\mathbf{V}_1^{(t)}$  y  $v_2^{(t)}$ . Las iteraciones de las variables dual son desarrolladas en los pasos 4 – 5. Primero, un paso como gradiente hacia delante es desarrollado, seguido por la inclusión de la estructura del algoritmo mediante las funciones no-suaves  $g_i$  a través de la aplicación de sus operadores proximales. Estos producen un paso hacia atrás, como subgradiente. El operador de proximidad de la función  $g_1$  que es la norma  $\ell_1$  que promueve la dispersión en la matriz, resuelve un umbral suave

$$\begin{aligned} \text{prox}_{g_1}(\mathbf{X})_{ij} &= \begin{cases} \mathbf{X}_{ij} - \lambda & \mathbf{X}_{ij} \geq \lambda \\ 0 & |\mathbf{X}_{ij}| \leq \lambda = (\text{sign}(\mathbf{X}_{ij}) \max(|\mathbf{X}_{ij}| - \lambda, 0))_{ij} \\ \mathbf{X}_{ij} + \lambda & \mathbf{X}_{ij} \leq -\lambda \end{cases} \\ &= \mathbf{T}_\lambda(\mathbf{X}). \end{aligned} \quad (4.21)$$

La fidelidad de los datos es forzada al desarrollar la proyección sobre la bola  $\ell_2$  de tamaño  $\epsilon$  definido de acuerdo a la distribución del ruido. El proximal es expresado como

$$\text{prox}_{g_2}(x) = P_C(x) = \begin{cases} \epsilon \frac{x-y}{\|x-y\|_2} + y & \|x-y\|_2 > \epsilon \\ x & \|x-y\|_2 \leq \epsilon. \end{cases} \quad (4.22)$$

Después de la actualización de las variables duales  $\mathbf{V}_1^{(t)}$  y  $v_2^{(t)}$ , éstas son utilizadas en el paso 7 para actualizar la variable primal  $\mathbf{X}^{(t)}$ , la matriz de correlación de interés. Este es

desarrollado a través de un paso similar forward-backward, y el algoritmo utiliza los pasos de actualización  $\tau$  y  $\sigma_1, \sigma_2$  para iterativamente actualizar la solución.

El operador de proximidad de la función indicador de las matrices simétricas con diagonal semi-definida positiva  $f$  es definido como la proyección

$$\begin{aligned} \text{prox}_f(\mathbf{X}) &= P_S = P_D(P_{\mathbb{S}^n}(\mathbf{X})) \\ \text{donde } P_{\mathbb{S}^n}(\mathbf{X}) &= \frac{\mathbf{X}^T + \mathbf{X}}{2}, \text{ y} \\ P_D(\mathbf{X}) &= \begin{cases} \mathbf{X}_{ii} & \text{si } \mathbf{X}_{ii} \geq 0 \\ 0 & \text{si } \mathbf{X}_{ii} \leq 0 \end{cases} \end{aligned} \quad (4.23)$$

La convergencia del algoritmo se garantiza si los parámetros de actualización satisfacen (Onose et al., 2016)

$$\tau(\sigma_1 + \sigma_2 \|\mathcal{L}\|_s^2) < 1 \quad (4.24)$$

además, el parámetro de configuración  $\gamma$  deber ser positivo.

#### 4.4.2. Complejidad computacional

Una eficiente implementación del algoritmo toma ventajas de la división de los datos y la paralelización implícita del algoritmo, por tanto una arquitectura con memoria compartida sería deseable para el proceso de muestro mediante los *sketch*, y una arquitectura distribuida sería eficiente para la comunicación con los nodos de datos dado el bajo costo de comunicación.

Por otra parte, la mayoría de operaciones realizadas en el algoritmo son proporcionales a  $n$ , dado que la variable de interés es  $\mathbf{X}^{(t)} \in \mathbf{R}^{n \times n}$ . Las más demandante operaciones están relacionadas con los operadores  $\mathcal{L}$  y  $\mathcal{L}^\dagger$ . El operador  $\mathcal{L}$  es la suma de los operadores  $\mathcal{A}_1$  y  $\mathcal{A}_2$ , y ambos presentan una complejidad  $O(n^3)$  ya que la mayor cantidad de operaciones la representa el producto de los subbloques de los vectores de muestreo y las matrices  $\mathbf{D}$  y  $\mathbf{B}$  de  $n \times n$ , respectivamente. Los proximal  $\mathbf{T}$  y  $P_D$  tienen complejidad  $O(n^2)$ , y el proximal  $P_C$  complejidad  $O(n)$ .

En resumen, la complejidad computacional del algoritmo propuesto será la máxima complejidad de los pasos  $\max(n^3, n^2, n)$ , por tanto el algoritmo presenta una complejidad  $O(n^3)$ .

#### 4.4.3. Experimentos numéricos

Para observar el desempeño del Algoritmo 4, se crearon dos escenarios. Los parámetros se establecieron por ensayo cumpliendo siempre con (4.24).  $\tau = 0,49$ ,  $\sigma_1 = 10^{-3}$ ,  $\sigma_2 = 1/\|\mathcal{L}\|_s^2$ ,

$\gamma = 10^{-3}$  y  $\kappa = 10^{-1}$ . Como criterios de convergencia se utilizaron la variación relativa

$$\delta = \frac{\|\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}\|_2}{\|\mathbf{X}^{(t)}\|_2} \quad (4.25)$$

que para la mayoría de escenarios fue de  $10^{-3}$ , y el segundo criterio corresponde a la fidelidad de la solución con los datos originales

$$\|y - \mathcal{L}(\mathbf{X}^{(t)})\| \leq \epsilon \quad (4.26)$$

donde  $\epsilon = \sigma m$ .

Para lograr una comparación con los resultados de los capítulos anteriores, los procesos estacionarios que generamos obedecen a una distribución Poisson con parámetro  $\lambda = 15$ . Para todos los escenarios se realizaron 20 experimentos independientes donde las entradas de la matriz de censado  $\mathbf{A}_i$  son i.i.d. generadas desde una distribución Gaussiana  $(0, 1)$ .

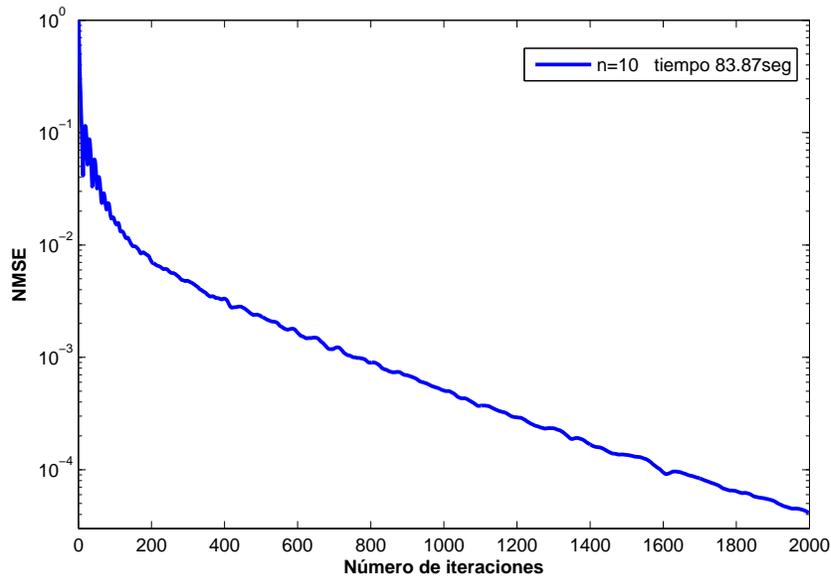
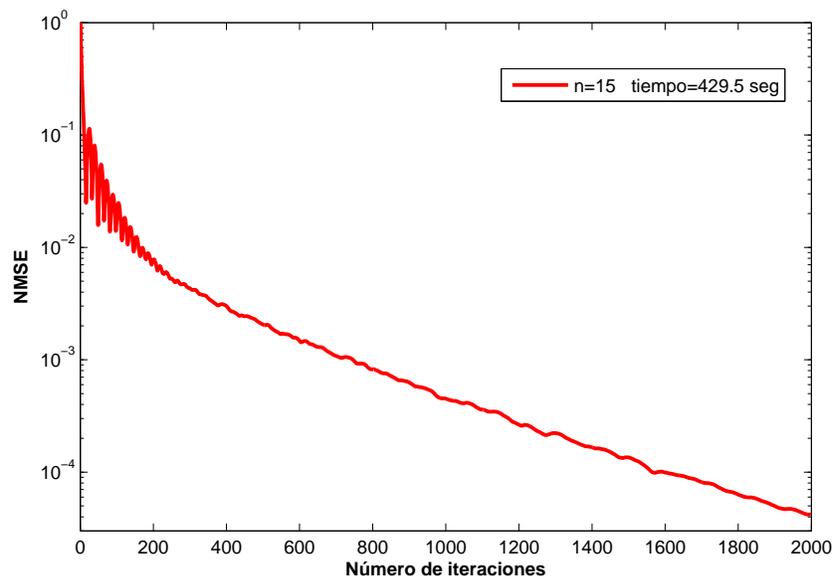
### Escenario 1

Este escenario analiza el error del algoritmo propuesto frente al número de iteraciones. Generamos un proceso aleatorio vectorial con dimensión  $n = 10$  y  $15$ , a continuación se ordena este vector de acuerdo a (3.12) y se calcula la matriz de correlación  $\mathbf{C}_z$  de  $49 \times 49$ , la cual utilizamos para obtener las mediciones cuadráticas. En este escenario los criterios de convergencia fueron eliminados para poder correr las 2000 iteraciones deseadas. El número de mediciones utilizadas fue de 55 y 120 respectivamente, que corresponde a la relación obtenida de  $n(n+1)/2$ .

En la figura 4.1(a) se observa como el NMSE va disminuyendo a medida que el número de iteraciones aumenta, también se aprecia que el número de iteraciones para obtener un error bajo es alto, lo que se ve compensado por la leve complejidad computacional que tienen las iteraciones. Dado que la dimensión es pequeña el tiempo necesario para alcanzar las 2000 iteraciones es de apenas 83,87 seg. Un comportamiento similar se observa para la dimensión  $n = 15$  (Figura 4.1(b)), y con un tiempo requerido de 429 seg para alcanzar las 2000 iteraciones, lo cual es comprensible dado que la dimensión es mayor por tanto los pasos del algoritmo toman más tiempo.

