UNIVERSIDAD DE CANTABRIA

(Universidades de A Coruña, Oviedo, Cantabria, País Vasco y Zaragoza)

Programa Interuniversitario de Doctorado en Tecnologías de la Información y Comunicaciones en Redes Móviles

TESIS DOCTORADO

Técnicas de Estimación de Orden basadas en Subespacios en Sistemas MIMO Masivo

PHD THESIS

Subspace-based Order Estimation Techniques in Massive MIMO

Realizada por: Vaibhav Garg

Dirigida por: Ignacio Santamaría

Escuela de Doctorado de la Universidad de Cantabria

Santander 2021

UNIVERSIDAD DE CANTABRIA

Department of Communications Engineering

PhD Thesis

Subspace-based Order Estimation Techniques in Massive MIMO

Author: Vaibhav Garg

Supervisor: Ignacio Santamaría

Santander 2021

He who can listen to the music in the midst of noise can achieve great things.

—*Vikram Sarabhai*

Affiliations

Advanced Signal Processing Group
Department of Communications Engineering
University of Cantabria, Spain

This work was supported by the Ministerio de Ciencia e Innovación (MICINN) of Spain, under grants TEC2016-75067-C4-4-R (CARMEN) and BES-2017-080542.

Acknowledgements

First and foremost, I would like to convey my gratitude for my supervisor Prof. Ignacio Santamaría for his invaluable advice, continuous support and patience during research tenure. His extensive knowledge and intense experience were encouraging to me and helped me greatly in my research work and in my daily life as well. He kindly gave me ample freedom to work on the problems of my interest. He also put a great effort in formalizing those problems. He also showed a great patience while improving of my writings of technical version.

I am grateful to Prof. David Ramírez for inviting me to be a visiting scholar at Department of Signal Theory and Communications of Universidad Carlos III de Madrid. It was a challenging time due to the restrictions enforced to curb the transmission of coronavirus; still he always guided me humbly in every possible way for the betterment of my research work. Furthermore, I was fortunate enough to collaborate with Prof. Alba Pagès-Zamora from Technical University of Catalonia. I am grateful for the technical knowledge she shared with me during numerous interesting discussions.

I extend my gratitude to Prof. Jesús Ibañez for helping me with various technical issues at the Grupo de Tratamiento Avanzado de Señal (GTAS). I thank Prof. Jesús Pérez, Prof. Luis Vielva, Prof. Javier Vía, and Prof. Steven for the excellent working atmosphere they have created at GTAS. I would also like to thanks my mates: Jacobo, Pere and Diego for supporting me during different phases of this journey.

I am blessed and take this opportunity to thank my parents for their love and support. They have been constant source of encouragement without which I would not have been able to make it this far. Without their tremendous understanding and encouragement in the past few years, it would have been impossible for me to complete my study. Finally, I thank all my friends in India and Spain to keep my moral up during this roller-coaster journey.

Vaibhav Garg

Santander, April 2021

Abstract

Order estimation, also known as source enumeration, is a classical problem in array signal processing which consists in estimating the number of signals received by an array of sensors. In the last decades, numerous approaches to this problem have been proposed. However, the need of working with large-scale arrays (like in massive MIMO systems), low signal-to-noise-ratios, and poor sample regime scenarios, introduce new challenges to order estimation problems. For instance, most of the classical approaches are based on information theoretic criteria, which usually require a large sample size, typically several times larger than the number of sensors. Obtaining a number of samples several times larger than the number of sensors is not always possible with large-scale arrays. In addition, most of the methods found in literature assume that the noise is spatially white, which is very restrictive for many practical scenarios.

This dissertation deals with the problem of source enumeration for large-scale arrays, and proposes solutions that work robustly in the small sample regime under various noise models. The first part of the dissertation solves the problem by applying the idea of subspace averaging. The input data are modeled as subspaces, and an average or central subspace is computed. The source enumeration problem can be seen as an estimation of the dimension of the central subspace. A key element of the proposed method is to construct a bootstrap procedure, based on a newly proposed discrete distribution on the manifold of projection matrices, for stochastically generating subspaces from a function of experimentally determined eigenvalues. In this way, the proposed subspace averaging (SA) technique determines the order based on the eigenvalues of an average projection matrix, rather than on the likelihood of a covariance model, penalized by functions of the model order. The proposed SA criterion is especially effective in high-dimensional scenarios with low sample support for uniform linear arrays in the presence of white noise.

In Chapter 3, the proposed SA method is extended for: i) non-white noises, and ii) non-uniform linear arrays. The SA criterion is sensitive with the chosen dimension of extracted subspaces. To solve this problem, we combine the SA technique with a majority vote approach. The number of sources is detected for increasing dimensions of the SA technique and then a majority vote is applied to determine the final estimate. Further, to extend SA for arrays with arbitrary geometries, the SA is combined with a sparse reconstruction (SR) step. In the first step, each received snapshot is approximated by a sparse linear combination of the rest of snapshots. The SR problem is regularized by the logarithm-based surrogate of the ℓ_0-norm and solved using a majorization-minimization approach. Based on the SR solution, a sampling mechanism is proposed in the second step to generate a collection of subspaces, all of which

approximately span the same signal subspace. Finally, the dimension of the average of this collection of subspaces provides a robust estimate for the number of sources.

The second half of the dissertation introduces a completely different approach to the order estimation for uniform linear arrays, which is based on matrix completion algorithms. Chapter 4 discusses the problem of order estimation in the presence of noise whose spatial covariance structure is a diagonal matrix with possibly different variances. The diagonal terms of the sample covariance matrix are removed and, after applying Toeplitz rectification as a denoising step, the signal covariance matrix is reconstructed by using a low-rank matrix completion method adapted to enforce the Toeplitz structure of the sought solution. The proposed source enumeration criterion is based on the Frobenius norm of the reconstructed signal covariance matrix obtained for increasing rank values. The proposed method performs robustly for both small and large-scale arrays with few snapshots.

Finally, an approach to work with a reduced number of radio–frequency (RF) chains is proposed in Chapter 5. The receiving array relies on antenna switching so that at every time instant only the signals received by a randomly selected subset of antennas are downconverted to baseband and sampled. Low-rank matrix completion (MC) techniques are then used to reconstruct the missing entries of the signal data matrix to keep the angular resolution of the original large-scale array. The proposed MC algorithm exploits not only the low- rank structure of the signal subspace, but also the shift-invariance property of uniform linear arrays, which results in a better estimation of the signal subspace. In addition, the effect of MC on DOA estimation is discussed under the perturbation theory framework. Further, this approach is extended to devise a novel order estimation criterion for missing data scenario. The proposed source enumeration criterion is based on the chordal subspace distance between two sub-matrices extracted from the reconstructed matrix after using MC for increasing rank values. We show that the proposed order estimation criterion performs consistently with a very few available entries in the data matrix.

Resumen

La estimación de orden, también conocida como enumeración de fuentes, es un problema clásico en el procesamiento de señal en array que consiste en estimar el número de señales recibidas por un conjunto de sensores. En las últimas décadas, se han propuesto numerosas formas de enfocar este problema. Sin embargo, la necesidad de trabajar con arrays de gran tamaño (como, por ejemplo, en sistemas MIMO masivos), bajas relaciones señal a ruido y escenarios con pocas muestras plantean nuevos desafíos para los problemas de estimación de orden. La mayoría de los enfoques clásicos se basan en criterios de teoría de la información, que suelen requerir un tamaño de muestra grande, generalmente varias veces mayor que el número de sensores. Obtener un número de muestras varias veces mayor que la cantidad de sensores no siempre es posible con arrays de gran tamaño. Además, la mayoría de los métodos propuestos en la literatura asumen que el ruido es espacialmente blanco, lo cual resulta muy restrictivo para muchos escenarios prácticos.

