

DISEÑO DE ARREGLOS Y PROCESAMIENTO DE SEÑALES EN SISTEMAS DE RADAR

Tesis de Doctorado

SEBASTIAN PAZOS

Presentada ante la Facultad de Ingeniería de la Universidad Nacional de La Plata
como requisito para la obtención del grado académico de

DOCTOR EN INGENIERIA

Dirección de tesis:

Director: Dr. Martín Hurtado

Codirector: Carlos H. Muravchik

Jurado de tesis:

Dr. Leonardo Rey Vega

Dr. Danilo Velis

Dr. Isidoro Vaquila

Fecha de defensa oral y pública: 27/03/2015

Resumen

El radar es un sistema de sensado remoto que utiliza técnicas de procesamiento estadístico de señales para obtener información de interés de la señal recibida. Una de las dificultades que se encuentra en los radares es que la señal recibida suele ser débil debido a la gran distancia entre el radar y los objetos. Por lo cual, el sistema debe ser capaz de operar en condiciones de baja relación de señal a ruido. Otro problema importante consiste en que la señal de interés suele encontrarse oscurecida por la reflexiones producidas por el ambiente y por la posible existencia de señales de interferencia. Con el avance de la tecnología, la tendencia en la construcción de radares es reemplazar los sistemas mecánicos de orientación del haz por sistemas electrónicos basados en arreglos de antenas. La cantidad de antenas del arreglo y su geometría determinan la resolución angular y la capacidad de filtrar espacialmente señales no deseadas. Por otro lado, la complejidad y el costo de un arreglo de antenas crece con el número de elementos. Por lo tanto, un diseño debe considerar en maximizar el desempeño del arreglo sin incrementar el número de sensores.

Para afrontar estos desafíos, buscando maximizar el desempeño del sistema de radar, en esta tesis se analizan sus diferentes componentes y se desarrollan métodos para el diseño de arreglos de sensores y procesamiento de señales de radar en el marco de formulaciones ralas.

En primer lugar se considera el diseño de arreglos robustos, definiendo qué tipo de configuración espacial de los elementos del arreglo de sensores es óptima para maximizar el desempeño en la estimación de señales y minimizar el número de elementos. Se analizan diferentes configuraciones clásicas y se optimizan mediante algoritmos genéticos en base a restricciones relacionadas con formulaciones ralas. En

segundo lugar se realiza un análisis del desempeño de clases de algoritmos basados en formulaciones ralas para arreglos de sensores. Con él se busca saber qué tipo de algoritmos resultan eficaces para permitir una buena estimación tanto de la dirección de arribo de la señal de interés como de la información transmitida por ella en un escenario con fuertes interferencias. Finalmente, debido a la gran cantidad de datos que manejan estos sistemas de tiempo real, se desean optimizar los tiempos de cómputo de los algoritmos de estimación aprovechando la rareza intrínseca de las señales de radar. Para ello se desarrollan métodos basados en modelos ralos para reducir significativamente, respecto de sistemas convencionales, la cantidad necesaria de datos y adicionalmente reducir el error de estimación.

Agradecimientos

A las instituciones que financiaron la realización de esta tesis: el Consejo Nacional de Investigaciones Científicas y Técnicas y la Universidad Nacional de La Plata.

Al Instituto LEICI por proporcionar el lugar y material de trabajo.

A todo el grupo de Procesamiento Estadístico de Señales del LEICI, Mariano, Juan Ignacio, Juan Pablo, Leandro, Nicolás y Martín, por tanto intercambio de charlas e ideas.

A mis directores Martín Hurtado y Carlos Muravchik, a quienes no sólo agradezco la dirección de mi doctorado, sino también la confianza, el apoyo y la amistad brindadas desde un principio.

A mis hijos nacidos durante este período, Felipe y Sofía.

Y por sobre todo a Victoria.

Índice general

1. Introducción	1
1.1. Resultados preexistentes	5
1.2. Contribuciones originales	7
1.3. Organización de la tesis	9
2. Algoritmos basados en modelos raros	11
2.1. Introducción	11
2.2. Clases de algoritmos	13
2.2.1. Algoritmos de relajación	13
2.2.2. Algoritmos greedy	14
2.2.3. Algoritmos probabilísticos	15
2.3. Múltiples vectores de mediciones	16
2.4. Algoritmo de estimación ESBL	17
2.4.1. Modelo	17
2.4.2. Algoritmo	18
2.5. Modelo raro con fusión de mediciones	23
2.5.1. Introducción	23
2.5.2. Modelo	24
2.5.3. Desempeño	24
2.6. Discusión y conclusiones	32
3. Arreglos no uniformes lineales	33
3.1. Introducción	33

3.2. Modelo	34
3.3. Configuraciones de arreglos	36
3.4. Estimador perfecto	37
3.5. Arreglos optimizados	39
3.5.1. Arreglo optimizado 1. Sin restricción	40
3.5.2. Arreglo optimizado 2. Distancia mínima restringida	40
3.5.3. Arreglo optimizado 3. Distancia mínima y lóbulos laterales restringidos	41
3.6. Desempeño de los arreglos	41
3.7. Discusión y conclusiones	53
4. Estimación de señales en modelos ralos con interferencias	55
4.1. Introducción	55
4.2. Arreglos de sensores	56
4.3. Representación rala	57
4.4. Análisis de desempeño	58
4.5. Discusión y conclusiones	65
5. Proyección óptima en modelos ralos con interferencia	67
5.1. Introducción	67
5.2. Formulación del problema	68
5.2.1. Modelo	68
5.2.2. Presentación del problema	70
5.3. Solución	70
5.3.1. Función de costo	70
5.3.2. Solución óptima restringida	73
5.3.3. Selección automática de E	74
5.4. Análisis de desempeño	76
5.4.1. Valor fijo de E	76
5.4.2. Selección automática de E	78
5.4.3. Robustez	81
5.5. Datos reales de radar	84

5.5.1. Objetivo estático	84
5.5.2. Objetivo sintético dinámico	88
5.6. Discusión y conclusiones	88
6. Conclusiones	91
Apéndice A. Algoritmo EM	95
Apéndice B. Algoritmo ESBL	97
B.1. Block trace	97
B.2. Obtención de los estimadores de los parámetros	97
B.3. Obtención del test de detección	100
Apéndice C. Radar polarimétrico	103
C.1. Introducción	103
C.2. Modelado	104
C.3. Representación rala	106
Apéndice D. Demostración del óptimo λ_i	109

Lista de abreviaturas

ADORE	Automatic double relaxation
BCS	Bayesian compressive sensing
CoSaMP	Compressive sampling matching pursuit
CS	Compressive sensing
EM	Expectation-Maximization
ESBL	Enhanced sparse Bayesian learning
MSE	Error cuadrático medio (Mean square error)
FOCUSS	Focal undetermined system solver
MC	Monte Carlo
MFOCUSS	MMV-FOCUSS
MIMO	Multiple input multiple output
MLE	Estimador de máxima verosimilitud
MMV	Multiple measurement vectors
MOMP	MMV-Orthogonal matching pursuit
OMP	Orthogonal matching pursuit
RIP	Restricted Isometry Property
RVM	Relevance vector machine
SBL	Sparse Bayesian learning
SIR	Relación señal a interferencia
SNR	Relación señal a ruido
StOMP	Stagewise orthogonal matching pursuit

Capítulo 1

Introducción

En sus inicios la palabra radar se gestó como un acrónimo, RADAR, *detección y localización por radio* (*radio detection and ranging*). Hoy en día la tecnología es tan ubicua que es un sustantivo común de la lengua española. La historia del radar se extiende a los primeros días de la teoría electromagnética moderna. En 1886 Hertz demostró la reflexión de las ondas electromagnéticas, y en 1900 Tesla describió el concepto de usarlas para detección y medición de velocidad. Finalmente, en 1904 el ingeniero alemán Hülsmeyer patentó su dispositivo “telemobiloscope” para evitar colisiones entre barcos basado en reflexión de ondas de radio. En los años siguientes el desarrollo de sistemas de radar fue impulsado principalmente por necesidades militares incluyendo vigilancia, navegación y guiado de misiles para vehículos de tierra, aire y agua. Hoy en día los sistemas de radar cubren un mayor rango de aplicaciones. Una de las más comunes es la del radar de tránsito para verificar velocidades máximas. Otra aplicación es la de radar meteorológico usado por los noticieros cuya extensión son sistemas mucho más sofisticados para ser usados en predicción y monitoreo del tiempo a gran escala e investigación atmosférica. Otra aplicación que afecta a un gran número de gente son los sistemas de control de tráfico aéreo usadas para guiar aviones comerciales en ruta y en vecindades de aeropuertos. La aviación también utiliza radares para el control de altitud y evitar clima severo y pronto podrá ser reutilizado para asistir el aterrizaje con baja visibilidad. El radar es comúnmente utilizado en barcos para evitar colisiones y detección de boyas, y hoy en

día esta empezando a ser usado en automóviles con el mismo objetivo. Finalmente, radares espaciales y aerotransportados son herramientas importantes para relevar mapas de la topología terrestre y características ambientales como las condiciones del agua y de los hielos, condiciones forestales, utilización del suelo y polución.

El radar es un sistema de sensado remoto que utiliza técnicas de procesamiento estadístico de señales para obtener información de interés de la señal recibida [1]. Los radares convencionales son sistemas activos que operan transmitiendo energía en forma de ondas electromagnéticas y recibiendo las señales reflejadas por la región y el objeto iluminado. Las reflexiones del ambiente, usualmente referidas como *clutter* [2], son consideradas como fuentes de interferencia que reducen el desempeño del sistema. La mayoría de los radares pueden clasificarse como de detección, de seguimiento o de relevamiento de imágenes. En este trabajo el énfasis está puesto en las técnicas de adquisición de señales y de reducción de interferencias necesarias para un buen desempeño de estas tareas.

El problema fundamental del radar es la detección de un objeto o un fenómeno físico. Esto requiere determinar si la señal de salida del receptor en un determinado instante representa el eco de un objetivo o sólo ruido. La calidad del radar puede cuantificarse mediante diferentes figuras de mérito, dependiendo de la función a ser considerada. Al analizar el desempeño en la detección los parámetros fundamentales son la probabilidad de detección P_D y la probabilidad de falsa alarma P_{FA} . Si otros parámetros del sistema se mantienen fijos, aumentar la P_D requiere aceptar una mayor P_{FA} , y las combinaciones viables están determinadas por estadísticos de la señal y la interferencia, especialmente la relación señal a interferencia (signal to interference ratio, SIR). En el caso de haber múltiples objetivos en el campo de visión del radar, deben considerarse la resolución y los lóbulos laterales para el desempeño del radar en detección. Esto se debe a que si los objetivos están muy cercanos pueden enmascarse y presentarse al radar como uno sólo. La resolución en rango y los lóbulos laterales en rango están determinados por la forma de onda y en ángulo por el patrón de la antena.

El objetivo principal del procesamiento de señales en radar es mejorar las figuras de mérito mencionadas. Este procesamiento tiene varias cualidades únicas que se

diferencian de otros campos de procesamiento de señales. Los radares utilizados en este trabajo son coherentes, por lo cual las señales recibidas, una vez demoduladas a banda base serán complejas en vez de reales. Por otro lado las señales de radar poseen un alto rango dinámico desde decenas de decibeles a casos extremos de 100 dB. Por lo tanto, es usual el uso de control de ganancia y el control de lóbulos laterales es crítico para evitar el enmascaramiento de señales débiles. Por otra parte, el SIR en general es muy bajo, incluso debajo de los 0 dB.

En muchas aplicaciones, características usuales de la señal medida, como ser tiempo, frecuencia y orientación no son suficientes para discriminar al objetivo del clutter. Por ejemplo, cuando ambos son estáticos o se mueven a una velocidad similar, el desplazamiento Doppler de frecuencia no puede ser usado para detectar el objetivo. Adicionalmente, en escenarios de propagación de ángulo rasante, la reflexión de la tierra no puede ser removida por filtrado espacial. La polarización de la señal provee información adicional que puede ser explotada para mejorar la detección del objetivo y el desempeño en el seguimiento.

Las antenas poseen un rol importante en determinar la sensibilidad y la resolución angular del radar. Se utiliza un amplio rango de tipos de antena. Los más comunes son reflectores parabólicos, antenas lentes y arreglos de antenas en fase (Phased Arrays). Éstos últimos, utilizados en esta tesis, están compuestos por una colección de antenas individuales denominados elementos del arreglo. Estos elementos son típicamente dipolos idénticos u otro tipo simple de antena con patrones de radiación anchos. En general, los radares convencionales utilizan un arreglo de sensores lineal cuyos elementos se encuentran separados a una distancia constante, denominado arreglo uniforme lineal (ULA). Como la complejidad y el costo de un arreglo de antenas crece con el número de elementos, un diseño debe considerar en maximizar el desempeño del arreglo sin incrementar el número de sensores.

El procesamiento de señales de radar requiere una representación de los datos medidos para poder procesarlos y aplicar algoritmos de estimación necesarios para extraer la información deseada. Una representación posible es mediante modelos raros, donde se trabaja sobre un espacio definido por un diccionario sobrecompleto de forma tal que sólo unos pocos componentes de la señal son distintos de cero. Adicio-

nalmente puede considerarse que las variables son aleatorias obteniendo un modelo ralo estocástico. Si se considera un modelo lineal, éste está dado por $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{w}$, donde \mathbf{y} son las mediciones, \mathbf{X} es la matriz de sensado o diccionario, \mathbf{b} es la señal de interés rala y \mathbf{w} es ruido. El objetivo del algoritmo de estimación es hallar \mathbf{b} dado el modelo y las mediciones, denominado problema inverso.

Buscar una solución rala de un modelo lineal es un problema de optimización no convexo ya que está restringido a minimizar el número de componentes no-nulos de la señal, llamado norma ℓ_0 . Un procedimiento común para evitar esta dificultad consiste en utilizar la norma ℓ_p para $p \in (0, 1]$ como una medida alternativa de la rareza. Ejemplos de este tipo de algoritmo son: *focal undetermined system solver* (FOCUSS) [3], *basis pursuit denoising* (BPDN) [4], *Dantzig selector* [5], y ℓ_1 -SVD [6]. Otra clase de algoritmos son los que utilizan búsqueda voraz (*greedy pursuit*) aproximando iterativamente el soporte de la solución, como ser *orthogonal matching pursuit* (OMP) [7], y sus variantes [8,9]. Adicionalmente, existe otra clase de métodos para resolver el problema inverso ralo que utilizan modelos y herramientas estadísticas, como ser *sparse bayesian learning* (SBL) [10]. Una variante de este último se denomina *enhanced sparse bayesian learning* (ESBL) [11]. Cada iteración de este algoritmo consta de dos pasos, en el primero se utiliza el algoritmo de esperanza-maximización (EM) para obtener los parámetros del modelo, y como esta solución no es rala, en el segundo paso se introduce un test de decisión que, aplicado en forma iterativa, garantiza la rareza eliminando gradualmente los componentes estadísticamente no significativos de la solución.

En esta tesis se estudian distintos aspectos que influyen en el desempeño de los sistemas de radar utilizando modelos ralos para la representación de las señales involucradas. En primer lugar se estudia el efecto de la configuración espacial de los elementos del arreglo de antenas. En segundo lugar se estudia desempeño de distintos algoritmos basados en modelos ralos para la determinación de la dirección de arribo de señales de interés en un ambiente de alta interferencia. Finalmente se diseña un método para comprimir las mediciones en un sistema de radar y adicionalmente mejorar su desempeño frente a interferencias.

Cabe mencionar que si bien este proyecto se enfoca en el problema de radar, los

métodos de análisis de arreglos que proponemos desarrollar pueden ser utilizados en otras áreas, tales como muestreo de señales o análisis espectral. De la misma forma, los algoritmos de procesamiento de datos pueden extenderse para ser aplicados en imágenes biomédicas, exploración sísmica y ensayos no-destructivos de materiales.

1.1. Resultados preexistentes

Una configuración común para un arreglo de sensores es el arreglo lineal uniforme, donde la distancia inter-elemento es constante. Para satisfacer la condición de muestreo de Nyquist, la distancia máxima entre elementos no debe ser mayor a media longitud de onda de la portadora λ . Esto se traduce en una apertura máxima del arreglo para un número de sensores fijo, que limita su capacidad de localización y resolución de fuentes de señal. Para alcanzar mayores aperturas y mejor resoluciones, se puede utilizar un arreglo no uniforme, donde la distancia inter-elemento es variable. Una posible estrategia para el diseño del arreglo no uniforme es distribuir los sensores de forma aleatoria. Este problema fue analizado desde un punto de vista probabilístico para determinar diferentes propiedades del arreglo como ser niveles de lóbulos laterales, ancho de haz, y ganancia [12]. Recientemente se propuso la aplicación de modelos ralos y técnicas de compressive sensing para el estudio de arreglos aleatorios [13, 14] y para el diseño de arreglos no uniformes [15]. Este problema también está relacionado con la estimación espectral a partir de datos muestreados no uniformemente [16].

El objetivo de los modelos ralos es representar una señal como una combinación lineal de funciones básicas llamadas átomos, correspondientes a un diccionario [4]. En el caso específico de arreglos de sensores, este diccionario puede estar dado por la respuesta del arreglo para cada posible ángulo de arribo [6]. Asumiendo que existen pocas fuentes de señal, los datos de salida del arreglo posee una representación rala. La teoría de compressive sensing afirma que, bajo estas condiciones, el número de sensores y datos necesarios para resolver el problema inverso puede ser reducido significativamente sin afectar el desempeño del sistema [17, 18]. Estas técnicas ya han sido usadas en diferentes áreas de procesamiento de señal, incluyendo procesamiento

de imágenes, bioingeniería, comunicaciones y sensado remoto.

Debido a la rareza de la señal de interés \mathbf{b} , las mediciones pueden comprimirse sin perder información. Una forma de reducir la cantidad de datos a procesar en este modelo es proyectando las mediciones sobre un espacio de dimensión reducida que contenga la información necesaria para reconstruir la señal de interés. Esta operación se representa mediante una matriz de proyección \mathbf{A} . Trabajos anteriores analizaron la construcción de esta matriz de proyección a través de diferentes estrategias y en diferentes escenarios. En el caso de reconstrucción sin ruido, [19] introdujo un algoritmo iterativo basado en *shrinkage*, denominado proyección optimizada, que minimiza una medida promedio de la coherencia mutua del diccionario efectivo \mathbf{AX} . El artículo [20] construyó la matriz de proyección basado en *Multidimensional Scaling* que fuerza a las distancias entre pares de átomos del diccionario efectivo \mathbf{AX} a ser lo más parecidas posible a los átomos del diccionario original \mathbf{X} . Similarmente, en [21] el objetivo del algoritmo es que la matriz de Gram $(\mathbf{AX})^H(\mathbf{XA})$ sea lo más cercana posible a la matriz identidad. Construcciones más específicas fueron introducidas en [22] basadas en teoría de codificación sobre grafos ofreciendo garantías de reconstrucción; y en [23] basadas en códigos BCH ofreciendo una construcción de matrices de muestreo binarias.

Extendiendo el escenario para incluir fuentes de interferencia y ruido, [24] analizó la aplicación de radar MIMO y propuso dos métodos de construcción de la matriz de proyección, el primero minimiza una combinación lineal de la coherencia de la matriz de sensado y de la inversa de la SIR; y el segundo fuerza una estructura en la matriz de proyección para mejorar la SIR manteniendo una coherencia similar a la de matrices Gaussianas. Otros trabajos que abordaron aplicaciones específicas impusieron estructura adicional a la matriz de proyección resultando en rareza a bloques. En [25] se usó un marco ajustado equiangular (equiangular tight frame) en lugar de la matriz de identidad para aproximar la matriz de Gram del diccionario equivalente; y [26] minimizó una suma pesada de la coherencia inter-bloque y sub-bloque. Adicionalmente, [27] extendió los resultados de [21] para señales pertenecientes a un modelo de mezclas Gaussianas.

Más relacionado con nuestro trabajo es el realizado en [28] que utilizó un modelo

similar, cubriendo señal de interés, interferencia y ruido pero usando un proceso de dos pasos. Primero aplicando una proyección sobre el espacio ortogonal a la interferencia y luego resolviendo el problema inverso con una matriz Gaussiana aleatoria mediante compressive sensing. Similarmente, [29] minimiza la información mutua entre la señal y las mediciones pero aplica la matriz de proyección solamente a la señal y no la interferencia.

1.2. Contribuciones originales

Las contribuciones originales de esta tesis están relacionados con los distintos componentes del sistema de radar. Se enfocan en analizar la influencia del arreglo de sensores y el tipo de algoritmos y métodos utilizados en el procesamiento de las señales recibidas.

En primer lugar se realiza una modificación del modelo ralo para mejorar su desempeño. Usualmente, al tener un grupo de mediciones, se utiliza para todas ellas la misma matriz de sensado. Sin embargo, si durante esas mediciones la señal de interés a estimar mantiene sus valores (intervalo de coherencia), el método propone utilizar una matriz de sensado distinta para cada instantánea. Este cambio conjuntamente con el algoritmo de estimación basado en modelos ralos permite disminuir el error en la estimación de los parámetros. Los resultados presentados en esta tesis aún no han sido publicados.

En segundo lugar, con el fin de analizar la configuración óptima de los elementos de un arreglo de sensores se proponen diferentes diseños. En particular se estudian tres tipos de configuraciones, un arreglo estructurado o virtual con ubicaciones prefijadas de los elementos, un arreglo aleatorio y un arreglo aleatorio restringido con una distancia mínima entre elementos. Adicionalmente se propone optimizar estos arreglos mediante algoritmos genéticos con el fin de mejorar su desempeño en distintos rangos de niveles de SNR. Para altos niveles de SNR se propone minimizar el error de estimación de un estimador perfecto u oráculo y para bajos SNR se agregan restricciones que por un lado disminuyan los lóbulos laterales del patron de radiación del arreglo, y por otro lado, mediante una distancia minima interelemento,

no permitan elementos del arreglo muy cercanos. Se analiza el desempeño de estos arreglos lineales no uniformes utilizando diferentes clases de algoritmos basados en modelos ralos para estimar distintas fuentes de señal y su dirección de arribo. Se proveen simulaciones del desempeño de cada configuración para diferentes valores de apertura del arreglo y número de fuentes de señal. Parte de los resultados se presentaron en el congreso regional *ARGENCON 2014* [30] y se publicaron en *IEEE Latin American Transactions* [31]. Se están finalizando sus detalles para publicarlo en una revista internacional del área (Probablemente *IEEE Transactions on Signal Processing*).

En tercer lugar se analiza el comportamiento de algoritmos basados en modelos ralos cuando son aplicados a estimación de señales y sus direcciones de arribo con arreglos de sensores. En particular se estudia el caso en el que existen fuentes de interferencia que afectan la correcta estimación de la señal de interés. Se introducen los conceptos de arreglos de sensores y los fundamentos del algoritmo ESBL aplicado específicamente a esta situación. Este escenario es propicio para la utilización de ese algoritmo, ya que explota la información de la interferencia y así disminuye sus efectos no deseados. Se compara su desempeño con distintos algoritmos en cuanto al error en la estimación de las señales de interés y sus direcciones de arribo y en el tiempo de cómputo utilizado. Se encuentra que el algoritmo presentado obtiene una buena relación de compromiso entre el tiempo de cómputo y los errores de estimación. Parte de los resultados se presentaron en el congreso regional *RPIC 2013* [32] y fueron publicados en *IEEE Antennas and Wireless Propagation Letters* [33].

Finalmente se propone un enfoque Bayesiano para el diseño óptimo de la matriz de proyección. Se elige una matriz que minimiza el error de estimación de los parámetros de interés. Esto se logra con la minimización de la traza de la covarianza posterior, resultando en una distribución posterior altamente concentrada. Para evitar el alto costo de optimizar cada elemento de la matriz de proyección, se elige una solución que pertenece a un espacio restringido. Se muestra que el diseño propuesto para la matriz de proyección combinado con el algoritmo ESBL reduce tanto el tiempo de cómputo como el error de estimación. Hasta donde se tiene conocimiento, no hay otro algoritmo para resolver el problema ralo inverso que separe la fuente de

contaminación en interferencia y ruido blanco. Como consecuencia, los métodos ralos convencionales no aprovechan al máximo el procedimiento propuesto para el diseño de la matriz. Parte de los resultados se presentaron en el congreso regional *RPIC 2011* [34] y en el congreso internacional *CAMPSAP 2011* [35]. Fueron enviados a *IEEE Transactions on Signal Processing* y ya pasaron una primera revisión.

1.3. Organización de la tesis

En el siguiente Capítulo se presenta una breve introducción a las formulaciones ralas y a los algoritmos clásicos para resolver problemas inversos que las involucren. Se detallan tres clases diferentes de este tipo de algoritmos con ejemplos de cada una. Dentro de éstos se encuentra el algoritmo ESBL que discrimina la interferencia del ruido blanco. Adicionalmente, aquí también se describe una propuesta de la formulación rala para mejorar el desempeño modificando la matriz de muestreo entre instantáneas.

En el Capítulo 3 se definen los distintos tipos de configuraciones de arreglos de sensores no uniformes, aleatorio, estructurado y restringido. Se comparan estas configuraciones con el arreglo uniforme mostrando sus mejorías y se optimiza una de las configuraciones en función de la densidad de potencia espacial de las fuentes de señal.

En el Capítulo 4 se estudia el escenario de señales inmersas en interferencias que inciden en un arreglo de sensores. Se muestra el comportamiento de algoritmos ralos para la estimación de esas señales y de sus direcciones de arribo.

En el Capítulo 5 se describe el método de optimización de la matriz de proyección para la estimación de señales de modelos ralos. Se muestran simulaciones revelando su desempeño para algoritmos ralos y la robustez del método frente a imprecisiones en los parámetros estadísticos estimados. Adicionalmente se ejemplifica su uso en una aplicación real de radar polarimétrico.

Finalmente se concluye la tesis en el Capítulo 6 donde se discuten los resultados obtenidos y se presentan posibles líneas de trabajo futuro.

Notación

Se adopta la siguiente notación. Una variable escalar se denota x , un vector es \mathbf{x} y una matriz es \mathbf{X} . Los vectores $\mathbf{X}_{.j}$ y $\mathbf{X}_{i.}$ representan la columna j y la fila i de la matriz \mathbf{X} , respectivamente. El escalar \mathbf{X}_{ij} es la componente ij de la matriz \mathbf{X} , y x_i es la componente i del vector \mathbf{x} . La matriz identidad de tamaño P es \mathbf{I}_P . El transpuesto conjugado de un vector o matriz compleja es \mathbf{x}^H y \mathbf{X}^H . La traza de una matriz cuadrada es $\text{tr}(\mathbf{X})$ y la traza a bloques es $\text{btr}(\mathbf{X})$ definida en el Apéndice B.1. El operador \otimes denota al producto de Kronecker. $\|\mathbf{X}\|_F$ es la norma de Frobenius de la matriz \mathbf{X} .

El vector aleatorio $\mathbf{x} \sim \mathcal{CN}_P(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ es un vector Gaussiano complejo circularmente simétrico de dimensión P con media $\boldsymbol{\mu}$ y covarianza $\boldsymbol{\Sigma}$. La variable aleatoria escalar $x \sim \mathcal{X}_n^2$ posee distribución Chi cuadrada con n grados de libertad.

Capítulo 2

Algoritmos basados en modelos ralos

2.1. Introducción

Las representaciones de señales ralas a partir de diccionarios sobrecompletos han ganado relevancia en un gran número de campos de aplicación [3, 4, 36]. Estas representaciones ralas permiten resolver problemas lineales inversos de gran importancia en el procesamiento de señales. Ejemplos de aplicaciones incluyen tratamiento de imágenes EEG/MMG [37–39], eliminación de ruido de señal [40, 41], muestreo sub-Nyquist [42–44], estimación y ecualización de canal [45–48], estimación espectral [49], estimación de dirección de arribo [50], aproximación funcional [51–53], cancelación de eco [54, 55], estimación y detección en radar [56–58] y radar MIMO [59, 60].

Una de las formas más comunes de representación rala es en base a un modelo de regresión lineal.

$$\mathbf{y} = \mathbf{X}\mathbf{b} \tag{2.1}$$

La señal observada \mathbf{y} puede ser así representada a través de un diccionario \mathbf{X} , que contiene, en sus columnas, diferentes formas de onda llamadas átomos [4]. Los coeficientes de regresión \mathbf{b} definen como combinar esos átomos para generar la señal. La representación será rala si sólo algunos de esos coeficientes de \mathbf{b} son distintos de cero.

El problema raro inverso consistirá en estimar \mathbf{b} en base a las mediciones \mathbf{y} . Esto quiere decir que el objetivo es minimizar alguna métrica basada en el residuo $\mathbf{y} - \mathbf{X}\mathbf{b}$, o error en la estimación de \mathbf{b} dado que este último es raro.

$$\min_{\mathbf{b}} f(\mathbf{y} - \mathbf{X}\mathbf{b}) \quad \text{tal que} \quad \mathbf{b} \text{ raro} \quad (2.2)$$

Por lo tanto, dada una señal $\mathbf{y} \in \mathbb{C}^N$, se busca el vector de coeficientes $\mathbf{b} \in \mathbb{C}^M$ que posea el menor número de coeficientes distintos de cero tal que $\mathbf{X}\mathbf{b} = \mathbf{y}$. Formalmente,

$$\min_{\mathbf{b}} \|\mathbf{b}\|_0 \quad \text{tal que} \quad \mathbf{y} = \mathbf{X}\mathbf{b} \quad (2.3)$$

Usualmente la cantidad de mediciones disponibles será mucho menor que el número de átomos del diccionario. Luego, el problema resulta ser subdeterminado y no puede ser resuelto por métodos clásicos. Sin embargo, si el vector de coeficientes de regresión es raro es posible resolver el problema inverso, aún cuando el soporte sea desconocido [17,18]. Lamentablemente, encontrar el soporte de este vector es un problema NP, ya que sería necesario probar todos los potenciales soportes y quedarse con el que minimice el error y la rareza. Este procedimiento de minimizar la norma ℓ_0 con restricciones resulta en un problema de optimización combinatorio, que no es factible aún para vectores de tamaño reducido [18]. En consecuencia, se propusieron algoritmos computacionalmente realizables que intentan aproximarse a la solución correcta de este problema inverso mediante diferentes técnicas.

Algunos de estos algoritmos fueron también extendidos para resolver el problema de la señal contaminada por ruido descrito por el siguiente modelo,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{w} \quad (2.4)$$

donde \mathbf{w} representa el ruido aditivo del modelo. En este caso dependiendo del nivel de ruido, la señal recuperada por los algoritmos previos no es rara. Sin embargo muchos de sus coeficientes son estadísticamente despreciables, y unos pocos de ellos son significativos. Para obtener una señal realmente rara donde la mayoría de sus coeficientes resulten ser estrictamente cero, algunos algoritmos añaden una etapa de

purga en el procedimiento para hacer nulos los coeficientes despreciables y mejorar la estimación de la señal verdadera.

Una forma de garantizar que se pueda reconstruir la señal en base a las mediciones dadas, es que la matriz de medición \mathbf{X} de tamaño $M \times N$ cumpla la condición **RIP** (Restricted Isometry Property). Esta condición establece que esa matriz obedece la hipótesis de isometría restringida. Dada la submatriz \mathbf{X}_T , $T \subset \{1, \dots, N\}$ de tamaño $M \times |T|$, obtenida al extraer de \mathbf{X} las columnas correspondientes a los índices en T , en [61] se define la constante de isometría S -restringida δ_S de \mathbf{X} que es la menor cantidad que cumple

$$(1 - \delta_s)\|c\|_{\ell_2}^2 \leq \|\mathbf{X}_T c\|_{\ell_2}^2 \leq (1 + \delta_s)\|c\|_{\ell_2}^2 \quad (2.5)$$

para todos los subconjuntos T con $|T| \leq S$ y secuencias de coeficientes $\{c_j\}_{j \in T}$. Esta propiedad requiere que cada conjunto de columnas con cardinalidad menor a S se comporte aproximadamente como un sistema ortonormal. En ese mismo artículo se muestra que si S cumple que $\delta_S + \delta_{2S} + \delta_{3S} < 1$ entonces resolviendo el problema (2.3) se puede recuperar cualquier señal \mathbf{b} cuyo soporte cumpla $|T| \leq S$.

Estos algoritmos de estimación pueden agruparse en diferentes clases de acuerdo a las bases teóricas usadas para su funcionamiento. Existen algoritmos que buscan minimizar la norma ℓ_1 del error como una alternativa a la norma ℓ_0 , los algoritmos de búsqueda voraz (greedy pursuit) y los algoritmos basados en un marco probabilístico.

2.2. Clases de algoritmos

2.2.1. Algoritmos de relajación

Estos algoritmos relajan la función de costo en el problema de minimización reemplazando la norma ℓ_0 con la norma ℓ_p para algún valor de $p \in (0, 1]$. Un ejemplo de estos algoritmos es el clásico *focal underdetermined system solver* (FOCUSS) [3] que utiliza cuadrados mínimos pesados iterativos (IRLS) para resolver la norma ℓ_p . En general otros métodos utilizan la norma ℓ_1 , que resulta en un problema de optimización convexo garantizando una única solución global. Esta solución puede alcanzarse mediante programación lineal con un alto costo computacional. Por

ejemplo, *basis pursuit* (BP) resuelve el problema que no considera ruido [4].

$$\min_{\mathbf{b}} \|\mathbf{b}\|_1 \quad \text{tal que} \quad \mathbf{X}\mathbf{b} = \mathbf{y} \quad (2.6)$$

Existen otros métodos de relajación que consideran ruido en el modelo, como ser *basis pursuit denoising* (BPDN) [4] también conocido como LASSO [62]. Este algoritmo utiliza programación cuadrática para hallar la solución óptima al problema

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1 \quad (2.7)$$

donde λ es un parámetro de regularización donde su valor, aún cuando existen varios procedimientos para elegirlo, influirá en la determinación una solución correcta. Sin embargo, una forma más natural de utilizar un parámetro de regularización aparece en el método conocido como *Dantzig selector* (DS) [5]. Este método también tiene en cuenta el error en las mediciones y su principal ventaja es que utiliza solamente programación lineal para hallar la solución del siguiente planteo,

$$\min_{\mathbf{b}} \|\mathbf{b}\|_1 \quad \text{tal que} \quad \|\mathbf{X}^H(\mathbf{y} - \mathbf{X}\mathbf{b})\|_\infty \leq \epsilon. \quad (2.8)$$

Nótese que el parámetro de regularización ϵ en este caso está relacionado con el nivel de ruido o cuantización.

2.2.2. Algoritmos greedy

Los algoritmos greedy se basan en aproximaciones iterativas de los coeficientes de la señal y su soporte, ya sea identificando el soporte de la señal hasta satisfacer un criterio de convergencia, o alternativamente obteniendo una mejor estimación de la señal rara en cada iteración al tratar de tener en cuenta la diferencia con los datos medidos. En cada iteración calcula el residuo entre las observaciones y la solución anterior, y actualiza el soporte, basándose en el átomo del diccionario que mejor se correlacione con el residuo. Puede demostrarse que algunos algoritmos greedy pueden alcanzar el desempeño de los métodos de relajación.

Ejemplos básicos de estos algoritmos son el algoritmo original *matching pursuit* (MP) [63] y *orthogonal matching pursuit* [7] aplicado a modelos raros, que pueden

obtener una solución moderadamente aceptable rápidamente. El algoritmo OMP comienza encontrando la columna de \mathbf{X} mayormente correlacionada con las mediciones \mathbf{y} . Luego, se repite esta operación correlacionando las columnas con el residuo de señal $\mathbf{r} = \mathbf{X}\mathbf{b} - \mathbf{y}$, que se obtiene al restar la contribución de la estimación parcial de la señal del vector original de mediciones. Por lo tanto el algoritmo OMP empieza con un modelo “vacío” y construye el modelo de señal un átomo a la vez, en cada paso añade al modelo el átomo más importante de todos los que aún no fueron tenidos en cuenta. Por otra parte, el algoritmo de relajación Basis Pursuit, previamente descrito, empieza con un modelo “completo” y luego iterativamente mejora este modelo completo dejando de lado átomos que no contribuyen al modelo y los intercambia por nuevos átomos que aporten información más útil. Entonces se dice que OMP es un algoritmo que construye átomos hacia arriba y BP es un algoritmo que intercambia átomos hacia abajo [64].

Variantes de estos algoritmos mejoran la estimación utilizando distintos tipos de ajustes, como ser *stagewise orthogonal matching pursuit* (StOMP) [65], y *compressive sampling matching pursuit* (CoSaMP) [66].