### Escenario 2

El escenario 2 compara el NMSE del algoritmo 4 con el algoritmo 2. Generamos un proceso aleatorio vectorial con dimensión  $n = 7$ . A continuación, se ordena este vector de acuerdo a (3.12) y computamos la matriz de correlación  $\mathbf{C}_z$  de  $49 \times 49$ , la cual utilizamos para obtener las mediciones (3.3). En orden para encontrar la solución al problema convexo (3.26) utilizamos el paquete software de optimización CVX (Grant y Boyd, 2014) para el algoritmo 2.

(a) NMSE para la estimacion del algoritmo 4(primal-dual) para  $n = 10$ .(b) NMSE para la estimacion del algoritmo 4(primal-dual) para  $n = 17$ .**Figura 4.1:** NMSE vs el número de iteraciones para las dimensiones  $n = 10$  y  $n = 17$ .

De la Figura 4.2(a) se observa que el algoritmo propuesto obtiene niveles de error similares con un número inferior de mediciones respecto al estimador referente de Chen et al. De la Figura 4.2(b) se aprecia que el algoritmo primal dual requiere un número de mediciones mayor que el requerido por el algoritmo 2, esta sería la compensación por la disminución de la carga computacional obtenida al utilizar el algoritmo primal-dual basado en iteraciones

forward-backward.

### Escenario 3

En este escenario estudiamos el comportamiento del algoritmo propuesto para diferentes dimensiones. Generamos un proceso aleatorio vectorial con distribución Poisson con parámetro  $\lambda = 15$  para las dimensiones  $n = 6, 7, 10, 15$  y  $50$ . Cada vector es ordenado de acuerdo a (3.12), obteniendo las matrices de correlación  $\mathbf{C}_z$  de  $49 \times 49$ ,  $100 \times 100$ ,  $225 \times 225$  y  $2500 \times 2500$  respectivamente, las cuales son usadas para obtener las mediciones (3.3).

De la Figura 4.3 se observa el desempeño del NMSE para las diferentes dimensiones para el algoritmo 4, se nota que a pesar de lograr un bajo NMSE en comparación al algoritmo 2 (Figura 3.9), para obtener el mismo NMSE se hace necesario un ligero incremento en el número de mediciones, el cual estaría dentro del rango aceptable.

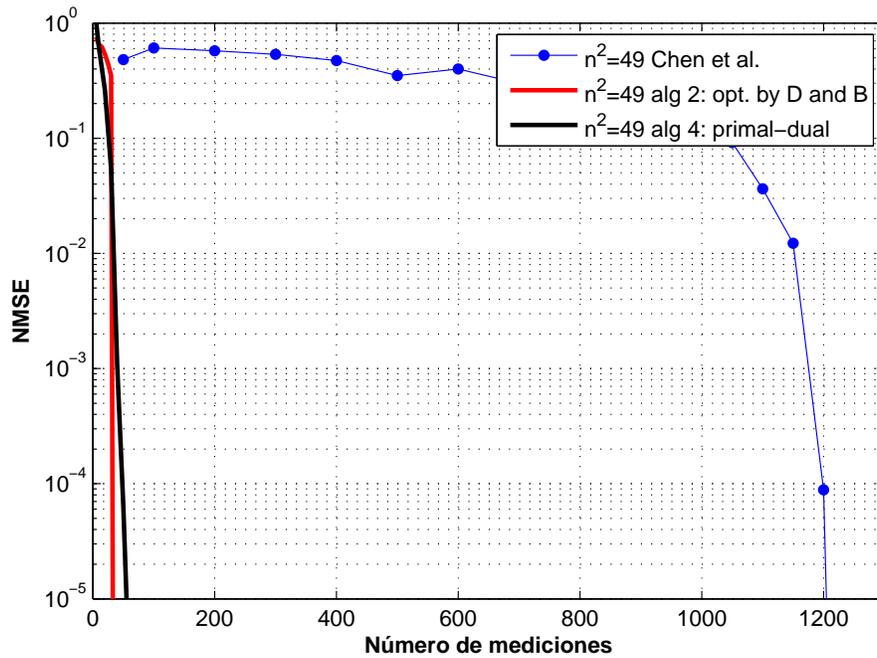
### Escenario 4

El escenario 4 estudia la robustez del algoritmo propuesto frente al ruido. Para las dimensiones  $n = 10$  y  $15$  perturbamos el vector de mediciones con ruido uniforme entre  $[-\sigma, \sigma]$ . La Figura 4.4 y la Figura 4.5 muestra el NMSE de la estimación de la matriz de correlación como una función del número de mediciones  $m$ , para  $\sigma = 10^{-2}$  y  $10^{-1}$  y  $\epsilon = \sigma m$  dado en (3.26).

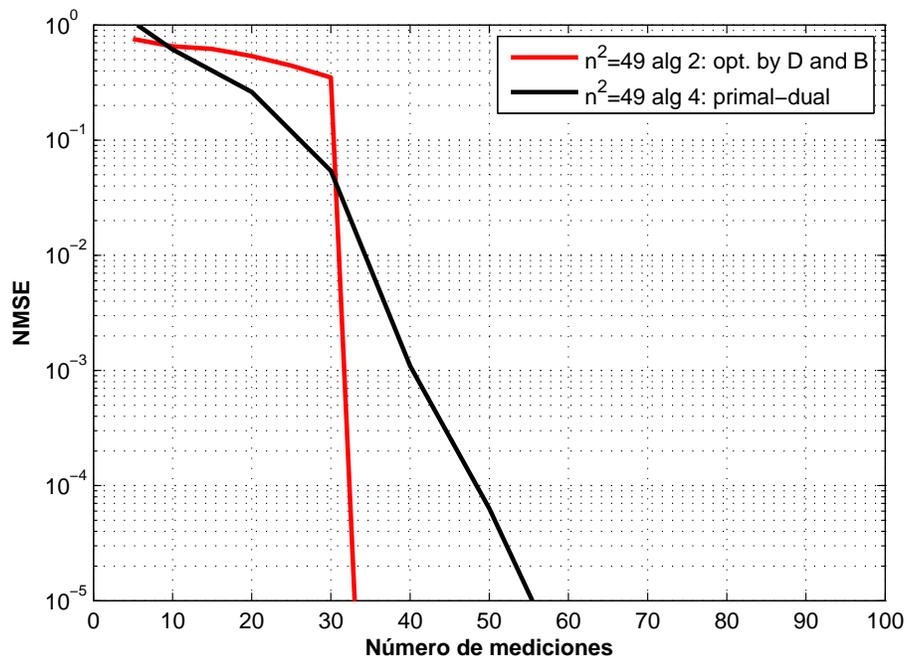
De la Figura 4.4 se observa que el algoritmo 2 basado en CVX alcanza un valor de NMSE más bajo que el algoritmo 4 basado en primal-dual. Como era de esperarse el mismo comportamiento es observado para la dimensión  $n = 15$ , el cual se puede apreciar en la Figura 4.5.

## 4.5. Conclusiones

En este capítulo se presentó el diseño de un algoritmo basado en métodos primal-dual y forward-backward que extiende el algoritmo 2 desarrollado en el capítulo 3. La estructura de este nuevo algoritmo permite una alta paralelización y escalabilidad para manejar un gran volumen de datos. Se mostró que el desempeño obtenido por el nuevo algoritmo es similar al algoritmo 2 pero con una carga computacional mucho menor, superando la gran limitante del algoritmo 2 que al utilizar un solucionador genérico requería de una carga computacional alta, por tanto el nuevo algoritmo podría ser utilizados en aplicaciones que hacen uso de datos de grandes dimensiones y requieren de algoritmos eficientes computacionalmente.

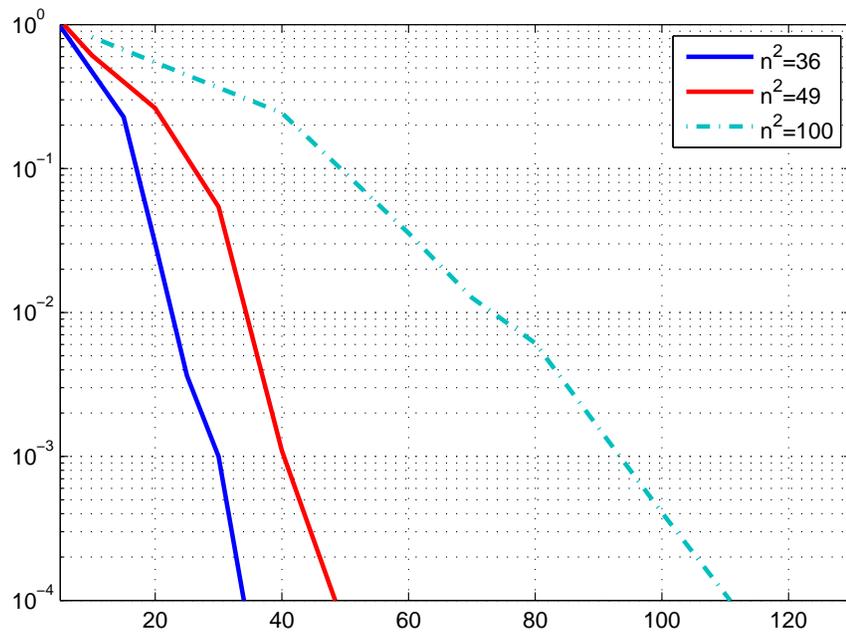
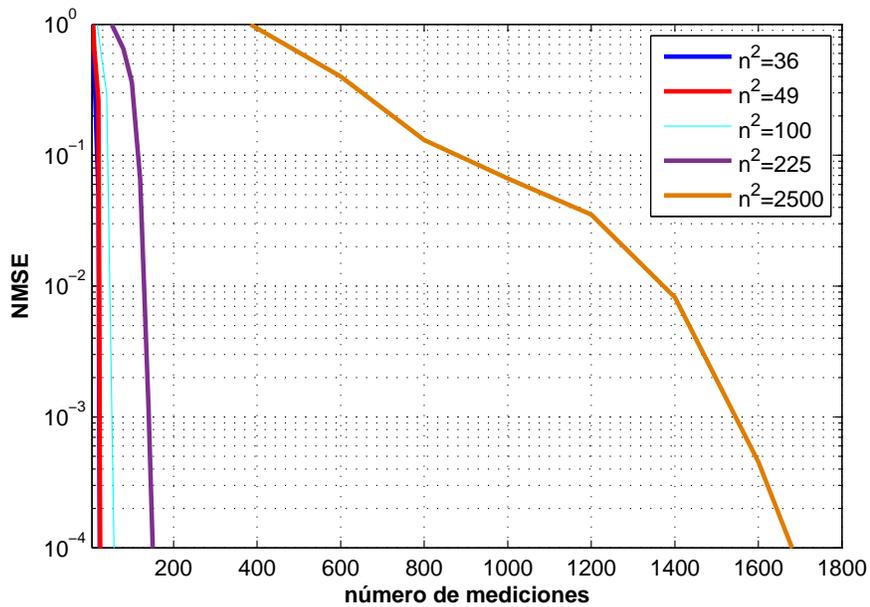


(a) NMSE para los estimadores: Chen et al y los algoritmos propuesto 2(cvx) y 4(primal-dual), para  $n = 7$ .