Esta tesis trata el problema de la enumeración de fuentes para arrays de gran escala y propone soluciones que funcionan de manera robusta en regímenes de pocas muestras bajo la presencia de varios modelos de ruido. La primera parte de la tesis resuelve este problema aplicando la idea del promediado de subespacios, que consiste en modelar los datos de entrada como subespacios y calcular un subespacio central o promedio. El problema de la enumeración de fuentes puede verse entonces como una estimación de la dimensión del subespacio central. Un elemento clave del método propuesto es la construcción de un procedimiento bootstrap, basado en una distribución discreta propuesta recientemente sobre la variedad de las matrices de proyección, para generar subespacios estocásticamente a partir de una función de autovalores determinados experimentalmente. De esta manera, la técnica de promediado de subespacios (SA) propuesta determina el orden basándose en los valores propios de una matriz de proyección promedio, en vez de en la verosimilitud de un modelo de covarianza, penalizado por funciones del orden del modelo. El criterio de SA propuesto es especialmente eficaz en escenarios de alta dimensionalidad con poco soporte de muestra para arrays lineales y uniformes en presencia de ruido blanco.

En el capítulo 3, el método SA propuesto se extiende para: i) ruidos no blancos y ii) arrays lineales no uniformes. El criterio SA es sensible a la dimensión escogida para los subespacios extraídos. Para resolver este problema, se combina la técnica SA con un enfoque de voto mayoritario. El número de fuentes es detectado para dimensiones crecientes de la técnica SA y, posteriormente, se aplica un voto mayoritario para determinar la estima final. Además, para extender SA a arrays con geometrías arbitrarias, este se combina con un paso de reconstrucción dispersa (SR). En el primer paso, cada snapshot recibido se aproxima mediante una combinación

lineal dispersa del resto de snapshots. El problema SR se regulariza mediante la sustitución basada en logaritmos de la norma ℓ_0 y se resuelve mediante un enfoque de Mayorización-Minimización. Basado en la solución SR, se propone un mecanismo de muestreo en el segundo paso para generar una colección de subespacios, todos los cuales abarcan aproximadamente el mismo subespacio de señal. Finalmente, la dimensión del promedio de esta colección de subespacios proporciona una estimación robusta del número de fuentes.

La segunda mitad de la tesis presenta un enfoque completamente diferente para la estimación de orden aplicada a arrays lineales y uniformes, que se basa en algoritmos matrix completion. El capítulo 4 analiza el problema de la estimación del orden en presencia de ruido cuya estructura de covarianza espacial es una matriz diagonal con varianzas posiblemente diferentes. Los términos diagonales de la matriz de covarianza de la muestra se eliminan y, después de aplicar la rectificación de Toeplitz como paso de eliminación de ruido, la matriz de covarianza de la señal se reconstruye utilizando un método matrix-completion de bajo rango adaptado para reforzar la estructura de Toeplitz de la solución buscada. El criterio de enumeración de fuentes propuesto está basado en la norma de Frobenius de la matriz de covarianza de la señal reconstruida obtenida para valores de rango crecientes. El método propuesto presenta un buen rendimiento tanto para arrays de pequeña como de gran escala con pocos snapshots.

Por último, en el capítulo 5 se realiza una propuesta para trabajar con un número reducido de cadenas de radiofrecuencia (RF). El array receptor está basado en la conmutación de antenas de modo que en cada instante de tiempo solo las señales recibidas por un subconjunto de antenas seleccionado al azar se convierten a banda base y se muestrean. Después, se utilizan técnicas matrix-completion (MC) de bajo rango para reconstruir las entradas faltantes de la matriz de datos de la señal para conservar la resolución angular del array de gran tamaño original. El algoritmo MC propuesto no solo aprovecha la estructura de bajo rango del subespacio de la señal, sino también la propiedad de invariancia de desplazamiento de los arrays lineales y uniformes, lo que da como resultado una mejor estimación del subespacio de la señal. Además, se analiza el efecto de MC en la estimación de DOA dentro del marco de la teoría de perturbaciones. Además, se amplía este enfoque para diseñar un nuevo criterio de estimación de orden para escenarios en los que falten datos. El criterio de enumeración de fuentes propuesto se basa en la distancia cordal entre dos sub-matrices extraídas a partir de la matriz reconstruida después de usar MC para aumentar los valores del rango. Por último, se demuestra que el criterio de estimación de orden propuesto funciona de manera consistente cuando hay muy pocas entradas disponibles en la matriz de datos.

Notation and Acronyms

Used Notation

a	Scalar (lowercase)
$\mathbf{a}$	Column vector (lowercase boldface)
$\mathbf{A}$	Matrix (uppercase boldface)
$\mathbf{A}(i,j)$	Entry in ith row and jth column of matrix $\mathbf{A}$
a^*	Complex conjugate of a complex number a
$\mathbf{A}^T$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian of matrix $\mathbf{A}$
$\mathbf{a}_i^H$	ith row of $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$
$\mathrm{rank}(\mathbf{A})$	Rank of matrix $\mathbf{A}$
$\lvert a \rvert$	Absolute value of a
$\mathrm{Re}(a)$	Real part of complex a
$\mathrm{Im}(a)$	Imaginary part of complex a
$\lVert \mathbf{A} \rVert_F$	Frobenius norm of matrix $\mathbf{A}$
$\lVert \mathbf{A} \rVert_*$	Nuclear norm of matrix $\mathbf{A}$
$\lVert \mathbf{A} \rVert_\infty$	Infinity norm of matrix $\mathbf{A}$
$\lVert \mathbf{a} \rVert_0$	ℓ_0-norm of vector $\mathbf{a}$
$\lVert \mathbf{a} \rVert_2$	ℓ_2-norm of vector $\mathbf{a}$
$A \succ B$	Matrix $\mathbf{A} - \mathbf{B}$ is positive semidefinite
$\mathbb{R}^{M \times N}$	Space of $M \times N$ real matrices
$\mathbb{C}^{M \times N}$	Space of $M \times N$ complex matrices
$\mathbb{C}^n$	Space of n-dimensional complex vectors
$\langle \mathbf{A} \rangle$	A subspace of the complex vector space $\mathbb{C}^n$ spanned by the unitary frame $\mathbf{A}$
$\dim(\langle \mathbf{A} \rangle)$	Dimension of $\langle \mathbf{A} \rangle$
$d_{geo}(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle)$	Geodestic or intrinsic distance between $\langle \mathbf{U} \rangle$ and $\langle \mathbf{V} \rangle$
$d(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle)$	Extrinsic distance between $\langle \mathbf{U} \rangle$ and $\langle \mathbf{V} \rangle$
$\mathrm{blkdiag}(\mathbf{A}_1, \mathbf{A}_2)$	Block diagonal matrix which matrices $\mathbf{A}_1, \mathbf{A}_2$ as diagonal
$\mathrm{diag}(a_1 \dots a_N)$	Diagonal matrix with elements $a_1 \dots a_N$ along the main diagonal
$\mathbf{I}_n$	Identity matrix of size n
$\mathbf{P}_\mathbf{A}$	Orthogonal projection onto $\langle \mathbf{A} \rangle$
$\mathcal{CN}(0,1)$	Complex Gaussian distribution with zero mean and unit variance
$\mathbf{x} \sim \mathcal{CN}_n(\mathbf{0}, \mathbf{R})$	Complex Gaussian vector in $\mathbb{C}^n$ with zero mean and covariance $\mathbf{R}$