2.2.3. Algoritmos probabilísticos

La mayoría de los métodos probabilísticos para la reconstrucción de señales raras son Bayesianos y explotan información a priori de los coeficientes de regresión para regularizar el problema subdeterminado [67]. Estas técnicas poseen un desempeño superior cuando el modelo probabilístico es una representación razonable del proceso físico que genera las observaciones. Aún más, la formulación Bayesiana no sólo provee una estimación puntual de los coeficientes sino que además estima su distribución a posteriori. Uno de los métodos de esta clase es *sparse Bayesian learning* (SBL). Este método fue desarrollado para *relevance vector machine* (RVM) [68] en el campo de machine learning y luego adaptado para selección de bases a partir de diccionarios sobrecompletos [10]. Luego fue aplicado para resolver el problema inverso de mediciones comprimidas y renombrado *Bayesian Compressive Sensing* (BCS) [69]. El algoritmo SBL supone que los coeficientes de regresión son variables aleatorias independientes con distribuciones Gaussianas de media cero. La

varianza de estos coeficientes son hiperparámetros que se obtienen de las observaciones maximizando la función de verosimilitud, en general a través del algoritmo de esperanza-maximización (EM). Esta función de verosimilitud alcanza su máximo global para la solución más rara, y sus máximos locales también son raros [10]. Este algoritmo posee una etapa de constricción (shrinkage), en la cual los hiperparámetros irrelevantes se anulan, que posee una elevada complejidad y un alto consumo de tiempo. El método BCS logra reducir este tiempo de cómputo adoptando un algoritmo rápido basado en RVM a cambio de una pérdida en el desempeño [70].

Existen otras formulaciones de algoritmos probabilísticos. Por ejemplo, *automatic double relaxation* (ADORE) utiliza selección de modelos para determinar la rareza de la señal y luego la compara con un umbral para asegurarse de que la señal tenga la rareza estimada [71]. Por otra parte, el algoritmo *fast Bayesian matching pursuit* (FBMP) utiliza un promediado de modelos en vez de selección de modelos, considerando una mezcla de Gaussianas como la distribución a priori de los coeficientes desconocidos [72].

Finalmente, algunos de los algoritmos greedy y de relajación poseen una interpretación Bayesiana en la cual las distribuciones a priori poseen alguna distribución que promueva la rareza de la señal [73, 74].

2.3. Múltiples vectores de mediciones

Una extensión del problema inverso que utiliza el modelo raro (2.4) surge cuando se dispone de varias instantáneas (snapshots) de las mediciones \mathbf{y} correspondientes a distintos tiempos de adquisición. Adicionalmente se supone que estos vectores pueden ser representados por los mismos átomos del diccionario \mathbf{X} , compartiendo así un perfil común de rareza. Por ejemplo si se poseen D instantáneas, se puede escribir el siguiente modelo extendido,

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{W}, \quad (2.9)$$

donde cada una de las matrices, \mathbf{Y} , \mathbf{B} y \mathbf{W} , se construye con los respectivos vectores como sus columnas. Por ejemplo $\mathbf{B} = [\mathbf{B}_{.1}, \dots, \mathbf{B}_{.D}]$. Suponiendo entonces que los

datos pueden ser representados por los mismos elementos del diccionario, la matriz \mathbf{B} posee un número bajo de filas distintas de cero. El objetivo es estimar las columnas de \mathbf{B} bajo la condición de un perfil común de rareza y estacionariedad de la interferencia y el ruido durante las diferentes instantáneas.

Esta formulación es una representación realista de múltiples realizaciones de una señal contaminada por ruido. Este nuevo problema se conoce como una representación rala de *múltiples vectores de mediciones* (MMV) [75] o *aproximación rala simultánea* [76]. Algunos de los algoritmos mencionados previamente fueron extendidos para múltiples realizaciones y adicionalmente para manipular números complejos. Estos algoritmos son MOMP [77], MFOCUSS [78], MSBL [79], y los métodos Rx-Penalty y Rx-Error [80].

2.4. Algoritmo de estimación ESBL

Dentro de la clase de algoritmos probabilísticos bayesianos se encuentra el algoritmo ESBL, cuyo funcionamiento se describe brevemente a continuación. Para detalles específicos y desarrollo de sus fundamentos se refiere al lector al artículo [11].

2.4.1. Modelo

Este algoritmo extiende el modelo (2.4) para considerar el siguiente modelo lineal con P factores de interferencia independientes.

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{p=1}^P \mathbf{Z}_p \mathbf{u}_p + \mathbf{w}, \quad (2.10)$$

donde \mathbf{y} es el vector de mediciones de tamaño M , $\mathbf{b} = (b_1, b_2, \dots, b_N)$ es el vector de tamaño N correspondiente a la señal de interés, \mathbf{u}_p es el vector de tamaño Q que representa a cada una de las interferencias, \mathbf{X} y \mathbf{Z}_p son matrices de regresores y \mathbf{w} es el ruido. Como ya se mencionó, la matriz \mathbf{X} es el diccionario cuyas columnas son los átomos, y la función del vector \mathbf{b} es elegir algunos de éstos para formar la señal deseada. Se supone que las matrices \mathbf{X} y \mathbf{Z}_p son conocidas y complejas, que

la interferencia \mathbf{u}_p y el ruido \mathbf{w} poseen una distribución gaussiana de media cero con covarianzas Σ_u y $\sigma\mathbf{I}_M$ respectivamente, y existe independencia estadística entre ellos.

$$\begin{aligned}\mathbf{w} &\sim \mathcal{CN}_K(\mathbf{0}, \sigma\mathbf{I}_K), \\ \mathbf{u}_p &\sim \mathcal{CN}_Q(\mathbf{0}, \Sigma_u),\end{aligned}$$

Los datos medidos también puede escribirse en forma condensada como

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{w}, \quad (2.11)$$

donde \mathbf{u} es el vector de tamaño P que agrupa las interferencias y $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_P]$. Considerando D observaciones independientes, la distribución de los datos medidos para cada instantánea resulta

$$\mathbf{y}_{d|b} \sim \mathcal{CN}_M(\mathbf{X}\mathbf{b}, \mathbf{Z}(\mathbf{I}_P \otimes \Sigma_u)\mathbf{Z}^H + \sigma\mathbf{I}_M),$$

y el modelo puede reescribirse como un MMV,

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z}\mathbf{U} + \mathbf{W}, \quad (2.12)$$

donde las matrices se construyen tomando los correspondientes vectores de (2.11) como sus columnas.

Se considera que el vector \mathbf{b} es ralo, cuya *raleza* es $\|\mathbf{b}\|_0 = S \ll N$. De acuerdo a la formulación de SBL, a este vector se le asigna una distribución gaussiana de media cero y covarianza $\Sigma_b = \text{diag}(\boldsymbol{\beta})$. El vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)$ contiene los hiperparámetros que controlan qué átomos del diccionario se encuentran activos.

$$\mathbf{b} \sim \mathcal{CN}_N(\mathbf{0}, \Sigma_b). \quad (2.13)$$

El objetivo del algoritmo es hallar la solución al problema inverso subdeterminado de estimar \mathbf{b} dado que $N > M$ y que Σ_b , Σ_u y σ son desconocidos.

2.4.2. Algoritmo

Cada iteración del algoritmo posee dos pasos. El primer paso consiste en determinar el estimador de máxima verosimilitud de Σ_b , Σ_u y σ , mediante el algoritmo

EM. Luego de calcular las estimaciones de los parámetros, se obtiene una estimación puntual de \mathbf{b} aplicando la media posterior. La estimación del vector \mathbf{b} no será rala, $\|\hat{\mathbf{b}}\|_0 = N$. Sin embargo, el algoritmo EM generará unos pocos coeficientes significativos y varios despreciables cuando \mathbf{b} sea ralo. Por lo tanto se necesita generar una regla de decisión para eliminar estos últimos. En el segundo paso se utiliza un test estadístico para decidir si un componente del vector \mathbf{b} está presente o no en cada iteración del algoritmo EM. El umbral necesario para el test de decisión está determinado por la probabilidad de falsa alarma del estimador P_{FA} .

Paso 1. Estimación de parámetros

En este paso de la iteración se utiliza el algoritmo EM para estimar los parámetros del modelo. De acuerdo al Apéndice A los datos observados \mathbf{y} y los datos ocultos $\mathbf{z} = [\mathbf{b}^T, \mathbf{u}^T]^T$ generan los datos completos $\mathbf{x} = [\mathbf{y}^T, \mathbf{b}^T, \mathbf{u}^T]^T$. En este caso los parámetros desconocidos son $\boldsymbol{\theta} = \{\boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_u, \sigma\}$.

En base al modelo estadístico descrito previamente, la covarianza de los datos resulta ser

$$\boldsymbol{\Sigma}_y = \mathbf{X}\boldsymbol{\Sigma}_b^{(t)}\mathbf{X}^H + \mathbf{Z}(\mathbf{I}_P \otimes \boldsymbol{\Sigma}_u^{(t)})\mathbf{Z}^H + \sigma^{(t)}\mathbf{I}_N, \quad (2.14)$$

y la covarianza muestral de ellos cuando se tienen D instantáneas,

$$\tilde{\boldsymbol{\Sigma}}_y = \frac{1}{D} \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^H. \quad (2.15)$$

En base a estas covarianzas se define $\mathbf{C}_y = \boldsymbol{\Sigma}_y^{-1} \tilde{\boldsymbol{\Sigma}}_y \boldsymbol{\Sigma}_y^{-1} - \boldsymbol{\Sigma}_y^{-1}$ y $\mathbf{C}_u = \mathbf{I}_P \otimes \boldsymbol{\Sigma}_u$ para simplificar la notación. Se muestra en el Apéndice B que la estimación de los parámetros provista por el algoritmo EM resulta

$$\sigma^{(t+1)} = \frac{1}{M} \text{tr} \left(\sigma^{(t)} \mathbf{I}_M + (\sigma^{(t)})^2 \mathbf{C}_y \right), \quad (2.16)$$

$$\boldsymbol{\Sigma}_u^{(t+1)} = \frac{1}{P} \text{btr} \left(\mathbf{C}_u^{(t)} + \mathbf{C}_u^{(t)} \mathbf{Z}^H \mathbf{C}_y \mathbf{Z} \mathbf{C}_u^{(t)} \right), \quad (2.17)$$

$$\beta_n^{(t+1)} = \beta_n^{(t)} + (\beta_n^{(t)})^2 \mathbf{X}_n^H \mathbf{C}_y \mathbf{X}_n. \quad (2.18)$$

Si se tiene un modelo lineal $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{w}$, donde \mathbf{X} es una matriz conocida de tamaño $M \times N$, $\mathbf{b} \sim \mathcal{CN}_N(0, \boldsymbol{\Sigma}_b)$ y $\mathbf{w} \sim \mathcal{CN}_M(0, \sigma \mathbf{I}_M)$, entonces \mathbf{y} es compleja

Gaussiana con media cero y covarianza $\Sigma_y = \mathbf{X}\Sigma_b\mathbf{X}^H + \Sigma_w$. Adicionalmente la distribución posterior de \mathbf{b} también es compleja Gaussiana con media y covarianza,

$$\boldsymbol{\mu}_{b|y} = \Sigma_b\mathbf{X}^H\Sigma_y^{-1}\mathbf{y} \quad (2.19)$$

$$\Sigma_{b|y} = \Sigma_b - \Sigma_b\mathbf{X}^H\Sigma_y^{-1}\mathbf{X}\Sigma_b. \quad (2.20)$$

Con esta información y las estimaciones de los parámetros, se calcula la estimación puntual de \mathbf{b}

$$\hat{\mathbf{b}} = \Sigma_b\mathbf{X}^H\Sigma_y^{-1}\mathbf{y}. \quad (2.21)$$

que como ya se mencionó, no cumple con la condición de rareza. Por lo tanto se aplica un test de detección para purgar sus componentes estadísticamente insignificantes.

Paso 2. Detección de componentes

Para determinar si una componente del vector de señales \mathbf{b} está presente o no en cada iteración del algoritmo, se define el siguiente test de hipótesis,

$$\begin{cases} \mathcal{H}_0 : & \beta_n = 0 \\ \mathcal{H}_1 : & \beta_n \neq 0 \end{cases} \quad (2.22)$$

para $n = 1, \dots, N$. En la iteración $t+1$ del algoritmo EM, asumiendo que los valores estimados de $\boldsymbol{\beta}$, σ , y Σ_u son los verdaderos, se construye un test de Neyman-Person basado en los datos completos \mathbf{x} . Considerando las definiciones de los pasos del algoritmo EM, un estadístico es $\mathcal{Q}(\beta_n^{(t+1)}, \boldsymbol{\theta}^{(t)})$, la función maximizada en el paso **M**. Preservando sólo el término que depende de los datos observados, se muestra en el apéndice B que el test estadístico es

$$\mathcal{T}^{(n)} = \mathbf{t}_n^H \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\tilde{\Sigma}_y}} \mathbf{t}_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha, \quad (2.23)$$

donde $\mathbf{t}_n = \Sigma_y^{-1}\mathbf{X}_n$.

Un aspecto importante en el diseño de un algoritmo de detección es poder fijar el valor del umbral α con el fin de alcanzar determinados resultados, por ejemplo la probabilidad de falsa alarma. Para ello se observa que la matriz $\tilde{\Sigma}_y$ posee una

distribución Wishart compleja de dimensión M con D grados de libertad y una matriz de covarianza $\mathbf{\Sigma}_y/D$, es decir $\tilde{\mathbf{\Sigma}}_y \sim \mathcal{CW}_M(\mathbf{\Sigma}_y/D, D)$ [81]. Por lo tanto, el estadístico de detección está distribuido de acuerdo a

$$\text{Bajo } \mathcal{H}_0 : T_0^{(n)} = \frac{2D\mathcal{T}^{(n)}}{\mathbf{t}_n^H \tilde{\mathbf{\Sigma}}_{y_0} \mathbf{t}_n} \sim \chi_{2D}^2 \quad (2.24)$$

$$\text{Bajo } \mathcal{H}_1 : T_1^{(n)} = \frac{2D\mathcal{T}^{(n)}}{\mathbf{t}_n^H \tilde{\mathbf{\Sigma}}_{y_1} \mathbf{t}_n} \sim \chi_{2D}^2 \quad (2.25)$$

donde $\tilde{\mathbf{\Sigma}}_{y_0}$ y $\tilde{\mathbf{\Sigma}}_{y_1}$ son las covarianzas de los datos bajo las respectivas hipótesis, relacionadas por la siguiente ecuación

$$\tilde{\mathbf{\Sigma}}_{y_1} = \tilde{\mathbf{\Sigma}}_{y_0} + \beta_n^{(t)} \mathbf{X}_{.n} \mathbf{X}_{.n}^H \quad (2.26)$$

Para determinar el desempeño de este detector se calcula la probabilidad de falsa alarma

$$P_{FA} = \mathcal{P}(T_0^{(n)} > \alpha_{FA}; \mathcal{H}_0) = 1 - \mathcal{F}_{\chi_{2D}^2}(\alpha_{FA}) \quad (2.27)$$

donde α_{FA} es el umbral requerido para alcanzar un determinado nivel de falsa alarma. Este umbral puede entonces calcularse como

$$\alpha_{FA} = \mathcal{F}_{\chi_{2D}^2}^{-1}(1 - P_{FA}). \quad (2.28)$$

donde \mathcal{F} representa la función de distribución acumulada. Finalmente, el algoritmo debe verificar si la siguiente desigualdad es válida para decidir si la componente n -ésima está activa.

$$\frac{2D\mathbf{X}_{.n}^H \mathbf{\Sigma}_y^{-1} \tilde{\mathbf{\Sigma}}_y \mathbf{\Sigma}_y^{-1} \mathbf{X}_{.n}}{\mathbf{X}_{.n}^H \mathbf{\Sigma}_y^{-1} \mathbf{X}_{.n} (1 - \beta_n \mathbf{X}_{.n}^H \mathbf{\Sigma}_y^{-1} \mathbf{X}_{.n})} < \alpha_{FA}. \quad (2.29)$$

El funcionamiento del algoritmo se resume en la figura 2.1. Es importante recalcar que el umbral de detección α_{FA} no depende de la covarianza $\mathbf{\Sigma}_u$, ni de σ , ni tampoco de las matrices \mathbf{X} o \mathbf{Z} . Por lo tanto, $T_0^{(n)}$ posee la propiedad de nivel de falsa alarma constante (CFAR). Adicionalmente, se destaca que en ningún momento se debe indicar al algoritmo ESBL la cantidad de señales presentes, es decir la *raleza* de la señal de interés. Por otra parte, este algoritmo es el único de los mencionados que utiliza la información de la interferencia, es decir que discrimina el ruido estructurado del ruido blanco permitiendo una mejor estimación del soporte de la señal de interés.

- Inicialización: fijar $i = 0$ y

$$\begin{aligned}\tilde{\Sigma}_y &= D^{-1}\mathbf{Y}\mathbf{Y}^H \\ \Sigma_b^{(0)} &= \mathbf{I}_N \\ \Sigma_u^{(0)} &= \mathbf{I}_P \otimes \Sigma_{u0} \\ \sigma^{(0)} &= \sigma_0\end{aligned}$$

- Iteración: Incrementar i una vez y realizar los pasos siguientes

- Paso EM:

$$\begin{aligned}\beta_n^{(i)} &= \beta_n^{(i-1)} + (\beta_n^{(i-1)})^2 \mathbf{X}_{\cdot n}^H \mathbf{C}_y \mathbf{X}_{\cdot n} \\ \sigma^{(i)} &= M^{-1} \text{tr} \left(\sigma^{(i-1)} \mathbf{I}_M + (\sigma^{(i-1)})^2 \mathbf{C}_y \right) \\ \Sigma_u^{(i)} &= P^{-1} \text{btr} \left(\mathbf{C}_u^{(i-1)} + \mathbf{C}_u^{(i-1)} \mathbf{Z}^H \mathbf{C}_y \mathbf{Z} \mathbf{C}_u^{(i-1)} \right)\end{aligned}$$

- Paso de actualización:

$$\begin{aligned}\Sigma_b^{(i)} &= \text{diag}(\beta_1^{(i)}, \dots, \beta_N^{(i)}) \\ \mathbf{C}_u^{(i)} &= \mathbf{I}_P \otimes \Sigma_u^{(i)} \\ \Sigma_y^{(i)} &= \mathbf{X} \Sigma_b^{(i)} \mathbf{X}^H + \mathbf{Z} \mathbf{C}_u^{(i)} \mathbf{Z}^H + \sigma^{(i)} \mathbf{I}_M \\ \mathbf{C}_y &= (\Sigma_y^{(i)})^{-1} \tilde{\Sigma}_y (\Sigma_y^{(i)})^{-1} - (\Sigma_y^{(i)})^{-1}\end{aligned}$$

- Paso de poda: Fijar $\beta_n^{(i)} = 0$ si

$$\frac{2D \mathbf{X}_{\cdot n}^H \Sigma_y^{-1} \tilde{\Sigma}_y \Sigma_y^{-1} \mathbf{X}_{\cdot n}}{\mathbf{X}_{\cdot n}^H \Sigma_y^{-1} \mathbf{X}_{\cdot n} (1 - \beta_n \mathbf{X}_{\cdot n}^H \Sigma_y^{-1} \mathbf{X}_{\cdot n})} < \alpha_{FA}.$$

- Regla de parada: Iterar si

$$\text{Error} = \sum_{n=1}^N [\beta_n^{(i)} - \beta_n^{(i-1)}]^2 > \text{Tol}$$

- Salida: Luego de i iteraciones,

$$\hat{B} = \Sigma_b^{(i)} \mathbf{X}^H (\Sigma_y^{(i)})^{-1} \mathbf{Y}$$

Figura 2.1: Pseudo-código del algoritmo ESBL.

2.5. Modelo ralo con fusión de mediciones

2.5.1. Introducción

En algunas aplicaciones se puede tener control sobre el diccionario \mathbf{X} , es decir que no está restringido al escenario o equipo de medición y entonces puede elegirse arbitrariamente para mejorar el desempeño del sistema [21, 24]. En particular, si la señal de interés \mathbf{b} se mantiene constante durante un número de instantáneas, una posibilidad es elegir una matriz distinta para cada una. Intuitivamente, este procedimiento permite mejorar la estimación ya que una mala elección del diccionario en una instantánea puede ser compensado por un diferente diccionario en otra instantánea. Formalmente esto quiere decir que los diccionarios elegidos en distintas instantáneas sean aproximadamente ortogonales.

Dependiendo de la aplicación, una opción es elegir aleatoriamente un diccionario \mathbf{X} cuyas componentes tienen una distribución Gaussiana, y por lo tanto diccionarios de distintas instantáneas serán prácticamente ortogonales con una alta probabilidad. Este caso es estudiado en la siguiente Sección.

Otra opción es tener un diccionario inicial \mathbf{X} y formar matrices \mathbf{X}_d para cada instantánea a partir de un subconjunto de columnas o átomos del diccionario original. Por ejemplo, un escenario particular donde sucede esto es cuando se tiene un arreglo de una gran cantidad de sensores y acceder a la información de todos ellos al mismo tiempo es muy costoso, ya sea por energía o velocidad de respuesta. En este caso se puede elegir el acceso a una cantidad menor de ellos y en base a esa información tratar de estimar la señal de interés. Si se accede siempre a los mismos sensores, todas las mediciones estarán expuestas a los mismos riesgos, pero si en cada instante de muestreo se accede a distinto conjunto de sensores entonces se incrementa la robustez. La desventaja de utilizar una menor cantidad de sensores a la totalidad de ellos es que se pierde desempeño en la estimación de la señal de interés.

A continuación se describe el modelo ralo a usar en este escenario, y se muestran simulaciones del desempeño del algoritmo para las dos aplicaciones mencionadas.

2.5.2. Modelo

En este caso el modelo a utilizar se basa en (2.4) pero la matriz de mediciones \mathbf{X} se considera distinta para cada instantánea.

$$\mathbf{y}_d = \mathbf{X}_d \mathbf{b} + \mathbf{w}_d, \quad (2.30)$$

donde \mathbf{y}_d es el vector de mediciones de tamaño M , $\mathbf{b} = (b_1, b_2, \dots, b_N)$ es el vector de tamaño N correspondiente a la señal de interés, \mathbf{X}_d es la matriz de regresores para la instantánea d y \mathbf{w}_d es el ruido. En este caso como \mathbf{X} difiere entre instantáneas, no se trata de un modelo MMV, sin embargo es equivalente a tener una única matriz $\bar{\mathbf{X}} = [\mathbf{X}_1^T, \mathbf{X}_2^T, \dots, \mathbf{X}_D^T]^T$, y apilar los vectores \mathbf{y}_d y \mathbf{w}_d para $d = 1, \dots, D$.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_D \end{bmatrix}$$

$$\bar{\mathbf{y}} = \bar{\mathbf{X}} \mathbf{b} + \bar{\mathbf{w}} \quad (2.31)$$

donde $\bar{\mathbf{y}}$ y $\bar{\mathbf{w}}$ son vectores de tamaño MD , y $\bar{\mathbf{X}}$ es una matriz de tamaño $MD \times N$. Para resolver este problema inverso se vuelve a utilizar el algoritmo ESBL de la sección anterior con el modelo (2.31). Ese modelo usa las matrices apiladas y supone que se tiene una sola instantánea.

2.5.3. Desempeño

Para mostrar las ventajas de utilizar una matriz distinta en cada instantánea e ilustrar el desempeño se realizan simulaciones comparando el modelo (2.4) con (2.30). Los resultados aquí presentados corresponden a promedios de $MC=1000$ realizaciones independientes de Monte Carlo, en cada una de las cuales se genera aleatoriamente el soporte de la señal \mathbf{b} y el ruido aditivo. Se define la potencia de la señal como $P_b = \|\mathbf{b}\|_2^2$, la potencia de ruido como $P_w = N\sigma$ y la relación señal a ruido como $SNR = 10 \log_{10}(P_b/P_w)$. Para resolver el problema inverso y hallar la estimación de \mathbf{b} se utiliza el algoritmo ESBL, con una probabilidad de falsa alarma

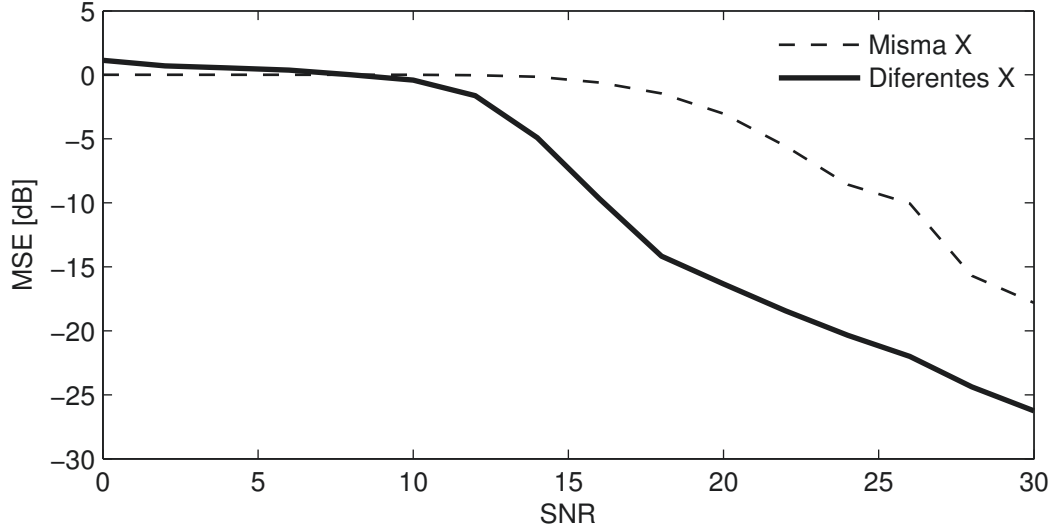


Figura 2.2: Error en la estimación de \mathbf{b} para diferentes niveles de SNR, utilizando $D=10$ instantáneas y una raleza $S=1$.

de $P_{FA}=10^{-6}$ y en este caso no se hace uso de la interferencia, ($\mathbf{Z}=0$). Para evaluar el desempeño del algoritmo se utiliza como métrica el error cuadrático medio en la estimación de \mathbf{b} , $\text{MSE} = \sum_{i=1}^{MC} \|\hat{\mathbf{b}}_i - \mathbf{b}\|_2^2 / \text{MC}$.

\mathbf{X}_d aleatoria

En primer lugar se analiza el desempeño de este algoritmo para el caso en que las matrices \mathbf{X}_d son aleatorias. Se adquieren $D=10$ instantáneas de un vector de mediciones de tamaño $M=5$ a partir de una señal de interés \mathbf{b} de tamaño $N=100$ cuya raleza es $S=1$. Para cada realización de Monte Carlo se generan las D matrices \mathbf{X}_d donde cada componente de ellas se genera con una distribución Gaussiana de media cero y varianza 1, $[\mathbf{X}_d]_{ij} \sim \mathcal{CN}(0, 1)$, luego normalizadas mediante sus normas de Frobenius.

En la Figura 2.2 se muestra el error cuadrático medio de la estimación de \mathbf{b} para distintas relaciones señal a ruido SNR. Se aprecia en la Figura que para obtener un error de 0,03 puede utilizarse una potencia de señal 10 veces menor al usar distintas matrices en cada instantánea. Sin embargo, esta ganancia en potencia de señal no

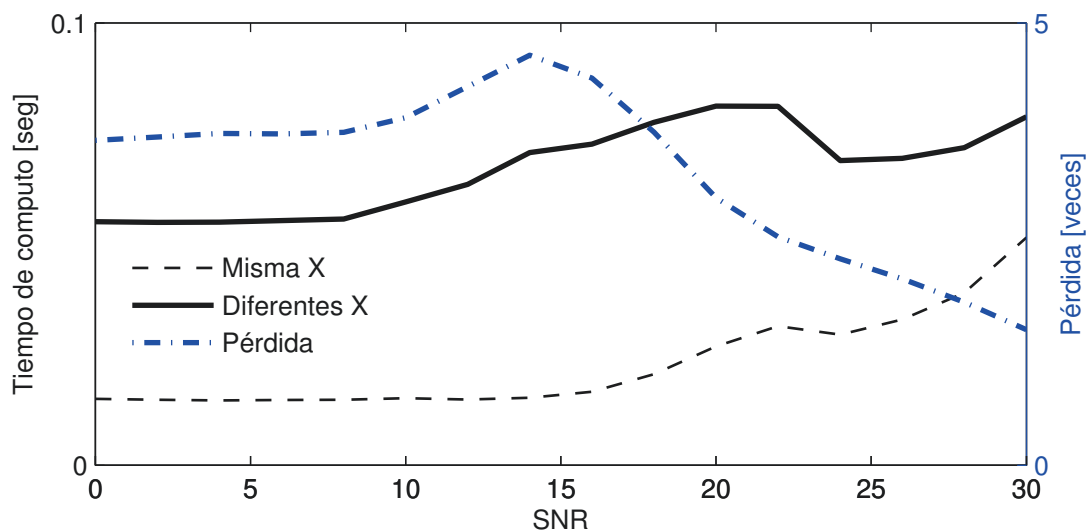


Figura 2.3: Tiempo de cómputo para las dos opciones de elección de \mathbf{X}_d y su relación (pérdida), para diferentes niveles de SNR, utilizando $D = 10$ instantáneas y una rareza $S=1$.

es gratuita, el costo de este beneficio es un mayor tiempo de cómputo del algoritmo de estimación, como se muestra en la Figura 2.3. Allí se muestran los tiempos de cómputo para las dos opciones de elección de \mathbf{X}_d . Si se elige siempre la misma matriz, puede usarse el modelo original donde el diccionario \mathbf{X} es de tamaño $M \times ND$ y el costo del algoritmo al invertir Σ_y será de $\mathcal{O}(M^3)$. Por otro lado, cuando se elige una matriz aleatoria en cada instantánea, la matriz resultante $\bar{\mathbf{X}}$ es de tamaño $MD \times N$ por lo tanto en este caso el costo es $\mathcal{O}((MD)^3)$. Este último alcanza a ser de 2 a 5 veces mayor, relación ilustrada en la misma figura como pérdida, y es el costo que tiene la ganancia obtenida en el desempeño ilustrada en la Figura previa.

En la Figura 2.4 se muestra el error para distinta cantidad de instantáneas D , con una relación señal a ruido fija de 15 dB. Se observa que a medida que se incrementa el número de muestras, la ganancia (diferencia entre los dos MSE) crece rápidamente para pocas instantáneas y luego se mantiene prácticamente constante. Por un lado, esto indica que aumentar indefinidamente el número de instantáneas con distintos diccionarios no aporta más información. Esto puede explicarse intuitivamente porque

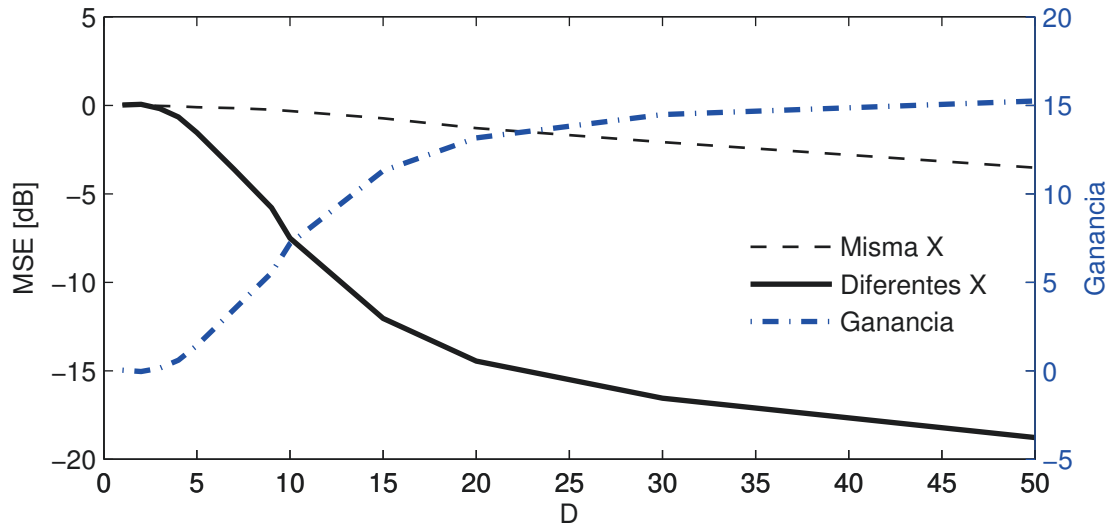
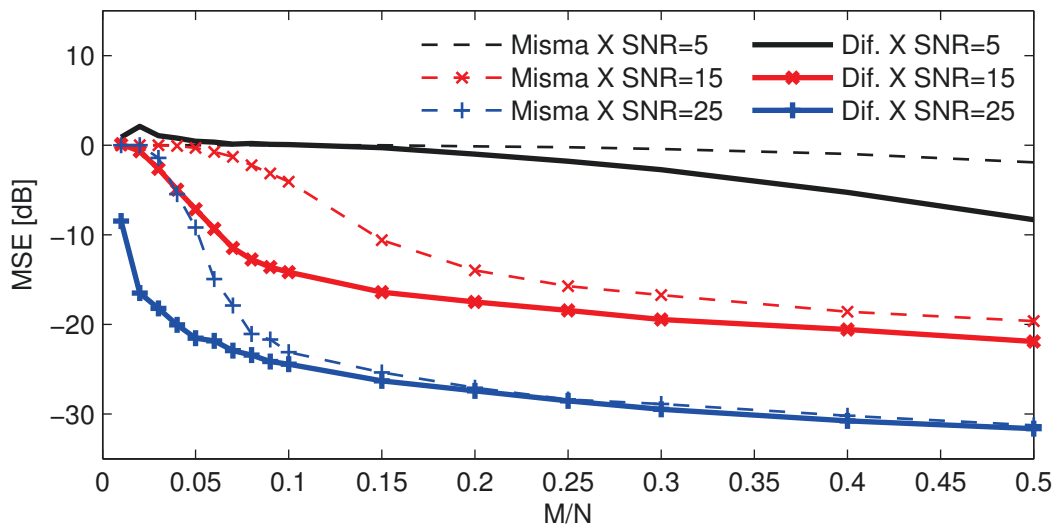


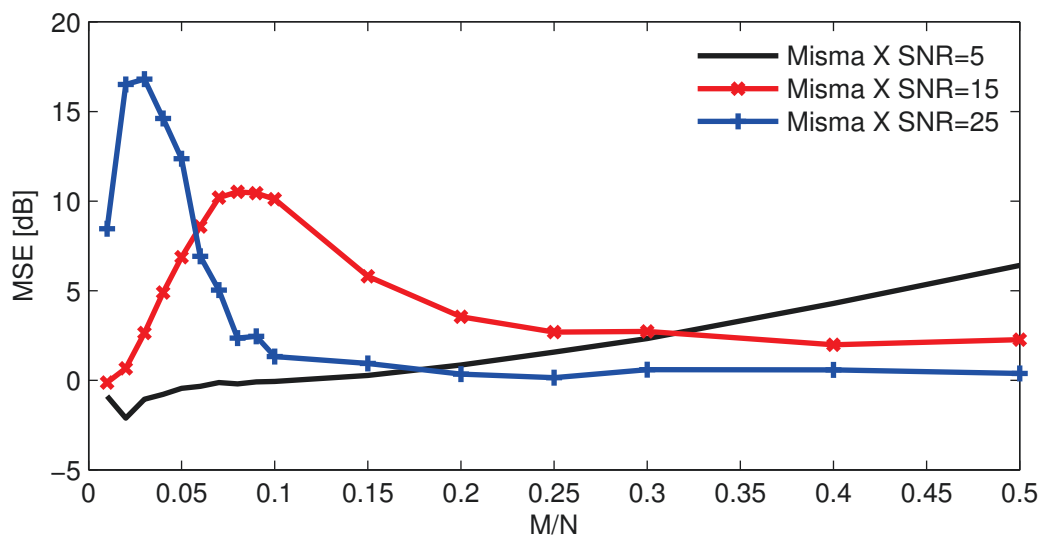
Figura 2.4: Error en la estimación de \mathbf{b} para diferentes cantidades de instantáneas disponibles D , para un $SNR=15$ dB y una raleza $S=1$.

al ir agregando distintos diccionarios una vez que se forma una base del espacio de señal, los nuevos diccionarios son combinaciones lineales de ellos y no aportan información sustancial para la resolución del problema inverso. Por otro lado, este comportamiento nos indica que se puede obtener una gran ganancia con sólo unas pocas instantáneas. Esta es una gran ventaja ya que el tiempo de cómputo del algoritmo crece con la cantidad de instantáneas que se tomen.

En la Figura 2.5 se analiza el desempeño para distinta cantidad de mediciones M . En la Figura 2.5a, se muestra el error en la estimación de ambos métodos para 3 niveles de SNR, 5, 15 y 25 dB. Es claro que no siempre se puede conseguir una ganancia implementando el muestreo aleatorio propuesto. En el caso en que se tenga control sobre el total de mediciones a tomar en cada instantánea, la Figura 2.5b permite cuantificar este resultado. Si el nivel de trabajo para la SNR es alto, con pocas mediciones se puede obtener una gran ganancia. A medida que disminuye el nivel de SNR, se debe utilizar una mayor cantidad de mediciones para que el método propuesto brinde una ventaja comparativa.



(a)



(b)

Figura 2.5: (a) Error en la estimación de \mathbf{b} y (b) ganancia para diferentes cantidades de mediciones M disponibles en cada instantánea, con tres niveles de SNR, 5, 15 y 25 dB y una raleza $S=1$.