(b) NMSE para los algoritmos propuesto 2(cvx) y 4(primal-dual), para  $n = 7$ .

**Figura 4.2:** Comparación del NMSE del algoritmo 2, el algoritmo 4(primal-dual) y el estimador de Chen et al.

(a) NMSE para las dimensiones  $n = 6, 7, 10$  utilizando el algoritmo 4 (primal-dual).(b) NMSE para las dimensiones  $n = 6, 7, 10, 15$  y  $50$  utilizando el algoritmo 4 (primal-dual).**Figura 4.3:** NMSE del estimador primal-dual para diferentes dimensiones.

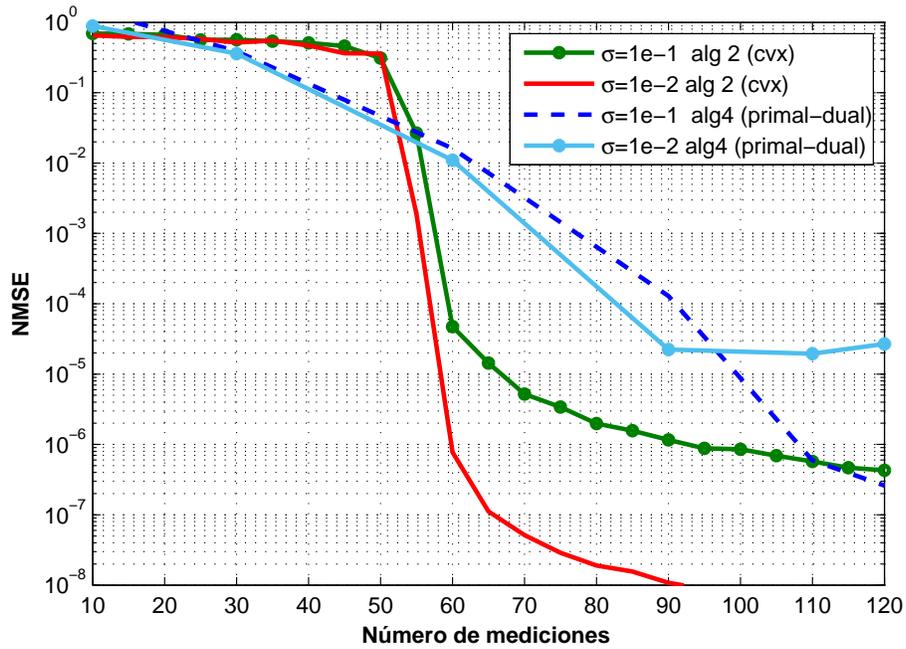


Figura 4.4: NMSE para la dimensión  $n=10$  con diferentes niveles de ruido.

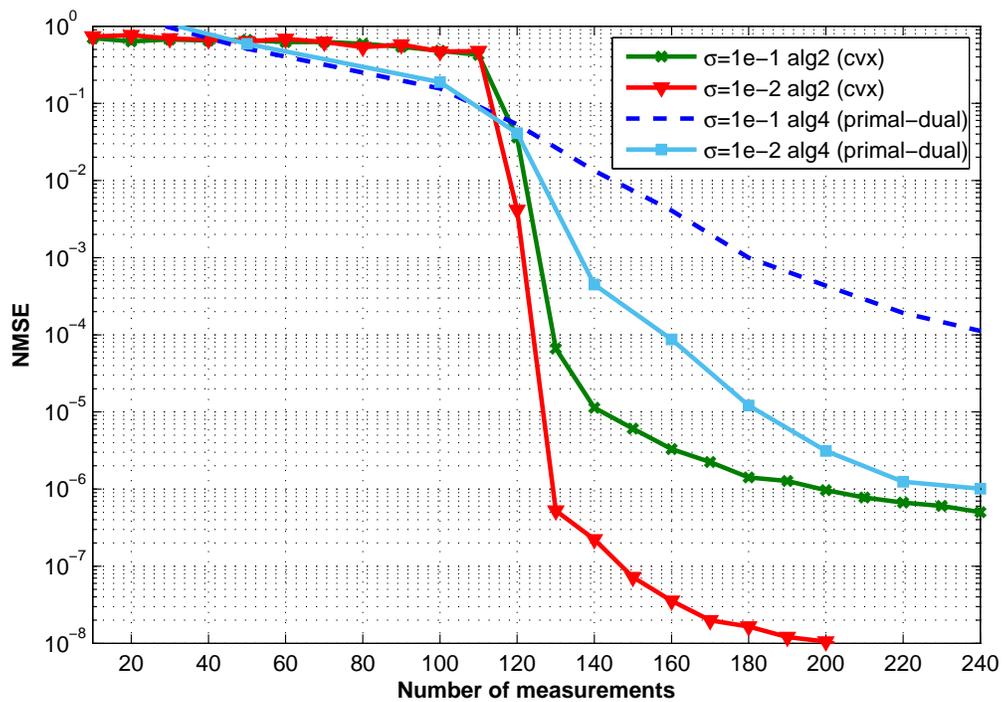


Figura 4.5: NMSE para la dimensión  $n=15$  con diferentes niveles de ruido.

# Capítulo 5

## Estimación de la Matriz de Correlación No-estacionaria a Partir de Mediciones Cuadráticas

En este capítulo se explora la estimación de la matriz de correlación de datos ordenados en tiempo y rango los cuales presentan un comportamiento no-estacionario. Basado en el modelo de estimación desarrollado en el capítulo 3, se muestra que es posible una estimación con una reducción en el número de recursos necesarios y con igual precisión que el modelo de referencia.

### 5.1. Estructura de la matriz de correlación

Dada la presencia en diferentes aplicaciones de señales con un comportamiento no-estacionario, buscamos basados en el modelo consolidado de estimación (3.3.2) lograr incluir información de la estructura de la matriz de correlación para lograr el diseño de un programa de optimización que permita recuperarla pero con una complejidad menor. Primero, recordando la estructura general de la matriz

$$C_z = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} \quad (5.1)$$

y dada la condición de simetría de la matriz de correlación, esta puede ser reescrita como

$$C_z = \begin{bmatrix} D_1 & B_{12} & \cdots & B_{1n} \\ B_{12}^T & D_2 & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{1n}^T & B_{2n}^T & \cdots & D_n \end{bmatrix}. \quad (5.2)$$

A su vez, esta matriz puede ser expresada en función de la submatrices que la conforman, es decir

$$\mathbf{C}_z = \begin{bmatrix} \mathbf{D}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{D}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D}_n \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{B}_{12} & \cdots & 0 \\ \mathbf{B}_{12}^T & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \mathbf{B}_{(n-1)n} \\ 0 & \cdots & \mathbf{B}_{(n-1)n}^T & 0 \end{bmatrix} \quad (5.3)$$

es decir

$$\mathbf{C}_z = \mathcal{D}_n + \mathcal{B}_{12} + \cdots + \mathcal{B}_{(n-1)n} \quad (5.4)$$

donde  $\mathcal{D}_n$  representa la matriz en bloques diagonal, con las submatrices  $\mathbf{D}_{ii}$  en la posición  $ii$ , y  $\mathcal{B}_{ij}$  representa la matriz en bloques de ceros excepto en las posición  $ij$  ( $\mathbf{B}_{ij}$ ) e  $ji$  ( $\mathbf{B}_{ij}^T$ ).

## 5.2. Estimador para el caso no-estacionario

Retomando el operador lineal general

$$\mathcal{A}(\mathbf{C}) : \mathcal{C} \rightarrow \{\langle \mathbf{A}_i, \mathbf{C} \rangle\}_{i=1}^m \quad (5.5)$$

e incluyendo la información de la estructura de la matriz no estacionaria (5.4), se tiene

$$\begin{aligned} \{\mathbf{a}_i^T \mathbf{C}_z \mathbf{a}_i\}_{i=1}^m &= \{\mathbf{a}_i^T (\mathcal{D}_n + \mathcal{B}_{12} + \cdots + \mathcal{B}_{(n-1)n}) \mathbf{a}_i\}_{i=1}^m \\ &= \{\mathbf{a}_i^T (\mathcal{D}_n) \mathbf{a}_i + \mathbf{a}_i^T (\mathcal{B}_{12}) \mathbf{a}_i + \cdots + \mathbf{a}_i^T (\mathcal{B}_{(n-1)n}) \mathbf{a}_i\}_{i=1}^m \end{aligned} \quad (5.6)$$

donde el paso dos se da por la linealidad del operador. Recordando el muestreo en bloques (3.32) desarrollado anteriormente

$$\mathbf{a}_i = [\mathbf{a}_{i1}^T, \mathbf{a}_{i2}^T, \dots, \mathbf{a}_{in}^T]^T. \quad (5.7)$$

se puede analizar cada producto independientemente, así para el caso de las matrices diagonales se tiene

$$\begin{aligned} \mathbf{a}_i^T (\mathcal{D}_n) \mathbf{a}_i &= [\mathbf{a}_{i1}^T, \mathbf{a}_{i2}^T, \dots, \mathbf{a}_{in}^T] \begin{bmatrix} \mathbf{D}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{D}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D}_n \end{bmatrix} \begin{bmatrix} \mathbf{a}_{i1} \\ \mathbf{a}_{i2} \\ \vdots \\ \mathbf{a}_{in} \end{bmatrix} \\ &= \mathbf{a}_{i1}^T \mathbf{D}_1 \mathbf{a}_{i1} + \cdots + \mathbf{a}_{in}^T \mathbf{D}_n \mathbf{a}_{in}. \end{aligned} \quad (5.8)$$