$\mathbb{S}(q,n)$	Complex Stiefel manifold of orthonormal q-frames in $\mathbb{C}^n$
$\mathbb{G}(q,n)$	Complex Grassman manifold of q-dimensional linear subspaces of the n-dimensional complex vector space $\mathbb{C}^n$
$\mathbb{P}(q,n)$	Set of all projection matrices of rank q
$\mathbb{U}(q)$	Set of all unitary matrices of rank-q
$\mathcal{D}(\mathbf{U},\alpha)$	Distribution with orientations $\mathbf{U}$ and concentrations $\boldsymbol{\alpha}$
$\mathcal{U}(a,b)$	Uniformly distributed random variables between a and b
$\mathcal{T}$	Set of Toeplitz matrices
$\mathfrak{D}$	Set of diagonal complex matrices

General symbol use and conventions

$\mathbf{A}$	Steering matrix
$\mathbf{c}_n$	Sparse representation for nth snapshot
$\mathbf{E}$	Noise component of data matrix
$\mathbf{H}_s$	Unknown multiple input multiple output channel
K	Number of sources/signals
$\hat{k}$	Estimated number of sources
$\hat{k}_{BIC}$	Estimated order via Bayesian information criterion
$\hat{k}_{MDL}$	Estimated order via minimum description length method
$\hat{k}_{NE}$	Estimated order via the method proposed by Nadakuditi and Edelman
$\hat{k}_{SA}$	Estimated order via subspace averaging method
$\hat{k}_{SA-MajVote}$	Estimated order via subspace averaging with majority vote approach
$\hat{k}_{SIMC}$	Estimated order via shift-invariant matrix completion (SIMC) method
$\hat{k}_{SSA}$	Estimated order via sparse subspace averaging method
$\hat{k}_{SDRP}$	Estimated order via source detection as regression problem (SDRP) method
$\hat{k}_{TMC}$	Estimated order via Toeplitz matrix completion method
$\hat{k}_{VTRS}$	Estimated order via variance of transformed rotational submatrix (VTRS) method
k_{max}	Dimension of the extracted subspace
L	Dimension of the subarrays in subspace averaging method
L_s	Number of sensors sampled for each snapshot
M	Number of sensors in an array
N	Number of snapshots
$\overline{\mathbf{P}}$	Average projection matrix
P_Ω	Projection operator that sets to zero the missing entries and leaves the observed ons unchanged
P_s	Percentage of missing data
$\mathbf{Q}$	Rotation matrix
$\mathbf{R}$	Covariance matrix
$\hat{\mathbf{R}}$	Sample covariance matrix
$\mathbf{R}_s$	Signal covariance matrix
$\mathbf{R}_n$	Noise covariance matrix
$\mathbf{R}_{ML}$	Maximum likelihood estimate for a covariance matrix with the required structure
$s_k[n]$	Vector of complex gain for kth source and nth snapshot
$\mathbf{s}[n]$	nth signal vector
$\mathbf{e}[n]$	nth noise vector

$\mathbf{X}$	Data matrix
$\mathbf{X}_m$	Data matrix with missing entries
$\mathbf{X}_s$	noiseless signal component of data matrix
$\mathbf{x}[n]$	nth snapshot
$\mathbf{X}_n$	Data matrix $\mathbf{X}$ after removing the nth snapshot
θ_k	Direction-of-arrival for kth source (electrical angle)
Δ_θ	Electrical angel separation among the sources
λ_m	mth eigenvalue of sample covariance matrix
σ^2	Noise variance
σ_m^2	Noise variance at the mth sensor
σ_s^2	Common signal variance
ν_k	number of free-adjusted real parameters
Ω	Set of observed indices
η	Tolerance parameter for matrix completion algorithms
μ	Regularization parameter for matrix completion algorithms
μ_s	Regularization parameter for sparse reconstruction process
δ	Small constant to avoid numerical issues
ϵ_d	Parameter to control the spatial non-whiteness of the noise with diagonal covariance matrix
ρ_c	Correlation coefficient for correlated noise
$\mathcal{L}(\cdot)$	Langevin

Acronyms

AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
CCA	Canonical Correlation Analysis
CRLB	Cramer-Rao Lower Bound
DOA	Directions Of Arrival
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Technique
iid	Independent and Identically Distributed
LASSO	Least Absolute Shrinkage and Selection Operator
LS	Least Square
LS-MDL	Linear Shrinkage - Minimum Description Length
MC	Matrix Completion
MDL	Minimum Description Length
MDL-unc	Minimum Description Length for Uncorrelated Noises (When noise covariance matrix is diagonal with different variances)
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MM	Majorization-Minimization
NE	The method proposed by Nadakuditi and Edelman
NULA	Non-Uniform Linear Array
OSE	Optimal Subspace Estimation
PCA	Principal Component Analysis
RF	Radio-Frequency
RMSE	Root Mean Square Error
SA	Subspace Averaging
SCM	Sample Covariance Matrix
SDRP	Source Detection as Regression Problem
SIMC	Shift-Invariant Matrix Completion
SNR	Signal-to-Noise Ratio
SR	Sparse Reconstruction
SSA	Sparse Subspace Averaging
TMC	Toeplitz Matrix Completion
ULA	Uniform Linear Array
VTRS	Variance of Transformed Rotational Submatrix

Contents

Chapter 1
Introduction

Detecting the number of signals received by an array of sensors is an important problem in many applications of statistical signal processing such as radar and wireless communications to biomedical and geophysical signal processing [Trees, 2002, Scharf, 1991]. This is an order estimation problem, also known as source enumeration in the array processing literature. A number of approaches to this problem have been proposed in the last decades based on information theoretic criteria such as Akaike's information criterion (AIC) [Akaike, 1974], minimum description length (MDL) criterion [Rissanen, 1978, Wax and Kailath, 1985], and Bayesian information criterion (BIC) [Schwarz, 1978, Lu and Zoubir, 2013], all of which are functions of the eigenvalues of the sample covariance matrix (SCM). A conventional approach to order determination is to compute likelihoods for a sequence of rank-k plus diagonal covariance models for multivariate normal measurements [Anderson, 1963], [Wax and Kailath, 1985], and then to penalize likelihoods for large ranks k. The resulting formulas use sums and products of sub-dominant eigenvalues of a sample covariance matrix in what amount to tests of whiteness of the trailing sequence of eigenvalues. In these methods, the scale of the rank-k and diagonal components are implicitly estimated in the estimation of the covariance model. Given a sufficiently large sample size (at least several times the number of sensors), and assuming that the noise is spatially white and Gaussian, these methods perform satisfactorily and provide accurate estimates for the number of sources.

The need of large bandwidths in 5G networks motivates to operate in mm-Wave bands, which require large-scale antenna arrays to compensate for the path loss [Rappaport et al., 2015, Rappaport et al., 2013]. Indeed, research in wireless communication systems has been shifted towards the use of large antenna arrays as in massive multiple-input multiple-output (MIMO) systems [Rusek et al., 2013]. In this scenario, achieving the sample size several times larger than number of sensors is often not possible. This forces to work in the so-called small sample regime [Nadakuditi and Edelman, 2008], where the number of snapshots, N, is in the order of the number of sensors, M, a regime in which the performance of above mentioned methods degrade drastically. Based on recent results from random matrix theory, several methods have been proposed for source enumeration in the small sample regime [Nadakuditi and Edelman, 2008, Huang and So, 2013, Huang et al., 2016]. Nevertheless, all these methods also assume white noise and hence provide in gen-

eral poor results in other noise conditions such as spatially correlated noise or noise with diagonal covariance matrix with unknown elements.