\mathbf{X}_d basada en arreglos de sensores

En segundo lugar se analiza el caso cuando las matrices \mathbf{X}_d poseen una estructura particular, detallada en el Capítulo 3. Se tiene un arreglo de $V = 60$ sensores

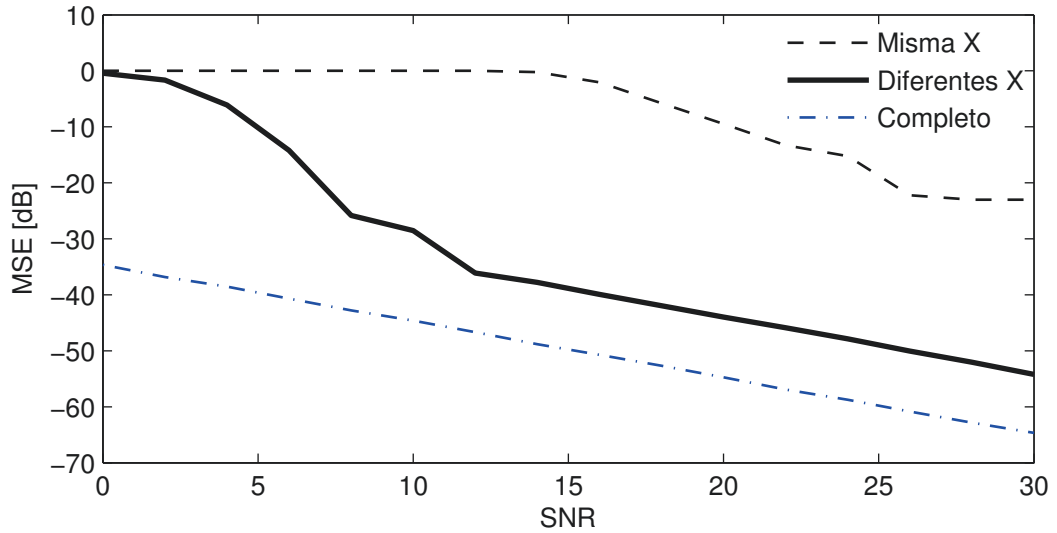


Figura 2.6: Error en la estimación de \mathbf{b} para diferentes SNR, utilizando $D = 10$ instantáneas y una raleza $S = 1$, para una aplicación de arreglo de sensores.

espaciados uniformemente a $0,2\lambda$ en forma lineal. Para este arreglo se genera la matriz \mathbf{X} de tamaño $V \times N$ con $V > M$ cuyas columnas son los vectores variedad del arreglo. Esta matriz se corresponde con el diccionario inicial de la aplicación. Se adquieren $D = 10$ instantáneas de un vector de mediciones de tamaño $M = 5$, que se obtiene al aplicar la matriz \mathbf{X}_d al vector de señal \mathbf{b} . Esa matriz se genera a partir de M filas distintas del diccionario original. Nuevamente se estudiarán dos opciones, la primer manteniendo la misma matriz $\mathbf{X}_d = \mathbf{X}_1$ en todas las instantáneas y la segunda tomando una distinta para cada instantánea. En esta aplicación, con el fin de realizar una justa comparación, los arreglos deben tener la misma apertura, por lo tanto la primera y la última fila de \mathbf{X} , correspondientes al primero y al último elemento del arreglo de sensores, forman parte de la matriz \mathbf{X}_d . Se genera una señal de interés \mathbf{b} de tamaño $N = 100$ cuya raleza es $S = 1$.

En la Figura 2.6 se muestra el error cuadrático medio de la estimación de \mathbf{b} para distintas relaciones señal a ruido SNR. En esta aplicación la ganancia es aún mayor, para obtener un error de 0,03 puede utilizarse una potencia de señal 30 veces menor al usar distintos sensores para cada instantánea. Nuevamente, esta ganancia

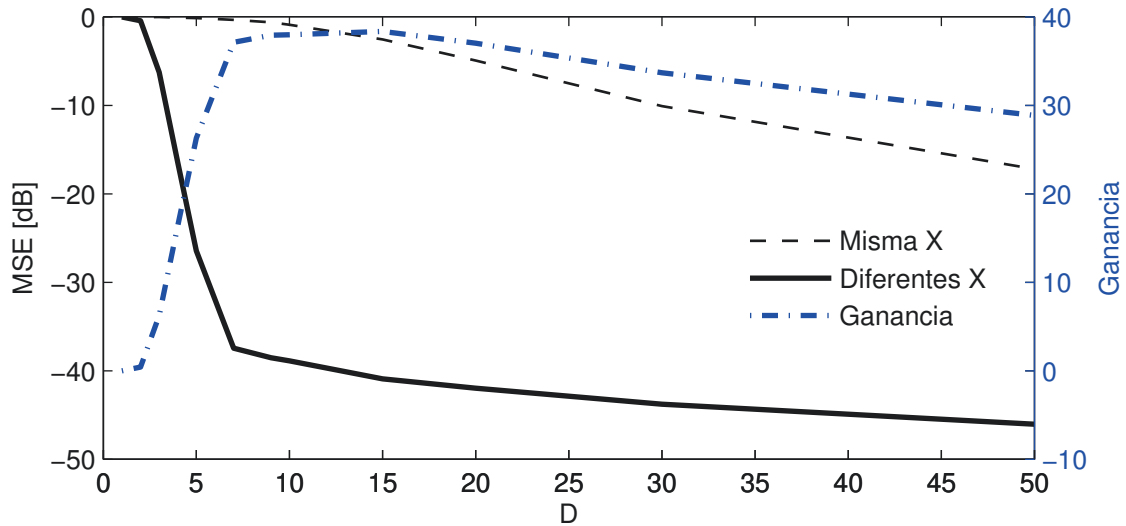
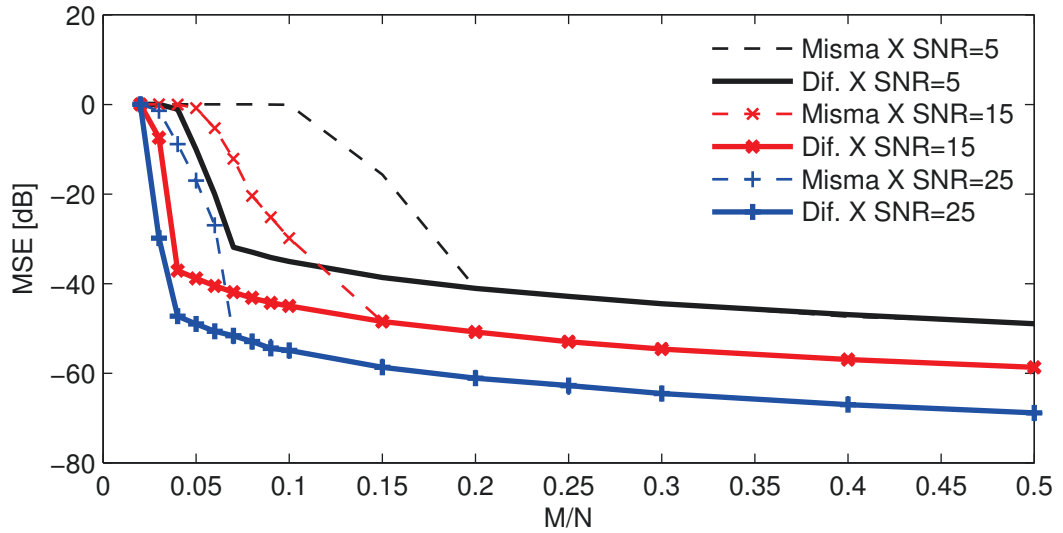


Figura 2.7: Error en la estimación de \mathbf{b} para diferentes cantidades de instantáneas disponibles, para un SNR= 15 dB y una rareza $S = 1$, para una aplicación de arreglo de sensores.

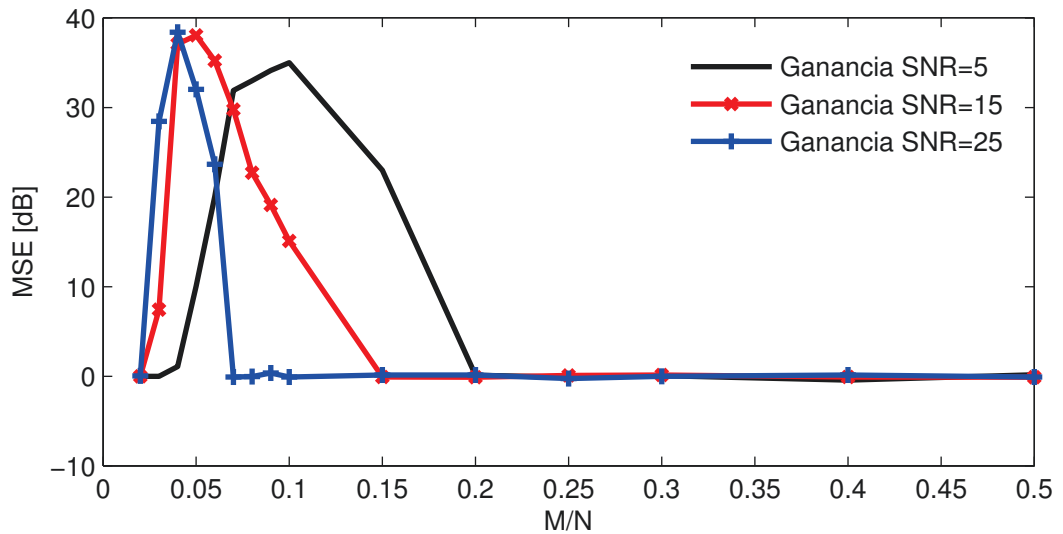
en potencia de señal sufre el costo de un mayor tiempo de cómputo del algoritmo de estimación. También se muestra, como referencia, el desempeño al utilizar todos los sensores (el arreglo completo) en cada instantánea.

En la Figura 2.7 se muestra el error para distinta cantidad de instantáneas D , con una relación señal a ruido fija de 15 dB. En este caso se alcanza rápidamente la saturación. Para este ejemplo con sólo 5 a 10 instantáneas se obtiene prácticamente la ganancia máxima.

Finalmente, en la Figura 2.8 se muestra el error para distinta cantidad de sensores M utilizados en cada instantánea, con tres niveles de SNR, 5, 15 y 25 dB. En esta aplicación, se obtiene una ganancia importante en un amplio rango de nivel de SNR. Adicionalmente, esta ganancia se logra utilizando un bajo número de sensores. En base a estos resultados la aplicación podría estimar el nivel de SNR y decidir la cantidad de sensores a medir.



(a)



(b)

Figura 2.8: (a) Error en la estimación de \mathbf{b} y (b) ganancia para diferentes cantidades de mediciones M disponibles en cada instantánea, con tres niveles de SNR, 5, 15 y 25 dB y una rareza $S=1$, para una aplicación de arreglo de sensores.

2.6. Discusión y conclusiones

En este capítulo se introdujeron las representaciones ralas, base de la tesis. En base a estas representaciones ralas se definió el problema inverso cuya resolución permite estimar señales que puedan ser descritas por estas representaciones. Se describieron distintas clases de algoritmos para la resolución del problema, identificando sus ventajas y desventajas.

Entre esos algoritmos se encuentra el algoritmo iterativo ESBL cuyo funcionamiento se describió en detalle ya que es usado ampliamente en la tesis y es objeto de estudio en el Capítulo 4 para una aplicación estimación de dirección de mediante un arreglo de sensores.

Finalmente se introdujo una modificación del modelo ralo para optimizar su utilización en aplicaciones capaces de realizar varias mediciones con diferentes sensores, dentro de un intervalo de coherencia de la señal de interés. Al tener varias instantáneas, se propuso utilizar un diccionario distinto para cada una de ellas y compararlo con el procedimiento clásico de utilizar siempre el mismo. Se mostró que esta elección mejora el desempeño con el costo de un mayor tiempo de cómputo del algoritmo de estimación.

Capítulo 3

Arreglos no uniformes lineales

3.1. Introducción

Un arreglo de antenas consiste en un número de sensores ubicados en una determinada posición relativa cuyo fin es el de adquirir muestras espacio temporales del campo electromagnético circundante. En contraste con una sola antena, el arreglo provee información adicional para mejorar la relación señal a ruido, resolver diferentes señales y cancelar interferencias [82]. Una importante aplicación de los arreglos de sensores es la caracterización del campo electromagnético incidente con el fin de determinar el número de fuentes, sus ubicaciones y las señales transmitidas [83]. La complejidad y el costo de un arreglo de sensores aumenta con la cantidad de sus elementos. Por lo tanto, un buen diseño debe considerar en maximizar el desempeño del arreglo sin incrementar el número de sensores.

En este capítulo se presentan tres configuraciones de arreglos lineales de sensores. Se analiza su desempeño comparándolos con un arreglo ULA con el mismo número de sensores y apertura. Se evalúa la capacidad de estos arreglos para reconstruir una señal rala aplicando diferentes técnicas de compressive sensing. Mediante ejemplos numéricos se muestra que la capacidad de los arreglos aleatorios mejora si se agrega una restricción mínima a la distancia entre sensores. Adicionalmente, se analizan arreglos optimizados mediante algoritmos genéticos que cumplen diferentes requerimientos para una buena estimación de la señal de interés. Mediante simulaciones se

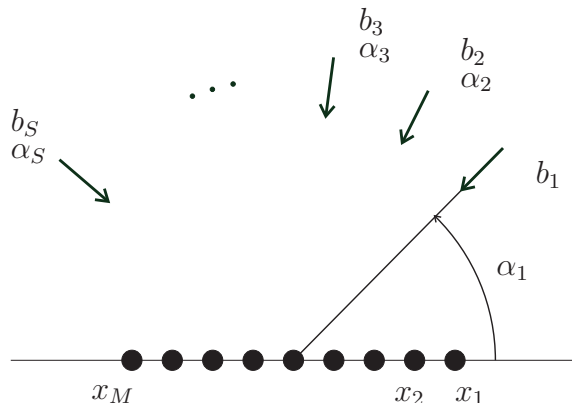


Figura 3.1: S señales incidentes a un arreglo de M sensores.

analiza el desempeño de estas configuraciones bajo distintas condiciones del entorno.

El capítulo se organiza como sigue. En la siguiente sección se presenta el modelo para arreglos de sensores. En la Sección 3.3, se definen las diferentes configuraciones de arreglos bajo estudio. En la Sección 3.4 se define un estimador de referencia para el algoritmo ESBL. En la Sección 3.5 se describe el método para diseñar algoritmos óptimos basados en algoritmos genéticos. Se analizan los resultados de simulaciones comparando estas configuraciones en la Sección 3.6. Finalmente, se concluye y se discuten futuros trabajos en la Sección 3.7.

3.2. Modelo

Se considera un arreglo lineal compuesto por M sensores y S señales incidentes de banda angosta provenientes de los ángulos $\alpha_1, \alpha_2, \dots, \alpha_S$, mostrado en la Figura 3.1. La envolvente compleja de la señal a la salida del arreglo es [84]

$$\mathbf{y} = \mathbf{V}(\boldsymbol{\alpha})\mathbf{b} + \mathbf{w}, \quad (3.1)$$

donde \mathbf{y} es el vector de mediciones de tamaño M , $\mathbf{V}(\boldsymbol{\alpha}) = [\mathbf{v}(\alpha_1), \mathbf{v}(\alpha_2), \dots, \mathbf{v}(\alpha_S)]$ es la matriz cuyas columnas corresponden a cada una de los vectores variedad del arreglo $\mathbf{v}(\alpha_i)$ de tamaño M para la dirección de arribo α_i , \mathbf{b} es un vector determinístico desconocido de tamaño S que representa las muestras de la envolvente compleja de las señales transmitidas, y $\mathbf{w} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$ corresponde al ruido y

errores de modelado. La estructura del vector variedad del arreglo para un ángulo α está dada por

$$\mathbf{v}(\alpha) = \begin{pmatrix} e^{j\frac{2\pi}{\lambda}x_1 \cos(\alpha)} \\ e^{j\frac{2\pi}{\lambda}(x_2-x_1) \cos(\alpha)} \\ \vdots \\ e^{j\frac{2\pi}{\lambda}(x_M-x_1) \cos(\alpha)} \end{pmatrix}, \quad (3.2)$$

donde x_1, x_2, \dots, x_M son las coordenadas de ubicación física de los elementos del arreglo, y λ es la longitud de onda de la señal incidente.

A partir de las mediciones anteriores, se busca generar una representación rala que se ajuste al modelo (2.9). En primer lugar se discretiza el espacio de ángulos de arriba en N componentes, $\theta_1, \theta_2, \dots, \theta_N$ formando una grilla. Usualmente se tiene que $N > M > S$, resultando en una matriz \mathbf{X} de $M \times N$ cuyas columnas corresponden a vectores variedad del arreglo evaluado en esos ángulos, que serán los átomos del diccionario [4],

$$\mathbf{X} = \begin{pmatrix} e^{j\frac{2\pi}{\lambda} \cos(\theta_1)x_1} & \dots & e^{j\frac{2\pi}{\lambda} \cos(\theta_N)x_1} \\ \vdots & \ddots & \vdots \\ e^{j\frac{2\pi}{\lambda} \cos(\theta_1)(x_M-x_1)} & \dots & e^{j\frac{2\pi}{\lambda} \cos(\theta_N)(x_M-x_1)} \end{pmatrix}. \quad (3.3)$$

Luego, el vector \mathbf{b} de tamaño N se convierte en ralo con S componentes distintas de cero, iguales al número de fuentes. Este problema puede extenderse para el caso de múltiples vectores de medición (MMV) descrito previamente. Si se considera una secuencia de D instantáneas, los datos de salida son

$$\mathbf{Y} = \mathbf{XB} + \mathbf{W}, \quad (3.4)$$

donde las matrices \mathbf{Y} , \mathbf{B} , y \mathbf{W} se construyen utilizando los respectivos vectores de (3.1) como sus columnas. Asumiendo que las ubicaciones de las fuentes son constantes durante el tiempo de observación, la matriz \mathbf{B} tiene un número bajo de filas distintas de cero. El objetivo es estimar las columnas de \mathbf{B} con la condición de que existe un perfil de raleza común.

3.3. Configuraciones de arreglos

Para poder resolver el problema inverso se necesita conocer las ubicaciones x_1, x_2, \dots, x_M de los elementos del arreglo. En este capítulo presentamos tres posibles arreglos lineales no uniformes. Como línea base se considera el arreglo uniforme con una distancia constante interelemento d y una apertura $L = d(M - 1)$. Para una comparación justa, cada arreglo posee el primer elemento en $x_1 = 0$ y el último en $x_M = L$; así todas las configuraciones poseen la misma cantidad de sensores M y la misma apertura L .

La primera configuración, denominada arreglo afinado (thinned), se basa en un ULA de K elementos con $K > M$. Las K posiciones están distribuidas uniformemente en la misma apertura L , resultando en una menor distancia inter-elemento $L/(K - 1)$. El arreglo no uniforme resulta de elegir aleatoriamente M de las posibles K ubicaciones, sin repetición. La Figura 3.2b muestra un ejemplo de esta configuración, donde los círculos blancos son las posiciones equiespaciadas y los círculos negros corresponden a los sensores separados irregularmente.

La segunda configuración, denominada arreglo aleatorio, se ilustra en la Figura 3.2c. En ésta la posición de los sensores se elige aleatoriamente según una distribución uniforme, $x_i \sim \mathcal{U}(0, L)$.

Finalmente, en la Figura 3.2d se muestra la tercera configuración analizada, denominada arreglo restringido. Los sensores también se ubican aleatoriamente en la apertura L . Sin embargo, existe una cota inferior d_{min} para la distancia mínima entre elementos.

El desempeño de estos arreglos se analiza en la Sección 3.6. En general el arreglo restringido por una distancia mínima interelemento es el que presenta las mejores cualidades para distintas aperturas. Sin embargo, este desempeño puede mejorarse si se realiza una optimización posterior del arreglo. Para poder plantear este problema de optimización se necesita una función de costo o métrica del desempeño. En la siguiente Sección se describe una función de costo propuesta, con buenos resultados para altos valores de SNR, y a continuación se formula el problema de optimización y se describen extensiones de éste que incluyen restricciones adecuadas para mejorar

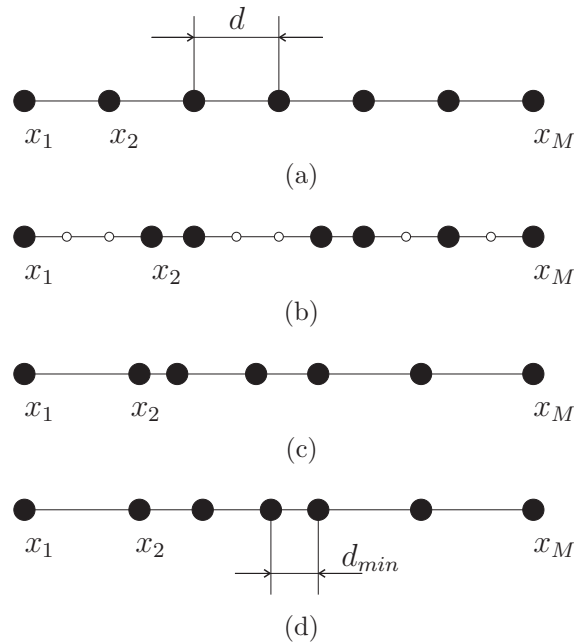


Figura 3.2: Diferentes configuraciones para una arreglo de M elementos con la misma apertura L . (a) Arreglo lineal uniforme, (b) arreglo afinado, (c) arreglo aleatorio, y (d) arreglo restringido.

aún más el desempeño.

3.4. Estimador perfecto

Para cuantificar el desempeño de las distintas configuraciones de arreglos presentadas se propone una métrica basada en un estimador perfecto u oráculo. Este estimador supone conocido el soporte de cada una de las columnas de \mathbf{B} y los parámetros estadísticos del modelo Σ_b y σ . Aunque el estimador es ideal y por lo tanto no realizable, provee una referencia contra la cual comparar el desempeño de los estimadores reales implementados por los algoritmos basados en modelos rales descritos en el Capítulo 2. Sabiendo que la distribución a posteriori de \mathbf{b} esta dada por la ecuación (2.20), el error cuadrático medio producido por este estimador

perfecto es

$$\begin{aligned}
\text{MSE}_p &= \frac{1}{D} \sum_{n=1}^N \sum_{d=1}^D \mathbf{E} \left\{ |\widehat{\mathbf{B}}_{nd} - \mathbf{B}_{nd}|^2 \mid \mathbf{Y} \right\} \\
&= \text{tr} (\boldsymbol{\Sigma}_{b|y}) \\
&= \text{tr} (\boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_S \mathbf{X}_S^H \boldsymbol{\Sigma}_y^{-1} \mathbf{X}_S \boldsymbol{\Sigma}_S) \\
&= \text{tr} (\boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_S \mathbf{X}_S^H (\mathbf{X}_S \boldsymbol{\Sigma}_S \mathbf{X}_S^H + \sigma \mathbf{I}_S)^{-1} \mathbf{X}_S \boldsymbol{\Sigma}_S) \\
&= \text{tr} (\boldsymbol{\Sigma}_S^{-1} + \mathbf{X}_S^H \sigma^{-1} \mathbf{I}_M \mathbf{X}_S)^{-1} \tag{3.5}
\end{aligned}$$

donde el conjunto \mathbb{S} está compuesto por los índices de las señales incidentes, generando así a partir de $\boldsymbol{\Sigma}_b$ la matriz $\boldsymbol{\Sigma}_S$ de $S \times S$ y a partir del diccionario \mathbf{X} la matriz \mathbf{X}_S de $M \times S$.

El objetivo de aumentar la apertura de las antenas a través de distancias interelemento mayores a la media longitud de onda es incrementar la capacidad del arreglo en discriminar señales cercanas. Por lo tanto, para este análisis, se suponen dos señales incidentes y se evalúa el desempeño cuando sus ángulos de arribo se asemejan. Esto se traduce en que la raleza de la señal de interés es $S = 2$, pudiendo representarse por dos columnas contiguas del diccionario $\mathbf{v}_1 = \mathbf{v}(\theta_1)$, $\mathbf{v}_2 = \mathbf{v}(\theta_2)$. Adicionalmente, sin perder generalidad, se supone que la potencia de estas señales es la misma, $\boldsymbol{\Sigma}_S = \beta \mathbf{I}_2$. Luego el error del estimador perfecto utilizado para resolver estas dos señales resulta,

$$\begin{aligned}
\text{MSE}_p &= \left(\frac{\mathbf{I}_S}{\beta} + \frac{\mathbf{X}_S^H \mathbf{X}_S}{\sigma} \right)^{-1} \\
&= \left[\begin{pmatrix} \frac{1}{\beta} & 0 \\ 0 & \frac{1}{\beta} \end{pmatrix} + \frac{1}{\sigma} \begin{pmatrix} M & \mathbf{v}_1^H \mathbf{v}_2 \\ \mathbf{v}_2^H \mathbf{v}_1 & M \end{pmatrix} \right]^{-1} \\
&= \frac{2\sigma \left(\frac{\sigma}{\beta} + M \right)}{\left(\frac{\sigma}{\beta} + M \right)^2 - \mu_{12}^2}, \tag{3.6}
\end{aligned}$$

donde $\mu_{12} = \|\mathbf{v}_1^H \mathbf{v}_2\|$ depende exclusivamente de las distancias entre elementos del

arreglo y de la separación entre señales.

$$\begin{aligned}\mu_{12}^2 &= \sum_{i=1}^M \sum_{j=1}^M \exp\left(j \frac{2\pi}{\lambda} (\cos(\theta_1) - \cos(\theta_2))(x_i - x_j)\right) \\ &= \sum_{i=1}^M \sum_{j=1}^M \exp\left(j \frac{2\pi \Delta\theta}{\lambda} (x_i - x_j)\right),\end{aligned}\quad (3.7)$$

donde $\Delta\theta = (\cos(\theta_1) - \cos(\theta_2))$. Si se considera una grilla uniforme para el parámetro $u = \cos(\theta)$ usualmente utilizado en la teoría de arreglos de sensores [84], el valor de $\Delta\theta$ puede tomarse como constante para señales contiguas. Este tipo de grilla permite tener mayor resolución en ángulos cercanos a $\pi/2$ (broadside) a costa de una menor resolución en ángulos rasantes al arreglo (grazing). En este caso $\Delta\theta = 2/M$ y

$$\mu_{12}^2 = \sum_{i=1}^M \sum_{j=1}^M \cos\left(\frac{4\pi}{M\lambda}(x_i - x_j)\right). \quad (3.8)$$

Esta cantidad podría aproximarse mediante una simple función de la varianza muestral, como se sugiere en [16]. Allí se busca minimizar su valor al maximizar la varianza de las posiciones de los elementos. Sin embargo para altos valores de apertura L esta aproximación es muy pobre y no resulta útil su uso. Esto implica que no se podrá trabajar con la varianza muestral como proxy y se tendrá que usar μ_{12} en el análisis.

Por otra parte, μ_{12} está directamente relacionada con la coherencia mutua del diccionario \mathbf{X} , definida como $\mu = \max_{k,l} |\mathbf{v}_k^H \mathbf{v}_l|$ [85]. La coherencia mutua es otra métrica comúnmente utilizada en modelos malos para caracterizar los diccionarios y en base a su valor poder garantizar determinado desempeño y capacidad de resolución del problema inverso.

3.5. Arreglos optimizados

Se propone un algoritmo genético para la minimización del estimador perfecto. Este tipo de algoritmos comienza con un numero de posibles configuraciones de arreglos, como población inicial. En este caso se utilizan diferentes realizaciones del arreglo restringido y del arreglo aleatorio. Estas configuraciones son luego filtradas

utilizando posibles restricciones al compararlas con respecto a ciertos umbrales. A cada una de estas configuraciones que pasa los umbrales se le asigna un valor de aptitud asociado con el MSE_p . Luego se realiza una recombinación o cruzamiento dándole más chances a las configuraciones con mejor aptitud. Adicionalmente se generan mutaciones de las configuraciones generando una mayor diversidad en la población de arreglos. Luego de estas operaciones se seleccionan los descendientes que posean la mejor aptitud y se añaden a la población para volver a repetir el proceso con la siguiente generación. Esto se repite hasta alcanzar un número suficiente de iteraciones en el cual la población ya no cambie y permite hallar la configuración del arreglo que minimice el MSE_p cumpliendo las restricciones [86].

3.5.1. Arreglo optimizado 1. Sin restricción

En primer lugar se propone simplemente minimizar el MSE_p , equivalente a minimizar (3.8), considerando las restricciones básicas de las posiciones de los elementos del arreglo,

$$\min_{\mathbf{x}} \sum_{i=1}^M \sum_{j=1}^M \cos \left(j \frac{4\pi}{M\lambda} (x_i - x_j) \right) \quad (3.9)$$

$$\text{tal que } x_1 = 0, \quad x_M = L, \quad x_1 < x_2 < \dots < x_{M-1} < x_M$$

3.5.2. Arreglo optimizado 2. Distancia mínima restringida

En base al buen desempeño del arreglo restringido, ilustrado en la siguiente Sección, se propone agregar esta restricción a la optimización fijando una distancia mínima entre sensores.

$$\min_{\mathbf{x}} \sum_{i=1}^M \sum_{j=1}^M \cos \left(j \frac{4\pi}{M\lambda} (x_i - x_j) \right) \quad (3.10)$$

$$\text{tal que } x_1 = 0, \quad x_M = L, \quad x_1 < x_2 < \dots < x_{M-1} < x_M$$

$$|x_i - x_j| > d_{min} \quad \forall \quad 1 \leq i, j \leq M$$

3.5.3. Arreglo optimizado 3. Distancia mínima y lóbulos laterales restringidos

Basándose en el clásico diseño de arreglos de sensores se puede agregar una restricción para minimizar los lóbulos laterales, error grosero del arreglo [87, 88]. Se desea minimizar la coherencia entre distintos vectores variedad $\mathbf{v}(\theta)^H \mathbf{v}(\phi)$. El ángulo de arribo de la señal θ se encuentra en un rango Θ y el ángulo ϕ pertenece al rango Φ donde se desea minimizar los lóbulos laterales. En principio, Θ abarca todo el espacio de ángulos de arribo de 180 grados, pero si se posee información a priori sobre la potencial dirección de arribo de la señal de interés, este rango puede ser mucho menor. Por otro lado, Φ deberá excluir el lóbulo principal.

$$\min_{\mathbf{x}} \sum_{i=1}^M \sum_{j=1}^M \cos \left(j \frac{4\pi}{M\lambda} (x_i - x_j) \right) \quad (3.11)$$

$$\text{tal que } x_1 = 0, \quad x_M = L, \quad x_1 < x_2 < \dots < x_{M-1} < x_M$$

$$x_i - x_j > d_{min} \quad \forall 1 \leq i, j \leq N$$

$$\max_{\theta, \phi} \left| \sum_{i=1}^M \exp \left(j \frac{2\pi x_i^2}{\lambda} (\cos(\theta) - \cos(\phi)) \right) \right| < a \quad \forall \theta \in \Theta, \phi \in \Phi$$

donde $0 < a < M$ representa la restricción en el nivel de los lóbulos secundarios del patrón del arreglo para los distintos ángulos de arribo.

3.6. Desempeño de los arreglos

En las simulaciones se utiliza un arreglo de $M = 30$ elementos, con $N = 100$ átomos en el diccionario, equivalente a 100 ángulos de arribo (DOA). Se procesan $D = 10$ instantáneas y el arreglo afinado se basa en un ULA de $K = 60$ elementos. La distancia interelemento d y la apertura L se normalizan con la longitud de onda λ . Se utiliza el error cuadrático medio por instantánea como medida de desempeño,

$$\text{MSE} = \frac{1}{D} \frac{1}{MC} \sum_{i=1}^{MC} \left\| \mathbf{B} - \hat{\mathbf{B}}^{(i)} \right\|_F^2, \quad (3.12)$$

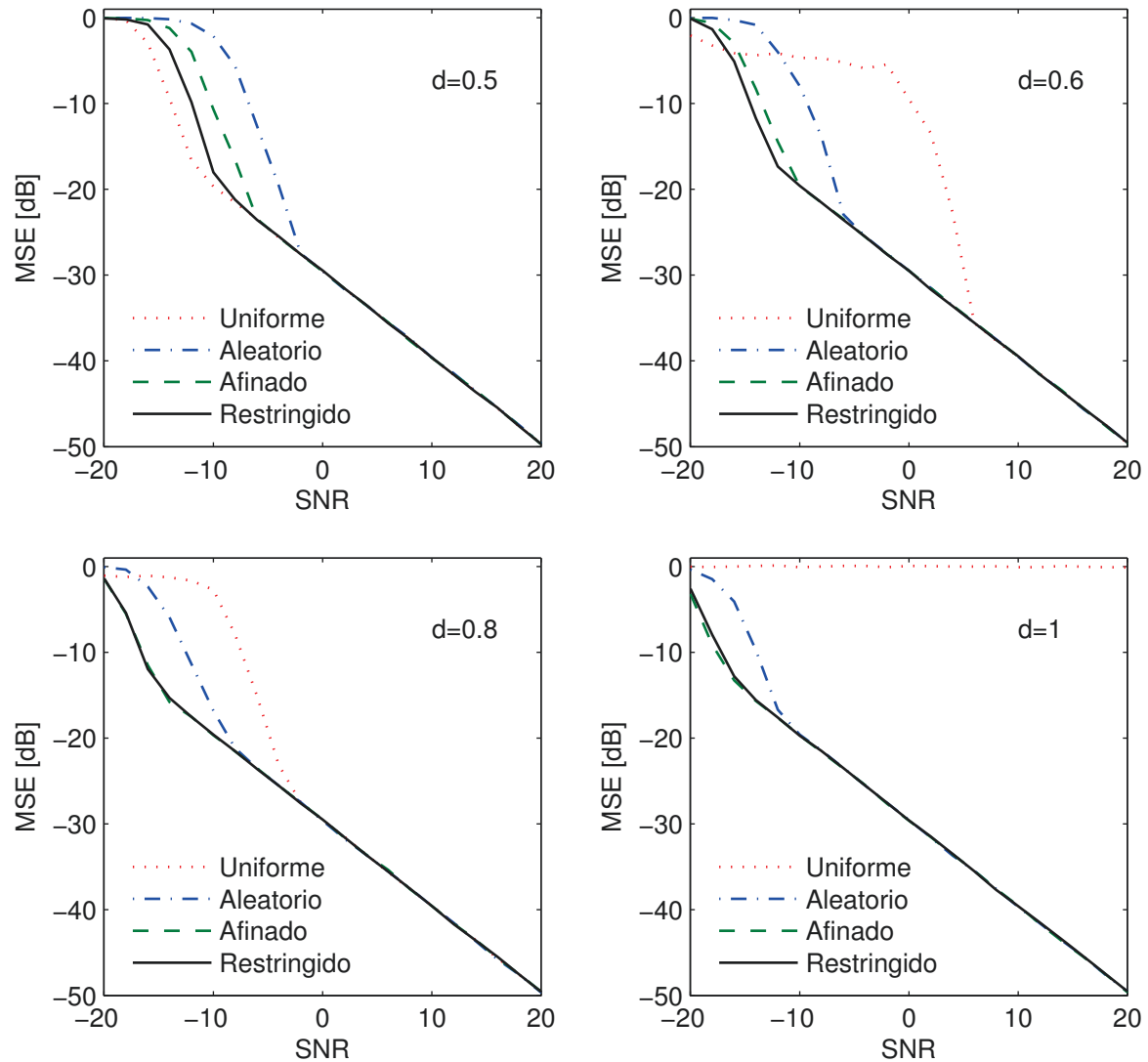


Figura 3.3: Desempeño de las distintas configuraciones del arreglo utilizando el algoritmo ESBL para valores de distancia interelemento d de 0.5, 0.6, 0.8 y 1.

donde $\hat{\mathbf{B}}^{(i)}$ representa la estimación de \mathbf{B} para la corrida i -ésima de Monte Carlo de las cuales se realizan $MC = 500$. Para cada realización Monte Carlo, se generan aleatoriamente los arreglos y las ubicaciones de las fuentes de señal.