Ahora, analizando el producto de las matrices  $\mathcal{B}_{lp}$  se tiene

$$\begin{aligned} \mathbf{a}_i^T (\mathcal{B}_{lp}) \mathbf{a}_i &= [\mathbf{a}_{i1}^T, \mathbf{a}_{i2}^T, \dots, \mathbf{a}_{in}^T] \begin{bmatrix} 0 & \cdots & \mathbf{B}_{lp} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{B}_{lp}^T & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a}_{i1} \\ \mathbf{a}_{i2} \\ \vdots \\ \mathbf{a}_{in} \end{bmatrix} \\ &= \mathbf{a}_{ip}^T \mathbf{B}_{lp}^T \mathbf{a}_{il} + \mathbf{a}_{il}^T \mathbf{B}_{lp} \mathbf{a}_{ip}, \end{aligned} \quad (5.9)$$

por tanto, el operador lineal esta expresado como la suma del operador actuando sobre las submatrices

$$\begin{aligned} \left\{ \mathbf{a}_i^T \mathbf{C}_z \mathbf{a}_i \right\}_{i=1}^m &= \left\{ \sum_{l=1}^n \mathbf{a}_{il}^T \mathbf{D}_{il} \mathbf{a}_{il} + \sum_l \sum_{p<l}^n \mathbf{a}_{ip}^T \mathbf{B}_{lp} \mathbf{a}_{il} + \sum_l \sum_{p>l}^n \mathbf{a}_{il}^T \mathbf{B}_{lp} \mathbf{a}_{ip} \right\}_{i=1}^m \\ &= \mathcal{L}_B(\mathbf{C}_z). \end{aligned} \quad (5.10)$$

el problema de optimización general (Chen et al., 2015) para recuperar la matriz de correlación de datos ordenados para el caso de un proceso no-estacionario

$$\begin{aligned} &\text{minimize } \|\mathbf{C}_z\|_1 \\ &\quad \mathbf{C}_z \in \mathbb{S}^n \\ &\text{sujeto a } \|\mathbf{y} - \mathcal{L}(\mathbf{C}_z)\|_1 \leq \epsilon \\ &\quad \text{diag}(\mathbf{C}_z) \succeq 0 \end{aligned} \quad (5.11)$$

puede ser replanteado como

$$\begin{aligned} &\text{minimize } \|\mathbf{C}_z\|_1 \\ &\quad \mathbf{D}_i \quad i=1, \dots, n \in \mathbb{D}^n \quad \mathbf{B}_{ij} \quad i, j=1, \dots, n \quad i < j \in \mathbb{B}^n \\ &\text{sujeto a } \|\mathbf{y} - \mathcal{L}_B(\mathbf{C}_z)\|_1 \leq \epsilon \\ &\quad \text{diag}(\mathbf{D}_i) \succeq 0 \\ &\quad \text{diag}(\mathbf{B}_{ij}) = 0 \end{aligned} \quad (5.12)$$

donde  $\mathbf{C}_z$  es la matriz formada por los submatrices  $\mathbf{D}_i$  y  $\mathbf{B}_{ij}$ . Con este nuevo programa de optimización se busca recuperar las submatrices de la triangular superior ya que la triangular inferior está conformada por sus transpuestas, y como estas submatrices son de dimensión  $n \times n$  contrario a la matriz  $n^2 \times n^2$ , este programa deberá tener un menor consumo de recursos.

Para determinar el número de mediciones necesarias, primero se observa que se tienen  $n(n-1)/2$  matrices  $\mathbf{B}_{ij}$  las cuales tienen  $n^2$  grados de libertad, y  $n$  matrices  $\mathbf{D}_i$  las cuales tienen  $n$  grados de libertad. En total se tendría un número de mediciones alrededor de  $n^3(n-1)/2 + n^2$  que es muy similar a las  $n^2(n^2+1)/2$  mediciones requeridas por el estimador general.

### 5.2.1. Resultados numéricos

Para estudiar el desempeño del estimador se tomaron 7 señales ECG de la base de datos MIT-BIH (Goldberger et al., 1992), dada su cualidad de no-estacionariedad. La dimension de cada señal ECG es de  $\mathbf{R}^{1000000}$ , por tanto todo el conjunto de datos es de  $7 \times 1000000$ . En la Figura 5.1 se observan las señales en toda su duración, mientras que en la Figura 5.2 se observan las señales en el intervalo de tiempo 1975 – 2000 segundos. Con nuestro estimador se demuestra que para recuperar la matriz de correlación no es necesario el almacenamiento de todas las realizaciones, sólo se requiere del vector de mediciones  $y \in \mathbf{R}^m$ , con  $m \ll 1000000$ . Para analizar el desempeño de los estimadores se diseñaron dos escenarios y se utilizó CVX (Grant y Boyd, 2014) como software de optimización.

#### Escenario 1

El escenario 1 compara el estimador propuesto (5.12) con el estimador general (5.11), para 3, 5 y 7 señales ECG. En la Figura 5.3 se observa la recuperación de la matriz de correlación para las dimensiones mencionadas, mostrando el NMSE como función del número de mediciones. Se aprecia una reducción en las mediciones requeridas para la recuperación de la matriz respecto al estimado de (Chen et al., 2015), así para la dimensión de  $n = 3$  la reducción es casi nula, mientras que para  $n = 5$  es de alrededor de 50 mediciones, y para  $n = 7$  es de 100 mediciones.

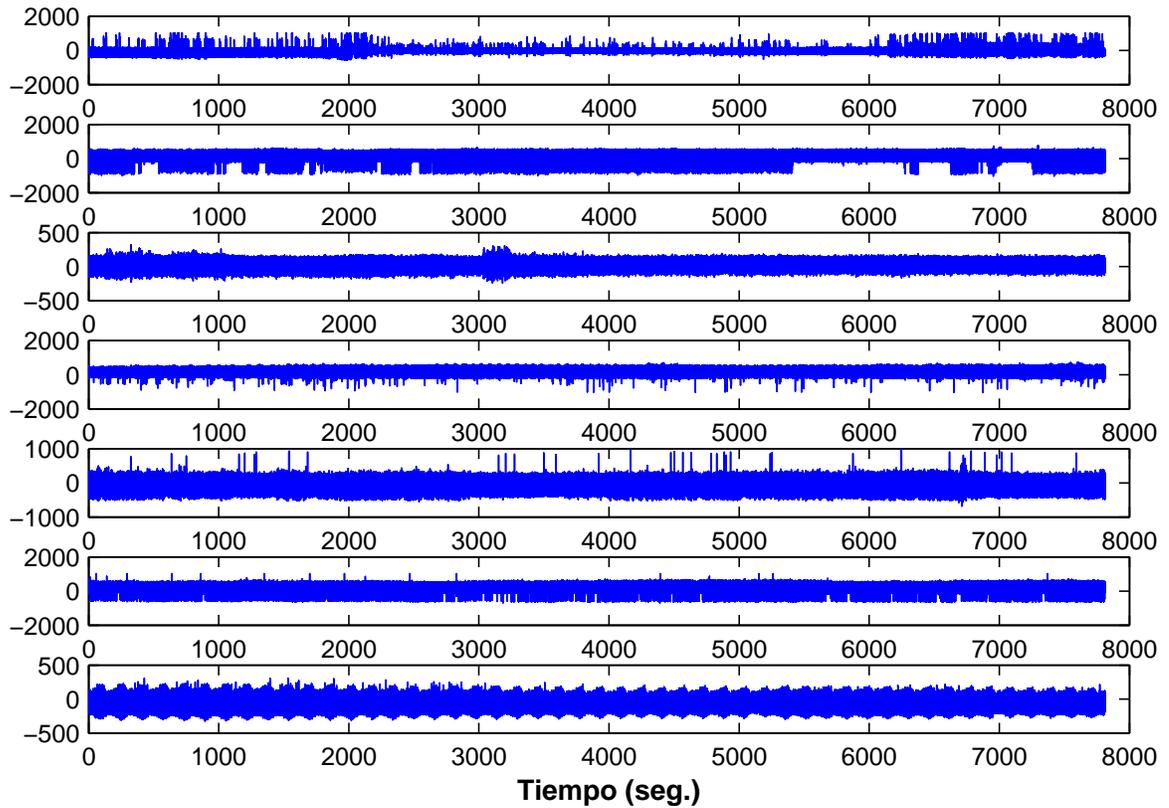
En resumen para recuperar la matriz de correlación sólo sería necesario almacenar un vector con  $m = 1100$  datos, contrario a tener que almacenar todas las realizaciones de las señales, es decir  $7 \times 1000000$ , por tanto se tiene una reducción del 99,98 % en el número de datos a ser almacenados. También se observa que la reducción es proporcional a la dimensión, resultado que coincide con el análisis de grados de libertad realizado anteriormente.

#### Escenario 2

En este escenario se estudia la robustez del estimador propuesto frente a los efectos adversos generados por el ruido. Para 5 y 7 señales ECG, se contaminó el vector de mediciones con ruido uniforme entre  $[-\sigma, \sigma]$ , con  $\sigma = 1e^{-1}$ , 1. La Figura 5.4 muestra el desempeño del NMSE frente al número de mediciones necesarias para  $n = 5$ , corroborándose que la reducción en las mediciones es poca para bajas dimensiones comparada con el estimador general.

En la Figura 5.5 se presenta el desempeño del estimador para la dimensión  $n = 7$ , apreciándose una reducción en las mediciones necesarias de un poco más de 100 por parte del estimador propuesto, corroborándose que el estimador propuesto para el caso no-estacionario presenta una reducción en el número de mediciones la cual es proporcional con la dimensión de los datos.

Por ultimo, en la Tabla 5.1 se aprecia la gran reducción obtenida en el consumo de la RAM, ya que para el estimador propuesto el incremento al pasar de una dimensión 5 ( $25 \times 25$ ) a



**Figura 5.1:** Señales ECG de la base MIT-BIH

7 ( $49 \times 49$ ) solo es de 350MB, mientras que el estimador de [Chen et al. \(2015\)](#) requiere de 12GB, por tanto se hace factible la adopción del estimador propuesto en escenarios donde el uso de los recursos sea de gran importancia.