When the noise is spatially correlated with an arbitrary unknown covariance matrix, source enumeration has been considered under different assumptions on the array geometry and the temporal correlation of the noise in [Zhang and Wong, 1993, Wu and Wong, 1994, Stoica and Cedervall, 1997, Jiang and Ingram, 2004]. In [Zhang and Wong, 1993] the authors assume that two well-separated sensor arrays are available and thus the noise spatial covariance matrix is block diagonal. The resulting test is based on the canonical correlations estimated from the sample coherence matrix. The method in [Jiang and Ingram, 2004] assumes that the signal is received by a uniform linear array (ULA) for which a property called shift invariance holds [Paulraj et al., 1986, Roy and Kailath, 1989, Li and Vaccaro, 1991], and propose an ad-hoc test based on the elements of the rotation matrix that relates the signal subspaces extracted from the two subarrays. These methods, however, require accurate estimates of the sample covariance or the coherence matrix and therefore their performance degrades when only a few snapshots are available. To alleviate this problem, [Song et al., 2016] applies a principal component analysis (PCA) rank-reduction pre-processing step before applying the Bartlettt-Lawley test [Bartlett, 1941, Lawley, 1959].

When the noise covariance matrix is diagonal with unknown elements, estimating the number of sources is equivalent to estimating the number of common factors in a multivariate factor analysis problem, for which several tests have been proposed in the statistics literature [Mardia et al., 1979, Lawley and Maxwell, 1971]. Algorithms to maximize the likelihood function for this problem can be found in [Pesavento and Gershman, 2001, Ramirez et al., 2011]. But again these methods perform poorly in the small sample regime.

In addition, performance of source enumeration methods degrade if some data is missing or with faulty sensors. This might happen when i) one or more sensors are damaged or ii) only few sensors are sampled intentionally to reduce the overall hardware cost. For this problem, matrix completion based methods can be used to reconstruct either the complete data matrix or the covariance matrix. Matrix completion (MC) techniques have been used in array signal processing recently for different applications such as denoising or reconstruction of missing data. But there is a need of more research in the problem of order estimation with missing data. A matrix completion based approach may be used when only a few sensors are sampled. By using a low rank matrix completion algorithm, it is possible reconstructing the complete data matrix, which could be used further for order estimation and direction of arrival applications. This idea is explored in this thesis.

In summary, source enumeration is still a challenging problem in scenarios such as; i) in small sample regime, ii) in the presence of non-white noises, and iii) in missing data scenario. This dissertation works towards finding solutions for the abovementioned problems. First, a subspace averaging (SA) based method is proposed for order estimation for ULA in small sample regime in the presence of white noise. By exploiting the shift invariance property of ULAs, the SA method estimates the low-dimensional signal subspace (and its dimension) as the average of a collection of

subspaces extracted from consecutive sub-arrays. The SA method is then extended for source enumeration for ULAs in the presence of non-white noises. The number of sources is detected for increasing dimensions of the SA technique and then a majority vote approach is applied to determine the final estimate. Furthermore, SA is combined with a sparse representation step in order to extend the applications of SA for non-uniform linear arrays.

The second half of the dissertation is devoted to develop source enumeration methods by using matrix completion algorithms. The signal covariance matrix is i) Toeplitz, ii) low rank, and iii) satisfies the shift invariance property. By exploiting these properties with matrix completion algorithms, a better estimate of the signal subspace is achieved. This part of the dissertation first develops a source enumeration method based on matrix completion by exploiting Toeplitz and low rank structures of the signal covariance matrix, which outperforms other methods for ULAs in the presence of a noise with diagonal noise covariance matrix. A denoising step of Toeplitz rectification is used as a pre-processing step. Further, this concept is extended to solve the problem of source enumeration in missing data scenario. Shift invariance property is exploited with matrix completion here, which provides a consistent solution even if percentage of available data is small. Note that for missing data scenario, we only consider the case of ULAs and white noise in this dissertation.

1.1 Order Estimation in Array Processing

Let us consider K narrowband signals impinging on a large, uniform, half-wavelength linear array with M antennas (cf. Fig. 1.1). The received signal is

$$\mathbf{x}[n] = [\mathbf{a}(\theta_1) \quad \cdots \quad \mathbf{a}(\theta_K)]\,\mathbf{s}[n] + \mathbf{e}[n] = \mathbf{A}\mathbf{s}[n] + \mathbf{e}[n], \tag{1.1}$$

where $\mathbf{s}[n] = \left[s_1[n], \ldots, s_K[n]\right]^T$ is the vector of complex gains $s_k[n]$ for $M \times 1$ complex array response $\mathbf{a}(\theta_k) = \left[1 \quad e^{-j\theta_k} \quad e^{-j\theta_k(M-1)}\right]^T$ to the kth source whose direction-of-arrival (DOA) θ_k is unknown; and $\mathbf{A} = [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_K)]$ is the steering matrix. In the case of narrowband sources, free space propagation, and a ULA with inter-element spacing d, the spatial frequency or electrical angle is

$$\theta_k = \frac{2\pi}{\lambda} d \sin(\phi_k),$$

where λ is the wavelength and ϕ_k is the direction-of-arrival (DOA). We will refer to θ_k as the DOA of source k for simplicity. Note that for a half-wavelength ULA $\theta_k = \pi \sin(\phi_k)$, and the spatial frequency varies between $-\pi$ and π when ϕ_k varies between $-\pi/2$ and $\pi/2$, with 0° being the broadside direction.

The signal and noise vectors are modeled as $\mathbf{s}[n] \sim \mathcal{CN}_K(\mathbf{0}, \mathbf{\Psi})$ and $\mathbf{e}[n] \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{R}_n)$, respectively. The dimensions are these: $\mathbf{x} \in \mathbb{C}^M, \mathbf{A} \in \mathbb{C}^{M\times K}, \mathbf{s} \in \mathbb{C}^K$, and $\mathbf{e} \in \mathbb{C}^M$. From the signal model (1.1), the theoretical covariance matrix is

$$\mathbf{R} = E\left[\mathbf{x}[n]\mathbf{x}^H[n]\right] = \mathbf{A}\mathbf{\Psi}\mathbf{A}^H + \mathbf{R}_n, \tag{1.2}$$

where $\mathbf{R}_n$ is an $M \times M$ noise covariance matrix. We assume that the K signals are uncorrelated, therefore $\mathbf{\Psi}$ is a $K \times K$ diagonal matrix. Unless stated otherwise, sources have equal power. We further assume there are N snapshots collected in the data matrix $\mathbf{X} = \begin{bmatrix}\mathbf{x}[1] & \dots & \mathbf{x}[N]\end{bmatrix}$, and the sample covariance matrix is

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}[n]\mathbf{x}^H[n] = \frac{1}{N}\mathbf{X}\mathbf{X}^H \tag{1.3}$$

The order estimation problem consists of estimating K from $\mathbf{X}$ or $\hat{\mathbf{R}}$. Throughout the dissertation we assume that the array is composed of a large number of antenna elements, and that the number of snapshots is in the order of number of antennas, that is, we operate in small sample regime.