En primer lugar se calcula el desempeño de los arreglos en función de la relación señal a ruido (SNR) para diferentes valores de la distancia interelemento d y por

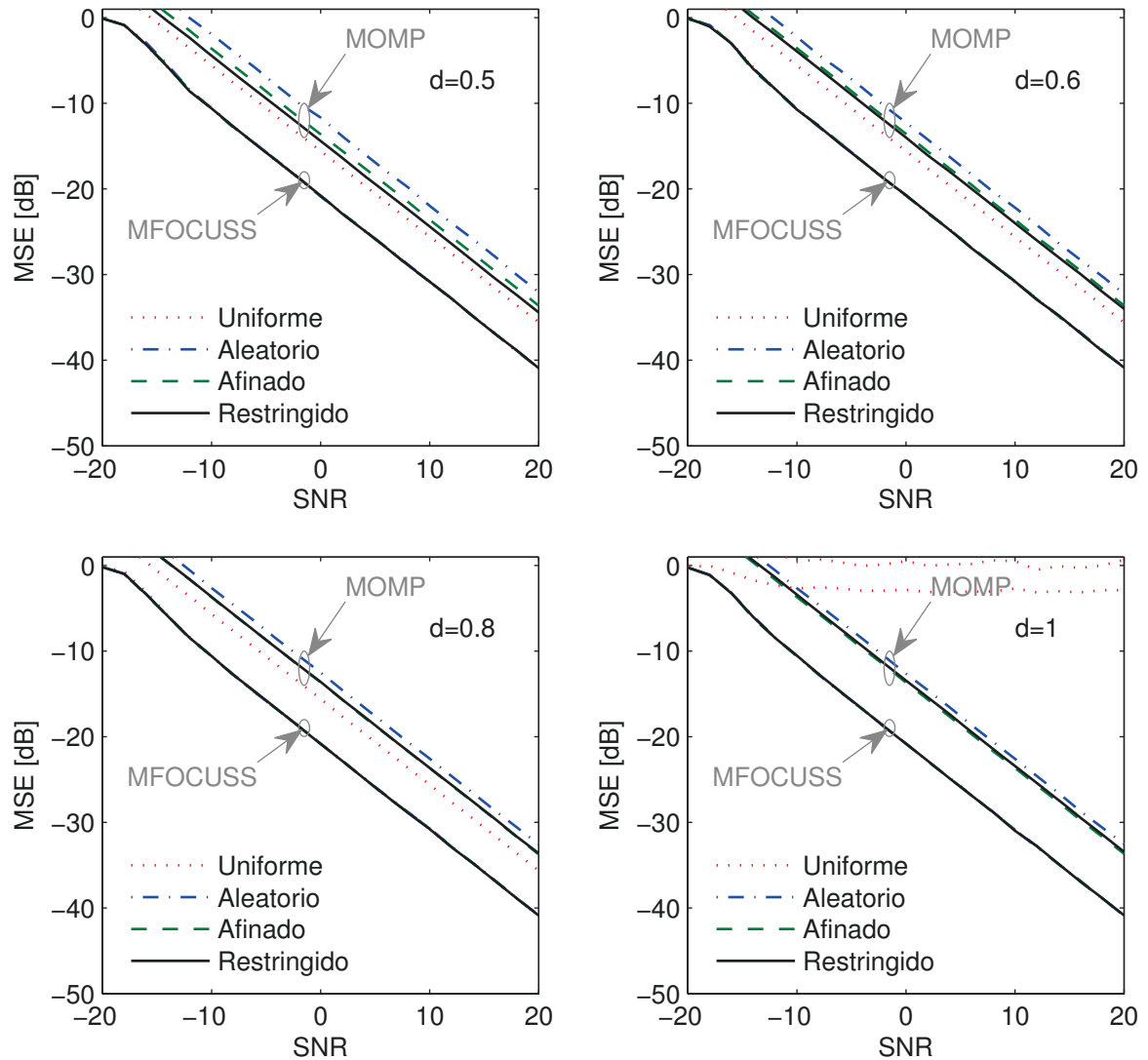


Figura 3.4: Desempeño de las distintas configuraciones del arreglo utilizando los algoritmos MOMP y MFOCUSS para valores de distancia interelemento d de 0.5, 0.6, 0.8 y 1.

consiguiente de apertura L . Se considera una sola fuente de señal, ($S = 1$) y una distancia mínima $d_{min} = 0,25$ para el arreglo restringido. La Figura 3.3 muestra los resultados cuando se aplica el algoritmo ESBL. Para $d = 0,5$, la mejor configuración es el arreglo uniforme, ya que se satisface la condición de Nyquist. Su desempeño empieza a degradarse cuando d aumenta, y falla completamente para una distan-

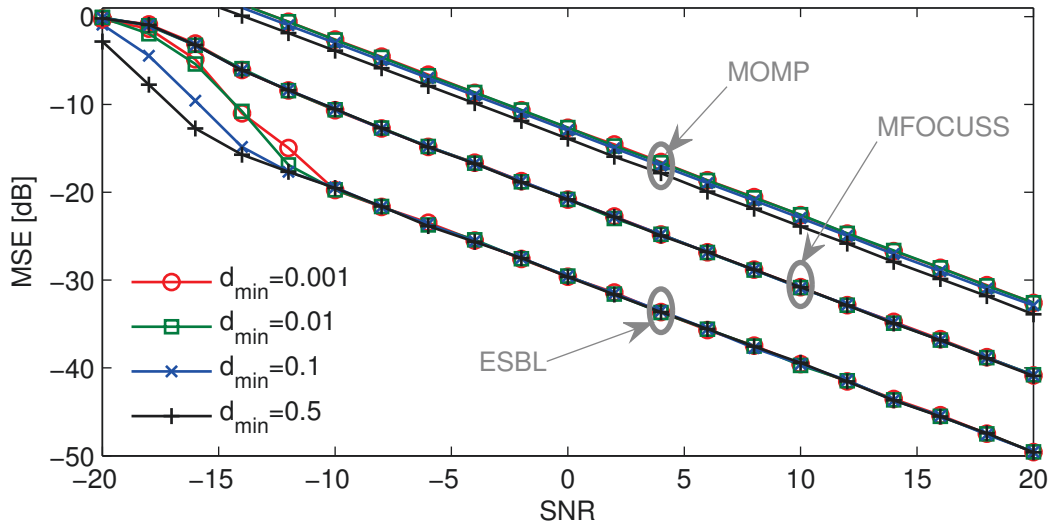


Figura 3.5: MSE para diferentes valores de distancia mínima interelemento usando el arreglo restringido y los algoritmos ESBL, MOMP and MFOCUSS.

cia $d = 1$. La Figura 3.4 muestra similares resultados para los algoritmos MOMP y MFOCUSS, respectivamente, excepto que no hay una diferencia notable entre configuraciones para MFOCUSS. Se nota que el mejor desempeño está dado por el algoritmo ESBL, como se esperaba, ya que está basado en un marco Bayesiano, el cual sobrepasa a los algoritmos greedy y de relajación [11, 69].

Como la configuración del arreglo restringido es la más interesante basada en las simulaciones previas, se fija el enfoque en sus parámetros. Se analiza el efecto de la distancia mínima interelemento d_{min} para una apertura fija de 29λ , equivalente a $d = 1$, y una única fuente de señal. La Figura 3.5 muestra que la sensibilidad del desempeño con respecto a d_{min} es muy baja para MOMP y prácticamente inexistente para MFOCUSS. En el caso de ESBL, a medida que se decrece la restricción, el arreglo se hace cada vez más similar al arreglo aleatorio y su desempeño se ajusta al de éste. Por el otro lado, para valores de d_{min} mayores a 0,5 el desempeño permanece constante. No existen indicios de desempeño pobre cuando se decrece la distancia mínima interelemento por debajo de $d_{min} = 0,5$, como se sugiere en [13, 14] para satisfacer la condición de RIP.

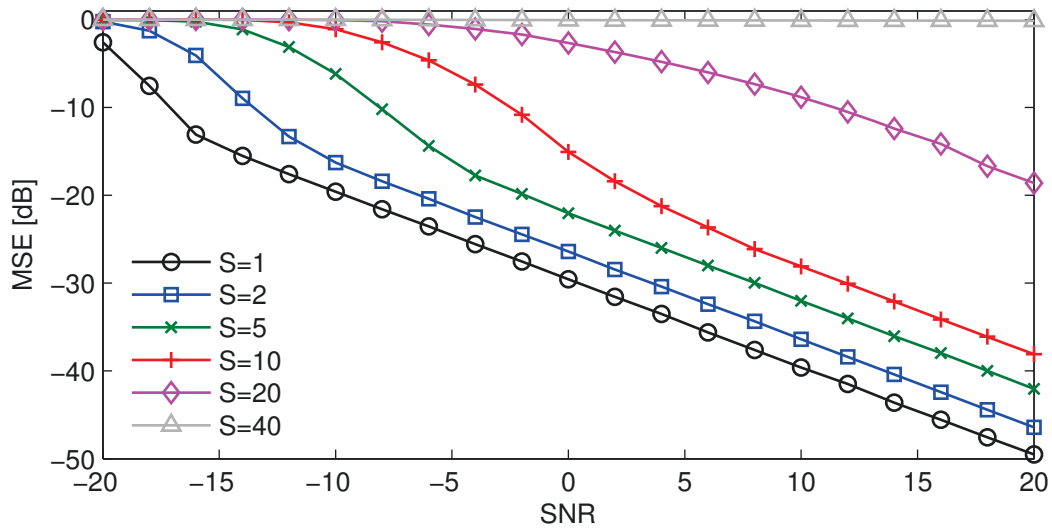
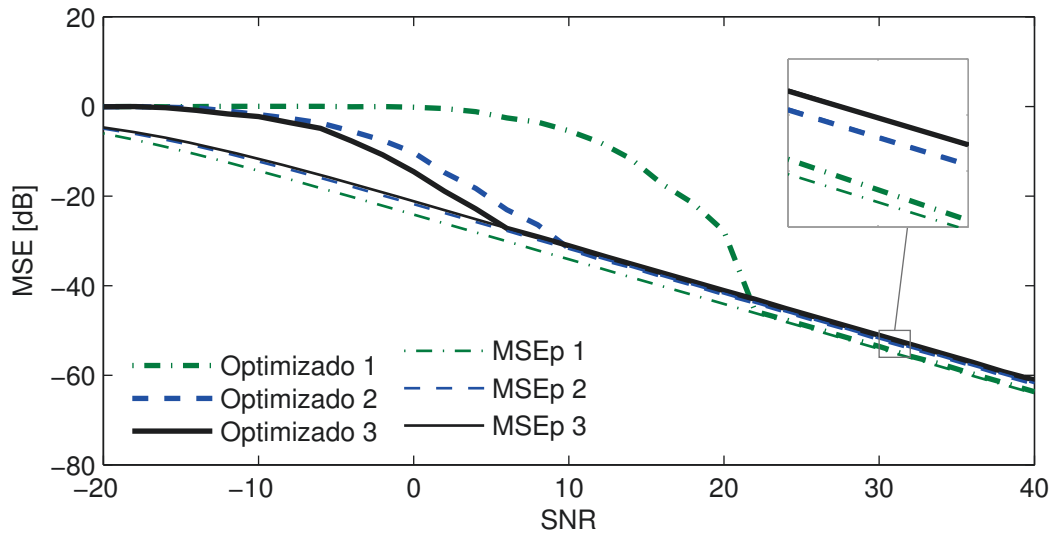


Figura 3.6: MSE para diferentes valores de raleza S usando el arreglo restringido.

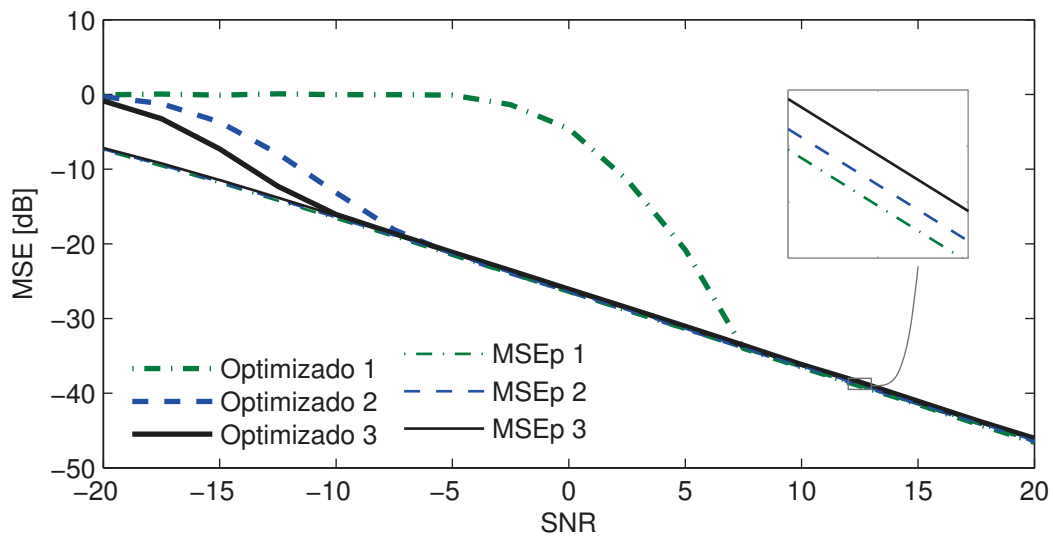
En la Figura 3.6 se muestra la degradación en el desempeño del arreglo restringido cuando se aumenta el número de fuentes de señal S (raleza de la señal) para una apertura fija $L = 29$, equivalente a $d = 1$, y con $d_{min} = 0,25$, utilizando el algoritmo ESBL. La degradación es monótona y se espera un desempeño razonable aún para un alto número de fuentes de señal.

En la Figura 3.7 se muestra el desempeño de los arreglos optimizados con las distintas restricciones. En este caso la raleza de la señal es $S = 2$ y se encuentran en ángulos contiguos del diccionario. Se analiza una apertura correspondiente a $d = 0,5$, mostrada en la Figura 3.7a, y otra correspondiente a $d = 1$ en la Figura 3.7b.

La primera optimización sólo busca minimizar μ_{12} que está relacionado con la coherencia mutua entre átomos contiguos del diccionario [85]. Esta medida, basada en el estimador perfecto, maximiza el desempeño en relaciones señal a ruido altas, como se ve en el detalle de la Figura 3.7a. Sin embargo, descuida el desempeño para relaciones señal a ruido bajas, como se observa claramente en la misma Figura. En segundo lugar se muestra el resultado cuando se agrega una restricción en la distancia mínima entre elementos, que en este caso es $d_{min} = 0,5$. Esta optimización logra una gran mejora en el desempeño a bajas SNR y el costo a pagar en altas SNR



(a)



(b)

Figura 3.7: Desempeño de los tres arreglos optimizados para dos aperturas correspondientes a (a) $d=0,5$ y (b) $d=1$.

es de aproximadamente 0.5 dB en el arreglo de $d=0,5$ y prácticamente imperceptible para $d=1$. Finalmente se agrega la restricción a los lóbulos laterales para un valor de $a=15$ (-3 dB) que logra un desempeño aún mejor a bajos SNR. Es interesante

notar que el desempeño alcanza el ideal cuando se tiene una buena relación señal a ruido.

En la Figura 3.8 se muestran los patrones del arreglo para los tres algoritmos optimizados para el caso $d = 1$. La primera optimización del MSE_p afina el lóbulo principal pero no optimiza los lóbulos laterales. La segunda optimización mejora los lóbulos laterales lejanos pero los secundarios adyacentes al lóbulo principal aún tienen una gran energía. La tercera optimización limita los lóbulos laterales a costa de ensanchar levemente el principal y aumentar los lóbulos lejanos, aunque aún dentro del límite propuesto. Adicionalmente se muestra la distribución de los sensores para cada configuración en la apertura. En la primera opción los elementos se ubican en dos grupos para aumentar μ_{12} pero sin la restricción, en cada grupo los elementos se encuentran muy cercanos, con el efecto de empobrecer la estimación a bajos SNR. En el segundo arreglo se siguen manteniendo los dos grupos pero los elementos ya no están tan pegados debido a la restricción de distancia mínima. Y en el tercer arreglo ya no es tan notable la existencia de dos grupos de sensores.

En la Figura 3.9 se compara el desempeño del arreglo optimizado según distancia mínima y lóbulos laterales restringidos con los arreglos previamente descritos para una señal de raleza $S = 2$. Se muestra el desempeño para distintas aperturas correspondientes a valores de $d = 1, 2, 4$ y 10 . Adicionalmente se incluye como referencia el desempeño de un arreglo uniforme de la misma cantidad N de elementos con distancia interelemento $d = 0,5$, que corresponde a una menor apertura. Esto muestra claramente el objetivo principal del trabajo que busca maximizar el desempeño del arreglo sin incorporar nuevos elementos. Para una apertura de $d = 1$ el arreglo optimizado es el que ofrece el mejor desempeño en todo el rango de SNR. A medida que aumenta la apertura esta ganancia es menor en bajas SNR, ya que todas las configuraciones tienen un comportamiento similar. Sin embargo, el arreglo restringido es el más robusto a estos cambios, ofreciendo un buen desempeño para las distintas aperturas.

En la Figura anterior puede también observarse que el arreglo afinado aumenta el error para $d = 2$ y aunque en menor medida algo similar le ocurre al arreglo optimizado para $d = 10$. La Figura 3.10 permite analizar este comportamiento en

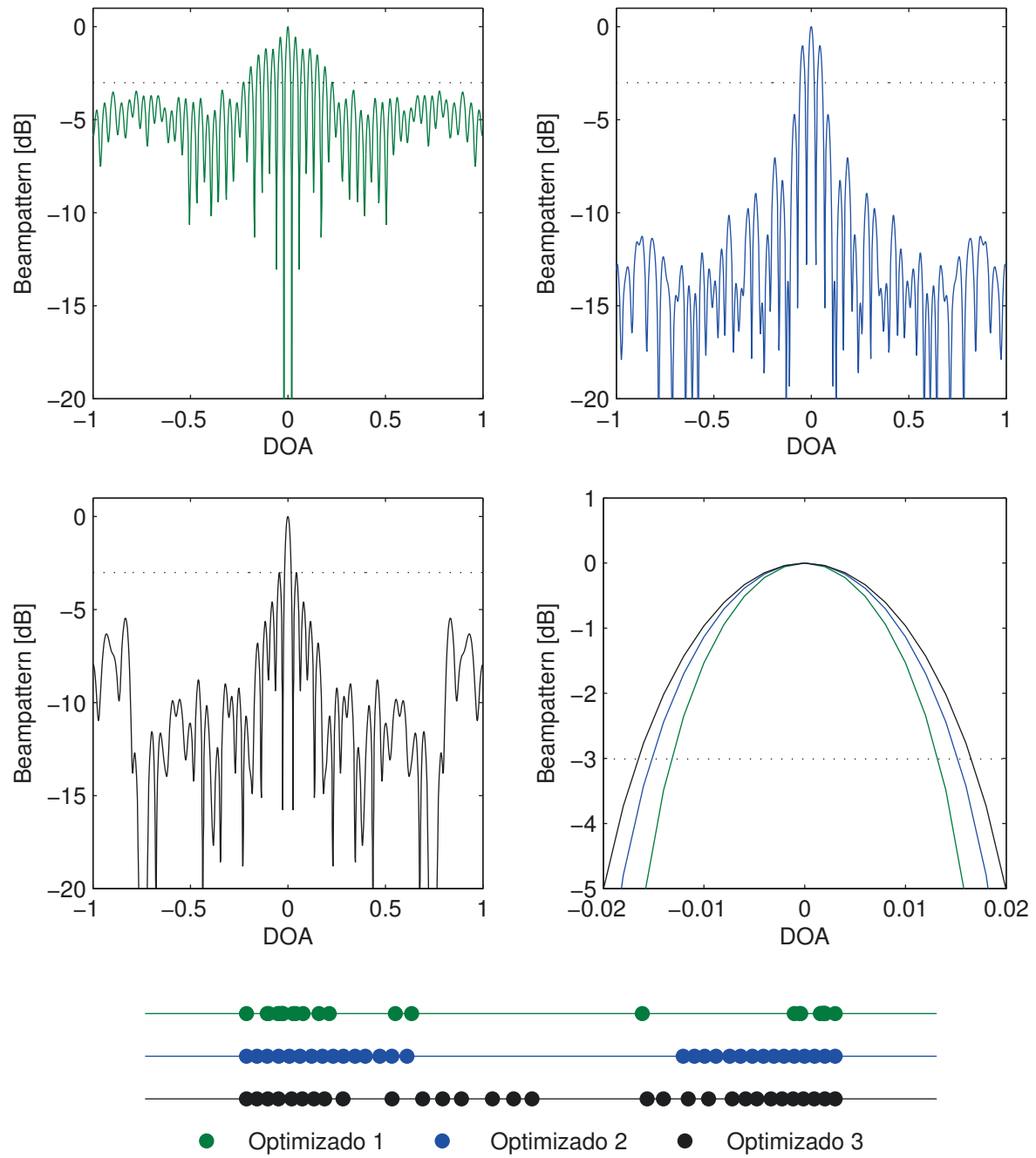


Figura 3.8: Espectros para de los distintos arreglos optimizados para $d=1$ y distribución de los elementos en la apertura.

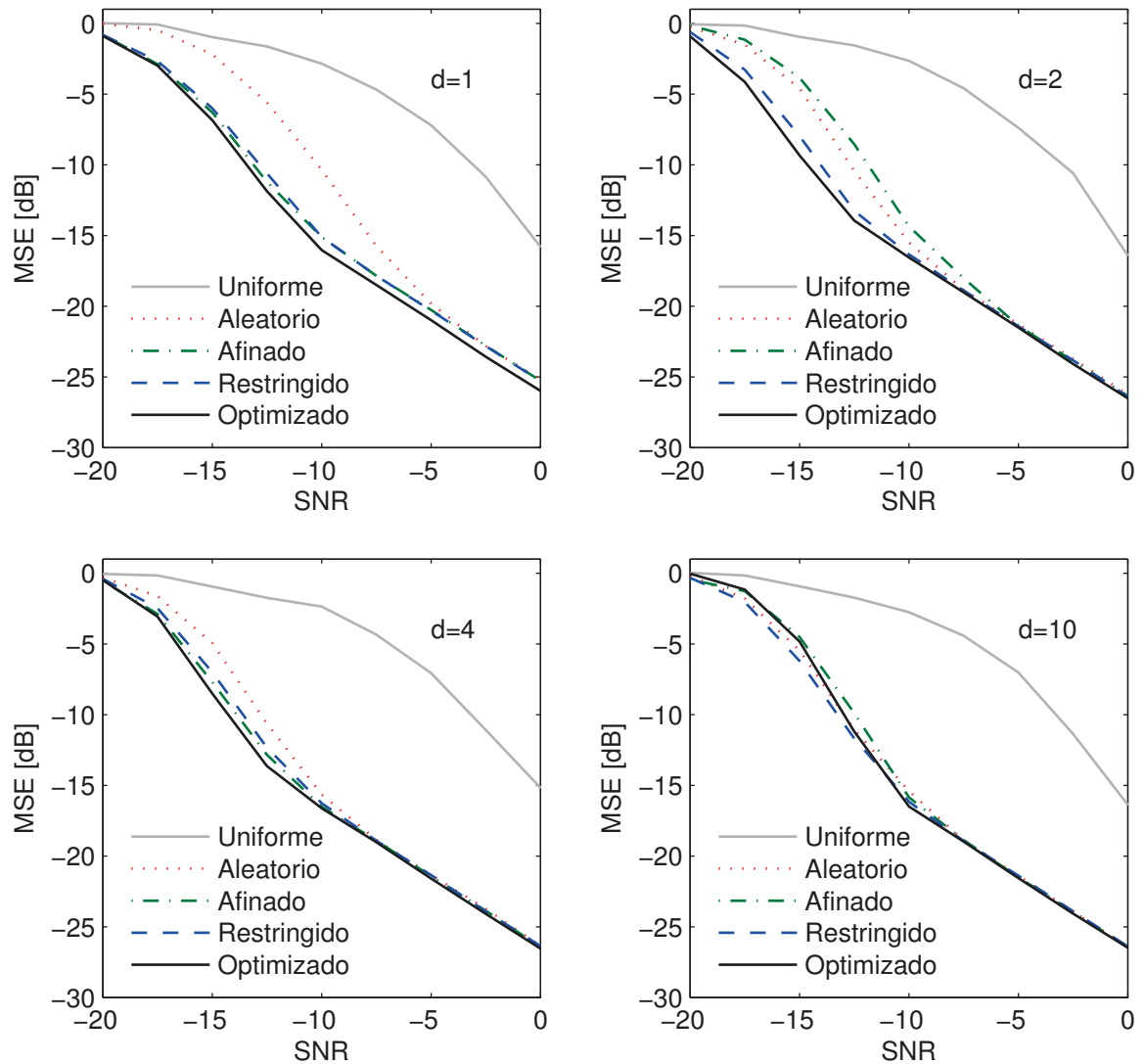


Figura 3.9: Desempeño de las tres configuraciones y las optimizaciones del arreglo para valores de distancia interelemento d de 1, 2, 4 y 10.

detalle.

Manteniendo el ULA de referencia de $K = 60$ elementos, la Figura 3.10a muestra este efecto para distintos valores de SNR sobre el arreglo afinado. Éste posee un comportamiento peculiar al incrementar la apertura, la mejora en su desempeño no es monótona. Al incrementar la distancia interelemento d , y consecuentemente su

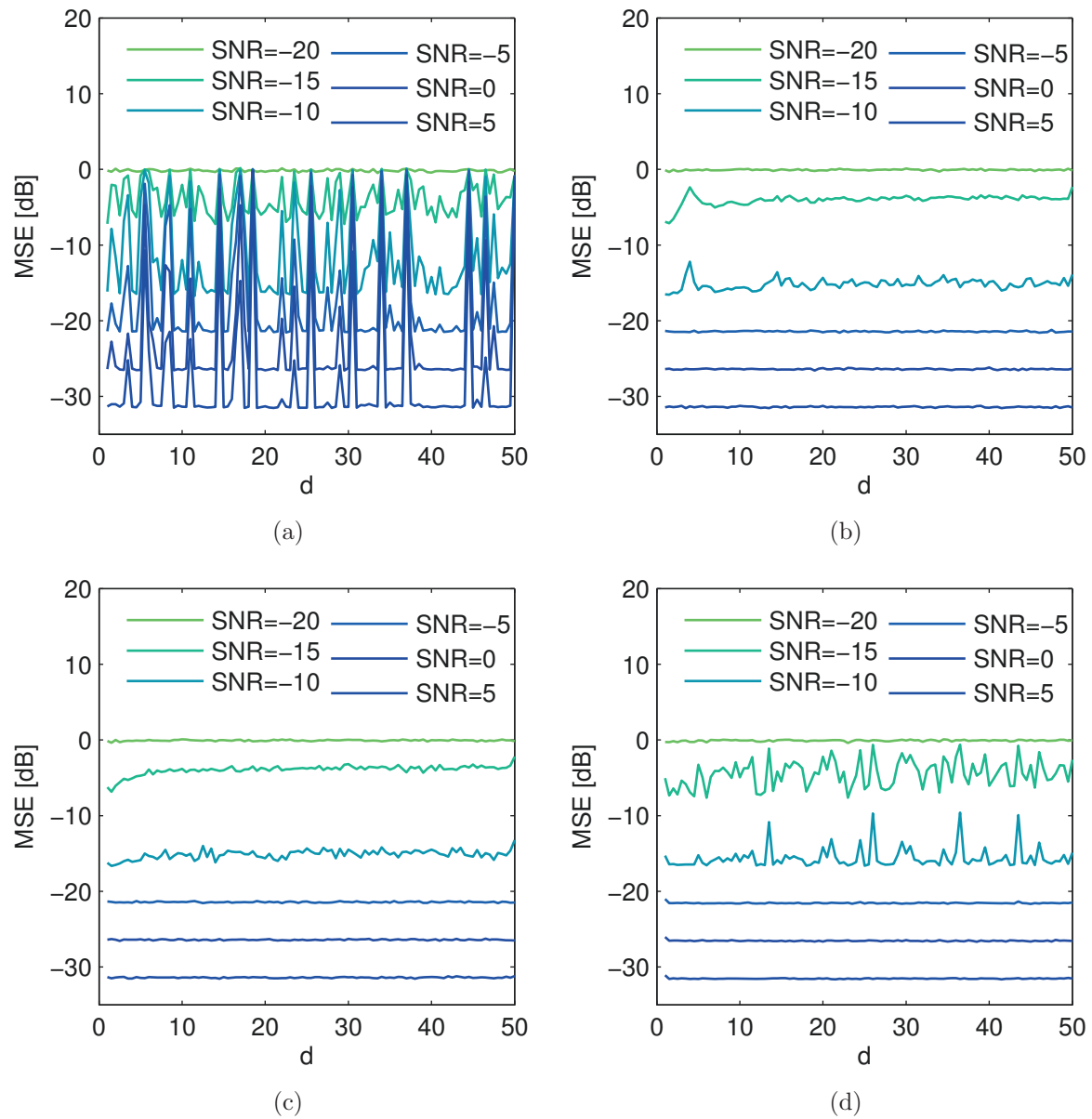


Figura 3.10: Variación del MSE al aumentar la apertura de los arreglos (a) afinado con $K=60$, (b) afinado con $K=2dN$, (c) restringido y (d) optimizado con ambas restricciones.

apertura, el MSE para cada SNR tiene saltos de desempeño pobre. Este efecto es debido a la estructura impuesta por el arreglo uniforme de referencia. Al cambiar este arreglo en un ULA de distancia interelemento $d=0,5$, correspondiente a $K=2dN$,

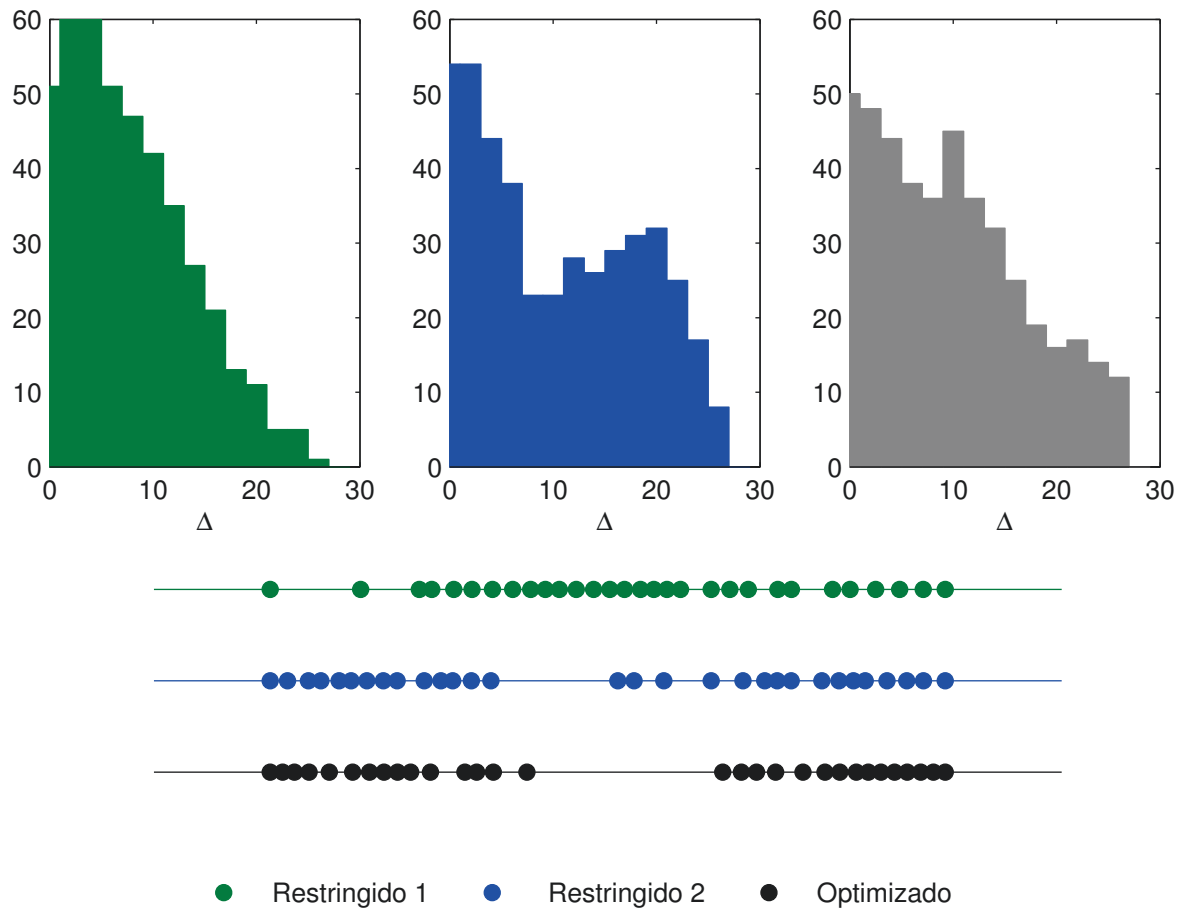


Figura 3.11: Histogramas de distancia interelemento Δ para dos realizaciones del arreglo restringido y una para el arreglo optimizado.

se muestra en la Figura 3.10b que este efecto desaparece. En las figuras Figura 3.10c y Figura 3.10d se muestra que el algoritmo restringido no posee este problema, y en el arreglo optimizado no es tan importante. Adicionalmente, estas figuras muestran que no es necesario aumentar la apertura más allá de la correspondiente unos pocos λ , para bajos SNR.

Finalmente, se analiza el desempeño de una realización específica del arreglo restringido y del arreglo óptimo con ambas restricciones. La Figura 3.11 muestra dos ejemplos del arreglo de $M = 30$ sensores para $d = 1$ y $d_{min} = 0,25$, y un arreglo optimi-

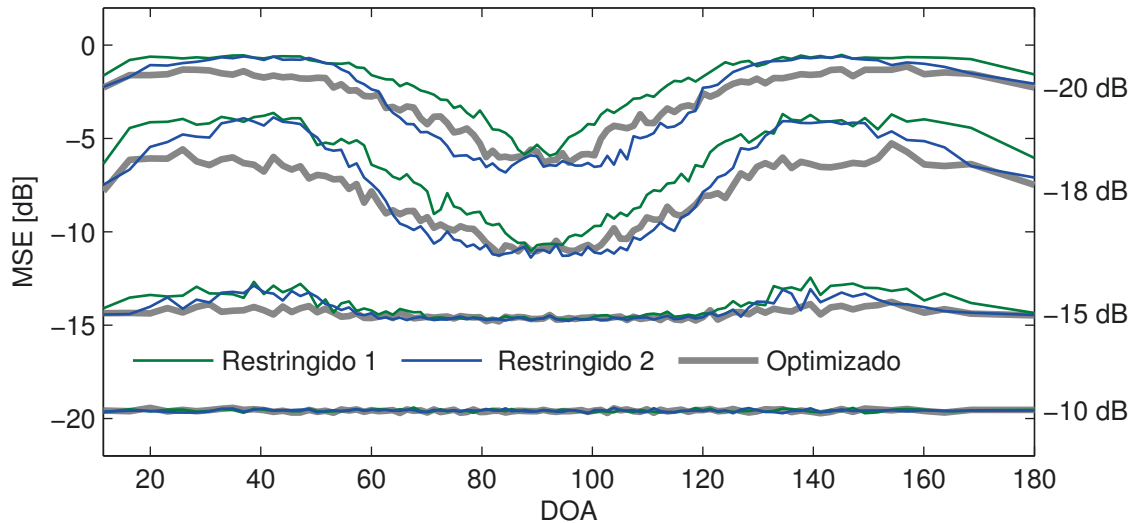


Figura 3.12: MSE en función del ángulo de arribo para valores de SNR entre -20 y -10 dB usando dos realizaciones del arreglo restringido y el arreglo optimizado.

zado. Adicionalmente se muestran los histogramas para las distancias interelemento reales $\Delta = |x_i - x_j|$ para cada una de estas realizaciones. Se aprecia que ambas configuraciones del arreglo restringido involucran diferentes perfiles de distancias y las ubicaciones físicas de los elementos se concentran en diferentes regiones.

Para valores de SNR altos, la Figura 3.12 muestra que el desempeño de estas tres realizaciones es prácticamente el mismo para todo el espacio de ángulos de arribo. Por otro lado, para bajos valores de SNR hay una diferente respuesta para señales que arriban de diferentes ángulos. El arreglo optimizado es el que posee una menor variación y el mejor comportamiento en todo el rango. Adicionalmente, la Figura 3.12 muestra que el arreglo restringido es robusto contra la eventual realización que define la posición de sus elementos. Se puede esperar obtener un desempeño similar sin importar la elección específica del arreglo restringido.

3.7. Discusión y conclusiones

El objeto del presente capítulo fue estudiar el impacto de la posición de los elementos del arreglo en su desempeño. En particular se analizaron arreglos no uniformes aplicando diferentes clases de algoritmos de compressive sensing para la estimación de direcciones de arribo de fuentes de señal y su información transmitida.

En primer lugar, el enfoque se centró en tres configuraciones: un arreglo estructurado con ubicaciones de los elementos prefijadas, un arreglo aleatorio y un arreglo aleatorio con una restricción en la distancia mínima interelemento.

En segundo lugar, se introdujeron arreglos optimizados que buscan minimizar el error de estimación. En base a simulaciones del desempeño que muestran que el arreglo restringido combinado con el algoritmo ESBL consiguen buenos resultados, se modificó la optimización para agregar el requisito de una distancia mínima entre sensores. Adicionalmente se incorporó una restricción para minimizar los lóbulos laterales del arreglo, permitiendo así un mejor desempeño. Se mostró que esta última optimización minimiza el error comparado con las otras configuraciones cuando se trabaja con valores de SNR bajos. Sin embargo, para altos valores de SNR es preferible el uso de la primera configuración, ya que alcanza un menor valor de error en la estimación.

También se mostró que cuando se aumenta la apertura, el desempeño de los los arreglos afinados y optimizados no tienen una degradación monótona, lo cual indica que dependiendo de la realización el aumentar la apertura no implica una mejora en el desempeño. Este efecto no ocurre con el arreglo restringido, por lo cual resulta más robusto en este sentido.

Finalmente se estableció que el desempeño del arreglo restringido es similar para diferentes realizaciones. Adicionalmente, para un escenario de altos valores de SNR, su comportamiento es constante para distintos ángulos de arribo de la señal de interés. Lo mismo ocurre con el arreglo optimizado, con la ventaja de que también posee una menor variabilidad en su desempeño para valores bajos de SNR.

En este trabajo se consideraron señales cuyos ángulos de arribo correspondían a ángulos presentes en el diccionario. Esto no se ajusta a la realidad, donde el ángulo

de arribo es arbitrario. En trabajos futuros se analizará el caso de señales cuyos ángulos de arribo se encuentren entre dos columnas del diccionario, conocido como discrepancia de grilla (grid mismatch) [89]. Por otra parte, este tipo de análisis puede extenderse a arreglos planos en dos y tres dimensiones, donde se incorporan más grados de libertad para para mejorar el desempeño del arreglo.