**Tabla 5.1:** Uso de la RAM por parte del estimador estándar y el estimador propuesto.

Dimensión ( $n$ )	5	7
Est. <a href="#">Chen et al. (2015)</a>	2,7GB	13GB
Est. propuesto	450MB	700MB

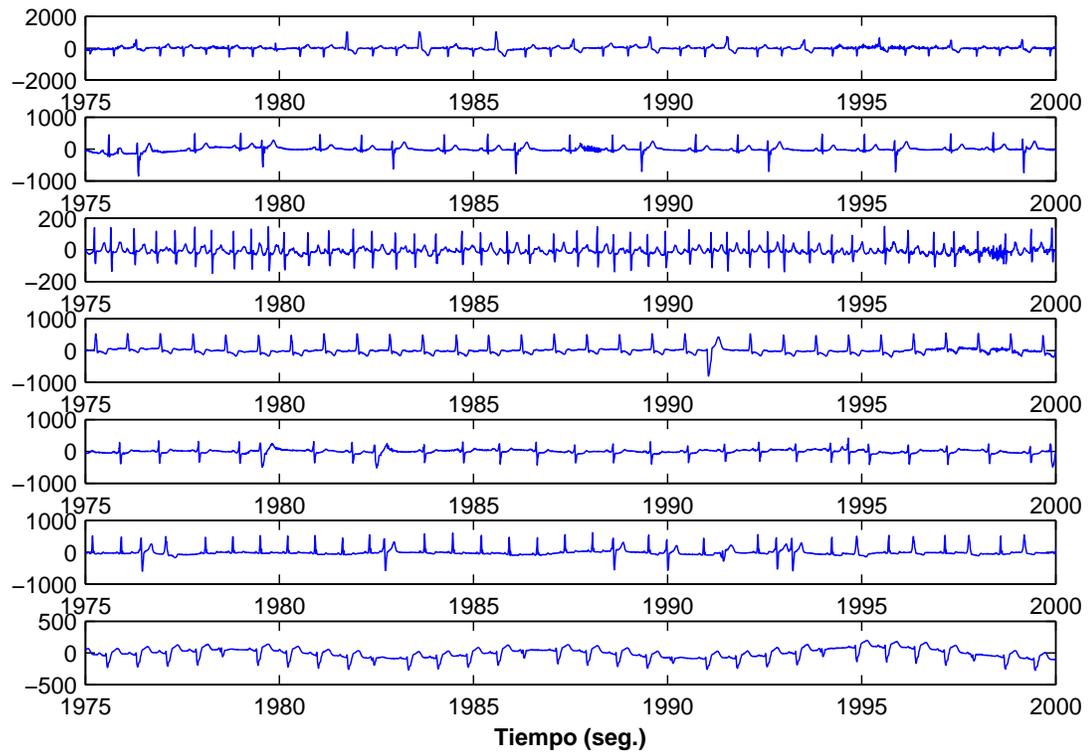


Figura 5.2: Señales ECG de la base MIT-BIH en el rango 1975 a 2000 segundos

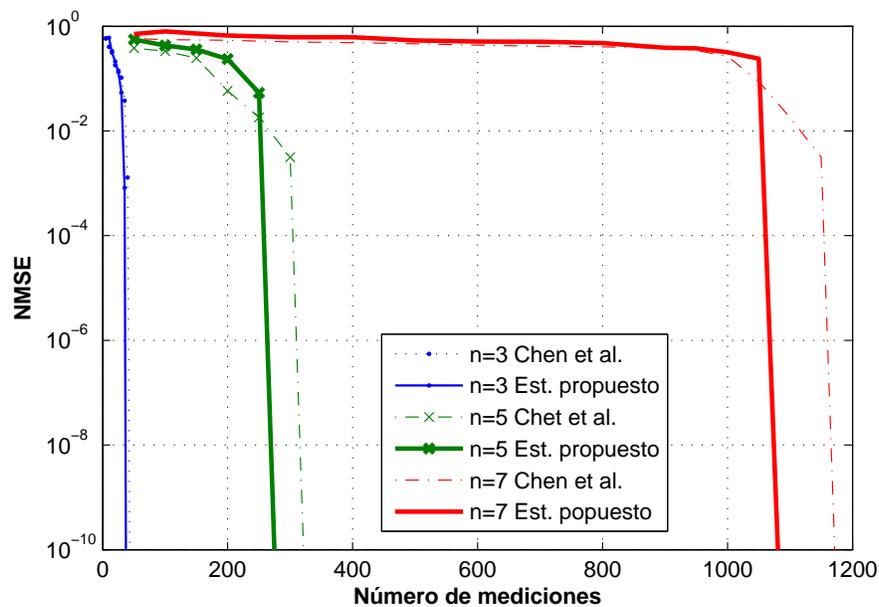
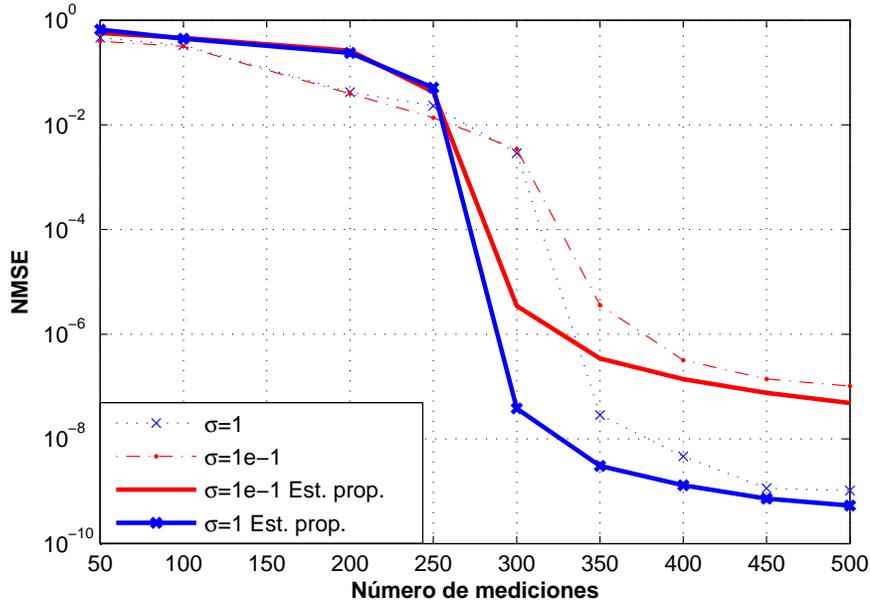
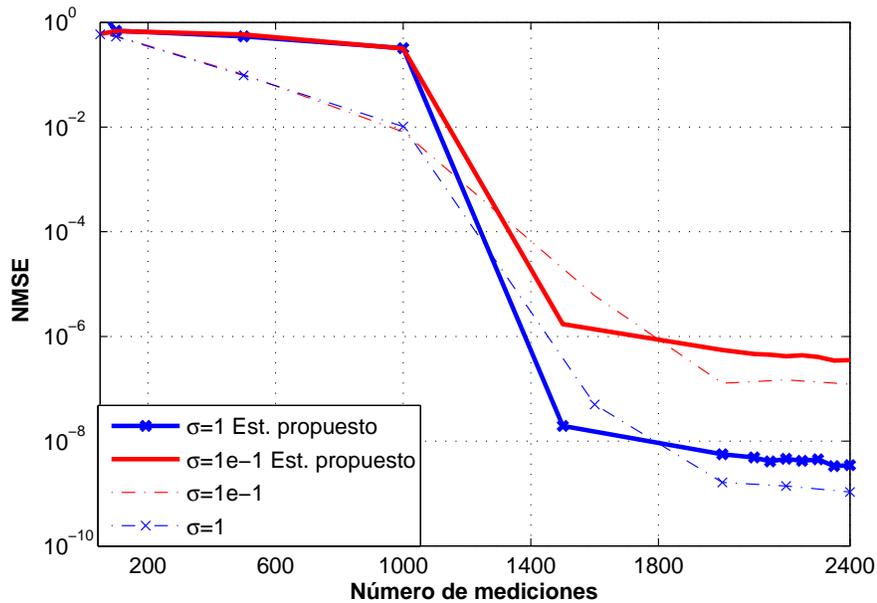


Figura 5.3: Comparación del NMSE entre el estimador general y el esquema propuesto para  $n = 3, 5, 7$ , caso no-estacionario.



**Figura 5.4:** Comparación del NMSE entre el estimador general y el esquema propuesto para  $n = 5$  y diferentes niveles de ruido, caso no-estacionario.



**Figura 5.5:** Comparación del NMSE entre el estimador general y el esquema propuesto para  $n = 7$  y diferentes niveles de ruido, caso no-estacionario.

# Capítulo 6

## Conclusiones y Trabajo Futuro

El tema central de esta tesis ha sido explorar la estructura de la matriz de correlación de datos que han sido ordenados y su estimación a partir de vector de mediciones con una dimensión mucho menor que la dimension original. Terminamos esta tesis con unas conclusiones de nuestros resultados y algunas direcciones futuras de investigación.

### 6.1. Conclusiones

Se han analizado las propiedades matemáticas de la medición cuadrática (Chen et al., 2015), que es una herramienta para la estimación de matrices de covarianza con estructura de bajo rango, dispersa, dispersa y bajo rango y Toeplitz, la cual mediante sketch cuadráticos captura la información de la matriz en un vector denominado vector de mediciones, a partir del cual utilizando un programa convexo es posible la estimación de la matriz. Así mismo se analizaron diferentes métodos que permiten la estimación de tales matrices con un número reducido de mediciones, observando que muchos de ellos se limitaban a un tipo específico de estructura para procesos estacionarios.

En esta tesis se investigó la estructura que presenta la matriz de correlación cuando el proceso aleatorio subyacente ha sido ordenado en tiempo y en rango, mostrando que cuando es estacionario la matriz depende de una submatriz diagonal y una submatriz con diagonal cero. Con base en esta estructura y usando los sketch como vectores de muestreo, se desarrolló dos nuevos algoritmos, el primero consistió en realizar un promedio luego de haber estimado la matriz, el segundo incluyó la información de la estructura dentro del programa de optimización lo que permitió reducir notablemente el número de mediciones necesarias para la estimación de la matriz, pasando de  $n^2(n^2 + 1)/2$  a  $n(n + 1)/2$ . Es de resaltar que ambos algoritmo mostraron un desempeño superior con y sin ruido uniforme frente al desarrollado por Chen et al. (2015).

A partir de la estructura en bloques descubierta en esta tesis, se desarrolló un muestro en bloques que permitió capturar la estructura de la matriz y que dio origen al algoritmo 2. También se diseñó un modelo de medición que permite una convergencia más rápida, de

especial interés cuando el número de realizaciones del proceso estacionario es limitada.

Dada la complejidad computacional cuando el proceso aleatorio es de grandes dimensiones, basado en métodos de división proximal, en especial en los métodos forward-backward y primal dual, desarrollamos un nuevo algoritmo escalable como extensión del algoritmo 2 desarrollado en nuestra tesis. Este nuevo algoritmo demostró ser eficiente computacionalmente ya que explota mejor las características de la estructura de la matriz al solucionar de forma iterativa y paralela una secuencia de subproblemas más fáciles, lo cual permite estimar la matriz de correlación aún cuando esta es de grandes dimensiones.

Por último, basados en el modelo de medición consolidado se desarrolló un estimador para el caso cuando los datos presenten un comportamiento no-estacionario. El desempeño fue similar al estimador general, pero con una notable reducción en la complejidad y en el uso de recursos para su estimación. Además, se logró obtener una reducción en el número de mediciones la cual es proporcional a la dimensión de los datos.

## 6.2. Limitaciones

Dada la limitación en recursos computacionales, los diferentes algoritmos fueron probados sobre un computador de escritorio con 16GB de memoria RAM y un proceso intel core i7 de séptima generación, lo cual nos limitó en las dimensiones a tener en cuentas en nuestros análisis. Otra limitante se relaciona con los datos utilizados en nuestras simulaciones, ya que muchas de ellas fueron realizadas utilizando datos sintéticos.