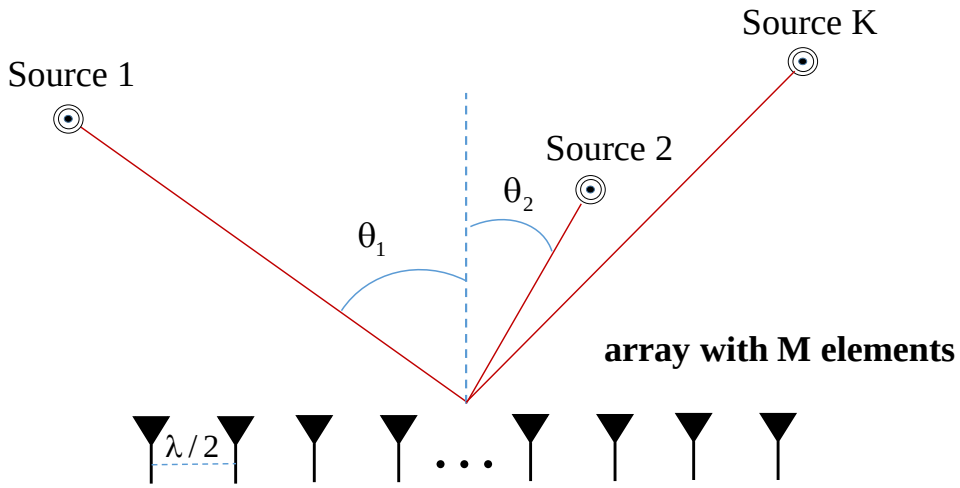


Figure 1.1: Source enumeration problem in large scale arrays: estimating the number of sources K in a ULA with a high number of antenna elements M.

Different noise models have been discussed in the literature, which can be classified as follow:

- **Noise Model 1:** We assume here that the noise is spatially white with the noise covariance matrix $\mathbf{R}_n = \sigma^2\mathbf{I}$, where σ^2 is the noise variance. Most of the order estimation methods are designed under this assumption.
- **Noise Model 2:** Uncorrelated noises across antennas so that the noise covariance matrix is diagonal with unknown elements along its diagonal

$$\mathbf{R}_n = \text{diag}\left(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2\right)$$

where σ_m^2 is the noise variance at the mth sensor. The noise variances are modeled as uniformly distributed independent random variables: $\sigma_m^2 \sim \mathcal{U}[\sigma^2(1-\epsilon_d), \sigma^2(1+\epsilon_d)]$, where σ^2 is a common noise variance and $0 \leq \epsilon_d \leq 1$ allows

us to control the spatial non-whiteness of the noise. Notice that for $\epsilon_d = 0$ the noise is spatially white with covariance matrix $\mathbf{R}_n = \sigma^2\mathbf{I}$

- **Noise Model 3:** We consider an exponentially correlated noise model whose noise covariance matrix is defined as $\mathbf{R}_n(i,j) = \rho_c^{|i-j|}$, where $0 \leq \rho_c \leq 1$ is the correlation coefficient. Like noise model 2, the noise is spatially white with $\mathbf{R}_n = \sigma^2\mathbf{I}$ for $\rho_c = 0$.

- **Noise Model 4:** Here we consider Gaussian noise with arbitrary unknown covariance matrix, that is, the noise covariance matrix is drawn from a complex Wishart distribution with M degrees of freedom and scale matrix proportional to $\mathbf{I}$. The noise covariance matrix for this noise model is $\mathbf{R}_n \succ 0$.

1.2 Prior Work

In this section we review some representative methods for order estimation. Most of these methods exploit random matrix results and are specifically designed to operate in the small sample regime. Unless stated otherwise, these methods are designed for a general linear array, not specifically a ULA. Further, most of them are functions of the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_M$ of the sample covariance matrix $\hat{\mathbf{R}}$. A brief description of the methods is presented here, which is divided into different categories, based on the assumptions of the noise model. These methods are also used for comparison in this dissertation.

1.2.1 Methods suitable for Noise Model 1

The following order estimation criteria for white noise can be used in small sample regime:

- **LS-MDL** criterion in [Huang and So, 2013]: The standard MDL method proposed by Wax and Kailath in [Wax and Kailath, 1985], based on a fundamental result of Anderson [Anderson, 1963], is

$$\hat{k}_{MDL} = \underset{0\leq k\leq M-1}{\operatorname{argmin}} (M-k)N\log\left(\frac{a(k)}{g(k)}\right) + \frac{1}{2}k(2M-k)\log N, \quad (1.4)$$

  where $a(k)$ and $g(k)$ are the arithmetic and the geometric mean, respectively, of the $M-k$ smallest eigenvalues of $\hat{\mathbf{R}}$. When the number of snapshots is smaller than the number of sensors or antennas ($N < M$), the sample covariance becomes rank-deficient and (1.4) can not be applied directly.

  The LS-MDL method proposed by Huang and So in [Huang and So, 2013] replaces the noise eigenvalues λ_i in the MDL criterion by a linear shrinkage, calculated as

$$\rho_i^{(k)} = \beta^{(k)}a(k) + (1-\beta^{(k)})\lambda_i, \quad i = k+1,\ldots,M,$$

where $\beta^{(k)} = \min(1, \alpha^{(k)})$, with

$$\alpha^{(k)} = \frac{\sum_{i=k+1}^{M} \lambda_i^2 + (M-k)^2 a(k)^2}{(N+1)\left(\sum_{i=k+1}^{M} \lambda_i^2 - (M-k)a(k)^2\right)}.$$

- **NE** criterion in [Nadakuditi and Edelman, 2008]: The method proposed by Nadakuditi and Edelman in [Nadakuditi and Edelman, 2008], which we refer to as the NE criterion, is given by

$$\hat{k}_{NE} = \underset{0 \le k \le M-1}{\operatorname{argmin}} \left\{ \frac{1}{2} \left(\frac{N t_k}{M} \right)^2 \right\} + 2(k+1),$$

where

$$t_k = \left[\frac{\sum_{i=k+1}^{M} \lambda_i^2}{a(k)^2 (M-k)} - \left(1 + \frac{M}{N}\right) \right] M.$$

- **BIC** method for large-scale arrays in [Huang et al., 2016]: The variant of the Bayesian Information Criterion (BIC) [Lu and Zoubir, 2013] for large-scale arrays proposed in [Huang et al., 2016] is

$$\hat{k}_{BIC} = \underset{0 \le k \le M-1}{\operatorname{argmin}} \, 2(M-k)N \log\left(\frac{a(k)}{g(k)}\right) + P(k, M, N),$$

where

$$P(k, M, N) = Mk \left(\log(2N) - \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{\lambda_i}{a(k)}\right) \right).$$

1.2.2 Method suitable for Noise Model 2

Source enumeration in the presence of noise whose noise covariance matrix is diagonal with different variances is still a challenging scenario for which not enough discussion is found in the literature. However, the MDL for this noise model can be developed from its general expression as follows:

- **MDL-unc:** The general expression of the MDL criterion is [Rissanen, 1978]

$$\hat{k}_{MDL} = \underset{k=0,\ldots,M-1}{\operatorname{argmin}} \left\{ -\log f(\mathbf{X}|\hat{\mathbf{R}}_{ML}) + \frac{1}{2}\nu_k \log N \right\}, \tag{1.5}$$

where $\mathbf{X}$ is the data matrix with collected snapshots, $\hat{\mathbf{R}}_{ML}$ denotes the Maximum Likelihood (ML) estimate for a covariance matrix with the required structure (low-rank plus diagonal) for a fixed order k, and $\nu_k = M + k(2M-k)$

is the number of free-adjusted real parameters [Vazquez-Vilar et al., 2011]. Although a closed-form ML estimate under non-iid noises is not possible, it can be obtained by using iterative algorithms [Joreskog, 1967, Pesavento and Gershman, 2001, Ramirez et al., 2011]. It is noted here that this method does not exploit any property of ULA, and should perform for non-uniform linear array as well.

1.2.3 Methods suitable for Noise Models 3 and 4

When the noise is spatially correlated in an arbitrary way or if noise is exponentially correlated, MDL or any other information theoretic criteria perform poorly. The only alternatives for this noise model in the literature either assume the availability of two well-separated subarrays such that the noise covariance matrix can be assumed to be block-diagonal [Zhang and Wong, 1993, Song et al., 2016], or exploit the shift invariance property of the signal subspace [Jiang and Ingram, 2004] of ULA. Recently a solution for source enumeration in the presence of colored noise is proposed in [Eguizabal et al., 2019], where source detection is modelled as a regression problem. In the following, we briefly review these approaches.