Capítulo 4

Estimación de señales en modelos ralos con interferencias

4.1. Introducción

En muchas aplicaciones que involucran arreglos de sensores se desea estimar una señal de interés incidente y su dirección de arribo. Lamentablemente suele haber señales interferentes que también inciden sobre el arreglo, incorporando errores en la estimación de la señal de interés. Asumiendo que la cantidad de fuentes de señales de interés es escasa, este escenario de señal más interferencia se corresponde con el problema de estimación de señales ralas, debido a que en el espacio de posibles ángulos de arribo, sólo unos pocos ángulos poseen energía de la señal, resultando así en una rareza espacial. En general la interferencia se debe a efectos de multicamino; por lo cual, se tiene información parcial sobre el rango posible de su dirección de arribo. En este capítulo se analiza el desempeño del algoritmo ESBL cuando se lo aplica a las mediciones obtenidas mediante un arreglo de sensores, en las condiciones previamente mencionadas. Se realiza una comparación de su desempeño con con algoritmos similares.

El presente capítulo se organiza de la siguiente manera. En la Sección 4.2 se provee una breve introducción a los arreglos de sensores utilizados. En la Sección 4.3 cómo aplicar el algoritmo ESBL para el caso de arreglos de sensores. En la Sec-

ción 4.4, se ilustra el desempeño del algoritmo, obteniéndose mejoras sobre otros métodos. Finalmente, se concluye en la Sección 4.5.

4.2. Arreglos de sensores

En este capítulo se utilizan arreglos lineales uniformes (ULA) como el descrito en la Figura 4.1, que consiste en M sensores colineales y equidistantes separados una distancia d . En este arreglo inciden K señales de banda angosta, de las cuales K_b son señales de interés, con direcciones de arribo $\alpha_1, \alpha_2, \dots, \alpha_{K_b}$, y K_u son señales de interferencia, con direcciones de arribo $\gamma_1, \gamma_2, \dots, \gamma_{K_u}$. Estas últimas se presume que provienen de un rango entre γ_I y γ_F . La señal a la salida de los sensores resulta ser [84]

$$\mathbf{y} = \sum_{i=1}^{K_b} \mathbf{v}(\alpha_i) b_i + \sum_{j=1}^{K_u} \mathbf{v}(\gamma_j) u_j + \mathbf{w}, \quad (4.1)$$

donde \mathbf{y} es el vector de mediciones de largo M , $\mathbf{v}(\alpha_i)$ es el vector variedad del arreglo de largo M para la dirección α_i , b_i es el escalar que representa la información transportada por la señal de interés i -ésima, $\mathbf{v}(\gamma_j)$ es el vector variedad del arreglo para la dirección γ_j , u_j es la información de la interferencia j -ésima, y \mathbf{w} es ruido blanco. La estructura del vector variedad de un ULA para un ángulo α es la descrita por la ecuación (3.2). Las mediciones a la salida de los sensores (4.1) también pueden expresarse equivalentemente en forma matricial

$$\mathbf{y} = \mathbf{V}_b(\boldsymbol{\alpha})\mathbf{b} + \mathbf{V}_u(\boldsymbol{\gamma})\mathbf{u} + \mathbf{w}, \quad (4.2)$$

donde \mathbf{V}_b es una matriz de $M \times K_b$ cuyas columnas son los vectores variedad para cada dirección de arribo de las señales, \mathbf{b} es el vector de tamaño K_b con la información de las señales de interés, \mathbf{V}_u es la matriz de $M \times K_u$ equivalente para la interferencia, y el vector \mathbf{u} de tamaño K_u contiene la información de cada interferencia. Finalmente, si se toman D instantáneas, situación que usaremos para analizar el desempeño, también puede escribirse en forma matricial como

$$\mathbf{Y} = \mathbf{V}_b\mathbf{B} + \mathbf{V}_u\mathbf{U} + \mathbf{W}, \quad (4.3)$$

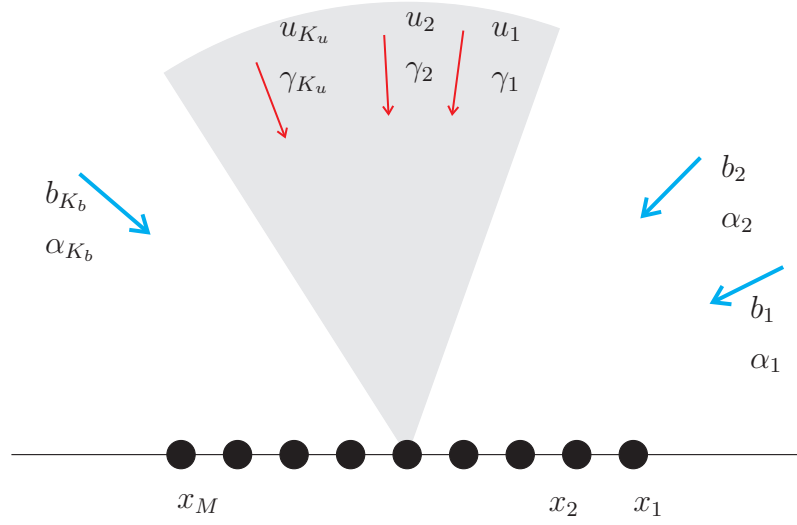


Figura 4.1: Arreglo de M sensores con K_b señales de interés y K_u interferencias incidentes.

donde las matrices \mathbf{Y} , \mathbf{B} , \mathbf{U} y \mathbf{W} se construyen usando los correspondientes vectores de (4.2) como sus columnas. Nótese que una fila de la matriz \mathbf{B} corresponde a muestras temporales de la envolvente compleja de la señal de interés.

4.3. Representación rala

El escenario de arreglo de sensores (4.3) puede ser convertido a una representación rala que se ajusta al modelo MMV (2.10). Como se describe en el capítulo previo, se discretiza el espacio de ángulos de arriba en N componentes. Así se genera el diccionario \mathbf{X} de $M \times N$, descrito por la ecuación 3.3, cuyas columnas corresponden a vectores variedad del arreglo evaluado en esos ángulos.

Los vectores variedad correspondientes a la interferencia serán columnas de la matriz \mathbf{X} , las que corresponden a los ángulos θ_i tales que $\gamma_I \leq \theta_i \leq \gamma_F$, formando la matriz \mathbf{Z} de tamaño $M \times P$,

$$\mathbf{Z} = (\mathbf{v}(\theta_I), \dots, \mathbf{v}(\theta_F)) = \begin{pmatrix} e^{j\frac{2\pi}{\lambda} \cos(\theta_I)x_1} & \dots & e^{j\frac{2\pi}{\lambda} \cos(\theta_F)x_1} \\ \vdots & \ddots & \vdots \\ e^{j\frac{2\pi}{\lambda} \cos(\theta_I)(x_M-x_1)} & \dots & e^{j\frac{2\pi}{\lambda} \cos(\theta_F)(x_M-x_1)} \end{pmatrix}. \quad (4.4)$$

Nótese que esta matriz representa los *posibles* ángulos de arribo de las interferencias, ya que contiene los vectores variedad de todos los ángulos discretizados dentro del rango. Sin embargo, el número real de señales de interferencia K_u es menor. En el escenario propuesto el algoritmo de estimación ESBL discrimina la señal de la interferencia a través de la información de dirección de arribo. Por lo tanto, el rango de ángulos $[\gamma_I, \gamma_F]$ de la interferencia no incluye la dirección de arribo de la señal. Otros problemas podrían considerar diferentes características de las señales para separar la interferencia de la señal de interés, como ser información sobre polarización o Doppler, pero ese escenario excede el alcance de esta tesis.

Por otra parte la información de la señal estará contenida en la matriz \mathbf{B} cuya fila i -ésima será distinta de cero si $\theta_i = \alpha_i$.

4.4. Análisis de desempeño

Para ilustrar el desempeño se propone el siguiente ejemplo para el escenario (4.3). Se utiliza un arreglo de $M = 20$ sensores con una distancia entre sensores $d = \lambda/2$ y una grilla de $N = 90$ ángulos posibles de arribo entre 0 y 180 grados. Se realizan D instantáneas donde el vector \mathbf{b} tiene una raleza $S = 2$ con dos señales que arriban desde $\alpha_1 = 120$ y $\alpha_2 = 160$ grados. Existen $K_u = 10$ interferencias que arriban aleatoriamente en el rango de $\theta_I = 20$ a $\theta_F = 100$, resultando en un rango con $P = 40$. Esto implica que una de las señales está cercana a la interferencia y la otra está prácticamente limpia. La potencia de las señales es $P_b = \|\mathbf{b}\|_2^2$ y la potencia de cada interferencia es $P_u = \|\boldsymbol{\Sigma}_u\|_F$. La potencia del ruido de mediciones es $\sigma = P_u/10$.

Para evaluar empíricamente el desempeño se realizaron 500 corridas de Monte Carlo. Se estima \mathbf{b} para distintas relaciones señal a interferencia definida como SIR = $10 \log_{10}(P_b/P_u)$ para una tasa de falsa alarma de $P_{FA} = 10^{-4}$. Se utilizaron los siguientes algoritmos para la comparación, MSBL, MFOCUSS, MOMP, ℓ_1 -SVD y MUSIC. Estos dos últimos no pueden estimar por sí solos el valor de la señal, sólo su espectro espacial (y consecuentemente las direcciones de arribo), por lo que se combinaron con el estimador Minimum Variance Distortionless Response (MVDR) [84] para realizar la estimación de la señal.

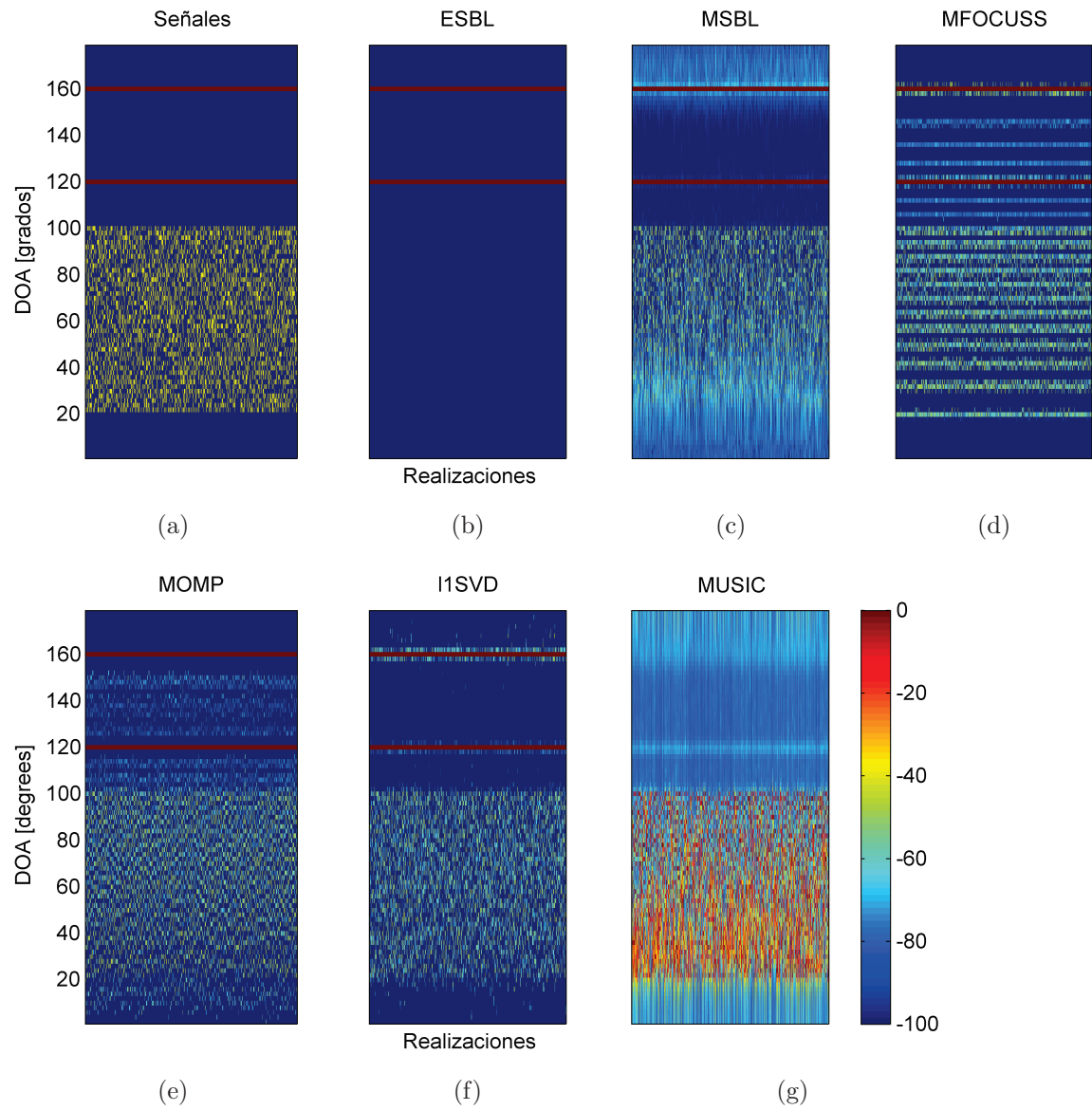


Figura 4.2: 500 Corridos de Monte Carlo del espectro espacial para dos fuentes con ángulos de arribo de 120 y 160 con 10 interferencias. (a) Valor verdadero de señal e interferencia, (b) ESBL, (c) MSBL (d) MFOCUSS (e) MOMP, (f) ℓ_1 -SVD y (g) MUSIC. (SIR=30 dB) usando una grilla de $N=90$ ángulos.

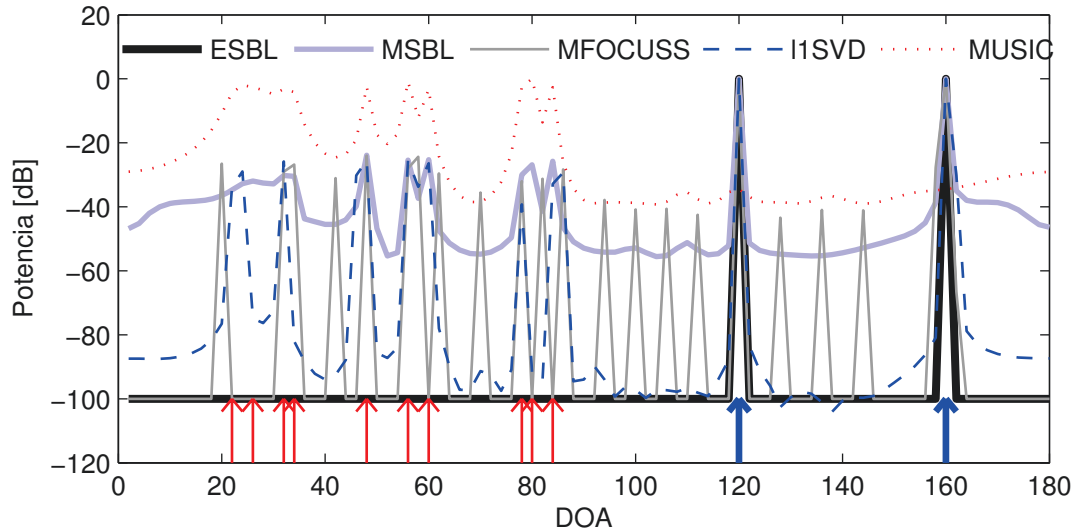


Figura 4.3: Ejemplo de un espectro espacial típico para dos fuentes con DOAs de 120 y 160 grados (flechas azules) y 10 interferencias en el rango de 20 a 100 grados (flechas rojas), usando los algoritmos MUSIC, ESBL, MSBL, ℓ_1 -SVD y MFOCUSS con un SIR=30 dB.

En la Figura 4.2 se muestran los espectros espaciales de las 500 corridas de Monte Carlo con una señal a interferencia SIR= 30 dB para los algoritmos mencionados. En la Figura 4.2a se muestran las señales originales en 120 y 160 grados y las interferencias que son aleatorias para cada corrida dentro del rango de 20 a 100 grados. En la Figura 4.2b se muestra el resultado del algoritmo ESBL que, al explotar la estructura de la interferencia y el conocimiento previo de que se encuentran en un rango determinado, puede recuperar los ángulos de arribo de ambas señales sin problemas. El resto de los métodos no pueden deshacerse de la interferencia y esto afectará en mayor o menor medida la estimación de la señal.

En la Figura 4.3 se muestra una realización del espectro espacial normalizado para algunos de los algoritmos analizados (ESBL, MSBL, MFOCUSS, ℓ_1 -SVD y MUSIC) en el escenario propuesto. Las direcciones de arribo de las señales de interés están marcadas con las flechas gruesas azules y las interferencias por las flechas finas rojas. Aquí puede apreciarse claramente que el espectro de ESBL es cero para todos los ángulos salvo para los de las señales, donde es máximo. MSBL también tiene

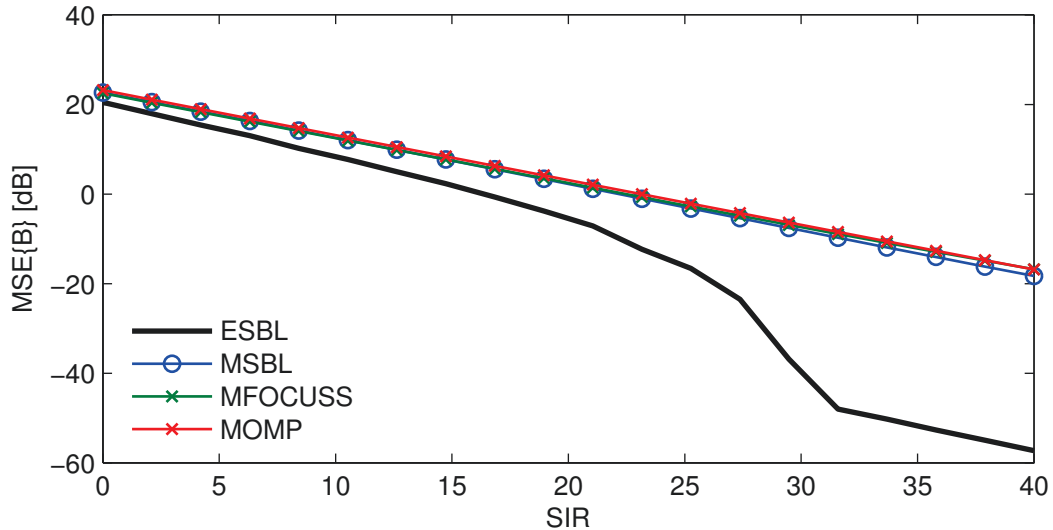


Figura 4.4: Error en la estimación completa del espectro espacial en función de SIR.

máximos en estos ángulos pero no tiene un gran rechazo de las interferencias. Lo mismo sucede con los algoritmos MFOCUSS y ℓ_1 -SVD.

Se analizan dos formas de evaluar o medir el desempeño de los algoritmos. Por un lado se quiere evaluar la capacidad del algoritmo de recuperar el espectro espacial correcto. Esto es equivalente a calcular el promedio sobre 500 simulaciones del error de estimación de la matriz \mathbf{B} , es decir

$$\text{MSE}\{\mathbf{B}\} = \frac{1}{MC} \sum_{i=1}^{MC} \|\mathbf{B} - \hat{\mathbf{B}}_i\|_F^2. \quad (4.5)$$

En la Figura 4.4 se muestra que utilizando esta medida el algoritmo ESBL es claramente el que presenta el mejor desempeño. Este resultado implica que el algoritmo ESBL recupera correctamente los ángulos de arribo de las señales de interés, sin interferencias. Para muchas aplicaciones ésta sería la medida más importante y decisiva para comparar los algoritmos.

En el caso de arreglos de sensores se desea saber el comportamiento específico en las direcciones de arribo de las señales, y recuperar la información transmitida por cada una de ellas correctamente. Por ello es que también se analiza el error en cada una de estas direcciones. En la Figura 4.5 se muestra el error en la estimación de la

señal proveniente de 120 grados, $\sum_{i=1}^{MC} \|\mathbf{B}(\alpha_1) - \widehat{\mathbf{B}}_i(\hat{\alpha}_1)\|_2^2/MC$, y en la Figura 4.6 el de la señal de 160 grados, $\sum_{i=1}^{MC} \|\mathbf{B}(\alpha_2) - \widehat{\mathbf{B}}_i(\hat{\alpha}_2)\|_2^2/MC$. El comportamiento del algoritmo ESBL con respecto a la señal cercana a la interferencia es similar a los otros algoritmos, lo que indica que en estos casos no habría una gran ganancia. Por otro lado, para la señal limpia proveniente de 160 grados el error en su estimación es mucho menor para el algoritmo ESBL. El desempeño del algoritmo ESBL en estas últimas dos figuras es mas cercano a los otros algoritmos porque esta medida de desempeño no refleja el error de elegir componentes espurios. Las Figuras 4.2 y 4.3 ilustran que el algoritmo ESBL posee un tasa muy baja de seleccionar componentes falsos.

Otra medida importante es el error en la estimación de la dirección de arribo de cada una de las señales de interés, $\|\hat{\alpha} - \alpha\|_2^2$. Para evaluar esta medida se toma un promedio de los errores en las 500 corridas Monte Carlo. Esto se muestra en las Figuras 4.7 y 4.8 para las señales de 120 y 160 grados respectivamente. A partir de un SIR de 15 dB, los ángulos de arribo son detectados sin error por casi todos los algoritmos, salvo MUSIC, y ello se indica con un 0 *virtual* en los ejes logarítmicos. Se puede apreciar que el algoritmo ℓ_1 -SVD es el que necesita una SIR menor para detectar correctamente las direcciones de arribo, pero lamentablemente no es bueno en la estimación de la señal como se mostró previamente.

También se desea evaluar el tiempo de cómputo para cada uno de los algoritmos de estimación para esta aplicación de arreglos de sensores en particular. Éstos pueden apreciarse en la Figura 4.9. El algoritmo MOMP es el más rápido pero su desempeño en la estimación de los parámetros de interés no es confiable. El algoritmo ESBL es similar a MSBL, MFOCUSS y MUSIC. El algoritmo ℓ_1 -SVD tiene una mayor complejidad computacional porque debe resolver el problema de relajación mediante una costosa programación cónica de segundo orden (SOCP [90]) implementada con el paquete CVX en Matlab [91,92]. La comparación se realizó en una computadora con un procesador INTEL CORE i3-540 con 8 GB de RAM.

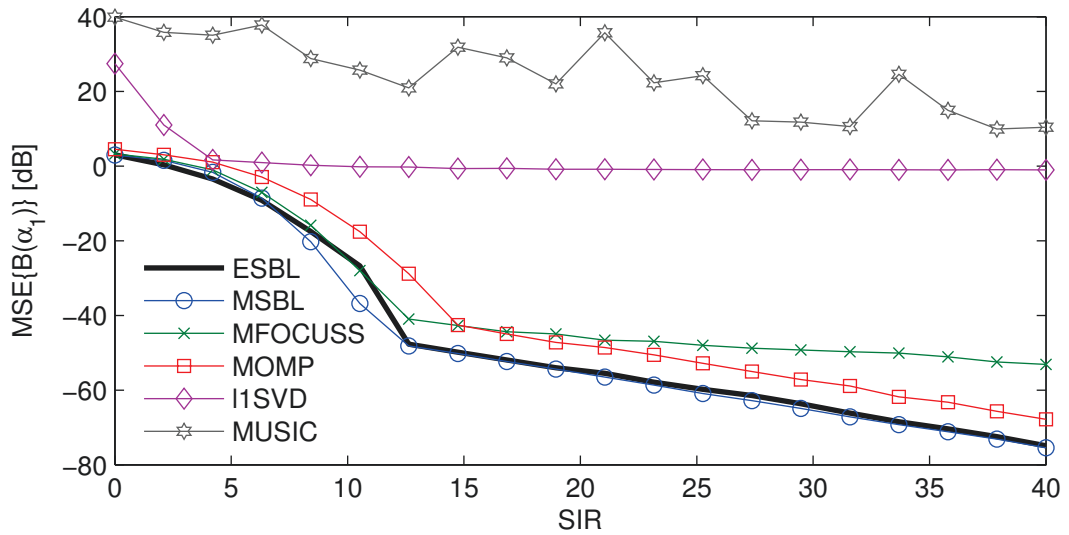


Figura 4.5: Error en la estimación de la señal de 120 grados.

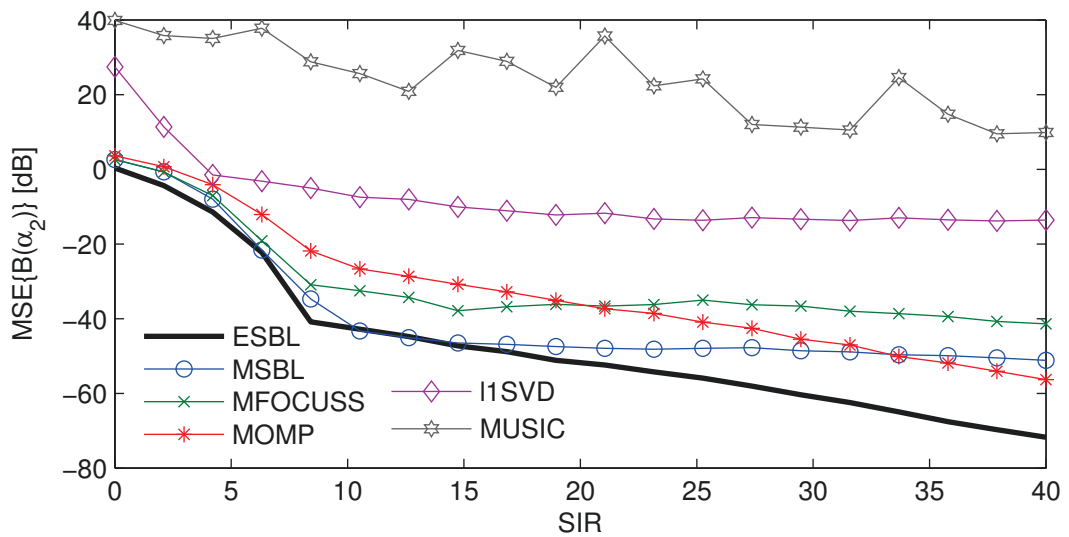


Figura 4.6: Error en la estimación de la señal de 160 grados.

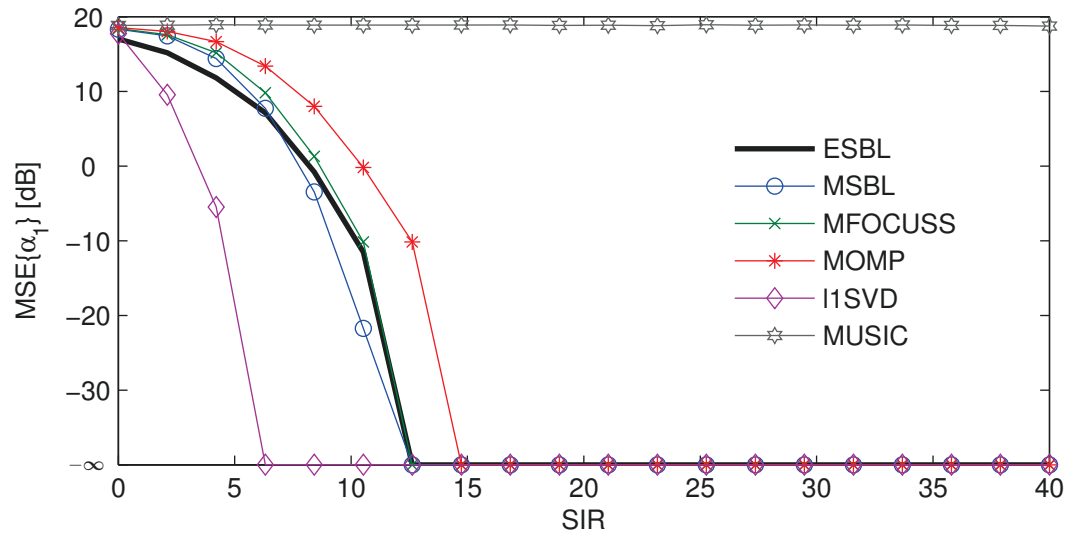


Figura 4.7: Error en la estimación de la dirección de arribo de la señal de 120 grados.

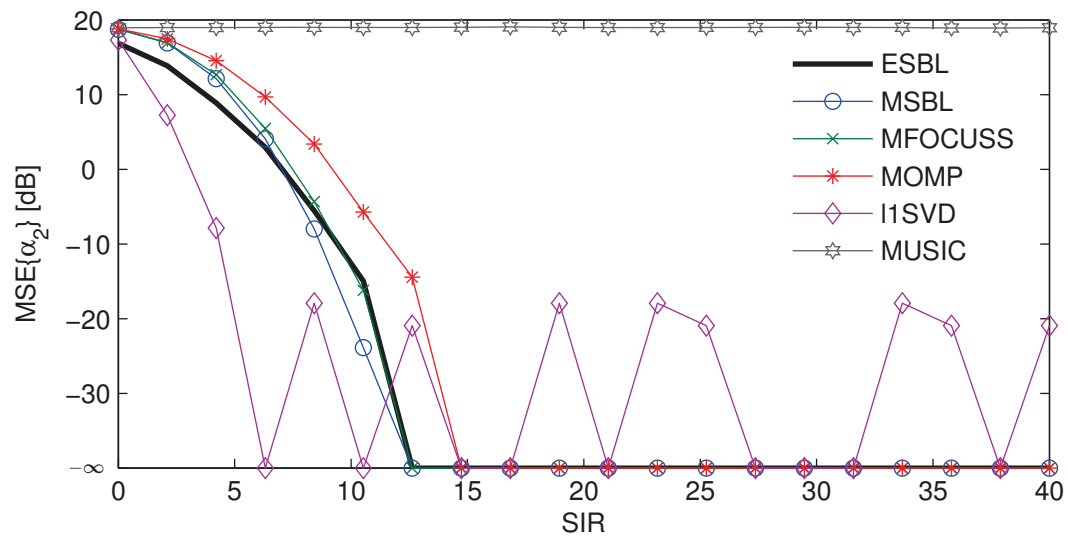


Figura 4.8: Error en la estimación de la dirección de arribo de la señal de 160 grados.

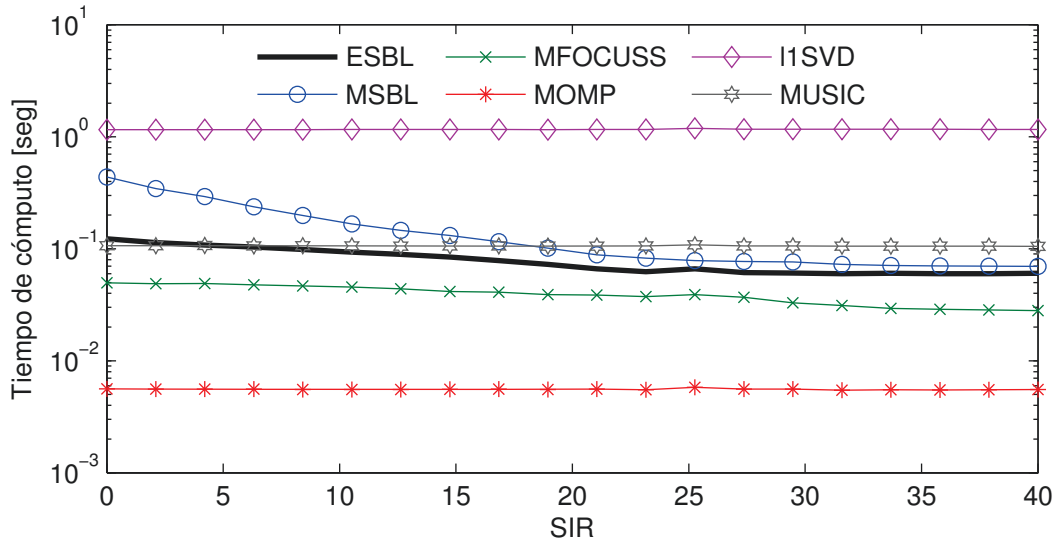


Figura 4.9: Tiempos de cómputo para los distintos algoritmos utilizados.

4.5. Discusión y conclusiones

Se estudió el desempeño de varios algoritmos de métodos raros para la aplicación específica de arreglo de sensores. En esta aplicación se busca estimar tanto la señal como su dirección de arribo, en presencia de fuentes de interferencia. En particular se analizó el algoritmo ESBL, introducido en un trabajo previo. Este algoritmo no requiere información de la cantidad de fuentes dado que anula los ángulos donde considera que no hay señal, mientras que los otros algoritmos deben buscar los picos mayores, pero es necesario saber a priori la cantidad de señales presentes K_b . Utilizando distintas medidas de desempeño se encontró que este algoritmo se comporta igual o mejor que otros algoritmos en varias condiciones, haciendo un mejor uso del tiempo de cómputo.

Capítulo 5

Proyección óptima en modelos ralos con interferencia

5.1. Introducción

En este capítulo se concibe un método para diseñar una matriz de proyección óptima basado en un enfoque Bayesiano. En este contexto, optimalidad se refiere al hecho de que la matriz resultante no solo comprime las mediciones sino que mejora los resultados de estimación al rechazar señales no deseadas. Se elige una matriz que minimiza el error de estimación de los parámetros de interés. Esto se logra con la minimización de la traza de la covarianza posterior, resultando en una distribución posterior altamente concentrada. Para evitar el alto costo de optimizar cada elemento de la matriz de proyección, se elige una solución que pertenece a un espacio restringido. Se muestra que el diseño propuesto para la matriz de proyección combinado con el algoritmo ESBL reduce tanto el tiempo de computo como el error de estimación. Hasta donde se tiene conocimiento, no hay otro algoritmo para resolver el problema ralo inverso que separe la fuente de contaminación en interferencia y ruido blanco. Como consecuencia, los métodos ralos convencionales no aprovechan al máximo el procedimiento propuesto para el diseño de la matriz.

En la siguiente sección se presenta el modelo de datos usado y se formula el problema de optimización. En la Sección 5.3 se analiza la función de costo para

convertir el problema a uno resoluble, se encuentra su solución cuando se agrega una restricción de energía, y se propone una solución subóptima libre de parámetros. En la Sección 5.4 se proveen diferentes simulaciones para analizar el desempeño del algoritmo propuesto. Se muestran resultados tanto para un valor fijo de la restricción de energía como para valores seleccionados automáticamente. Adicionalmente se analiza la robustez del método cuando los parámetros estadísticos del modelo no son estimados en forma precisa o exacta y se analiza el comportamiento del método para diferentes grados de rareza de la señal. En la Sección 5.5 se muestra una aplicación del método para datos reales de radar con un objetivo estático y luego se aplica el método para un objetivo dinámico sintético sobre clutter de radar real. Se concluye en la Sección 5.6.

5.2. Formulación del problema

5.2.1. Modelo

Se considera el siguiente modelo lineal libre de ruido,

$$\mathbf{y}_{nf} = \mathbf{X}\mathbf{b} + \sum_{p=1}^P \mathbf{Z}_p \mathbf{u}_p = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u}, \quad (5.1)$$

donde \mathbf{y}_{nf} es el vector de tamaño M de datos observados o mediciones libres de ruido, \mathbf{b} es el vector de tamaño N que representa la señal de interés, con $M < N$. Cada \mathbf{u}_p es un vector de tamaño Q representando una de las P fuentes de interferencia. \mathbf{X} and \mathbf{Z}_p son matrices de regresores lineales de tamaños $M \times N$ y $M \times Q$ respectivamente, $\mathbf{u} = [\mathbf{u}'_1, \dots, \mathbf{u}'_P]'$ y $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_P]$. En este trabajo se asume que la rareza $S = \|\mathbf{b}\|_0$ del vector \mathbf{b} es menor que M y por lo tanto puede ser comprimida aún más y menos mediciones pueden utilizarse para estimar \mathbf{b} . Esta operación puede representarse como una matriz de proyección \mathbf{A} de tamaño $K \times M$, donde K es tal que $S < K < M$.

Luego, el modelo para las mediciones comprimidas puede expresarse como

$$\mathbf{y} = \mathbf{A}\mathbf{X}\mathbf{b} + \mathbf{A}\mathbf{Z}\mathbf{u} + \mathbf{w}, \quad (5.2)$$

donde \mathbf{y} es el vector de tamaño K de mediciones comprimidas y \mathbf{w} representa al ruido introducido por el sistema y por errores de modelado. Una forma clásica de atacar el problema de adquisición comprimida es optimizar simultáneamente la matriz de proyección \mathbf{A} y el diccionario \mathbf{X} [21]. Sin embargo, en varias aplicaciones, el diccionario \mathbf{X} está restringido por consideraciones físicas y debe mantenerse fijo. Por lo tanto, sólo se considera la optimización de la matriz \mathbf{A} .