## 6.3. Trabajo Futuro

- Se observó que el uso de sketch para la construcción de modelos de medición mostró una gran adaptabilidad a las condiciones particulares del problema, un interesante tópico de investigación sería la estimación de matrices con estructura Toeplitz a bloques o matrices circulantes, ya que aparecen en diferentes aplicaciones.
- Un tema interesante es explorar algoritmos aleatorios que permitan una convergencia más rápida de los algoritmos desarrollados en esta tesis, ya que disminuiría la carga computacional y por tanto tendría un impacto notable en grandes datos.
- Sería interesante utilizar el modelo de medición desarrollado con datos reales de otras aplicaciones, como por ejemplo datos de radar.
- Otra gran estructura de interés práctico es la matriz de covarianza inversa o matriz de precisión, en especial cuando se imponen restricciones de dispersión (*sparsity*), ya que los ceros de esta matriz indican independencia condicional entre las variables, sería interesante adaptar el modelo de medición para la estimación de tal matriz.

- Sería interesante ver la posibilidad de extender el modelo para el caso de procesos no estacionarios a grandes dimensiones.
- Por último, dada la necesidad de determinar la matriz inversa de covarianza o matriz de precisión en muchas aplicaciones, sería interesante llevar ideas desarrolladas en esta tesis para buscar determinar eficientemente dicha matriz.

# Apéndice A

## Apéndice: Optimización convexa

Un problema de optimización tiene la forma

$$\underset{x}{\text{minimize}} f_0(x) \text{ sujeto a } f_i(x) \leq b_i, \quad i = 1, \dots, m \quad (\text{A.1})$$

el vector  $x$  representa la variable de optimización del problema, la función  $f_0 : \mathbf{R}^n \rightarrow \mathbf{R}$  es llamada función objetivo o función costo. Las restricciones de desigualdades  $f_i(x) \leq b_i$  y las correspondientes funciones  $f_i : \mathbf{R}^n \rightarrow \mathbf{R}, i = 1, \dots, m$  son las funciones de restricción, y las constantes  $b_i, i = 1, \dots, m$  son los límites o cotas para las funciones restricción.

El conjunto de puntos para los cuales las funciones objetivo y las funciones de restricción están definidas

$$D = \bigcap_{i=0}^m \text{dom} f_i \quad (\text{A.2})$$

es llamado el dominio del problema de optimización. Un punto  $x \in D$  se dice ser realizable si satisface las restricciones  $f_i(x) \leq b_i, i = 0, 1, \dots, m$ . Si existe al menos un punto realizable el problema de optimización se dice ser realizable o solucionable. El valor óptimo  $x^*$  del problema es definido como

$$x^* = \inf\{f_0(x) | f_i(x) \leq b_i, i = 1, \dots, m\} \quad (\text{A.3})$$

o en otras palabras,  $x^*$  es óptimo si tiene el valor de la función objetivo menor entre todos los vectores que satisfacen las restricciones, es decir, para cualquier  $y$  con  $f_1(y) \leq b_1, \dots, f_m(y) \leq b_m$  tenemos  $f_0(y) \geq f_0(x^*)$ .

En esta tesis hacemos especial interés en problemas de optimización caracterizados por formas particulares de las funciones objetivo y de restricción. Cuando estas funciones son lineales, es decir  $f_0, \dots, f_m$  satisfacen que para todo  $x, y \in \mathbf{R}^n$  y  $\alpha, \beta \in \mathbf{R}$

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y) \quad (\text{A.4})$$

el problema de optimización es llamado problema de optimización lineal. Cuando las funciones objetivo y restricción  $f_0, \dots, f_m$  son convexas

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \quad (\text{A.5})$$

para todo  $x, y \in \mathbf{R}^n$  y todo  $\alpha, \beta \in \mathbf{R}$  con  $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$ , el problema se le llama *problema de optimización convexo*. Se observa que convexidad es más general que linealidad, entonces la optimización convexa es una generalización de la programación lineal.

Una propiedad fundamental de los problemas de optimización convexa es que cualquier punto óptimo local es también un punto óptimo global.

# Apéndice B

## Apéndice: Pruebas de proposiciones Capítulo 3

### B.1. Operadores lineales

Sean  $X, Y \in \mathbb{D}^n$ , el operador  $\mathcal{A}_1$  es lineal ya que

$$\begin{aligned}\mathcal{A}_1(X + Y) &= \left\{ \sum_{k=1}^n \mathbf{a}_{ik}^T (X + Y) \mathbf{a}_{ik} \right\}_{i=1}^m \\ &= \left\{ \sum_{k=1}^n (\mathbf{a}_{ik}^T X \mathbf{a}_{ik} + \mathbf{a}_{ik}^T Y \mathbf{a}_{ik}) \right\}_{i=1}^m \\ &= \left\{ \sum_{k=1}^n \mathbf{a}_{ik}^T X \mathbf{a}_{ik} + \sum_{k=1}^n \mathbf{a}_{ik}^T Y \mathbf{a}_{ik} \right\}_{i=1}^m \\ &= \left\{ \sum_{k=1}^n \mathbf{a}_{ik}^T X \mathbf{a}_{ik} \right\}_{i=1}^m + \left\{ \sum_{k=1}^n \mathbf{a}_{ik}^T Y \mathbf{a}_{ik} \right\}_{i=1}^m \\ &= \mathcal{A}_1(X) + \mathcal{A}_1(Y).\end{aligned}\tag{B.1}$$

Sean  $X, Y \in \mathbb{B}^n$ , el operador  $\mathcal{A}_2$  es lineal ya que

$$\begin{aligned}\mathcal{A}_2(X + Y) &= \left\{ \sum_{j=1}^n \sum_{k \neq j} \mathbf{a}_{ik}^T (X + Y) \mathbf{a}_{ij} \right\}_{i=1}^m \\ &= \left\{ \sum_{j=1}^n \sum_{k \neq j} (\mathbf{a}_{ik}^T X \mathbf{a}_{ij} + \mathbf{a}_{ik}^T Y \mathbf{a}_{ij}) \right\}_{i=1}^m \\ &= \left\{ \sum_{j=1}^n \sum_{k \neq j} \mathbf{a}_{ik}^T X \mathbf{a}_{ij} + \sum_{j=1}^n \sum_{k \neq j} \mathbf{a}_{ik}^T Y \mathbf{a}_{ij} \right\}_{i=1}^m \\ &= \left\{ \sum_{j=1}^n \sum_{k \neq j} \mathbf{a}_{ik}^T X \mathbf{a}_{ij} \right\}_{i=1}^m + \left\{ \sum_{j=1}^n \sum_{k \neq j} \mathbf{a}_{ik}^T Y \mathbf{a}_{ij} \right\}_{i=1}^m \\ &= \mathcal{A}_2(X) + \mathcal{A}_2(Y)\end{aligned}\tag{B.2}$$

Por último, el operador  $\mathcal{L} : \mathbb{S}^n \rightarrow \mathbf{R}^m$  definido como

$$\mathcal{L} = \mathcal{A}_1 + \mathcal{A}_2 \quad (\text{B.3})$$

es lineal, ya que es la suma de dos operadores lineales.

## B.2. Operadores adjuntos

Para determinar el operador adjunto de un operador  $\mathcal{A} : X \rightarrow Y$ , recordemos que se debe satisfacer

$$\langle \mathcal{A}(\mathbf{D}), \mathbf{y} \rangle_Y = \langle \mathbf{D}, \mathcal{A}^\dagger(\mathbf{y}) \rangle_X \quad (\text{B.4})$$

donde el conjunto de partida en nuestro caso es  $\mathbf{R}^m$  y el de llegada es el conjunto de las matrices simétricas  $\mathbb{S}^n$ . El producto interno sobre  $\mathbf{R}^m$  esta dado por

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x} \quad (\text{B.5})$$

y el producto interno sobre las matrices simétricas viene dado por

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Tr}(\mathbf{X}\mathbf{Y}^*). \quad (\text{B.6})$$

Desarrollando el primer producto interno de (B.4) para el operador  $\mathcal{A}_1 : \mathbb{D}^n \rightarrow \mathbf{R}^m$ , se tiene

$$\begin{aligned} \langle \mathcal{A}_1(\mathbf{D}), \mathbf{y} \rangle_{\mathbf{R}^m} &= \mathbf{y}^* \left\{ \sum_{k=1}^n \mathbf{a}_{ik}^T \mathbf{D} \mathbf{a}_{ik} \right\}_{i=1}^m \\ &= \sum_{i=1}^m y_i \left( \sum_{k=1}^n \mathbf{a}_{ik}^T \mathbf{D} \mathbf{a}_{ik} \right) \end{aligned}$$

Para definir el operador adjunto recordemos que este debe ser un mapeo  $\mathcal{A}_1^\dagger : \mathbf{R}^m \rightarrow \mathbb{D}^n$ , entonces si se toma el operador igual a

$$\mathcal{A}_1^\dagger(\mathbf{y}) = \mathcal{T} \left( \sum_{i=1}^m y_i \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{a}_{ik}^T \right) \quad (\text{B.7})$$

donde  $\mathcal{T}$  diagonaliza la matriz. Y reemplazándolo en el segundo producto interno de (B.4)

$$\begin{aligned} \langle \mathbf{D}, \mathcal{A}_1^\dagger(\mathbf{y}) \rangle_{\mathbb{S}^n} &= \text{Tr}(\mathbf{D} \mathcal{A}_1^{\dagger T}(\mathbf{y})) \\ &= \text{Tr} \left( \mathbf{D} \cdot \mathcal{T} \left( \sum_{i=1}^m y_i \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{a}_{ik}^T \right) \right) \\ &= \text{Tr} \left( \mathcal{T} \left( \sum_{i=1}^m y_i \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{D} \mathbf{a}_{ik}^T \right) \right) \\ &= \sum_{i=1}^m y_i \sum_{k=1}^n \mathbf{a}_{ik}^T \mathbf{D} \mathbf{a}_{ik} \\ &= \langle \mathcal{A}_1(\mathbf{D}), \mathbf{y} \rangle_{\mathbf{R}^m}. \end{aligned} \quad (\text{B.8})$$

entonces el operador adjunto  $\mathcal{A}_1^\dagger$  satisface la condición del igualdad de (B.4).