- **CCA**: Let us divide the array into two equal-sized non-overlapping subarrays of dimension $M/2$ (we assume M even wlog). The composite vector for the received signal is written as $\mathbf{x}[n] = (\mathbf{x}_1[n]^T, \mathbf{x}_2[n]^T)^T$, where $\mathbf{x}_1[n]$ and $\mathbf{x}_2[n]$ are the signals received at each subarray. Assuming now that the noise vectors at the two subarrays are uncorrelated, the noise covariance matrix is block diagonal $\mathbf{R_n} = \text{blkdiag}(\mathbf{R}_{n1}, \mathbf{R}_{n2})$. Under these assumptions, it was shown in [Zhang and Wong, 1993] (see also [Santamaria et al., 2017]) that the maximum of the log-likelihood function in (1.5) can be written as

$$-\log f(\mathbf{X}|\hat{\mathbf{R}}_{ML}) = -N \sum_{i=k+1}^{M/2} \log(1-k_i^2), \tag{1.6}$$

  where k_i is the ith sample canonical correlation, that is the ith eigenvalue of the coherence matrix $\hat{\mathbf{C}} = \hat{\mathbf{R}}_{x_1x_1}^{-1/2}\hat{\mathbf{R}}_{x_1x_2}\hat{\mathbf{R}}_{x_2x_2}^{-1/2}$. On the other hand, the number of free parameters for this noise model is $\nu_k = 2k(M-k+1/2)$. Admittedly, this MDL criterion is obtained for a structured block-diagonal noise covariance matrix, which is different from the spatially correlated noise model. Nevertheless, it is used for comparison in this dissertation. This canonical correlation analysis (CCA) based criterion was combined with a PCA rank-reduction preprocessing step in [Song et al., 2016], to improve its performance in the small sample regime.

- **VTRS:** Jiang and Ingram proposed a method in [Jiang and Ingram, 2004] that uses the variance of transformed rotational submatrix (VTRS) as a criterion for the detection of number of sources for ULA. The VTRS criterion is

$$\hat{k}_{VTRS} = \underset{1\le k\le M-1}{\text{argmin}} \frac{||\boldsymbol{\Delta}_k||_F^2}{(M-k-1)k},$$

where $||\cdot||_F^2$ denotes the squared Frobenius norm and

$$\boldsymbol{\Delta}_k = \begin{bmatrix} \Psi_{vk+1,1} & \Psi_{vk+1,2} & \dots & \Psi_{vk+1,k} \\ \Psi_{vk+2,1} & \Psi_{vk+2,2} & \dots & \Psi_{vk+2,2} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{vM-1,1} & \Psi_{vM-1,2} & \dots & \Psi_{vM-1,k} \end{bmatrix}$$

is a submatrix of $\mathbf{\Psi_v}$, which is given as $\mathbf{E_y} = \mathbf{E_x}\mathbf{\Psi_v}$, where the matrices $\mathbf{E_x}$ and $\mathbf{E_y}$ contain the first $M-1$ and the last $M-1$ rows of the eigenvectors of $\hat{\mathbf{R}}$, respectively.

The VTRS criterion exploits the shift invariance property of ULA, and can be used with arbitrarily correlated noises. However, it is neither designed to work with small sample support nor designed to work with non-uniform linear arrays.

- **SDRP:** In [Eguizabal et al., 2019], source detection problem is modeled as a regression problem to detect the number of sources in the presence of colored noise. Information-theoretic criterion is used to determine the model order of the regression. In this dissertation we are denoting this method as source detection as regression problem (SDRP). The observed data is separated in two subsets of N_1 and N_2 snapshots such as $N_1 + N_2 = N$, and $\mathbf{X}_1 = \big[\mathbf{x}(1), \dots, \mathbf{x}(N_1)\big]$ and $\mathbf{X}_2 = \big[\mathbf{x}(N_1+1), \dots, \mathbf{x}(N_1+N_2)\big]$. Further, order estimation criterion is developed based on the Maximum Likelihood (ML) estimates of the model parameters, which is as follows:

$$\hat{k}_{SDRP} = \underset{k}{\operatorname{argmin}} \Big[(N_2 \log(|\hat{\boldsymbol{\Sigma}}_k|) + tr\Big(\hat{\mathbf{E}}^H (\hat{\boldsymbol{\Sigma}}_k^n)^{-1} \hat{\mathbf{E}}\Big) + \nu_k \Big],$$

and

$$\hat{\mathbf{E}} = \mathbf{X}_2 - \mathbf{P}_k \hat{\mathbf{B}}_k,$$

where $\mathbf{P}_k$ is the matrix of regressors containing the eigenvectors corresponding to the k largest eigenvalues of $\hat{\mathbf{R}}_{xx} = \frac{1}{N_1}\mathbf{X}_1\mathbf{X}_1^H$, and $\hat{\mathbf{B}}_k$ and $\hat{\boldsymbol{\Sigma}}_k$ are the constrained maximum likelihood estimates of the model parameters for model order k. Here, tr() denotes the trace of the matrix.

For the sample-poor case, when $N < M$, $\boldsymbol{\Sigma}_k$ is modelled as block-diagonal with block size of $s \times s$. This results in $d = \lfloor \frac{M}{s} \rfloor$ blocks, that is $\hat{\boldsymbol{\Sigma}}_k = \text{blkdiag}\Big(\hat{\boldsymbol{\Sigma}}_k^1, \dots \hat{\boldsymbol{\Sigma}}_k^d\Big)$, where $\lfloor \cdot \rfloor$ is the floor function. Based on this, the order estimation criterion for poor-sample case is reformulated as

$$\hat{k}_{SDRP} = \underset{k}{\operatorname{argmin}} \Big[\sum_{n=1}^{d} (N_2 \log(|\hat{\boldsymbol{\Sigma}}_k^n|) + tr\Big(\hat{\mathbf{E}}^H (\hat{\boldsymbol{\Sigma}}_k^n)^{-1} \hat{\mathbf{E}}\Big) + \nu_k \Big]$$

The reader is referred to [Eguizabal et al., 2019] for more details about this method. Note that SDRP does not exploit shift invariance property, therefore it can be used for non-uniform linear arrays as well.

1.3 Outline and Contributions

The technical content and contributions in this dissertation are structured in two parts, namely: order estimation via subspace averaging and order estimation via matrix completion. In addition, the final chapter summarizes the main conclusions and a brief reflection on the research lines to be considered as further work. In this section, we summarize the contents provided in each chapter:

- Chapter 1 introduces the dissertation in the context of the order estimation in statistical signal processing. The existing literature on order estimation is reviewed in this section, and the challenging scenarios, on which a study on source enumeration is still required are pointed out. Some already existing methods of order estimation are also briefly reviewed in this section.

- Part I discusses the order estimation criteria based on subspace averaging technique. First, subspace averaging method is designed for ULA and white noise scenario, and then this work is extended for non-white noises. Furthermore, subspace averaging for non-uniform linear arrays is also proposed.

  Chapter 2 addresses the problem of subspace averaging, with special emphasis placed on the question of estimating the dimension of the average. The results suggest that the enumeration of sources in a multi-sensor array, which is a problem of estimating the dimension of the array manifold, and as a consequence the number of radiating sources, may be cast as a problem of averaging subspaces. This point of view stands in contrast to conventional approaches, which cast the problem as one of identifying covariance models in a factor model. This chapter has produced the following publication:

  – [Garg et al., 2019] V. Garg, I. Santamaria, D. Ramirez, and L. L. Scharf, "Subspace averaging and order determination for source enumeration," IEEE Trans. Signal Process., vol. 67, pp. 3028–3041, 2019.