La motivación del procedimiento propuesto es aplicaciones de radar y sistemas de comunicaciones. como el descrito en la Sección 5.5 donde la información involucrada está representada típicamente por datos complejos. Luego, se asume que todas las variables aleatorias involucradas en el modelo de señal (5.2) poseen una distribución Gaussiana compleja circularmente simétrica. El caso de variables reales puede ser incluido dentro de los más generales complejos. Adicionalmente, existe independencia estadística entre las interferencias, entre las componentes del ruido y entre ellos. Las matrices de covarianza para \mathbf{b} , \mathbf{w} y \mathbf{u}_p son Σ_b , $\sigma\mathbf{I}_K$ y Σ_u , respectivamente. Luego, las distribuciones de densidad de probabilidad son

$$\begin{aligned}\mathbf{w} &\sim \mathcal{CN}_K(\mathbf{0}, \sigma\mathbf{I}_K), \\ \mathbf{u} &\sim \mathcal{CN}_{PQ}(\mathbf{0}, \mathbf{I}_P \otimes \Sigma_u), \\ \mathbf{b} &\sim \mathcal{CN}_M(\mathbf{0}, \Sigma_b).\end{aligned}$$

Como ya se menciona en el Capítulo 2, una generalización del problema anterior aparece cuando se realizan varios grupos de mediciones en el tiempo, cada una denominada instantánea. Luego, considerando una secuencia de D instantáneas, los datos de salida son

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{A}\mathbf{Z}\mathbf{U} + \mathbf{W}, \quad (5.3)$$

donde las matrices \mathbf{Y} , \mathbf{B} , \mathbf{U} , y \mathbf{W} se construyen usando los respectivos vectores de (5.2) como sus columnas. Asumiendo que los datos pueden ser representados por los mismos elementos del diccionario, la matriz \mathbf{B} posee un pequeño número de filas distintas de cero. El objetivo es estimar las columnas de \mathbf{B} dado que existe un perfil de rareza común a todas ellas. Todas las matrices de (5.3) son complejas.

5.2.2. Presentación del problema

El objetivo es hallar la matriz \mathbf{A} que minimiza el error de reconstrucción de \mathbf{b} , asumiendo que los parámetros estadísticos de la señal, la interferencia y el ruido son conocidos. Para poder minimizar este error se debe elegir una función de costo escalar que mida la dispersión de la distribución a posteriori. Suponiendo estadísticas Gaussianas, la matriz de covarianza está directamente relacionada con la denominada elipsoide de confianza que encierra los valores cuya probabilidad excede un cierto umbral. La longitud de los ejes de esta elipsoide son proporcionales a los autovalores de la matriz de covarianza [90, Cap. 7.5]. Una opción clásica es utilizar el determinante de la covarianza posterior $\Sigma_{b|y}$, que minimiza el volumen de la elipsoide de confianza. Otra opción es utilizar la traza de esa matriz, que minimiza la suma de la longitud de los ejes. Con esta última métrica, se obtiene una forma cerrada de la solución al problema. Adicionalmente, se obtiene una ganancia significativa en el desempeño, como se muestra mediante simulaciones en la Sección 5.4 y en una aplicación real de radar en la Sección 5.5. Por lo tanto, se adopta la traza de la covarianza posterior $\Sigma_{b|y}$, que conlleva a una distribución posterior altamente concentrada. Luego el problema puede presentarse como

$$\min_{\mathbf{A}} \text{tr}(\Sigma_{b|y}) \quad \text{tal que} \quad \|\mathbf{A}\|_{\text{F}}^2 = E, \quad (5.4)$$

donde la restricción se introduce para acotar la energía de la matriz de proyección \mathbf{A} .

5.3. Solución

5.3.1. Función de costo

Para poder hallar la solución óptima del problema (5.4), en primer lugar se analiza la función de costo para encontrar una buena expresión con la cual trabajar. A partir de (5.2) se obtiene la covarianza de las mediciones comprimidas $\Sigma_y = \mathbf{A}\Sigma\mathbf{A}^H + \sigma\mathbf{I}_K$ donde

$$\Sigma = \mathbf{X}\Sigma_b\mathbf{X}^H + \mathbf{Z}(\mathbf{I}_P \otimes \Sigma_u)\mathbf{Z}^H, \quad (5.5)$$

y luego la covarianza posterior es [93, Teorema 10.3]

$$\Sigma_{b|y} = \Sigma_b - \Sigma_b \mathbf{X}^H \mathbf{A}^H (\mathbf{A} \Sigma \mathbf{A}^H + \sigma \mathbf{I}_K)^{-1} \mathbf{A} \mathbf{X} \Sigma_b. \quad (5.6)$$

Aplicando la identidad de Woodbury [94, ec. 145]

$$\Sigma_{b|y} = \Sigma_b - \frac{\Sigma_b \mathbf{X}^H \mathbf{A}^H}{\sigma} \left[\mathbf{I}_K - \mathbf{A} \left(\mathbf{I}_M + \frac{\Sigma}{\sigma} \mathbf{A}^H \mathbf{A} \right)^{-1} \frac{\Sigma}{\sigma} \mathbf{A}^H \right] \mathbf{A} \mathbf{X} \Sigma_b. \quad (5.7)$$

Definiendo $\mathbf{G} = \mathbf{A}^H \mathbf{A}$, una matriz de tamaño $M \times M$ de rango K , y aplicando una de las identidades de Searle [94, ec. 154] se obtiene

$$\Sigma_{b|y} = \Sigma_b - \frac{\Sigma_b \mathbf{X}^H}{\sigma} \left(\mathbf{I}_M + \mathbf{G} \frac{\Sigma}{\sigma} \right)^{-1} \mathbf{G} \mathbf{X} \Sigma_b. \quad (5.8)$$

Como el problema de minimización (5.4) no puede ser resuelto fácilmente para una matriz \mathbf{A} arbitraria, inspirándose en el procedimiento introducido en [95], se restringe el espacio de soluciones \mathcal{A} a matrices \mathbf{A} con una estructura particular. Este espacio restringido se define a través de los autovectores de la matriz \mathbf{G} , que serán iguales a los autovectores de la matriz de covarianza Σ . Idealmente, los autovectores de \mathbf{G} deberían obtenerse a partir del espacio de señal mediante Σ_b . Sin embargo, como esta opción hace que el problema sea insoluble, se proyecta sobre el espacio de señal más interferencia. Esto resulta en una solución que ya no es óptima para el problema (5.4), y por lo tanto se tiene una matriz \mathbf{A} subóptima. En las siguientes secciones, se muestra mediante simulaciones y datos reales de radar que este procedimiento, a pesar de su suboptimalidad, provee una gran mejora en el desempeño comparado con una proyección aleatoria. En base a lo expuesto, se realiza la descomposición de Σ y \mathbf{G} ,

$$\Sigma = \mathbf{V} \mathbf{D} \mathbf{V}^H \quad \mathbf{G} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^H, \quad (5.9)$$

donde \mathbf{V} es la matriz de tamaño $M \times M$ de autovectores de Σ y $\mathbf{D} = \text{diag}\{d_1, \dots, d_M\}$ es la matriz diagonal de tamaño $M \times M$ de autovalores de Σ ; $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_K\}$ es la matriz de tamaño $K \times K$ de autovalores de \mathbf{G} y $\mathbf{\Gamma}$ es la matriz de tamaño $M \times K$ de autovectores de \mathbf{G} . A partir de la discusión anterior, los autovectores de \mathbf{G} corresponderán a un subconjunto de las columnas de \mathbf{V} , y se definen con la solución

subóptima del problema en la siguiente Sección. De aquí en más, por simplicidad y sin pérdida de generalidad, se asumirá que ese subconjunto son las primeras K columnas de \mathbf{V} . Dadas estas consideraciones, se puede reescribir (5.8) como

$$\begin{aligned}\boldsymbol{\Sigma}_{b|y} &= \boldsymbol{\Sigma}_b - \frac{\boldsymbol{\Sigma}_b \mathbf{X}^H}{\sigma} \left(\mathbf{I}_M + \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^H \frac{\boldsymbol{\Sigma}}{\sigma} \right)^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^H \mathbf{X} \boldsymbol{\Sigma}_b \\ &= \boldsymbol{\Sigma}_b - \frac{\boldsymbol{\Sigma}_b \mathbf{X}^H \boldsymbol{\Gamma}}{\sigma} \left[\mathbf{I}_M - \left(\mathbf{I}_K + \boldsymbol{\Lambda} \boldsymbol{\Gamma}^H \frac{\boldsymbol{\Sigma}}{\sigma} \boldsymbol{\Gamma} \right)^{-1} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^H \frac{\boldsymbol{\Sigma}}{\sigma} \boldsymbol{\Gamma} \right] \boldsymbol{\Lambda} \boldsymbol{\Gamma}^H \mathbf{X} \boldsymbol{\Sigma}_b.\end{aligned}\quad (5.10)$$

Dado que

$$\begin{aligned}\boldsymbol{\Gamma}^H \boldsymbol{\Sigma} \boldsymbol{\Gamma} &= \boldsymbol{\Gamma}^H \mathbf{V} \mathbf{D} \mathbf{V}^H \boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{I}_K & \mathbf{0} \end{pmatrix} \mathbf{D} \begin{pmatrix} \mathbf{I}_K \\ \mathbf{0} \end{pmatrix} \\ &= \text{diag}\{d_1, \dots, d_K\} \triangleq \tilde{\mathbf{D}},\end{aligned}\quad (5.11)$$

luego usando nuevamente una identidad de Searle [94, eq. 154]

$$\boldsymbol{\Sigma}_{b|y} = \boldsymbol{\Sigma}_b - \frac{\boldsymbol{\Sigma}_b \mathbf{X}^H \boldsymbol{\Gamma} \mathbf{Q} \boldsymbol{\Gamma}^H \mathbf{X} \boldsymbol{\Sigma}_b}{\sigma},\quad (5.12)$$

donde

$$\mathbf{Q} = \left[\mathbf{I}_K - \left(\mathbf{I}_K + \boldsymbol{\Lambda} \frac{\tilde{\mathbf{D}}}{\sigma} \right)^{-1} \boldsymbol{\Lambda} \frac{\tilde{\mathbf{D}}}{\sigma} \right] \boldsymbol{\Lambda}\quad (5.13)$$

es una matriz diagonal con elementos

$$[\mathbf{Q}]_{kk} = \frac{\lambda_k}{1 + \frac{d_k \lambda_k}{\sigma}}.\quad (5.14)$$

Con esta definición, la función de costo puede reescribirse como

$$\text{tr}(\boldsymbol{\Sigma}_{b|y}) = \text{tr}(\boldsymbol{\Sigma}_b) - \sum_{k=1}^K \frac{\lambda_k}{\sigma + d_k \lambda_k} t_k,\quad (5.15)$$

donde $t_k = \|v_k\|_2^2$ y $v_k = [\boldsymbol{\Sigma}_b \mathbf{X}^H \mathbf{V}]_{\cdot k}$ es la columna k de la matriz $\boldsymbol{\Sigma}_b \mathbf{X}^H \boldsymbol{\Gamma}$. Nótese que se ha transformado un problema de minimización sobre matrices de tamaño $K \times M$ a uno de minimización sobre vectores de tamaño K , que contiene los autovalores de \mathbf{G} .

5.3.2. Solución óptima restringida

Teniendo en cuenta las restricciones del espacio de soluciones mencionadas en la sección anterior, y el hecho de que $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}'\mathbf{A}) = \text{tr}(\mathbf{G}) = \sum_{k=1}^K \lambda_k$, se define un problema modificado descrito por

$$\min_{\mathbf{A} \in \mathcal{A}} \text{tr}(\boldsymbol{\Sigma}_{b|y}) \quad \text{tal que} \quad \sum_{k=1}^K \lambda_k = E, \quad \lambda_k \geq 0 \quad \forall k. \quad (5.16)$$

Se demuestra en el Apéndice D que el valor óptimo para los autovalores de \mathbf{G} para el problema (5.16) es

$$\lambda_k = \begin{cases} \frac{t_k}{d_k} \frac{\sum_{i=1}^L \sigma/d_i + E}{\sum_{i=1}^L t_i/d_i} - \frac{\sigma}{d_k}, & 1 \leq k \leq L \\ 0, & k > L \end{cases}. \quad (5.17)$$

donde L es el número de autovalores distintos de cero determinado en el proceso de optimización descrito en el Apéndice D. Finalmente, la matriz de proyección \mathbf{A} puede obtenerse a partir de \mathbf{G} usando los L autovalores óptimos y los correspondientes autovectores de $\boldsymbol{\Sigma}$,

$$\mathbf{A} = \text{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_L}\}[\mathbf{V}]'_{1:L}. \quad (5.18)$$

Es digno de mencionar que la restricción de energía E tiene un impacto en el valor de L y puede ser usada como un parámetro de ajuste que afecta el comportamiento del algoritmo.

Nótese que el procedimiento de optimización necesita los valores de $\boldsymbol{\Sigma}_b$, $\boldsymbol{\Sigma}_u$ y σ como parámetros de entrada. En aplicaciones reales sus valores verdaderos no son conocidos, sin embargo, pueden ser estimados aproximadamente a partir de instantáneas previas con algoritmos de estimación adecuados. En la sección 5.4.3 se analiza la sensibilidad de la optimización con respecto a las estimaciones de los parámetros.

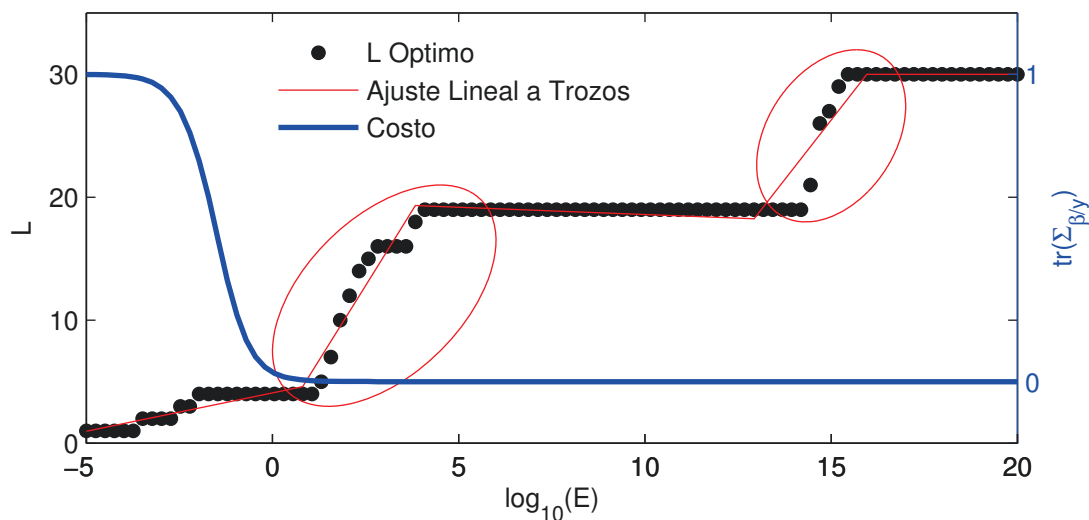


Figura 5.1: Valor óptimo de L para diferentes valores de la restricción de energía E , y su correspondiente costo.

5.3.3. Selección automática de E

Si la aplicación no requiere una restricción de energía E en las filas de la matriz, se tiene un nuevo grado de libertad. En esta sección se introduce un método heurístico para elegir un valor quasi-óptimo para esta restricción E y consecuentemente L , el número de autovalores usado.

La Figura 5.1 muestra los valores óptimos de L , obtenidos como solución del problema (5.16), cuando se varía la restricción E para un escenario típico del problema. El valor óptimo de esta restricción tiende a infinito a medida que E crece, esto se traduce en importantes errores numéricos en el algoritmo y por lo tanto un desempeño pobre. Puede apreciarse en la Figura que existen valores umbrales para la restricción E en los cuales el tamaño L de la matriz aumenta. Estos umbrales, denotados E_L , pueden calcularse como sigue. Para un valor fijo de L , fijar los autovalores $\lambda_{L+1}, \lambda_{L+2}, \dots, \lambda_K$ a zero y calcular el valor requerido de E necesario para anular λ_{L+1} , de (5.17)

$$E_L = \sum_1^L \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{t_{L+1}}} - 1 \right), \quad (5.19)$$

y consecuentemente, usando (5.19) en (5.17) para cada L , los autovalores requeridos λ_i son

$$\lambda_i = \text{máx} \left\{ 0, \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{t_{L+1}}} - 1 \right) \right\} \quad (5.20)$$

La expresión analítica de la función de costo resulta

$$\text{tr}(\mathbf{\Sigma}_{b|y}) = \text{tr}(\mathbf{\Sigma}_b) - \sum_{k=1}^L \frac{t_k}{d_k} + \frac{\sum_{k=1}^L \frac{\sigma \sqrt{t_k}}{d_k} \sum_{l=1}^L \frac{\sqrt{t_l}}{d_l}}{\sum_{l=1}^L \frac{\sigma}{d_k} + E}, \quad (5.21)$$

cuyo valor cuando $E \rightarrow 0$ es $\text{tr}(\mathbf{\Sigma}_b)$ y cuando $E \rightarrow \infty$ es $\text{tr}(\mathbf{\Sigma}_b) - \sum_{k=1}^L \frac{t_k}{d_k}$.

Adicionalmente, se ilustra en la Figura 5.1 el comportamiento de la función de costo (5.21) a medida que E aumenta, conduciendo a un valor óptimo infinito de E , como se mencionó anteriormente. Por lo tanto se tiene una relación de compromiso entre un L bajo, para comprimir la señal, y un alto L , para decrecer el valor de la función de costo. Este último resulta en valores muy altos E que introducen errores numéricos en el procedimiento.

Se introduce un método heurístico para resolver esta relación de compromiso. En la Figura 5.1 hay dos grupo de puntos, el primero tiene valores bajos de E y de L y el segundo grupo tiene valores demasiado altos de E que se quieren evitar. Entonces se desea elegir el valor más grande de E_L correspondiente al primer grupo. Para conseguir esto en forma automática se propone el uso de Multivariate Adaptive Regression Splines (MARS) [96] para ajustar una función lineal a trozos en estos puntos y tomar el punto de quiebre eligiendo el correspondiente valor de E . Los detalles de este procedimiento exceden el alcance de la presente tesis y pueden ser leídos de la referencia indicada. Los resultados de este procedimiento se muestran en la siguiente sección.

5.4. Análisis de desempeño

5.4.1. Valor fijo de E

Para calcular el desempeño empírico, se propone el siguiente ejemplo del modelo (5.2). La longitud de los vectores \mathbf{y} , \mathbf{b} y \mathbf{u}_p son $M=30$, $N=60$ y $Q=5$, respectivamente. Existen $P=3$ fuentes de interferencia y los datos \mathbf{y} son observados independientemente en $D=10$ instantes.

El uso de diccionarios aleatorios es frecuente en el procesamiento de señales ralas [10, 36, 97]. En muchas aplicaciones reales las mediciones pueden usualmente modelarse como aleatorias, por lo tanto estos diccionarios aleatorios representan una amplio rango de fenómenos y pueden utilizarse como un justo indicador del desempeño del procedimiento. Adicionalmente, un buen desempeño con un diccionario aleatorio sería un buen indicador de que uno puede esperar que el algoritmo no tendrá un pobre desempeño en otro caso [10]. Las matrices \mathbf{X} y \mathbf{Z} son generadas aleatoriamente una vez, al comienzo de la simulación, cuyas componentes poseen distribución $\mathcal{CN}(0, 1)$ y normalizadas con sus normas de Frobenius. El vector ralo \mathbf{b} posee una componente distinta de cero, luego $S=1$. La potencia de la señal es $P_{\mathbf{b}} = \|\mathbf{b}\|_2^2$. La covarianza de cada interferencia es $\Sigma_u = P_u \mathbf{I}_Q / \sqrt{Q}$. Luego, la potencia de interferencias es $P_u = \|\Sigma_u\|_F$. La potencia de ruido de medición es $\sigma = P_u / 10$. Para evaluar el desempeño empírico se realizaron 1000 corridas de simulaciones Monte Carlo donde el soporte de \mathbf{b} se define aleatoriamente para cada corrida. Se estima \mathbf{b} usando el algoritmo ESBL para diferentes relaciones señal a interferencia, definida como $\text{SIR} = 10 \log_{10}(P_{\mathbf{b}}/P_u)$ con una probabilidad de falsa alarma $P_{FA} = 10^{-3}$.

Para la compresión, se empieza con un valor de $K=30$ y una restricción de energía $E=500$. Los parámetros Σ_b , Σ_u y σ se asumen conocidos en las simulaciones, pero podrían ser aproximados usando datos de entrenamiento o estimaciones previas, como se muestra en la siguiente sección. Adicionalmente, se incluye el uso de una matriz aleatoria \mathbf{A} para comparar su desempeño con el de la optimización propuesta. En esta matriz aleatoria cada componente posee una distribución compleja Gaussiana $A_{ij} \sim \mathcal{CN}(0, 1)$ normalizada para satisfacer la restricción de energía y con las mismas dimensiones que la matriz optimizada para permitir una justa

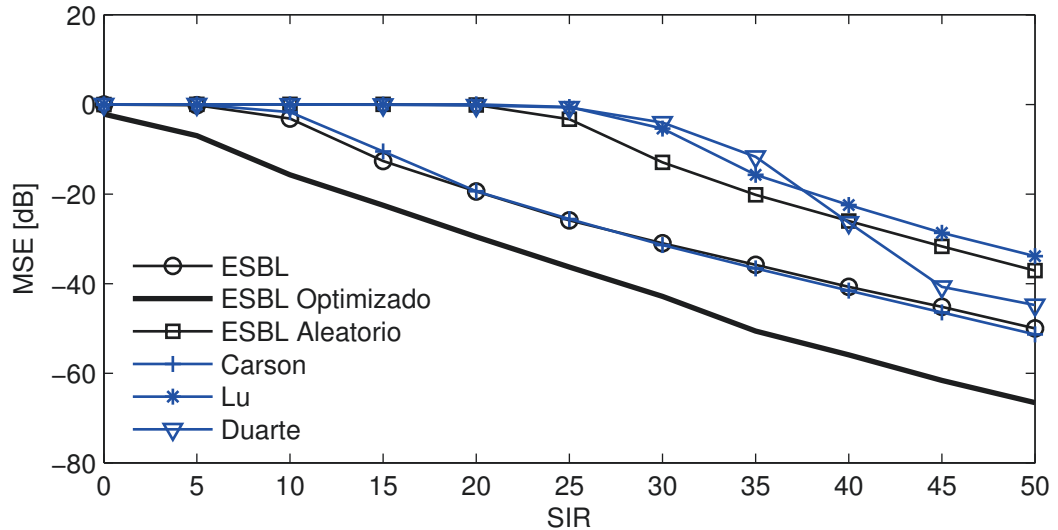


Figura 5.2: Error de estimación relativo en función de la SIR utilizando el algoritmo ESBL con un valor fijo de E sin proyección, con proyección óptima y con proyección aleatoria; y los algoritmos alternativos para el diseño de la matriz de proyección propuestos por Lu [28], Carson [29] y Duarte [21].

comparación.

Como se muestra en la Figura 5.2, usando el algoritmo ESBL con la proyección óptima se obtiene una ganancia de 12 dB para un error de estimación de 10^{-4} sobre el caso sin compresión. Esto se logra a pesar de haber usado un menor número de observaciones, entre 10 y 15, en vez de las 30 originales. Adicionalmente, se aprecia que el uso de una matriz cualquiera, en lugar de la óptima, empeora el error de estimación.

También se incluyen en la Figura 5.2 los desempeños para los algoritmos de Lu [28], Carson [29] y Duarte [21]. El algoritmo de proyección ortogonal de Lu se comporta casi como si fuera una matriz aleatoria, dado que el número de interferencias es muy alto comparado con las mediciones. Como se explica en su trabajo, la proyección ortogonal suprime tanto a la señal de interés como a la interferencia en este escenario. Mientras que el modelo (5.1) considerado en esta tesis aplica la matriz de proyección tanto a la señal de interés como a la interferencia, el algoritmo de Carson diseña la matriz de proyección basándose en el modelo $\mathbf{y} = \mathbf{A}\mathbf{X}\mathbf{b} + \mathbf{w}$.

esto quiere decir que solo aplica la matriz de proyección a la señal de interés. La no proyección de la interferencia resulta en un comportamiento similar al sin optimización. Finalmente el algoritmo de Duarte utiliza un modelo que no tiene en cuenta la interferencia y se comporta pobremente para bajas SIR. Mejora a medida que la SIR aumenta y la interferencia no es significativa pero no tanto como el algoritmo propuesto.

5.4.2. Selección automática de E

Para esta modificación se utilizan los mismos parámetros que en la sección anterior, pero en este caso la restricción de energía E se determina automáticamente por el procedimiento explicado anteriormente. Los resultados se muestran en la Figura 5.3, donde se analizan tres niveles de rareza S de la señal, desde 1 a 3. Para $S = 1$ se obtiene aún un mejor desempeño que fijar $E = 500$. Para mayores valores de rareza, $S = 2$ y $S = 3$, se pierde un poco de desempeño, pero se gana independencia de una posible pobre elección del parámetro E , objetivo de la modificación. Por otro lado, el costo de esta independencia, se traduce en un número mayor de mediciones a ser utilizadas, pero aún menor que las originales.

También se analiza el comportamiento de la matriz de proyección en combinación con otros algoritmos ralos. La Figura 5.4 muestra que el mejor desempeño se alcanza con la aplicación de la matriz de proyección al algoritmo ESBL. Además, como se muestra en la Figura 5.5, esta combinación obtiene la mayor ganancia sobre el caso sin proyección. Este es el resultado de ser el único algoritmo que explota información sobre la interferencia en el proceso de estimación.

En la Figura 5.6 se ilustra el tiempo requerido para cada método, con y sin compresión, para resolver el problema de estimación. Como las dimensiones de entrada para los algoritmos de estimación se reducen, estos operan más rápido resultando en un menor tiempo de cómputo. Aquí, aunque OMP es el algoritmo más rápido, su desempeño, mostrado en la Figura 5.4, no es aceptable. El tiempo adicional consumido en calcular los autovalores para el diseño de la matriz de proyección óptima \mathbf{A} es insignificante frente a estos tiempos.

La Figura 5.7 muestra el error de estimación para diferentes niveles de rareza S

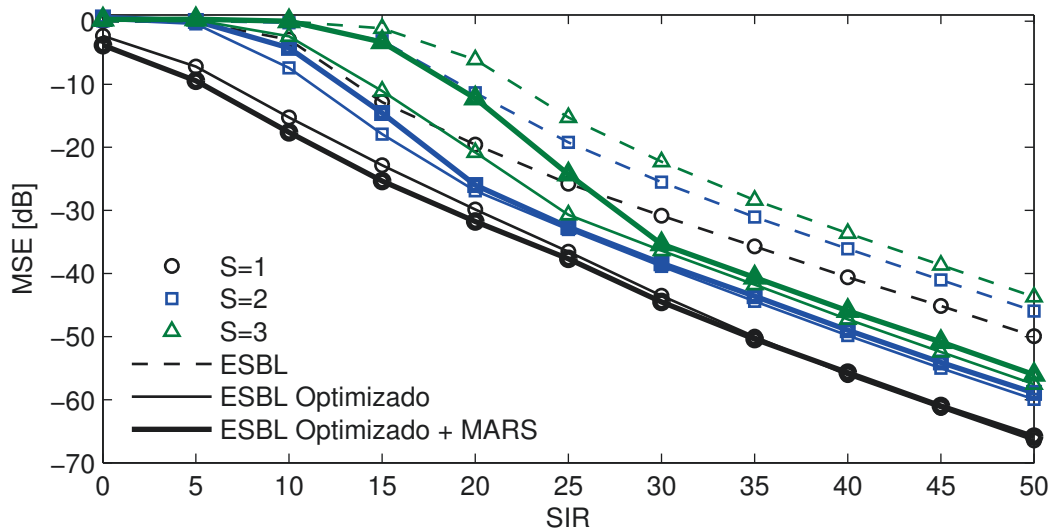


Figura 5.3: Error de estimación relativo en función de la SIR para distintos valores de raleza S , con un valor fijo y selección automática del parámetro E .

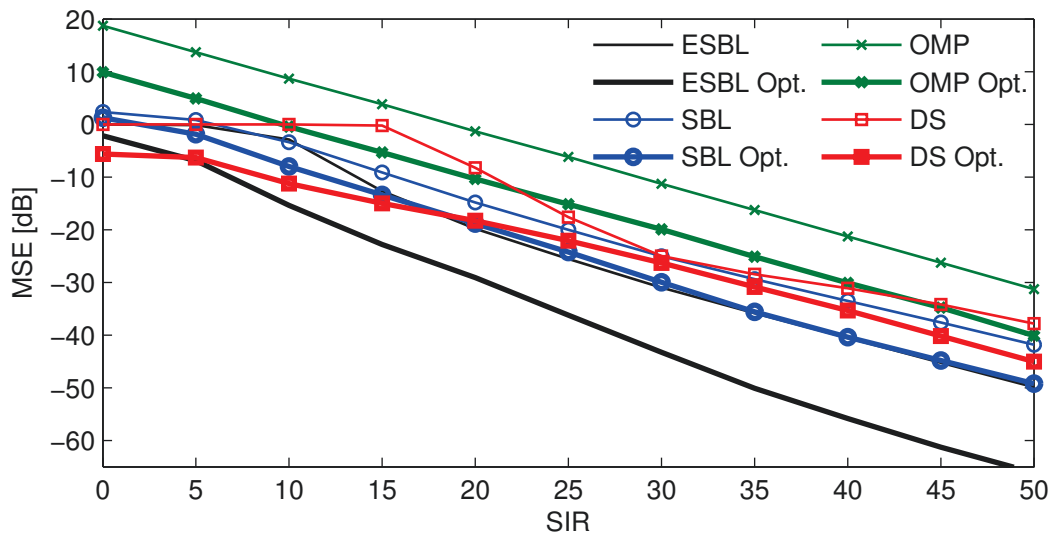


Figura 5.4: Error de estimación relativo, $\|\hat{\mathbf{b}} - \mathbf{b}\|_2^2 / \|\hat{\mathbf{b}}\|_2^2$, en función de la SIR usando selección automática del parámetro E .

de la señal. Como es esperado, cuando la raleza aumenta, el error de estimación aumenta. Para $S = 1$ se obtiene una ganancia considerable sobre el caso sin proyección,

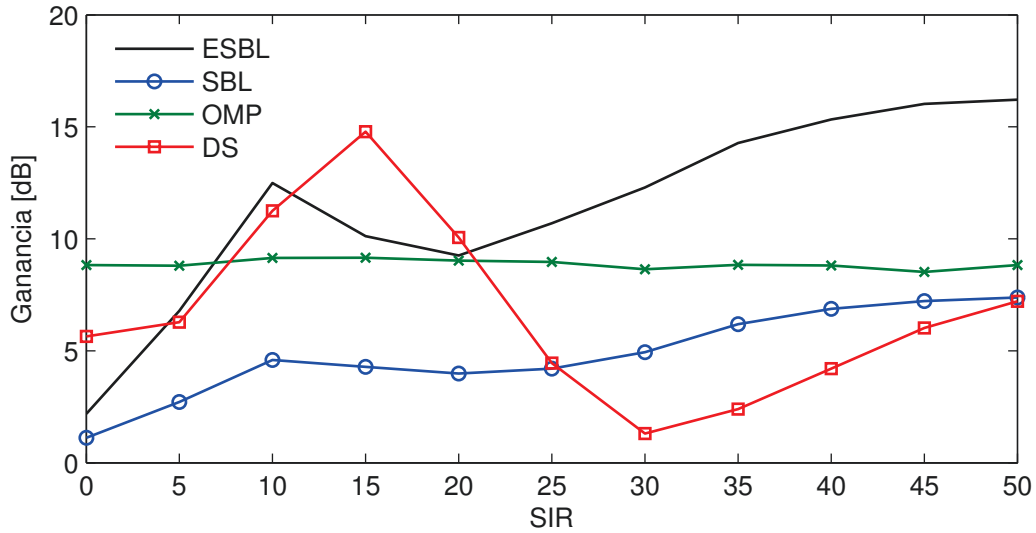


Figura 5.5: Ganancia al usar la matriz de proyección con diferentes algoritmos raros, para distintos valores de SIR y selección automática del parámetro E .

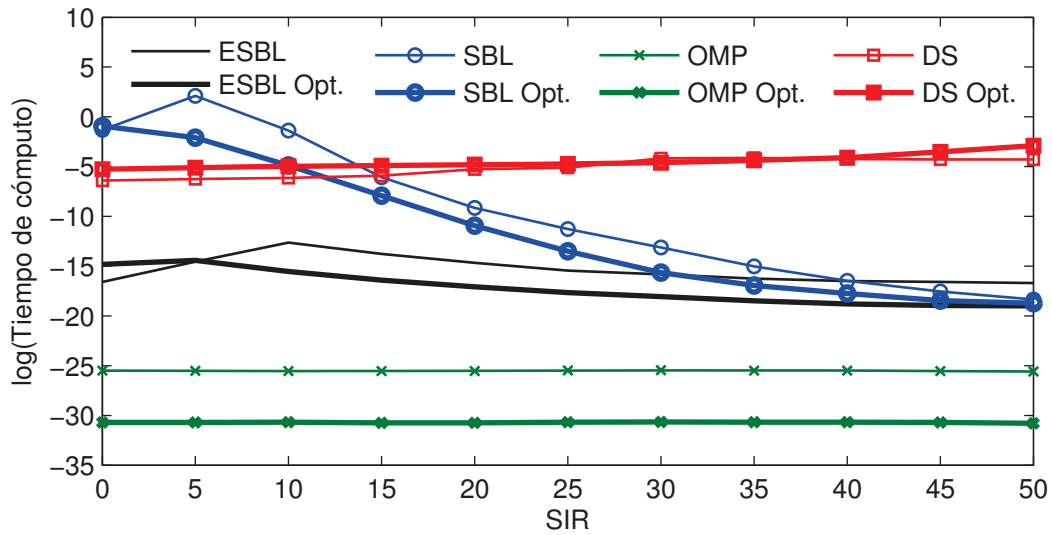


Figura 5.6: Tiempo de cómputo medio para los algoritmos con y sin optimización.

pero para una rareza de $S = 9$ esta ganancia se pierde.

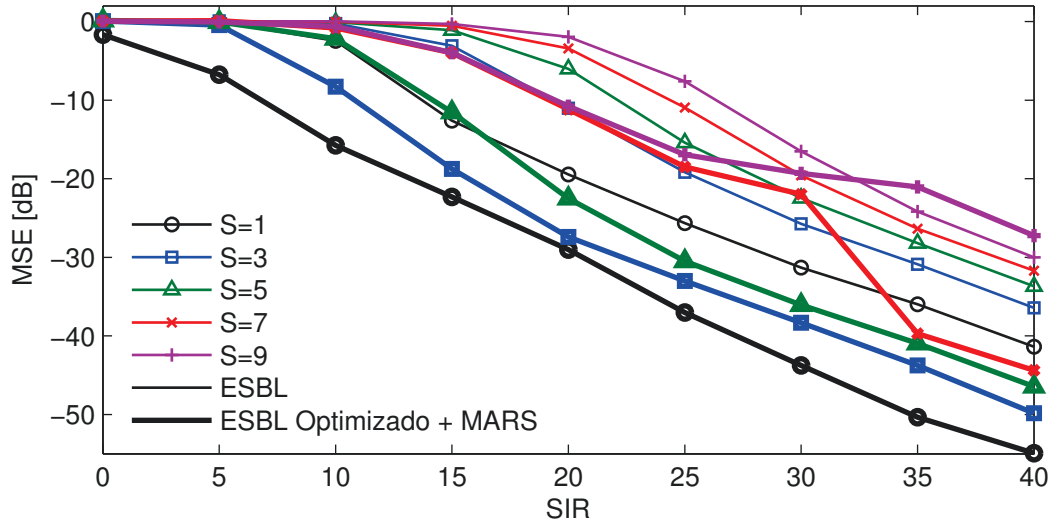


Figura 5.7: Error de estimación relativo en función de la SIR para distintos valores de la raleza S de la señal original.

5.4.3. Robustez

En la práctica, los parámetros estadísticos Σ_b , Σ_u , σ , no son conocidos y deben ser estimados. Esto resulta en errores de estimación que afectan la optimalidad del método y potencialmente degradación en su desempeño. Se analizan varios escenarios de posibles fuentes de discrepancias en las estimaciones.

En primer lugar se analiza la falta de precisión en la estimación del soporte de Σ_b , esto es que no esté tan concentrado en el soporte verdadero como debería. Esta estimación imprecisa se simula mediante una distribución gaussiana con una desviación estándar del valor verdadero en el rango de 0,5 a 1,5 elementos del soporte. Los resultados se muestran en la Figura 5.8 donde cada línea sólida corresponde a un valor diferente de desviación estándar. Puede apreciarse que todavía se tiene una ganancia importante sobre el caso sin proyección para este rango de valores. Existe un comportamiento no esperado en bajas SIR, donde se observa una ganancia resultado de la estimación imprecisa, pero se cree que este efecto se debe a la quasi optimalidad del procedimiento original.