Para obtener el operador adjunto del segundo operador  $\mathcal{A}_2 : \mathbb{B}^n \rightarrow \mathbf{R}^m$  el procedimiento es similar,

$$\begin{aligned} \langle \mathcal{A}_2(\mathbf{B}), \mathbf{y} \rangle_{\mathbf{R}^m} &= \mathbf{y}^* \left\{ \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik}^T \mathbf{B} \mathbf{a}_{ij} \right\}_{i=1}^m \\ &= \sum_{i=1}^m y_i \left( \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik}^T \mathbf{B} \mathbf{a}_{ij} \right). \end{aligned}$$

El operador adjunto debe ser un mapeo  $\mathcal{A}_2^\dagger : \mathbf{R}^m \rightarrow \mathbb{B}^n$ , entonces si se toma el operador igual a

$$\mathcal{A}_2^\dagger(\mathbf{y}) = \mathcal{N} \left( \sum_{i=1}^m y_i \left( \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik}^T \mathbf{a}_{ij} \right) \right), \quad (\text{B.9})$$

donde  $\mathcal{N}$  representa el operador que remueve la diagonal principal de la matriz. Reemplazándolo en el segundo producto interno de (B.4)

$$\begin{aligned} \langle \mathbf{B}, \mathcal{A}_2^\dagger(\mathbf{y}) \rangle_{\mathbb{B}^n} &= \text{Tr}(\mathbf{B} \mathcal{A}_2^{\dagger T}(\mathbf{y})) \\ &= \text{Tr} \left( \mathbf{B} \cdot \mathcal{N} \left( \sum_{i=1}^m y_i \left( \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik} \mathbf{a}_{ij}^T \right) \right) \right) \\ &= \text{Tr} \left( \mathcal{N} \left( \sum_{i=1}^m y_i \left( \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik} \mathbf{B} \mathbf{a}_{ij}^T \right) \right) \right) \\ &= \sum_{i=1}^m y_i \left( \sum_{j=1}^n \sum_{k \neq j}^n \mathbf{a}_{ik}^T \mathbf{B} \mathbf{a}_{ij} \right) \\ &= \langle \mathcal{A}_2(\mathbf{B}), \mathbf{y} \rangle_{\mathbf{R}^m} \end{aligned} \quad (\text{B.10})$$

por tanto el operador adjunto  $\mathcal{A}_2^\dagger$  satisface la condición del igualdad de (B.4).

# Bibliografía

- Abramovich, Y.I., Johnson, B.A., y Spencer, N.K. (2008a). Multivariate spectral reconstruction of stap covariance matrices: Toeplitz-block solution. In *2008 5th IEEE Sensor Array and Multichannel Signal Processing Workshop*, 229–233. doi:10.1109/SAM.2008.4606861.
- Abramovich, Y.I., Johnson, B.A., y Spencer, N.K. (2008b). Two-Dimensional Multivariate Parametric Models for Radar Applications #x2014;Part I: Maximum-Entropy Extensions for Toeplitz-Block Matrices. *IEEE Transactions on Signal Processing*, 56(11), 5509–5526. doi:10.1109/TSP.2008.929868.
- Afonso, M.V., Bioucas-Dias, J.M., y Figueiredo, M.A.T. (2011). An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems. *IEEE Transactions on Image Processing*, 20(3), 681–695. doi:10.1109/TIP.2010.2076294.
- Arce, G.R. (2004). *Nonlinear Signal Processing: A Statistical Approach*. Wiley-Interscience. Published: Hardcover.
- Argyriou, A., Evgeniou, T., y Pontil, M. (2008). Convex multi-task feature learning. *Machine Learning*, 73(3), 243–272. doi:10.1007/s10994-007-5040-8. URL <https://link.springer.com/article/10.1007/s10994-007-5040-8>.
- Arora, S., Hazan, E., y Kale, S. (2006). A Fast Random Sampling Algorithm for Sparsifying Matrices. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 272–279. Springer, Berlin, Heidelberg.
- Bahmani, S. y Romberg, J.K. (2015). Sketching for simultaneously sparse and low-rank covariance matrices. In *6th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015, Cancun, Mexico, December 13-16, 2015*, 357–360. doi:10.1109/CAMSAP.2015.7383810. URL <http://dx.doi.org/10.1109/CAMSAP.2015.7383810>.
- Bai, J. y Shi, S. (2011). Estimating High Dimensional Covariance Matrices and its Applications. *Annals of Economics and Finance*, 12(2), 199–215. URL <https://ideas.repec.org/a/cuf/journal/y2011v12i2p199-215.html>.

- Bickel, P.J. y Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6), 2577–2604. doi:10.1214/08-AOS600. URL <http://arxiv.org/abs/0901.3079>. ArXiv: 0901.3079.
- Bioucas-Dias, J.M., Cohen, D., y Eldar, Y.C. (2014). Covalsa: Covariance estimation from compressive measurements using alternating minimization. In *2014 22nd European Signal Processing Conference (EUSIPCO)*, 999–1003.
- Boutsidis, C., Drineas, P., y Magdon-Ismail, M. (2014). Near-Optimal Column-Based Matrix Reconstruction. *SIAM Journal on Computing*, 43(2), 687–717. doi:10.1137/12086755X. URL <http://epubs.siam.org/doi/abs/10.1137/12086755X>.
- Boutsidis, C., Mahoney, M., y Drineas, P. (2009). An Improved Approximation Algorithm for the Column Subset Selection Problem. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, Proceedings, 968–977. Society for Industrial and Applied Mathematics. URL <http://epubs.siam.org/doi/abs/10.1137/1.9781611973068.105>. DOI: 10.1137/1.9781611973068.105 DOI: 10.1137/1.9781611973068.105.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., y Eckstein, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.*, 3(1), 1–122. doi:10.1561/22000000016. URL <http://dx.doi.org/10.1561/22000000016>.
- Boyd, S. y Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK ; New York, 1 edition edition.
- Bryan, K. y Leise, T. (2013). Making Do with Less: An Introduction to Compressed Sensing. *SIAM Review*, 55(3), 547–566. doi:10.1137/110837681. URL <http://epubs.siam.org/doi/abs/10.1137/110837681>.
- Cai, T.T. y Yuan, M. (2012). Adaptive covariance matrix estimation through block thresholding. *The Annals of Statistics*, 40(4), 2014–2042. doi:10.1214/12-AOS999. URL <http://projecteuclid.org/euclid.aos/1351602535>.
- Cai, T. y Liu, W. (2011). Adaptive Thresholding for Sparse Covariance Matrix Estimation. *arXiv:1102.2237 [stat]*. URL <http://arxiv.org/abs/1102.2237>. ArXiv: 1102.2237.
- Candes, E.J., Eldar, Y., Strohmer, T., y Voroninski, V. (2011). Phase Retrieval via Matrix Completion. *arXiv:1109.0573 [cs, math]*. URL <http://arxiv.org/abs/1109.0573>. ArXiv: 1109.0573.
- Candès, E.J. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9), 589–592. doi:10.1016/j.crma.2008.03.014. URL <http://www.sciencedirect.com/science/article/pii/S1631073X08000964>.

- Candès, E.J., Strohmer, T., y Voroninski, V. (2013). PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Communications on Pure and Applied Mathematics*, 66(8), 1241–1274. doi:10.1002/cpa.21432. URL <http://onlinelibrary.wiley.com/doi/10.1002/cpa.21432/abstract>.
- Cevher, V., Becker, S., y Schmidt, M. (2014). Convex Optimization for Big Data: Scalable, randomized, and parallel algorithms for big data analytics. *IEEE Signal Processing Magazine*, 31(5), 32–43. doi:10.1109/MSP.2014.2329397.
- Chen, Y., Chi, Y., y Goldsmith, A.J. (2015). Exact and Stable Covariance Estimation From Quadratic Sampling via Convex Programming. *IEEE Transactions on Information Theory*, 61(7), 4034–4059. doi:10.1109/TIT.2015.2429594.
- Chi, Y. (2016). Kronecker covariance sketching for spatial-temporal data. In *2016 24th European Signal Processing Conference (EUSIPCO)*, 316–320. doi:10.1109/EUSIPCO.2016.7760261.
- Combettes, P.L. y Pesquet, J.C. (2011). Proximal Splitting Methods in Signal Processing. In H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, y H. Wolkowicz (eds.), *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, 185–212. Springer New York, New York, NY. URL [https://doi.org/10.1007/978-1-4419-9569-8\\_10](https://doi.org/10.1007/978-1-4419-9569-8_10). DOI: 10.1007/978-1-4419-9569-8\_10.
- Dasarathy, G., Shah, P., y Baraniuk, R.G. (2017). Sketched covariance testing: A compression-statistics tradeoff. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, 676–680. doi:10.1109/ACSSC.2017.8335428.
- Dasarathy, G., Shah, P., Bhaskar, B., y Nowak, R. (2015). Sketching Sparse Matrices, Covariances, and Graphs via Tensor Products. *Information Theory, IEEE Transactions on*, 61(3), 1373–1388. doi:10.1109/TIT.2015.2391251.
- Davenport, M.A. y Romberg, J. (2016). An Overview of Low-Rank Matrix Recovery From Incomplete Observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4), 608–622. doi:10.1109/JSTSP.2016.2539100.
- Davies, M.E. y Eldar, Y.C. (2012). Rank Awareness in Joint Sparse Recovery. *IEEE Transactions on Information Theory*, 58(2), 1135–1146. doi:10.1109/TIT.2011.2173722.
- Drineas, P. y Kannan, R. (2003). Pass Efficient Algorithms for Approximating Large Matrices. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '03, 223–232. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. URL <http://dl.acm.org/citation.cfm?id=644108.644147>.
- Drineas, P. y Zouzias, A. (2011). A note on element-wise matrix sparsification via a matrix-valued Bernstein inequality. *Information Processing Letters*, 111(8), 385–389. doi:10.1016/j.ipl.2011.01.010. URL <http://www.sciencedirect.com/science/article/pii/S0020019011000196>.