  Chapter 3 extends the SA technique for various scenarios. First, this chapter addresses the problem of source enumeration for ULAs in the challenging conditions of: i) large uniform arrays with few snapshots, and ii) non-white or spatially correlated noises with arbitrary correlation. To solve this problem, the SA technique is combined with a majority vote approach. Later, the SA technique is extended for arrays with arbitrary geometries. A sparse reconstruction step is combined with SA technique, which approximates each snapshot by a sparse linear combination of the rest of the snapshots. This sparse expansion is used to generate a collection of subspaces, which further used for averaging step of the SA technique. Based on this work, the following publications have been produced:

  – [Garg and Santamaria, 2019] V. Garg and I. Santamaria, "Source enumeration in non-white noise and small sample size via subspace averaging,"

in 27th European Signal Processing Conference (EUSIPCO), A Coruña, Spain, Sep. 2019.

- [Garg et al., 2021c] V. Garg, D. Ramirez and I. Santamaria,"Sparse subspace averaging for order estimation", Submitted to the IEEE Statistical Signal Processing (SSP) Workshop, July 2021.

- Part II introduces the matrix completion approaches to source enumeration. The properties of ULAs are exploited with low rank matrix completion technique in order to achieve better performance. This part of the dissertation first proposes a novel solution to source enumeration for ULAs in the presence of the noise whose noise covariance matrix is diagonal. Further, this approach is used to design a source enumeration method for the challenging scenario of missing data.

  Chapter 4 addresses the problem of source enumeration by an array of sensors in the presence of noise whose spatial covariance structure is a diagonal matrix with possibly different variances. To tackle this problem, in the first step a denoising step of Toeplitz rectification is applied. After that, the diagonal terms of the sample covariance matrix are removed, and the signal covariance matrix is reconstructed by using a low-rank matrix completion method. The proposed matrix completion method is tailored to exploit the Toeplitz structure of the signal covariance matrix. The proposed source enumeration criterion is based on the Frobenius norm of the reconstructed signal covariance matrix obtained for increasing rank values. This chapter has produced the following publication:

  - [Garg et al., 2020] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria, “Source enumeration via Toeplitz matrix completion,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 2020.

  Chapter 5 first discusses the possibility of working with a reduced number of radio-frequency chains. The receiving array relies on antenna switching so that at every time instant only the signals received by a randomly selected subset of antennas are downconverted to baseband and sampled. Low-rank matrix completion techniques are then used to reconstruct the missing entries of the signal data matrix to keep the angular resolution of the original large-scale array. The proposed MC algorithm exploits not only the low- rank structure of the signal subspace, but also the shift-invariance property of ULAs, which results in a better estimation of the signal subspace. Further, an order estimation technique for missing data scenario is introduced, which is based on the proposed matrix completion technique. This chapter has produced the following publications:

  - [Garg et al., 2021a] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria,“DOA estimation via shift-invariant matrix completion”. Signal Processing, volume 183, 2021.

  - [Garg et al., 2021b] V. Garg, A. Pagès-Zamora, and I. Santamaria,"Order estimation with missing data for massive MIMO systems", Submitted to the IEEE Signal Processing Letters, 2021.

- Finally, Chapter 6 presents the conclusion of the dissertation, including future research lines.

Part I

Order Estimation via Subspace Averaging

Chapter 2

Order Estimation via Subspace Averaging for Uniform Linear Arrays and White Noise

This chapter addresses the problem of subspace averaging, with special emphasis placed on the question of estimating the dimension of the average. The results suggest that the enumeration of sources in a multi-sensor array, which is a problem of estimating the dimension of the array manifold, and as a consequence the number of radiating sources, may be cast as a problem of averaging subspaces. This point of view stands in contrast to conventional approaches, which cast the problem as one of identifying covariance models in a factor model. We present a robust formulation of the proposed order fitting rule based on majorization–minimization algorithms. A key element of the proposed method is to construct a bootstrap procedure, based on a newly proposed discrete distribution on the manifold of projection matrices, for stochastically generating subspaces from a function of experimentally determined eigenvalues. In this way, the proposed subspace averaging (SA) technique determines the order based on the eigenvalues of an average projection matrix, rather than on the likelihood of a covariance model, penalized by functions of the model order. By means of simulation examples, we show that the proposed SA criterion is especially effective in high-dimensional scenarios with low sample support.

2.1 Introduction

In many applications of statistical signal processing, high-dimensional data exhibits a low dimensional structure that admits a subspace representation. In pattern recognition and machine learning applications, for instance, discriminative features are typically obtained after a principal component analysis (PCA) stage that selects a subspace to explain a large fraction of the variance in the original data [Burges, 2009]. In computer vision applications, the set of images under different illuminations can be represented by a low-dimensional subspace [Basri and Jacobs, 2003]. And subspaces appear also as invariant representations of signals geometrically deformed under a set of affine transformations [Hagege and Francos, 2016]. There are many

more applications where low-dimensional subspaces capture the intrinsic geometry of the problem, ranging from array processing [Scharf and Friedlander, 1994], motion segmentation [Vidal et al., 2008], subspace clustering [Vidal, 2011], spectrum sensing for cognitive radio [Ramirez et al., 2011], or noncoherent multiple-input multiple-output (MIMO) wireless communications [Gohary and Davidson, 2009, Zheng and Tse, 2002].

When input data are modeled as subspaces, possibly of different dimensions, a fundamental problem is to compute an average or central subspace and, more importantly, to determine the dimension of the average or prototype. When all subspaces have the same dimension, they are formally represented as points on the Grassmann manifold. Geodesic, or intrinsic, distances on the Grassmannian are measured based on the arc length (ℓ_2-norm of the vector of principal angles), and the average subspace according to this canonical distance metric is the Riemannian center of mass also known as the Karcher mean [Karcher, 1977]. The Karcher mean is typically found by iterative algorithms that map the subspaces to and from the tangent plane at a given point (using Exp and Log maps), which make them computationally costly [Turaga et al., 2011]. Another drawback of the intrinsic distance metric is that a unique optimal Karcher mean is not always guaranteed to exist [Marrinan et al., 2014].

As an alternative to the intrinsic mean of the manifold, Srivastava and Klassen proposed the extrinsic mean in [Srivastava and Klassen, 2002], which uses a chordal distance metric in the ambient vector space defined as the squared Frobenius norm of the difference between projection matrices. Unlike the intrinsic mean, the extrinsic mean of a collection of subspaces is always unique, it is easy to compute, and can be used for subspaces that have different dimensions and therefore live in a union of Grassmannians. For these reasons, as in [Santamaria et al., 2016], we focus on the extrinsic distance to compare subspaces that might not live in the same manifold.

We address the problem of determining the optimal order of the low-dimension average subspace that minimizes an extrinsic distance measure. The solution to this problem provides a simple order fitting rule based on thresholding the eigenvalues of the average projection matrix, and thus it is free of penalty terms or other tuning parameters commonly used by other model order estimation techniques. The proposed rule appears to be particularly well suited to problems involving high-dimensional data and low sample support, such as the determination of the number of sources with a large array of sensors: the so-called source enumeration problem [Zhu and Kay, 2018],[Huang et al., 2016],[Nadakuditi and Edelman, 2008].