En segundo lugar, se analiza el caso en que la estimación de Σ_b no es exacta,

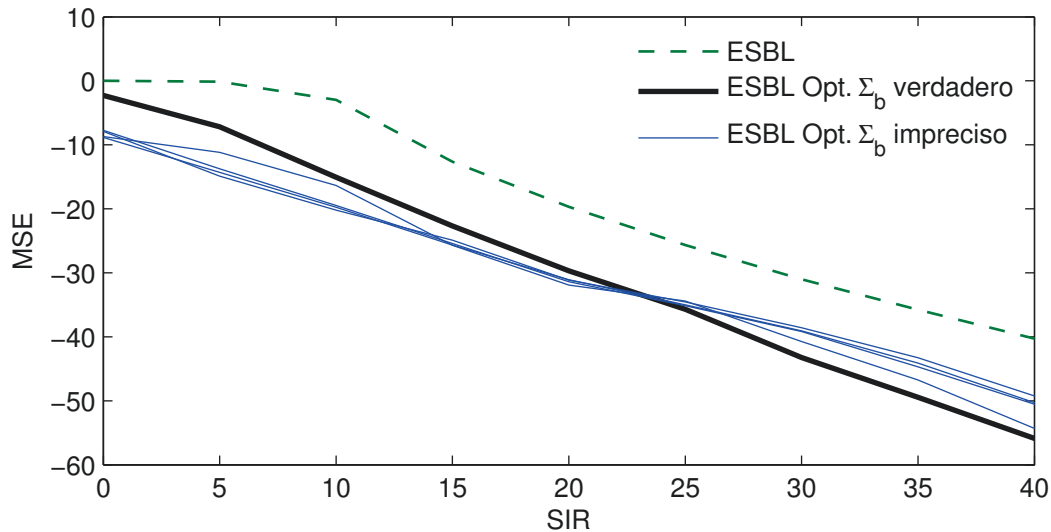


Figura 5.8: Error de estimación relativo en función de la SIR cuando la estimación de Σ_b no es precisa, con un rango de 0,5 a 1,5 para la desviación estándar.

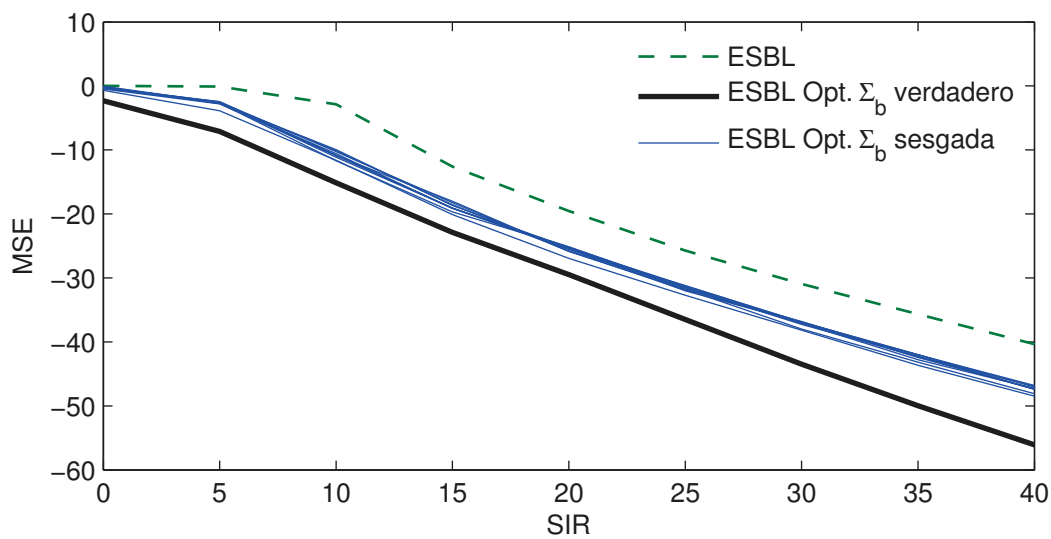


Figura 5.9: Error de estimación relativo en función de la SIR cuando la estimación de Σ_b no es exacta, con un rango de 1 a 6 para el sesgo.

esto es que el soporte Σ_b no es el correcto. Utilizando un sesgo de 1 a 6 elementos de soporte, los resultados mostrados en la Figura 5.9 indican una pequeña pérdida

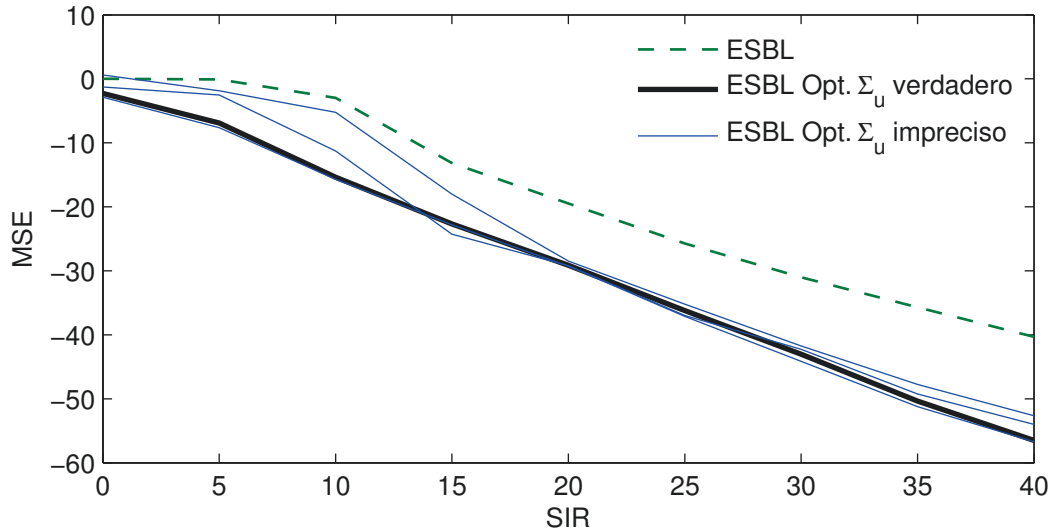


Figura 5.10: Error de estimación relativo en función de la SIR cuando la estimación de Σ_u no es precisa, con un rango de 0,1 a 1 para la desviación estándar.

en el desempeño cuando el soporte no es el verdadero, pero esta pérdida no depende fuertemente del tamaño del sesgo. Sin embargo, aún se obtiene un mejor desempeño que al no usar compresión.

En tercer lugar, se explora la falta de precisión en la estimación de Σ_u añadiendo a su valor verdadero una matriz Gaussiana afectada por un factor de escala. Para cada factor de escala, en el rango de 0,1 a 1, se obtiene una diferente línea sólida en la Figura 5.10. La degradación es grande para pequeños valores de SIR indicando que una estimación precisa de las propiedades de la interferencia es crítica en este régimen. Sin embargo, para niveles medio y altos de SIR se obtiene un desempeño aceptable.

Finalmente, se muestra en la Figura 5.11 que errores en la estimación de σ prácticamente no tienen efecto en el desempeño ya que los niveles de ruido son insignificantes contra la interferencia propuesta en estos escenarios.

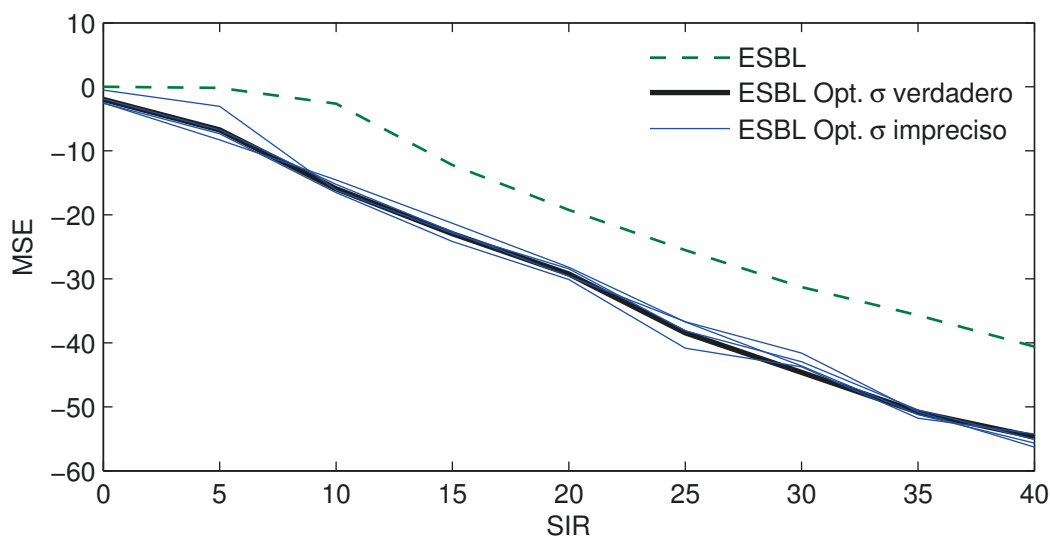


Figura 5.11: Error de estimación relativo en función de la SIR cuando la estimación de σ no es precisa, con un rango de 0,5 a 5 para la desviación estándar.

5.5. Datos reales de radar

5.5.1. Objetivo estático

Se ilustra el funcionamiento del método optimizado mediante el uso de datos de radar recogidos por el radar polarimétrico IPIX de la Universidad de McMaster [98]. Se procesó el grupo de datos stare1 grabado el 11 de Noviembre de 1993. Estos datos corresponden a una pelota envuelta en aluminio flotando en la superficie del mar en clima medio (altura de ola de 0.67 m y velocidad del viento de 21 km/h). Una explicación detallada sobre la forma de ajustar este radar al modelo (5.1), incluyendo el diseño de las matrices \mathbf{X} y \mathbf{Z} , se muestra en el Apéndice C en base al artículo [99].

Se genera el diccionario sobrecompleto permitiendo la presencia del objetivo en cada una de las 54 celdas de rango que forman la huella del radar. Cada objetivo se representado usando la descomposición de Krogager de la matriz de scattering [100]. Para cada celda de rango se consideran nueve componentes: una esfera, hélices izquierda y derecha, y seis dipolos con diferentes orientaciones. Esta configuración resulta en un vector de señal de tamaño $N = 486$. Por otro lado, Se tienen 4 polariza-

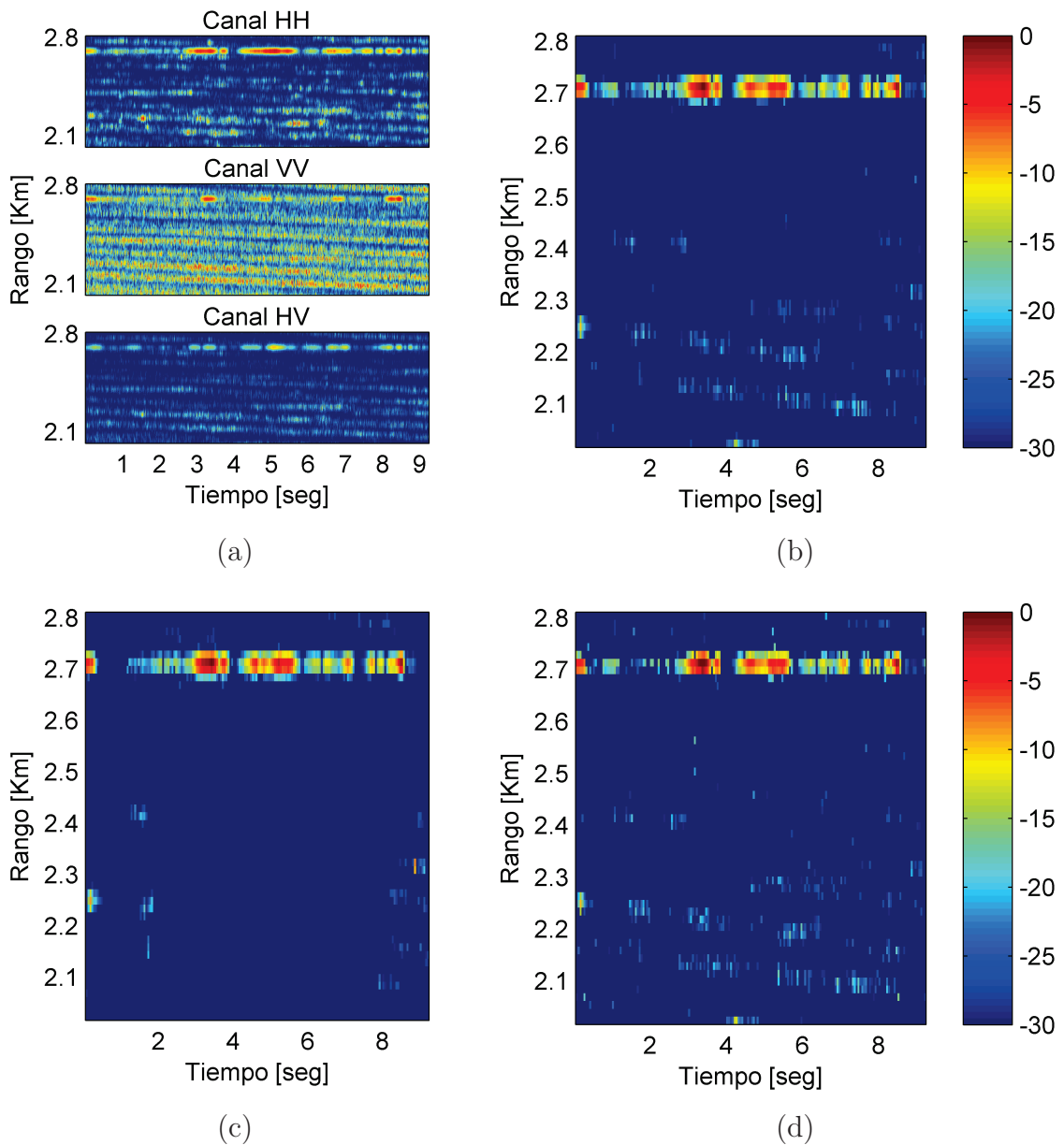


Figura 5.12: Imágenes de radar en dominios de rango y tiempo para (a) datos crudos originales para los distintos canales polarimétricos, (b) datos reconstruidos sin compresión, (c) datos reconstruidos con compresión óptima, y (d) datos reconstruidos usando el algoritmo de Duarte

ciones para cada celda de rango resultando en $M = 216$ mediciones. El valor inicial de la covarianza del clutter Σ_u y de la varianza del ruido σ se estiman a partir de la primera instantánea usando el MLE concentrado [101] y la covarianza de señal Σ_b se fija como la matriz identidad. Luego, las estimaciones resultantes del algoritmo ESBL correspondientes a las instantáneas previas son usadas como entrada para el procedimiento de optimización en las siguientes instantáneas. Cada estimación del algoritmo ESBL utiliza $D = 10$ instantáneas.

En primer lugar se muestra en la Figura 5.12a la energía para los diferentes canales polarimétricos de los datos crudos en escala logarítmica. Es posible ubicar el objetivo aproximadamente a 2.7 km, que corresponde a la celda de rango 47; adicionalmente, esta Figura muestra que el clutter marítimo genera reflexiones casi tan fuertes como el objetivo, especialmente en el canal VV. En la Figura 5.12b se muestra la imagen de radar reconstruida utilizando el algoritmo ESBL con una probabilidad de falsa alarma $P_{FA} = 10^{-5}$ sin aplicar la matriz de proyección. La respuesta del objetivo se mantiene prácticamente sin cambios; sin embargo la amplitud del clutter es ahora 20 dB más débil que el objetivo. En la Figura 5.12c se aplica la proyección óptima propuesta, resultando en un mejor rechazo del clutter, la mayor parte de él es al menos 30 dB más débil que el objetivo, como sugieren las simulaciones previas. También se procesaron los datos reales mediante los algoritmos de Lu [28], Carson [29] y Duarte [21], previamente mencionados en las simulaciones. Los primeros dos algoritmos fallaron en estimar correctamente el objetivo en la mayoría de las instantáneas, a pesar de que Carson utiliza información sobre la interferencia, y no fueron incluidos en las figuras. El mejor desempeño de estos algoritmos alternativos es alcanzado por el algoritmo de Duarte, mostrado en la figura 5.12d. Como este algoritmo no explota información previa de la interferencia, no puede rechazar tan eficientemente el clutter y degrada el objetivo cuando la señal es pobre.

Como se mencionó previamente, se comienza con $M = 216$ mediciones. El procedimiento toma un conjunto de 10 instantáneas, calcula la matriz de proyección óptima y la aplica a estas mediciones. Para este conjunto de datos específico, el tamaño del problema inverso se reduce entonces desde 216 hasta el rango de 20 a 120 mediciones comprimidas, como se muestra en la Figura 5.13. Nótese que se ob-

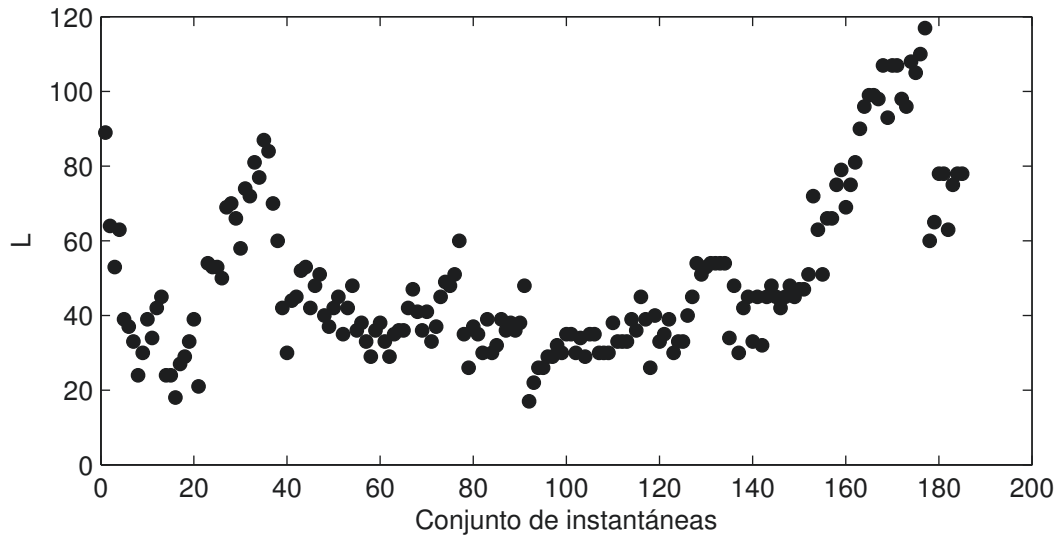


Figura 5.13: Número medio de mediciones comprimidas para cada grupo de instantáneas.

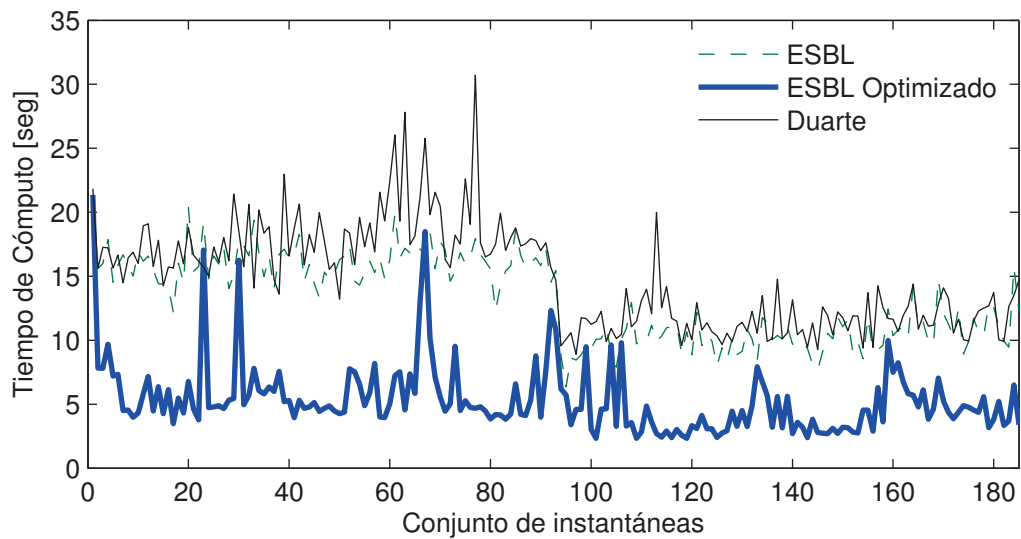


Figura 5.14: Comparación del tiempo de cómputo para el procesamiento de datos reales de radar.

tiene una mejor compresión y rechazo de clutter cuando la señal del objetivo es más fuerte. Esta reducción se traduce en un menor tiempo de cómputo para el algoritmo ESBL. La Figura 5.14 muestra que al aplicar la matriz de proyección óptima, el

tiempo de cómputo se reduce a $1/3$ en este caso.

5.5.2. Objetivo sintético dinámico

Se analiza el desempeño del algoritmo cuando hay dinámica presente en el escenario en oposición al caso previo de objetivo estático. Debido a la falta de datos reales de radar con objetivos dinámicos, se simula su comportamiento añadiendo un objetivo sintético sobre datos reales de clutter marítimo correspondientes al mismo radar polarimétrico IPIX. El grupo de datos utilizado es el stare4 grabado el 12 de Noviembre de 1993, que consiste sólo de clutter marítimo y sin un objetivo como en la sección anterior. El target añadido posee una matriz de scattering correspondiente a un diplano en el canal polarimétrico VV; se aleja a 22,5 m/s del radar durante 10 segundos de observación; y tiene una potencia aproximada de 10 dB sobre el clutter. La Figura 5.15a muestra los canales polarimétricos de los datos crudos para este escenario. La Figura 5.15a muestra la imagen reconstruida con el algoritmo ESBL sin compresión revelando evidente clutter en algunas regiones. El algoritmo propuesto provee un mejor rechazo del clutter indeseable en la escena, como se observa en la Figura 5.15c, aún cuando el objetivo es móvil y por lo tanto las estimaciones de su covarianza no son precisas. El algoritmo de Duarte tiene un comportamiento aceptable pero aún presenta altos niveles de clutter en algunas regiones de la imagen, como se observa en la Figura 5.15d utilizando un mayor tiempo de cómputo.

5.6. Discusión y conclusiones

En este capítulo, se introdujo un método para comprimir óptimamente las mediciones de un modelo lineal ralo con ruido estructurado. La compresión se obtiene proyectando la señal contaminada en un espacio de menor dimension. Para el diseño de esta matriz de proyección, se propone minimizar la traza de la covarianza condicional de la señal rala condicionada a los datos comprimidos. Este método resulta en un procedimiento de water-filling para la construcción de la matriz. El algoritmo presentado tiene un mejor desempeño que otros métodos actuales para el diseño de la matriz de proyección.

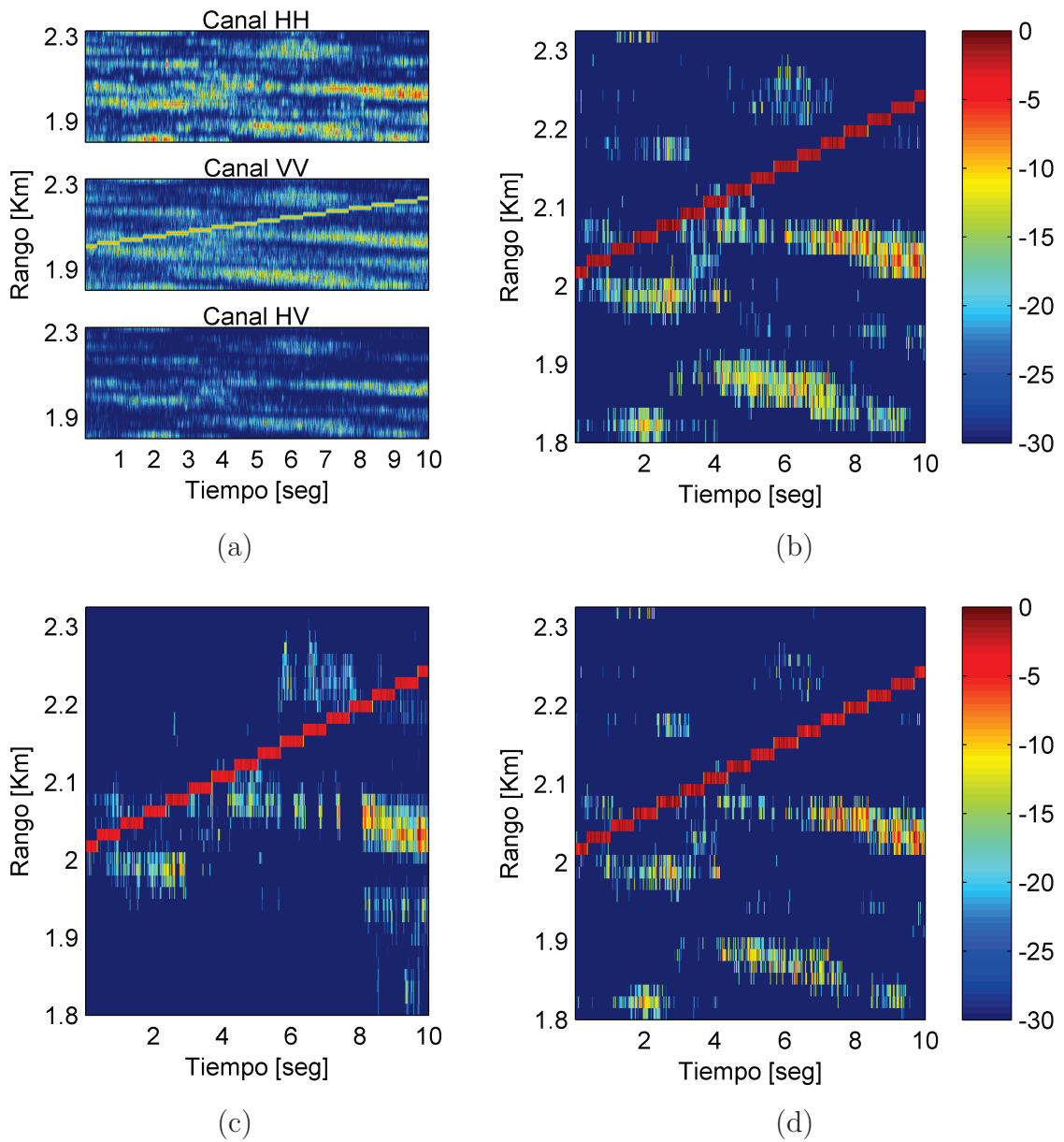


Figura 5.15: Imágenes de radar en dominios de rango y tiempo con un objetivo sintético y dinámico para (a) datos crudos originales para los distintos canales polarimétricos, (b) datos reconstruidos sin compresión, (c) datos reconstruidos con compresión óptima, y (d) datos reconstruidos usando el algoritmo de Duarte

Para demostrar el desempeño del algoritmo en un escenario de la vida real se procesaron datos de radar polarimétrico correspondientes a un objetivo inmerso en clutter marítimo intenso. En este caso el algoritmo mostró tener un mejor rechazo de clutter comparado con el caso sin compresión, aún cuando utiliza menor número de mediciones.

Capítulo 6

Conclusiones

En esta tesis se presentaron distintos aportes relacionados con los sistemas de radar y los arreglos de sensores utilizados en ellos.

Se evaluó el desempeño de arreglos no uniformes para la estimación de direcciones de arribo de fuentes de señal y la información transmitida. Se propusieron tres tipos de configuraciones de los elementos de los arreglos comparándolos a su vez con un arreglo uniforme, todos con la misma cantidad de elementos. Una configuración ubica sus elementos aleatoriamente en la apertura prefijada, otra los ubica aleatoriamente pero impone una restricción a la distancia mínima entre elementos y la última los ubica aleatoriamente en posibles posiciones prefijadas que resultan de un arreglo uniforme de mayor cantidad de sensores. Adicionalmente se estudiaron arreglos optimizados basados en estas configuraciones. Para la optimización se aplicaron algoritmos genéticos y restricciones relacionadas con los algoritmos basados en modelos raros. Una de las restricciones está inspirada en el buen desempeño del arreglo restringido que impone una distancia mínima entre sensores, y otra restricción está dada por el nivel en los lóbulos laterales del patrón de radiación del arreglo. Se realizaron simulaciones aplicando diferentes clases de algoritmos de compressive sensing siendo el algoritmo ESBL el que aporta una mejor estimación de las señales de interés, y para el cual se obtienen diferencias de comportamiento entre las diferentes configuraciones del arreglo. De las primeras configuraciones, se concluyó que el arreglo restringido consigue mejores resultados, aunque el arreglo optimizado es

levemente superior a éste. Adicionalmente se estableció que, para diversos escenarios, la realización específica del arreglo restringido es robusta frente a diferentes ángulos de arribo de las señales, aunque no constante. Finalmente, se encontró que esta configuración mantiene un desempeño aceptable para altos valores de rareza.

Volviendo a la aplicación de arreglos de sensores lineales uniformes, se estudió el desempeño en ella de varios algoritmos basados en métodos ralos. En este escenario se busca estimar nuevamente tanto la señal como su dirección de arribo, pero en presencia de fuentes de interferencia. Se formuló el problema de medición de señales mediante arreglos de sensores como un modelo ralo, creando el diccionario en base a los ángulos de arribo. Luego, se compararon varios algoritmos clásicos y en particular se analizó el algoritmo ESBL. Este algoritmo no requiere información de la cantidad de fuentes presentes dado que anula los ángulos donde considera que no hay señal, mientras que los otros algoritmos deben realizar una búsqueda exhaustiva para determinar de todos los picos mayores cuales corresponden a señales de interés, pero para ello es necesario saber a priori la cantidad de señales presentes, tanto de interés como interferencias. Mediante distintas medidas de desempeño se encontró que este algoritmo se comporta igual o mejor que otros algoritmos tanto en la estimación de la señal en sí como de su ángulo de arribo. Aún más, se mostró que este desempeño puede obtenerse con un menor tiempo de cómputo.

Se introdujo un método para disminuir la cantidad de datos que debe manejar el algoritmo de estimación. Se analizó en particular el escenario de un modelo lineal con ruido estructurado cuya particularidad es que es ralo. Esta característica permite comprimir las mediciones, manteniendo la información y consecuentemente el desempeño. La compresión se obtiene proyectando la señal contaminada en un espacio de menor dimensión. Para el diseño de esta matriz de proyección, se propuso minimizar la traza de la covarianza condicional de la señal rala condicionada a los datos comprimidos. Esta elección permitió obtener una estimación de los parámetros altamente concentrada en su valor verdadero. Se demostró que el diseño de la matriz óptima se basa en un procedimiento de water-filling inverso. Mediante simulaciones se mostró que el procedimiento presentado tiene un mejor desempeño que otros métodos actuales para el diseño de la matriz de proyección. Pudo observarse

que el método propuesto para la proyección, al hacer uso de información sobre la interferencia, provee la mayor ganancia al combinarlo con el algoritmo ESBL que también explota esa información. Adicionalmente, se mostró la robustez del método aún cuando la estimación de los parámetros estadísticos necesarios para el diseño de la matriz no es precisa. Finalmente, para demostrar el desempeño del algoritmo en un escenario de la vida real se procesaron datos de radar polarimétrico correspondientes a un objetivo inmerso en clutter marítimo intenso. Se analizaron los casos de un objetivo estático y uno dinámico. En estos caso el método propuesto mostró tener un mejor rechazo de clutter comparado con otros métodos de proyección. Aún más, utilizando menor número de mediciones, su desempeño es mucho mejor que el caso sin proyección.

Existen muchas líneas de investigación que pueden continuarse más allá de los resultados mostrados en esta tesis, así como resultados parciales que requieren mayor confirmación teórica y experimental.

En relación a las configuraciones de arreglos de sensores, el tipo de arreglos introducido puede extenderse a arreglos planos en dos y tres dimensiones, donde se incorporarían más grados de libertad para para mejorar el desempeño del arreglo.

Los algoritmos basados en modelos ralos pueden extenderse para incluir otras características de las señales recibidas, como ser la polarización o Doppler, y poder así tener más grados de libertad para poder discriminar entre las señales de interés y la interferencia o clutter.

En ambos trabajos de arreglos de sensores, se consideraron señales cuyos ángulos de arribo correspondían a ángulos presentes en el diccionario. Esto no se ajusta a la realidad, donde el ángulo de arribo puede ser arbitrario. En trabajos futuros se analizará el caso de señales cuyos ángulos de arribo se encuentren entre dos columnas del diccionario, conocido como discrepancia de grilla (grid mismatch) [89].

Finalmente, en cuanto a la proyección de datos, es de interés analizar cómo aplicarlo al seguimiento de objetivos. En este escenario los parámetros estadísticos deberían ser actualizados mediante algún modelo dinámico correspondiente al objetivo e incorporados en el procedimiento de diseño de la matriz de proyección. Esto permitiría una mejor robustez frente a imprecisiones en la estimación.

Apéndice A

Algoritmo EM

Los modelos estadísticos utilizados en esta tesis consisten en un conjunto de datos observados \mathbf{x} y otro de datos ocultos \mathbf{z} que se generan de acuerdo a los parámetros $\boldsymbol{\theta}$ específicos del modelo. Estas cantidades se encuentran relacionadas mediante la función de verosimilitud $\mathcal{P}(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta})$. El objetivo es hallar los valores de los parámetros para los cuales los datos observados y ocultos son los más probables, es decir que se busca maximizar esta función.

El algoritmo de esperanza-maximización (EM) es un método numérico para determinar este estimador de máxima verosimilitud (MLE) [102]. Es un procedimiento iterativo que garantiza un incremento en la verosimilitud en cada paso y convergencia al menos a un máximo local. Este algoritmo aumenta los datos observados \mathbf{y} con un conjunto de datos ocultos \mathbf{z} para producir un nuevo conjunto de datos \mathbf{x} , llamado datos completos, que simplifican el cálculo del MLE.

Con el objetivo de encontrar una estimación de los parámetros $\boldsymbol{\theta}$, cada iteración del algoritmo EM consiste en dos pasos: el paso de esperanza **E** y el paso de maximización **M**. El paso **E** calcula la esperanza de la verosimilitud logarítmica de los datos completos $\mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) = \ln p(\mathbf{x}; \boldsymbol{\theta})$ con respecto a la distribución condicional de los datos ocultos dados los datos observados y los parámetros $\boldsymbol{\theta}$ en su valor actual $\boldsymbol{\theta}^{(t)}$. Luego, el paso **M** actualiza la estimación del parámetro $\boldsymbol{\theta}$ maximizando la función calculada en el paso previo con respecto a $\boldsymbol{\theta}$. En resumen,

$$\text{Paso E} \quad \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = \mathbf{E}_{z|y, \boldsymbol{\theta}^{(t)}} \left\{ \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) \mid \mathbf{y}, \boldsymbol{\theta}^{(t)} \right\} \quad (\text{A.1})$$

$$\text{Paso M} \quad \boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) \quad (\text{A.2})$$

Es decir que en el primer paso los datos ocultos son estimados dados los datos observados y las estimaciones actuales de los parámetros. Luego, en el segundo paso, se utilizan esas estimaciones de los datos ocultos, como si fueran los valores verdaderos, para maximizar la función de verosimilitud.

Apéndice B

Algoritmo ESBL

B.1. Block trace

Dada una matriz \mathbf{A} de tamaño $(PQ \times PQ)$ cuyos bloques de tamaño $(Q \times Q)$, llamados A_{ij} , para $i, j = 1, \dots, P$; el operador traza bloque es [103]

$$\text{btr}(\mathbf{A}) = \sum_{p=1}^P \mathbf{A}_{pp}. \quad (\text{B.1})$$

Nótese que $\text{btr}(\mathbf{A})$ es una matriz de tamaño $(Q \times Q)$.