- Duffin, R.J. y Schaeffer, A.C. (1952). A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72, 341–366.
- Foucart, S. y Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, New York, edición: 2013 edition.
- Gandhi, P.P. y Kassam, S.A. (1991). Design and performance of combination filters for signal restoration. *IEEE Transactions on Signal Processing*, 39(7), 1524–1540. doi: 10.1109/78.134392.
- Ghashami, M., Liberty, E., Phillips, J., y Woodruff, D. (2016). Frequent Directions: Simple and Deterministic Matrix Sketching. *SIAM Journal on Computing*, 45(5), 1762–1792. doi:10.1137/15M1009718. URL <http://epubs.siam.org/doi/abs/10.1137/15M1009718>.
- Gilbert, A. y Indyk, P. (2010). Sparse Recovery Using Sparse Matrices. *Proceedings of the IEEE*, 98(6), 937–947. doi:10.1109/JPROC.2010.2045092.
- Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., y Stanley, H.E. (1992). The MIT-BIH Long Term Database. DOI: 10.13026/C2KS3F.
- Grant, M. y Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>.
- Gray, R.M. (2005). Toeplitz and Circulant Matrices: A Review. *Commun. Inf. Theory*, 2(3), 155–239. doi:10.1561/0100000006. URL <http://dx.doi.org/10.1561/0100000006>.
- Guan, Y., Lu, T., Zhang, Y., Wang, B., Li, X., y Xiong, Z. (2016). Efficient low-rank supported extreme learning machine for robust face recognition. In *2016 Visual Communications and Image Processing (VCIP)*, 1–4. doi:10.1109/VCIP.2016.7805436.
- Guerci, J.R. y Baranoski, E.J. (2006). Knowledge-aided adaptive radar at DARPA: an overview. *IEEE Signal Processing Magazine*, 23(1), 41–50. doi:10.1109/MSP.2006.1593336.
- Gutiérrez-Gutiérrez, J. y Crespo, P.M. (2012). Block Toeplitz Matrices: Asymptotic Results and Applications. *Foundations and Trends® in Communications and Information Theory*, 8(3), 179–257. doi:10.1561/0100000066. URL <http://www.nowpublishers.com/article/Details/CIT-066>.
- Hong, S., Wan, X., Cheng, F., y Ke, H. (2015). Covariance differencing-based matrix decomposition for coherent sources localisation in bi-static multiple-input multiple-output radar. *Sonar Navigation IET Radar*, 9(5), 540–549. doi: 10.1049/iet-rsn.2014.0193.

- Horn, R.A. y Johnson, C.R. (1994). *Topics in Matrix Analysis*. Cambridge University Press. Google-Books-ID: LeuNXB2bl5EC.
- Hoyos, J.P., Carrillo, R.E., Pazos, S., y Jojoa, P. (2017a). Correlation sketching for ordered data. In *2017 IEEE Latin-American Conference on Communications*.
- Hoyos, J.P. y Jojoa, P. (2017a). A review of low rank and sparse matrices estimators. *aceptado en Información Tecnológica*.
- Hoyos, J.P. y Jojoa, P. (2017b). A review of toeplitz and block toeplitz estimators. *a enviar a DYNA*.
- Hoyos, J.P. y Jojoa, P. (2017c). Transformación qsp en el dominio wavelet. In *Procesamiento Cuántico de Señales en la Teoría Wavelets*, chapter 5, 65–82. En edición por Editorial Universidad del Cauca, Popayan.
- Hoyos, J.P., Pazos, Carrillo, R., y E.Jojoa, P. (2018). Scalable splitting algorithm for correlation matrix estimation of ordered data. ——— *por enviar a IEEE Transactions on Signal Processing*.
- Hoyos, J.P., Pazos, S., Jojoa, P., y Arce, G. (2017b). Estimación de la matriz de correlación de datos ordenados utilizando sketches. In *2017 IEEE International XVII Workshop on Information Processing and Control*.
- Komodakis, N. y Pesquet, J.C. (2015). Playing with Duality: An overview of recent primal?dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6), 31–54. doi:10.1109/MSP.2014.2377273.
- Krim, H. y Viberg, M. (1996). Two decades of array signal processing research: the parametric approach. *IEEE Signal Processing Magazine*, 13(4), 67–94. doi:10.1109/79.526899.
- Krishna, T.G., Singh, I., y Giridhar, K. (2013). Null-space of block convolution matrix. In *2013 National Conference on Communications (NCC)*, 1–5. doi:10.1109/NCC.2013.6487951.
- Kwan, C. (2011). An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration. *Spreadsheets in Education (eJSiE)*, 4(3). URL <http://epublications.bond.edu.au/ejsie/vol4/iss3/6>.
- Ledoit, O. y Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5), 603–621. doi:10.1016/S0927-5398(03)00007-0. URL <http://www.sciencedirect.com/science/article/pii/S0927539803000070>.
- Ledoit, O. y Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411. doi:10.1016/S0047-259X(03)00096-4. URL <http://www.sciencedirect.com/science/article/pii/S0047259X03000964>.

- Ledoit, O. y Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2), 1024–1060. doi:10.1214/12-AOS989. URL <http://arxiv.org/abs/1207.5322>. ArXiv: 1207.5322.
- Leus, G. y Tian, Z. (2011). Recovering second-order statistics from compressive measurements. In *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 337–340. doi:10.1109/CAMSAP.2011.6136019.
- Liu, Z. y Vandenberghe, L. (2009). Interior-Point Method for Nuclear Norm Approximation with Application to System Identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3), 1235–1256. doi:10.1137/090755436. URL <http://epubs.siam.org/doi/abs/10.1137/090755436>.
- Omar, S.M. y Slock, D.T.M. (2008). Singular block Toeplitz matrix approximation and application to multi-microphone speech dereverberation. In *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 52–57. doi:10.1109/MMSP.2008.4665048.
- Onose, A., Carrillo, R.E., Repetti, A., McEwen, J.D., Thiran, J.P., Pesquet, J.C., y Wiaux, Y. (2016). Scalable splitting algorithms for big-data interferometric imaging in the SKA era. *Monthly Notices of the Royal Astronomical Society*, 462(4), 4314–4335. doi:10.1093/mnras/stw1859. URL <https://academic.oup.com/mnras/article/462/4/4314/2589458>.
- Oudin, M. y Delmas, J.P. (2008). Asymptotic generalized eigenvalue distribution of Toeplitz block Toeplitz matrices. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3309–3312. doi:10.1109/ICASSP.2008.4518358.
- Palmieri, F. y Boncelet, C.G. (1989). L1-filters—a new class of order statistic filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5), 691–701. doi:10.1109/29.17561.
- Parikh, N. y Boyd, S. (2014). Proximal Algorithms. *Foundations and Trends® in Optimization*, 1(3), 127–239. doi:10.1561/2400000003. URL <http://dx.doi.org/10.1561/2400000003>.
- Pedarsani, R., Lee, K., y Ramchandran, K. (2015). Sparse covariance estimation based on sparse-graph codes. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 612–619. doi:10.1109/ALLERTON.2015.7447061.
- Pesquet, J.C. y Repetti, A. (2014). A Class of Randomized Primal-Dual Algorithms for Distributed Optimization. *arXiv:1406.6404 [math]*. URL <http://arxiv.org/abs/1406.6404>. ArXiv: 1406.6404.
- Qiao, H. y Pal, P. (2015). Generalized Nested Sampling for Compressing Low Rank Toeplitz Matrices. *IEEE Signal Processing Letters*, 22(11), 1844–1848. doi:10.1109/LSP.2015.2438066.

- Qiao, H. y Pal, P. (2016). Finite sample analysis of covariance compression using structured samplers. In *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 1–5. doi:10.1109/SAM.2016.7569732.
- Qin, S., Zhang, Y.D., Amin, M.G., y Zoubir, A.M. (2017). Generalized Coprime Sampling of Toeplitz Matrices for Spectrum Estimation. *IEEE Transactions on Signal Processing*, 65(1), 81–94. doi:10.1109/TSP.2016.2614799.
- Romero, D., Ariananda, D.D., Tian, Z., y Leus, G. (2016). Compressive Covariance Sensing: Structure-based compressive sensing beyond sparsity. *IEEE Signal Processing Magazine*, 33(1), 78–93. doi:10.1109/MSP.2015.2486805.
- Romero, D. y Leus, G. (2013a). Compressive covariance sampling. In *2013 Information Theory and Applications Workshop (ITA)*, 1–8. doi:10.1109/ITA.2013.6502949.
- Romero, D. y Leus, G. (2013b). Wideband Spectrum Sensing From Compressed Measurements Using Spectral Prior Information. *IEEE Transactions on Signal Processing*, 61(24), 6232–6246. doi:10.1109/TSP.2013.2283473.
- Romero, D., López-Valcarce, R., y Leus, G. (2015). Compression Limits for Random Vectors with Linearly Parameterized Second-Order Statistics. *IEEE Transactions on Information Theory*, 61(3), 1410–1425. doi:10.1109/TIT.2015.2394784.
- Rothman, A.J., Levina, E., y Zhu, J. (2009). Generalized Thresholding of Large Covariance Matrices. *Journal of the American Statistical Association*, 104(485), 177–186. URL <http://www.jstor.org/stable/40591909>.
- Sahinoglou, H. y Cabrera, S.D. (1991). On phase retrieval of finite-length sequences using the initial time sample. *IEEE Transactions on Circuits and Systems*, 38(8), 954–958. doi:10.1109/31.85639.
- Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x. URL <http://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1948.tb01338.x/abstract>.
- Sun, H., Nallanathan, A., Wang, C.X., y Chen, Y. (2013). Wideband spectrum sensing for cognitive radio networks: a survey. *IEEE Wireless Communications*, 20(2), 74–81. doi:10.1109/MWC.2013.6507397.
- Tang, J., Golbabaee, M., y Davies, M. (2017). Exploiting the structure via sketched gradient algorithms. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 1305–1309. doi:10.1109/GlobalSIP.2017.8309172.
- Traganitis, P.A. y Giannakis, G.B. (2017). Distributed sketched subspace clustering for large-scale datasets. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 1–5. doi:10.1109/CAMSAP.2017.8313130.

- Tropp, J.A. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2), 1–230. doi:10.1561/22000000048. URL <http://dx.doi.org/10.1561/22000000048>.
- Tsiligkaridis, T. y Hero, A.O. (2013). Covariance Estimation in High Dimensions Via Kronecker Product Expansions. *IEEE Transactions on Signal Processing*, 61(21), 5347–5360. doi:10.1109/TSP.2013.2279355.
- Unser, M. y Tafti, P. (2014). *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, Cambridge, UK. 367 p.
- Vershynin, R. (2012). How Close is the Sample Covariance Matrix to the Actual Covariance Matrix? *Journal of Theoretical Probability*, 25(3), 655–686. doi:10.1007/s10959-010-0338-z. URL <https://link.springer.com/article/10.1007/s10959-010-0338-z>.
- Walden, A.T. y Schneider-Luftman, D. (2015). Random Matrix Derived Shrinkage of Spectral Precision Matrices. *IEEE Transactions on Signal Processing*, 63(17), 4689–4699. doi:10.1109/TSP.2015.2443726.
- Wang, M., Li, K., Wu, F., Lai, Y.K., y Yang, J. (2016). 3-D motion recovery via low rank matrix analysis. In *2016 Visual Communications and Image Processing (VCIP)*, 1–4. doi:10.1109/VCIP.2016.7805473.
- Yu, B., Ding, Y., Huang, X., y Wu, B. (2016). Corrupted old film sequences restoration using improved PatchMatch and low-rank matrix recovery. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, 295–300. doi:10.1109/ICALIP.2016.7846589.