B.2. Obtención de los estimadores de los parámetros

Para una de las instantáneas en particular, la función logarítmica de verosimilitud de los datos completos \mathbf{x} es

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) = \ln p(\mathbf{x}; \boldsymbol{\theta}). \quad (\text{B.2})$$

Luego, por el teorema de Bayes, $p(\mathbf{x}; \boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{b}, \mathbf{u}; \boldsymbol{\theta})p(\mathbf{u}|\mathbf{b}; \boldsymbol{\theta})p(\mathbf{b}; \boldsymbol{\theta})$, donde

$$p(\mathbf{y}|\mathbf{b}, \mathbf{u}; \boldsymbol{\theta}) = p(\mathbf{w}; \sigma), \quad (\text{B.3})$$

$$p(\mathbf{u}|\mathbf{b}; \boldsymbol{\theta}) = p(\mathbf{u}; \boldsymbol{\Sigma}_u), \quad (\text{B.4})$$

$$p(\mathbf{b}; \boldsymbol{\theta}) = p(\mathbf{b}; \boldsymbol{\beta}) = \prod_{n=1}^N p(b_n; \beta_n). \quad (\text{B.5})$$

Entonces,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) &= \ln p(\mathbf{w}; \sigma) + \ln p(\mathbf{u}; \boldsymbol{\Sigma}_u) + \sum_{n=1}^N \ln p(b_n; \beta_n) \\ &= \mathcal{L}(\mathbf{w}; \sigma) + \mathcal{L}(\mathbf{u}; \boldsymbol{\Sigma}_u) + \sum_{n=1}^N \mathcal{L}(b_n; \beta_n). \end{aligned} \quad (\text{B.6})$$

Nótese que los parámetros están desacoplados. Si se consideran las D observaciones independientes de los datos \mathbf{Y} , usando las variables del modelo MMV, \mathbf{B} , \mathbf{U} y \mathbf{W} ,

$$\mathcal{L}(\mathbf{B}_n; \beta_n) = -D \left(\ln \pi + \ln \beta_n + \beta_n^{-1} \frac{1}{D} \sum_{d=1}^D |\mathbf{B}_{nd}|^2, \right) \quad (\text{B.7})$$

$$\begin{aligned} \mathcal{L}(\mathbf{U}; \boldsymbol{\Sigma}_u) &= -D \left(PQ \ln \pi + \ln |\mathbf{C}_u| + \text{tr} \left(\frac{\mathbf{C}_u^{-1}}{D} \sum_{d=1}^D \mathbf{U}_{\cdot d} \mathbf{U}_{\cdot d}^H \right) \right) \\ &= -DP \left(Q \ln \pi + \ln |\boldsymbol{\Sigma}_u| + \text{tr} \left(\frac{\boldsymbol{\Sigma}_u^{-1}}{DP} \sum_{d=1}^D \sum_{p=1}^P [\mathbf{U}_{\cdot d}]_p [\mathbf{U}_{\cdot d}^H]_p \right) \right), \end{aligned} \quad (\text{B.8})$$

$$\mathcal{L}(\mathbf{W}; \sigma) = -DM \left(\ln \pi + \ln \sigma + \sigma^{-1} \text{tr} \left(\frac{1}{DM} \sum_{d=1}^D \mathbf{W}_{\cdot d} \mathbf{W}_{\cdot d}^H \right) \right), \quad (\text{B.9})$$

donde se usaron las propiedades del producto de Kronecker, $|\mathbf{C}_u| = |\mathbf{I}_P \otimes \boldsymbol{\Sigma}_u| = |\boldsymbol{\Sigma}_u|^P$ y $\mathbf{C}_u^{-1} = (\mathbf{I}_P \otimes \boldsymbol{\Sigma}_u)^{-1} = \mathbf{I}_P \otimes \boldsymbol{\Sigma}_u^{-1}$. Aplicando el valor esperado a (B.6) se obtiene la función \mathcal{Q} necesaria para el algoritmo EM,

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = \mathcal{Q}(\sigma, \boldsymbol{\theta}^{(t)}) + \mathcal{Q}(\boldsymbol{\Sigma}_u, \boldsymbol{\theta}^{(t)}) + \sum_{n=1}^N \mathcal{Q}(\beta_n, \boldsymbol{\theta}^{(t)}). \quad (\text{B.10})$$

Se debe maximizar esta última función para obtener los mejores estimadores de los parámetros. Como éstos están desacoplados se puede trabajar con cada término por

separado. Luego,

$$\mathcal{Q}(\beta_n, \boldsymbol{\theta}^{(t)}) = -D \left(\ln \pi + \ln \beta_n + \beta_n^{-1} \mathbf{E}_{\mathbf{B}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ S_{b_n} | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} \right), \quad (\text{B.11})$$

donde

$$S_{b_n} = \frac{1}{D} \sum_{d=1}^D |b_n|^2. \quad (\text{B.12})$$

En este caso es simple demostrar, igualando la derivada a cero, que la función $\mathcal{Q}(\beta_n, \boldsymbol{\theta}^{(t)})$ se maximiza cuando

$$\beta_n = \mathbf{E}_{\mathbf{B}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ S_{b_n} | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\}. \quad (\text{B.13})$$

La esperanza condicional es

$$\begin{aligned} \mathbf{E}_{\mathbf{B}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ S_{b_n} | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} &= \frac{1}{D} \sum_{d=1}^D \mathbf{E}_{\mathbf{B}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ |b_n|^2 | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} \\ &= \frac{1}{D} \sum_{d=1}^D \beta_n^{(t)} + (\beta_n^{(t)})^2 \mathbf{X}_{\cdot n}^H (\boldsymbol{\Sigma}_y^{-1} \mathbf{Y}_{\cdot d} \mathbf{Y}_{\cdot d}^H \boldsymbol{\Sigma}_y^{-1} - \boldsymbol{\Sigma}_y^{-1}) \mathbf{X}_{\cdot n} \\ &= \beta_n^{(t)} + (\beta_n^{(t)})^2 \mathbf{X}_{\cdot n}^H (\boldsymbol{\Sigma}_y^{-1} \mathbf{C}_y \boldsymbol{\Sigma}_y^{-1} - \boldsymbol{\Sigma}_y^{-1}) \mathbf{X}_{\cdot n}, \end{aligned} \quad (\text{B.14})$$

donde se utiliza el teorema 10.3 de [93] para el cálculo de la esperanza condicional. De esta manera se obtiene la actualización de la ecuación (2.18)

Por otra parte,

$$\mathcal{Q}(\boldsymbol{\Sigma}_u, \boldsymbol{\theta}^{(t)}) = -DP \left(Q \ln \pi + \ln |\boldsymbol{\Sigma}_u| + \text{tr} \left(\boldsymbol{\Sigma}_u^{-1} \mathbf{E}_{\mathbf{U}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ S_u | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} \right) \right), \quad (\text{B.15})$$

donde

$$S_u = \frac{1}{DP} \sum_{d=1}^D \text{btr}(\mathbf{U}_{\cdot d} \mathbf{U}_{\cdot d}^H). \quad (\text{B.16})$$

Siendo \mathbf{A} una matriz definida positiva de tamaño $Q \times Q$ y \mathbf{B} otra matriz definida positiva, entonces la desigualdad $\ln |\mathbf{A}| + \text{tr}(\mathbf{A}^{-1} \mathbf{B}) \geq Q + \ln |\mathbf{B}|$ es válida y alcanza la igualdad cuando $\mathbf{A} = \mathbf{B}$ [101]. Por lo tanto, la función $\mathcal{Q}(\boldsymbol{\Sigma}_u, \boldsymbol{\theta}^{(t)})$ se maximiza cuando

$$\boldsymbol{\Sigma}_u = \mathbf{E}_{\mathbf{U}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ S_u | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} \quad (\text{B.17})$$

siendo

$$\mathbf{E}_{\mathbf{U}|\mathbf{Y},\boldsymbol{\theta}^{(t)}} \left\{ S_u | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} = \frac{1}{P} \text{btr} \left(\mathbf{C}_u^{(t)} + \mathbf{C}_u^{(t)} \mathbf{Z}^H \mathbf{C}_y \mathbf{Z} \mathbf{C}_u^{(t)} \right) \quad (\text{B.18})$$

Se obtiene así la actualización del parámetro $\boldsymbol{\Sigma}_u$ de la ecuación (2.17).

Finalmente para el ruido, la función a maximizar es

$$\mathcal{Q} \left(\sigma, \boldsymbol{\theta}^{(t)} \right) = -DM \left(\ln \pi + \ln \sigma + \sigma^{-1} \mathbf{E}_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}} \left\{ S_w | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} \right), \quad (\text{B.19})$$

donde

$$S_w = \frac{1}{DM} \sum_{d=1}^D \text{tr} \left(\mathbf{W}_{\cdot d} \mathbf{W}_{\cdot d}^H \right). \quad (\text{B.20})$$

También en este caso se demuestra fácilmente que la función $\mathcal{Q} \left(\sigma, \boldsymbol{\theta}^{(t)} \right)$ se maximiza cuando

$$\sigma = \mathbf{E}_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}} \left\{ S_w | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\}, \quad (\text{B.21})$$

donde

$$\mathbf{E}_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}} \left\{ S_w | \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} = \frac{1}{M} \text{tr} \left(\sigma^{(t)} \mathbf{I}_M + (\sigma^{(t)})^2 \mathbf{C}_y \right). \quad (\text{B.22})$$

Se demuestra entonces la actualización del parámetro σ de la ecuación (2.16).

B.3. Obtención del test de detección

Se construye un test de Neyman-Person basado en los datos completos $\mathbf{C} = [\mathbf{Y}, \mathbf{B}, \mathbf{U}]$, donde el test logarítmico decide por \mathcal{H}_1 si

$$\mathcal{T}^{(n)} = L(\mathcal{H}_1; \mathbf{C}) - L(\mathcal{H}_0; \mathbf{C}) > \alpha \quad (\text{B.23})$$

donde $L(\mathcal{H}_1; \mathbf{C})$ y $L(\mathcal{H}_0; \mathbf{C})$ son las funciones logarítmicas de verosimilitud de \mathbf{C} bajo las hipótesis \mathcal{H}_1 y \mathcal{H}_0 respectivamente. La primera está definida en la ecuación (B.6) y la segunda es similar pero no posee el término correspondiente al parámetro β_n . Entonces,

$$\mathcal{T}^{(n)} = \ln p(\mathbf{B}_n; \beta_n^{(t)}) \quad (\text{B.24})$$

Un test equivalente resulta de aplicarle una función monótonamente creciente a la expresión anterior. Aplicamos la esperanza con respecto a la distribución condicional de los datos ocultos dados los datos observados con el valor de los parámetros $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$,

$$\mathcal{T}^{(n)} = \int \ln p(\mathbf{B}_n; \beta_i^{(t)}) p(\mathbf{C}/\mathbf{Y}, \boldsymbol{\theta}^{(t)}) d\mathbf{C} \quad (\text{B.25})$$

Considerando el Paso **E** definido en (A.1), se trata entonces de la función \mathcal{Q} evaluada en $\beta_n^{(t+1)}$, $\mathcal{Q}(\beta_n^{(t+1)}, \boldsymbol{\theta}^{(t)})$, que resulta ser la función maximizada en el paso **M** definido en (A.2). Combinando (B.11) y (B.13) el test es,

$$\mathcal{T}^{(n)} = -D \left(1 + \ln \pi + \ln \left(\mathbf{E}_{\mathbf{B}|\mathbf{Y}, \boldsymbol{\theta}^{(t)}} \left\{ \mathcal{L}(\boldsymbol{\theta}; \mathbf{C}) \mid \mathbf{Y}, \boldsymbol{\theta}^{(t)} \right\} \right) \right) > \alpha \quad (\text{B.26})$$

Quitando el signo, la función logarítmica y dejando el término que sólo depende de los datos a la izquierda se obtiene el test estadístico descrito en (2.23).

Apéndice C

Radar polarimétrico

C.1. Introducción

La orientación de los campos eléctrico y magnético en el plano perpendicular a la dirección de desplazamiento de la onda electromagnética, se denomina polarización de la onda. Los diversos estados de polarización de esta onda permiten capturar la misma información de un objetivo de distintas formas. Esto se denomina diversidad en polarización.

Los radares polarimétricos aprovechan la polarización de la onda transmitida para describir distintas propiedades del objetivo y del clutter presente en el terreno iluminado por el radar. Esto es importante cuando se desea detectar y seguir un objetivo que posee una pequeña sección radar equivalente (RCS) inmerso en un fuerte clutter.

Los sistemas de radar convencionales operan con la misma polarización en sus antenas para transmisión y recepción. Los radares polarimétricos, en cambio, transmiten y reciben ondas electromagnéticas con distinta polarización, permitiendo así obtener información polarimétrica del objetivo y el ambiente. Esta información puede ser la geometría, el material o la orientación del objetivo. El uso de esta información puede mejorar notablemente el desempeño del radar, sobre todo si las características usualmente usadas, como ser tiempo, frecuencia o dirección de arribo, no son suficientes para discriminar el objetivo del clutter.

C.2. Modelado

Se considera un radar mono-estático capaz de transmitir señales con polarizaciones arbitrarias pulso a pulso. Para poder identificar completamente los aspectos polarimétricos del objetivo, el tiempo de integración del radar (dwell) debe consistir en pulsos con diversidad en polarización. Por ejemplo, un radar polarimétrico convencional transmite una secuencia alternante de pulsos con polarización vertical y horizontal.

El campo electromagnético reflejado por el objetivo está relacionado con el campo incidente mediante la matriz de scattering \mathbf{S} ,

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad (\text{C.1})$$

donde las variables S_{11} y S_{22} son los coeficientes de scattering co-polar y S_{12} y S_{21} son los coeficientes cross-polares. Para el caso del radar mono-estático estos últimos son iguales.

El campo electromagnético transmitido \mathbf{f} puede expresarse en función de las bases polarimétricas del transmisor, $\mathbf{f} = [f_1, f_2]^T$. Por ejemplo cuando se tienen bases polarimétricas correspondientes a polarización horizontal y vertical, pueden representarse mediante dos ángulos,

$$\mathbf{f} = \begin{bmatrix} \cos(\alpha) & \text{sen}(\alpha) \\ -\text{sen}(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} \cos(\beta) \\ j \text{sen}(\alpha) \end{bmatrix}, \quad (\text{C.2})$$

donde α indica la orientación y β la elipticidad de la polarización. Otras bases polarimétricas usadas frecuentemente son polarización circular derecha e izquierda, y polarización oblicua izquierda y derecha.

Cuando el radar transmite el pulso j -ésimo, el campo reflejado por la celda de rango p -ésima resulta de la contribución del objetivo presente y del clutter. Este campo es medido por el receptor mediante R antenas, cada una con su polarización, obteniendo el vector de mediciones \mathbf{y}_{pj} de tamaño R del campo reflejado correspondiente al pulso j y a la celda p

$$\mathbf{y}_{pj} = \mathbf{G} (\mathbf{S}_p^o + \mathbf{S}_p^c) \mathbf{f}_j + \mathbf{w}_{pj}, \quad (\text{C.3})$$

donde \mathbf{G} es la matriz de tamaño $R \times 2$ representando las polarizaciones del arreglo de antenas del receptor, \mathbf{S}_p^o y \mathbf{S}_p^c son las matrices de scattering correspondientes al objetivo y al clutter, \mathbf{f}_j es el campo incidente correspondiente al pulso j y \mathbf{w} representa el error en la medición. Estas mediciones pueden agruparse para las distintas celdas de rango en

$$\mathbf{y}_j = \begin{bmatrix} \mathbf{y}_{1j} \\ \vdots \\ \mathbf{y}_{Pj} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \cdots 0 \\ \vdots \cdots \vdots \\ 0 \cdots \mathbf{G} \end{bmatrix} \left(\begin{bmatrix} \mathbf{S}_1^o \\ \vdots \\ \mathbf{S}_P^o \end{bmatrix} + \begin{bmatrix} \mathbf{S}_1^c \\ \vdots \\ \mathbf{S}_P^c \end{bmatrix} \right) \mathbf{f}_j + \begin{bmatrix} \mathbf{w}_{1j} \\ \vdots \\ \mathbf{w}_{Pj} \end{bmatrix} \quad (\text{C.4})$$

$$= (\mathbf{I}_P \otimes \mathbf{G}) (\mathbf{S}^o + \mathbf{S}^c) \mathbf{f}_j + \mathbf{w}_j, \quad (\text{C.5})$$

donde \mathbf{y}_j y \mathbf{w}_j son vectores de tamaño $2R$, y \mathbf{S}^o y \mathbf{S}^c son matrices de tamaño $2P \times 2$. Agrupando los datos para los distintos pulsos podemos reescribir las mediciones del receptor,

$$\mathbf{Y} = (\mathbf{I}_P \otimes \mathbf{G}) (\mathbf{S}^o + \mathbf{S}^c) \mathbf{F} + \mathbf{W}. \quad (\text{C.6})$$

donde \mathbf{Y} y \mathbf{W} son matrices de $RP \times J$, y la matriz $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_J]$, de tamaño $2 \times J$, representa la polarización de los pulsos transmitidos. Aplicando las propiedades del producto de Kronecker, $\text{vec}(\mathbf{A} + \mathbf{B}) = \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{B})$ y $\text{vec}(\mathbf{ADB}) = (\mathbf{B}^T \mathbf{A}) \text{vec}(\mathbf{D})$, los datos pueden apilarse en un vector de longitud $N = JRP$,

$$\mathbf{y} = \text{vec}(\mathbf{Y}) = (\mathbf{F}^T \otimes \mathbf{I}_P \otimes \mathbf{G}) \text{vec}(\mathbf{S}^o + \mathbf{S}^c) + (\mathbf{I}_J \otimes \mathbf{G}) \text{vec}(\mathbf{W}). \quad (\text{C.7})$$

Otra forma de describir los coeficientes de scattering es en forma vectorial

$$\mathbf{u}_p^o = \begin{bmatrix} S_{11} \\ S_{22} \\ S_{12} \end{bmatrix} \quad (\text{C.8})$$

Luego,

$$\text{vec}(\mathbf{S}^o) = \sum_{p=1}^P \mathbf{H}_p \mathbf{u}_p. \quad (\text{C.9})$$

donde \mathbf{H}_p son matrices de tamaño $(4P \times 3)$ que tienen componentes nulas excepto por los elementos

$$\begin{aligned} [\mathbf{H}_p]_{2p-1,1} &= 1 & [\mathbf{H}_p]_{2p,3} &= 1 \\ [\mathbf{H}_p]_{2(p+P),1} &= 1 & [\mathbf{H}_p]_{2(p+P),3} &= 1 \end{aligned} \quad (\text{C.10})$$

En forma similar se puede expresar $\text{vec}(\mathbf{S}^e)$ mediante \mathbf{u}_p^c . Luego, renombrando el vector de tamaño M , $\mathbf{w} = (\mathbf{I}_J \otimes \mathbf{G}) \text{vec}(\mathbf{W})$

$$\mathbf{y} = (\mathbf{F}^T \otimes \mathbf{I}_P \otimes \mathbf{G}) \sum_{p=1}^P \mathbf{H}_p (\mathbf{u}_p^o + \mathbf{u}_p^c) + \mathbf{w}. \quad (\text{C.11})$$

C.3. Representación rala

Para generar un diccionario sobrecompleto se considera la descomposición de Krogager de la matriz de scattering [100]. Esta descomposición se basa en que la matriz de scattering puede representarse por la combinación de tres formas canónicas: una esfera, un diplano y una hélice. Luego, la vectorización de la matriz de scattering del objetivo \mathbf{S}_p^o es

$$\mathbf{u}_p^o = \sum_{l=1}^L b_{pl} \mathbf{u}_{pl}. \quad (\text{C.12})$$

donde b_{pl} son los coeficientes de peso para cada uno de los objetivos canónicos y L es el número total de componentes incluyendo la esfera, las hélices izquierda y derecha y los diplanos (dos planos a 90 grados) para distintos ángulos de orientación ϕ . Los vectores de scattering para cada forma son:

$$\mathbf{u}_{\text{sphere}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_{\text{helix}} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ \pm j \end{bmatrix}, \quad \mathbf{u}_{\text{diplane}} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos \phi \\ -\cos \phi \\ \text{sen } \phi \end{bmatrix} \quad (\text{C.13})$$

Luego, el término correspondiente al objetivo puede descomponerse como

$$\begin{aligned}
(\mathbf{F}^T \otimes \mathbf{I}_P \otimes \mathbf{G}) \text{vec}(\mathbf{S}^o) &= \sum_{p=1}^P \sum_{l=1}^L (\mathbf{F}^T \otimes \mathbf{I}_P \otimes \mathbf{G}) \mathbf{H}_p \mathbf{u}_p^o \\
&= \sum_{m=1}^M \mathbf{X}_m b_m \\
&= \mathbf{X} \mathbf{b},
\end{aligned} \tag{C.14}$$

donde $M = PL$. El vector \mathbf{b} resulta ralo ya que sólo unas pocas celdas de rango se encuentran ocupadas por objetivos, y éstos pueden ser representados por pocas componentes fundamentales de las L posibles formas canónicas de polarización. Similarmente, para el clutter

$$\begin{aligned}
(\mathbf{F}^T \otimes \mathbf{I}_P \otimes \mathbf{G}) \text{vec}(\mathbf{S}^c) &= \sum_{p=1}^P (\mathbf{F}^T \otimes \mathbf{I}_P \otimes \mathbf{G}) \mathbf{H}_p \mathbf{u}_p^c \\
&= \sum_{p=1}^P \mathbf{Z}_p \mathbf{u}_p^c \\
&= \mathbf{Z} \mathbf{u}.
\end{aligned} \tag{C.15}$$

Combinando (C.14) y (C.15) en (C.7) se obtiene el modelo (2.11) utilizado por el algoritmo ESBL.

Apéndice D

Demostración del óptimo λ_i

Para resolver el problema (5.16), se introduce el lagrangiano

$$\mathcal{L} = \text{tr}(\mathbf{\Sigma}_{b|y}) + \alpha \left(E - \sum_{k=1}^K \lambda_k \right) + \sum_{k=1}^K \nu_k \lambda_k. \quad (\text{D.1})$$

Usando las condiciones de Karush-Kuhn-Tucker [90], en el punto óptimo $(\boldsymbol{\lambda}^*, \alpha^*, \boldsymbol{\nu}^*)$ se tiene, para $i = 1, \dots, K$,

$$\begin{aligned} \nabla \mathcal{L} &= \mathbf{0}, \\ \nu_i &\geq 0, \\ \lambda_i &\geq 0, \\ \nu_i \lambda_i &= 0, \\ E - \sum_{k=1}^M \lambda_k &= 0. \end{aligned} \quad (\text{D.2})$$

Es posible deshacerse de $\boldsymbol{\nu}$ utilizando la primera condición en el lagrangiano

$$\nu_i = \alpha - \frac{\partial \text{tr}(\mathbf{\Sigma}_{b|y})}{\partial \lambda_i}, \quad (\text{D.3})$$

luego las condiciones para la solución óptima se reducen a

$$\nabla \left(\text{tr}(\Sigma_{b|y}) + \alpha \left(E - \sum_{k=1}^K \lambda_k \right) + \sum_{k=1}^K \nu_k \lambda_k \right) = \mathbf{0}, \quad (\text{D.4})$$

$$\frac{\partial \text{tr}(\Sigma_{b|y})}{\partial \lambda_i} + \alpha \leq 0, \quad (\text{D.5})$$

$$\left(\frac{\partial \text{tr}(\Sigma_{b|y})}{\partial \lambda_i} + \alpha \right) \lambda_i = 0, \quad (\text{D.6})$$

$$\lambda_i \geq 0, \quad (\text{D.7})$$

$$\sum_{k=1}^K \lambda_k - E = 0. \quad (\text{D.8})$$

A continuación se necesita hallar la derivada parcial de la covarianza condicional

$$\begin{aligned} \frac{\partial \text{tr}(\Sigma_{b|y})}{\partial \lambda_i} &= \frac{\partial}{\partial \lambda_i} \left[\text{tr}(\Sigma_b) - \sum_{k=1}^K \frac{\lambda_k}{\sigma + d_k \lambda_k} t_k \right] \\ &= -\frac{t_i \sigma}{(\sigma + d_i \lambda_i)^2}. \end{aligned} \quad (\text{D.9})$$

En este punto se tienen dos condiciones para α . Si $\alpha < \frac{t_i}{\sigma}$ entonces por (D.5) $\lambda_i > 0$ y por (D.6), $\alpha = \frac{t_i \sigma}{(\sigma + d_i \lambda_i)^2}$. Luego

$$\lambda_i = \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{\sigma \alpha}} - 1 \right). \quad (\text{D.10})$$

Por otro lado, si $\alpha \geq \frac{t_i}{\sigma}$ entonces λ_i debe ser cero. Para demostrar esto, si se asume que λ_i es positivo, entonces

$$\alpha \geq \frac{t_i}{\sigma} > \frac{t_i}{\sigma} \frac{1}{\left(1 + \frac{d_i \lambda_i}{\sigma}\right)^2}, \quad (\text{D.11})$$

y

$$\left(\frac{t_i \sigma}{(\sigma + d_i \lambda_i)^2} - \alpha \right) \lambda_i \neq 0, \quad (\text{D.12})$$

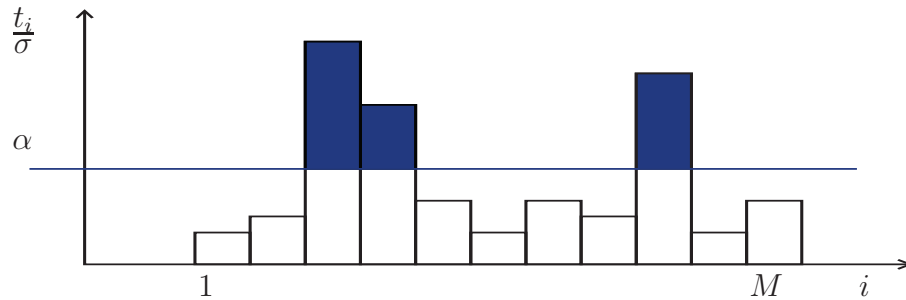


Figura D.1: Solución *inverse water-filling* para los autovalores de \mathbf{H} basada en la relación entre t_i y α .

que viola la optimalidad de la solución en la ecuación (D.6). Por lo tanto, λ_i debe ser cero. En resumen, los autovalores óptimos de \mathbf{H} son

$$\lambda_i = \max \left\{ 0, \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{\sigma\alpha}} - 1 \right) \right\}. \quad (\text{D.13})$$

Esta es una solución del tipo *inverse water-filling* donde la restricción de la energía se distribuye entre aquellos autovalores cuyo correspondientes valores de t_i son mayores que $\sigma\alpha$, como se muestra en la Figura. D.1. Asumiendo que los primeros L autovalores óptimos λ_i son distintos de cero, se utiliza (D.13) en (D.8) y se halla el multiplicador de Lagrange

$$\alpha = \frac{\left(\sum_{i=1}^L \sqrt{\frac{t_i}{\sigma}} \frac{\sigma}{d_i} \right)^2}{\left(\sum_{i=1}^L \frac{\sigma}{d_i} + E \right)^2}. \quad (\text{D.14})$$

Para poder hallar el valor de L , se asume que los T_i están en orden decreciente, luego se fija $L = K$ y se calcula α y λ_L . Si λ_L es negativo entonces el valor óptimo de L debe ser menor y se intenta con $L = K - 1$. Este proceso se repite hasta que λ_L sea positivo. Luego, el resto de los autovalores λ_i pueden calcularse con el valor correcto de α y L , de acuerdo a (D.13).

Bibliografía

- [1] M. I. Skolnik, *Introduction to Radar Systems*. McGraw-Hill, 1980.
- [2] S. Haykin, “Radar signal processing,” *IEEE ASSP Mag.*, vol. 2, no. 2, pp. 2–18, Apr 1985.
- [3] I. F. Gorodnitsky and B. D. Rao, “Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm,” *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [4] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.
- [5] E. Candes and T. Tao, “The Dantzig selector: Statistical estimation when p is much larger than n ,” *Ann. Stat.*, vol. 35, p. 2313, 2007.
- [6] D. Malioutov, M. Cetin, and A. S. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug 2005.
- [7] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.
- [8] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, “Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit,” Stanford, Tech. Rep., 2006. [Online]. Available: <http://www-stat.stanford.edu/~donoho/Reports>
- [9] D. Needell and J. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.
- [10] D. Wipf and B. Rao, “Sparse bayesian learning for basis selection,” *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [11] M. Hurtado, C. Muravchik, and A. Nehorai, “Enhanced sparse bayesian learning via statistical thresholding for signals in structured noise,” *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5430–5443, Nov. 2013.

-
- [12] Y. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Trans. Antennas Propag.*, vol. 12, no. 3, pp. 257–268, May 1964.
- [13] L. Carin, "On the relationship between compressive sensing and random sensor arrays," *IEEE Antennas Propag. Mag.*, vol. 51, no. 5, pp. 72–81, Oct. 2009.
- [14] L. Carin, D. Liu, and B. Guo, "Coherence, compressive sensing, and random sensor arrays," *IEEE Antennas Propag. Mag.*, vol. 53, no. 4, pp. 28–39, Aug. 2011.
- [15] G. Oliveri and A. Massa, "Bayesian compressive sampling for pattern synthesis with maximally sparse non-uniform linear arrays," *IEEE Trans. Antennas Propag.*, vol. 59, no. 2, pp. 467–481, Feb 2011.
- [16] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 575–585, Dec 2007.
- [17] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [18] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [19] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [20] V. D. M. Nhat, D. Vo, S. Challa, and S. Lee, "Efficient projection for compressed sensing," in *7th IEEE/ACIS Int. Conf. Comput. Inform. Sci.*, May 2008, pp. 322–327.
- [21] J. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
- [22] A. Tehrani, A. Dimakis, and G. Caire, "Optimal deterministic compressed sensing matrices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2013, pp. 5895–5899.
- [23] A. Amini and F. Marvasti, "Deterministic construction of binary, bipolar, and ternary compressed sensing matrices," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2360–2370, Apr. 2011.
- [24] Y. Yu, A. Petropulu, and H. Poor, "Measurement matrix design for compressive sensing-based mimo radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5338–5352, Nov. 2011.

- [25] S. Li, Z. Zhu, G. Li, L. Chang, and Q. Li, "Projection matrix optimization for block-sparse compressive sensing," in *Proc. Int. Conf. Signal Process., Commun. & Comput. (ICSPCC)*, Aug. 2013, pp. 1–4.
- [26] L. Zelnik-Manor, K. Rosenblum, and Y. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
- [27] J. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Adapted statistical compressive sensing: Learning to sense gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3653–3656.
- [28] Y. Lu, S. Hegler, C. Statz, A. Finger, and D. Plettemeier, "Improving detection performance of compressed sensing by orthogonal projection," in *Int. Radar Symp.*, vol. 1, Jun. 2013, pp. 107–112.
- [29] W. Carson, M. Chen, M. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imaging Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.
- [30] S. Pazos, M. Hurtado, and C. H. Muravchik, "DOA estimation using random linear arrays via compressive sensing," *IEEE Biennial Congress of Argentina ARGENCON*, pp. 219–222, Jun. 2014.
- [31] S. Pazos, M. Hurtado, and C. Muravchik, "Doa estimation using random linear arrays via compressive sensing," *IEEE Latin Amer. Trans.*, vol. 12, no. 5, pp. 859–863, Aug 2014.
- [32] S. Pazos, M. Hurtado, and C. H. Muravchik, "Métodos ralos para arreglos de sensores en presencia de interferencias," in *XV Reunión de Trabajo en Procesamiento de la Información y Control (RPIC2013)*, Sep. 2013, pp. 451–456.
- [33] S. Pazos, M. Hurtado, and C. Muravchik, "On sparse methods for array signal processing in the presence of interference," *IEEE Antennas Wireless Propag. Lett.*, vol. PP, no. 99, pp. 1–1, 2015.
- [34] S. Pazos, M. Hurtado, and C. H. Muravchik, "Optimal compression for sparse linear models," in *XIV Reunión de Trabajo en Procesamiento de la Información y Control (RPIC2011)*, Sep. 2011, pp. 631–635.
- [35] S. Pazos, M. Hurtado, C. Muravchik, and A. Nehorai, "Optimal sensing matrix for sparse linear models," in *IEEE Int. Workshop on Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, 2011, pp. 257–260.
- [36] B. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan 1999.

- [37] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm," *Electroencephalog. Clinical Neurophysiol.*, vol. 95, no. 4, pp. 231 – 251, 1995.
- [38] D. Wipf and S. Nagarajan, "A unified bayesian framework for meg/eeg source imaging," *NeuroImage*, vol. 44, no. 3, pp. 947 – 966, 2009.
- [39] N. von Ellenrieder, M. Hurtado, and C. Muravchik, "Electromagnetic source imaging for sparse cortical activation patterns," in *Proc. IEEE Int. Conf. Eng. in Medicine and Biology Soc. EMBC*, Sep. 2010, pp. 4316 –4319.
- [40] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept 2006.
- [41] M. Protter, I. Yavneh, and M. Elad, "Closed-form mmse estimation for signal denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3471–3484, July 2010.
- [42] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 520–544, Jan 2010.
- [43] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wide-band analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, April 2010.
- [44] A. Harms, W. U. Bajwa, and R. Calderbank, "Beating nyquist through correlations: A constrained random demodulator for sampling of sparse bandlimited signals," in *Proc. ICASSP*, May 2011, pp. 5968–5971.
- [45] D. Shutin and B. H. Fleury, "Sparse variational bayesian sage algorithm with application to the estimation of multipath wireless channels," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3609–3623, Aug 2011.
- [46] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.
- [47] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 927–937, Jun 1999.
- [48] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar 2002.

- [49] S. D. Cabrera and T. W. Parks, "Extrapolation and spectral estimation with iterative weighted norm modification," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 842–851, Apr 1991.
- [50] B. D. Jeffs, "Sparse inverse solution methods for signal and image processing applications," in *Proc. ICASSP*, vol. 3, May 1998, pp. 1885–1888 vol.3.
- [51] S. Chen and J. Wigger, "Fast orthogonal least squares algorithm for efficient subset model selection," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1713–1715, Jul 1995.
- [52] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.
- [53] R. E. Carlson and B. K. Natarajan, "Sparse approximate multiquadric interpolation," *Comput. Math. Applicat.*, vol. 27, no. 6, pp. 99 – 108, 1994.
- [54] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancellers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep 2000.
- [55] B. D. Rao and B. S., "Adaptive filtering algorithms for promoting sparsity," in *Proc. ICASSP*, vol. 6, April 2003, pp. VI-361–4 vol.6.
- [56] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, June 2009.
- [57] J. H. Ender, "On compressive sensing applied to radar," *Signal Process.*, vol. 90, no. 5, pp. 1402 – 1414, 2010.
- [58] L. Anitori, A. Maleki, M. Otten, R. G. Baraniuk, and P. Hoogeboom, "Design and analysis of compressed sensing radar detectors," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 813–827, Feb 2013.
- [59] Y. Yu, A. P. Petropulu, and H. V. Poor, "Mimo radar using compressive sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 146–163, Feb 2010.
- [60] S. Gogineni and A. Nehorai, "Target estimation using sparse modeling for distributed mimo radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5315–5325, Nov 2011.
- [61] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.
- [62] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Ser. B*, vol. 58, pp. 267–288, 1996.
- [63] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.
- [64] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

- [65] D. L. Donoho, Y. Tsaig, I. Drori, and J. luc Starck, “Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit,” Tech. Rep., 2006.
- [66] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Computat. Harmon. Anal.*, vol. 26, no. 3, pp. 301 – 321, 2009.
- [67] E. G. Larsson and Y. Selen, “Linear regression with a sparse parameter vector,” *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 451–460, Feb 2007.
- [68] M. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.
- [69] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.
- [70] M. E. Tipping, A. Faul, J. J. T. Avenue, and J. J. T. Avenue, “Fast marginal likelihood maximisation for sparse bayesian models,” in *Proc. Int. Workshop Artif. Intell. Statist.*, 2003, pp. 3–6.
- [71] K. Qiu and A. Dogandzic, “Sparse signal reconstruction via ecme hard thresholding,” *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4551–4569, Sept 2012.
- [72] P. Schniter, L. Potter, and J. Ziniel, “Fast bayesian matching pursuit,” in *Proc. Inf. Theory Appl. Workshop*, Jan 2008, pp. 326–333.
- [73] B. Rao, K. Engan, S. Cotter, J. Palmer, and K. Kreutz-Delgado, “Subset selection in noise based on diversity measure minimization,” *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, March 2003.
- [74] M. Elad and I. Yavneh, “A plurality of sparse representations is better than the sparsest one alone,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4701–4714, Oct 2009.
- [75] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec 2006.
- [76] A. Rakotomamonjy, “Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms,” *Signal Process.*, vol. 91, no. 7, pp. 1505–1526, 2011.
- [77] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. part I: Greedy pursuit,” *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [78] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

- [79] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul 2007.
- [80] J. Tropp, "Algorithms for simultaneous sparse approximation. part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.
- [81] D. Maiwald and D. Kraus, "Calculation of moments of complex wishart and complex inverse wishart distributed matrices," *IEE Proc. Radar, Sonar and Navig.*, vol. 147, no. 4, pp. 162–168, Aug 2000.
- [82] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.
- [83] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [84] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*. John Wiley & Sons, 2002.
- [85] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.
- [86] J. R. Koza, *Genetic Programming*. MA: MITPress, 1992.
- [87] M. Gavish and A. J. Weiss, "Array geometry for ambiguity resolution in direction finding," *IEEE Trans. Antennas Propag.*, vol. 44, no. 6, pp. 889–895, Jun 1996.
- [88] V. Roy, S. Chepuri, and G. Leus, "Sparsity-enforcing sensor selection for doa estimation," in *IEEE Int. Workshop on Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec 2013, pp. 340–343.
- [89] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.
- [90] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.
- [91] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [92] —, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.
- [93] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

-
- [94] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Technical University of Denmark, Oct. 2008, version 20081110.
- [95] D. Fuhrmann, “One-step optimal measurement selection for linear gaussian estimation problems,” in *Proc. Int. Waveform Diversity Design Conf.*, Jun. 2007, pp. 224–227.
- [96] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.
- [97] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization,” *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar 2003.
- [98] S. Haykin, C. Krasnor, T. J. Nohara, B. W. Currie, and D. Hamburger, “A coherent dual-polarized radar for studying the ocean environment,” *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 1, pp. 189–191, 1991.
- [99] M. Hurtado, N. von Ellenrieder, C. Muravchik, and A. Nehorai, “A sparse modeling for polarimetric radar,” in *IEEE Workshop Stat. Signal Process.*, 2011.
- [100] E. Krogager, “New decomposition of the radar target scattering matrix,” *Electron. Lett.*, vol. 26, pp. 1525–1527, 1990.
- [101] P. Stoica and A. Nehorai, “On the concentrated stochastic likelihood function in array signal processing,” *Circuits, Syst., Signal Process.*, vol. 14, no. 5, pp. 669–674, 1995.
- [102] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *J. Roy. Stat. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1977. [Online]. Available: <http://dx.doi.org/10.2307/2984875>
- [103] A. Edelman and H. Murakami, “Polynomial roots from companion matrix eigenvalues,” *Math. Comput.*, vol. 64, no. 4, pp. 763–776, 1995.