In multi-sensor signal processing, temporal snapshots are typically used to estimate a second-order spatial covariance matrix. The eigenvalues of this sample covariance matrix are used for detection and localization, using methods that are inspired variations on factor analysis. But the use of these eigenvalues for source enumeration is fraught with difficulties, as they scale with source powers and background noise levels, and this fact conflicts with the fact that the dimension of an array manifold is invariant to scale. So the fundamental problem is to extract from a sample covariance matrix a scale invariant subspace. In summary, our approach to find this subspace is this; We replace a large sensor array by an overlapping sequence of subarrays,

as inspired by [Vaccaro and Ding, 1993, Vaccaro, 2017], and extract a collection of subspaces from the measurements made in each such subarray. These subspaces are then averaged, and from this average the dimension of the array manifold is determined. We propose a method for extracting subspaces based on a bootstrap sampling for subspaces drawn from a distribution determined by eigenvalues. This method is quite different in philosophy from previously published methods, based on asymptotic formulas for the distribution of eigenvalues

The order fitting rule for subspace averaging was first published in [Santamaria et al., 2016], and an unrefined application to source enumeration was presented in [Santamaria et al., 2018]. In this thesis, these works are extended and refined to provide a common framework for subspace averaging and its application to source enumeration. The main contributions of this chapter may be summarized as follows:

- We consider continuous and discrete distributions on the manifold of projection matrices as underlying distributions from which the measured collection of subspaces is a random draw. From this standpoint, the eigenvalues of the average projection matrix admit a probabilistic interpretation that enables a better understanding of the proposed order estimation rule.

- We propose a robust formulation of the problem to account for outliers within the set of measured subspaces. The standard extrinsic mean distance is replaced by a smooth concave function such as the ℓ_1 norm or the Huber loss function that limits the effect of subspaces far away from the average. Majorization-minimization (MM) algorithms [Sun et al., 2017] are then used to find the minimizer of this robust distance measure, which, in turn, provides a robust order fitting rule.

- The application of SA techniques to source enumeration in [Santamaria et al., 2018] is enhanced by including a sampling mechanism to generate random subspaces based on the eigenstructure of the sample covariance matrix. When exploited jointly with the shift invariance property of uniform linear arrays (ULAs), this random sampling scheme enhances the performance of SA in high-resolution scenarios in comparison to the preliminary results presented in [Santamaria et al., 2018]. Further, the method is proven to provide a consistent estimate of the number of sources as the number of samples or the signal-to-noise-ratio grow.

The structure of the chapter is as follows. In Section 2.2 we derive the order estimation rule for the average subspace using an extrinsic distance measure. As reported in [Marrinan et al., 2014, Santamaria et al., 2016], an average of projection matrices, not itself a projection matrix, is the key quantity summarizing all information needed to solve this problem. We also present in this section a robust version of the SA problem and solve it using MM algorithms. Section 2.3 reviews uniform and non-uniform distributions on the Grassmannian, and proposes a new discrete distribution motivated by interpreting the eigenvalues of the average projection matrix as

probabilities. The application of the SA technique to source enumeration in large uniform linear arrays in the presence of white noise is discussed in detail in Section 2.4. Section 2.5 evaluates the performance of the order fitting rule through numerical simulations, paying special attention to the application to source enumeration. Finally, the main conclusions are summarized in Section 2.6.

2.2 Order Estimation via Subspace Averaging

2.2.1 Distances between Subspaces

Let us consider two subspaces $\langle \mathbf{V} \rangle \in \mathbb{G}(q_V, n)$ and $\langle \mathbf{U} \rangle \in \mathbb{G}(q_U, n)$, where, $\mathbb{G}(q, n)$ denotes complex Grassman manifold of q-dimensional linear subspaces of the n-dimensional complex vector space $\mathbb{C}^n$. Let $\mathbf{V} \in \mathbb{C}^{n \times q_V}$ be a matrix whose columns form a unitary basis for $\langle \mathbf{V} \rangle$. Then $\mathbf{V}^H \mathbf{V} = \mathbf{I}_{q_V}$, and $\mathbf{P_V} = \mathbf{V}\mathbf{V}^H$ is the idempotent orthogonal projection onto $\langle \mathbf{V} \rangle$. Notice that $\mathbf{P_V}$ is a unique representation of $\langle \mathbf{V} \rangle$, whereas $\mathbf{V}$ is not unique, because if $\mathbf{G}$ is an arbitrary unitary $q_V \times q_V$ matrix, then $\mathbf{VG}$ will be another representation of $\langle \mathbf{V} \rangle$ with orthonormal columns. In a similar way, we define $\mathbf{U}$ and $\mathbf{P_U}$ for the subspace $\langle \mathbf{U} \rangle$.

To measure the distance between two subspaces we need the concept of principal angles, which is introduced in the following definition [Golub and Van Loan, 1996].

Definition 2.1. *Let $\langle \mathbf{V} \rangle$ and $\langle \mathbf{U} \rangle$ be subspaces of $\mathbb{C}^n$ whose dimensionality satisfy $\dim(\langle \mathbf{V} \rangle) = q_V \geq \dim(\langle \mathbf{U} \rangle) = q_U \geq 1$. The principal angles $\theta_1, \ldots, \theta_{q_U} \in [0, \pi/2]$ between $\langle \mathbf{V} \rangle$ and $\langle \mathbf{U} \rangle$ are defined recursively by*

$$\begin{aligned} \cos(\theta_k) = \max_{\mathbf{u} \in \langle \mathbf{U} \rangle} \max_{\mathbf{v} \in \langle \mathbf{V} \rangle} & Re(\mathbf{u}^H \mathbf{v}) = Re(\mathbf{u}_k^H \mathbf{v}_k) \\ \textit{subject to} \quad & ||\mathbf{u}|| = ||\mathbf{v}|| = 1, \\ & \mathbf{u}^H \mathbf{u}_i = 0, \quad i = 1, \ldots, k-1, \\ & \mathbf{v}^H \mathbf{v}_i = 0, \quad i = 1, \ldots, k-1, \end{aligned}$$

for $k = 1, 2, \ldots, q_U$, where $\mathrm{Re}(\cdot)$ denotes the real part of the complex number.

Assume that $\mathbf{U}$ and $\mathbf{V}$ are unitary bases for the two subspaces. Then the singular values of $\mathbf{U}^H \mathbf{V}$ are $\left(\cos(\theta_1), \ldots, \cos(\theta_{q_U})\right)$ [Björk and Golub, 1973]. The principal angles induce several distance metrics, from which the most widely used are the geodesic or intrinsic distance [Edelman et al., 1998, Karcher, 1977]

$$d_{geo}\left(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle\right)^2 = \sum_{r=1}^{q_U} \theta_r^2,$$

and the extrinsic distance, which is given by the Frobenius norm of the difference between the respective projection matrices [Srivastava and Klassen, 2002,

Marrinan et al., 2014],

$$
\begin{aligned}
d\left(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle\right)^2 &= \frac{1}{2} \left\| \mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}} \right\|_F^2 \\
&= \frac{1}{2} \left(q_V + q_U - 2 \sum_{r=1}^{q_U} \cos(\theta_r)^2 \right) \\
&= \frac{|q_V - q_U|}{2} + \sum_{r=1}^{q_U} \sin(\theta_r)^2. \qquad (2.1)
\end{aligned}
$$

The second term in the right hand side of (2.1) measures the chordal distance defined by the principal angles, whereas the first term accounts for projection matrices of different ranks.

There are arguments in favor of the extrinsic distance (2.1), among them, its uniqueness and its computational simplicity in contrast to the intrinsic distance that needs to compute the singular values of $\mathbf{V}^H\mathbf{U}$. Also, the extrinsic distance is related to the squared error in resolving the standard basis for the ambient space, $\{\mathbf{e}_i\}_{i=1}^n$, onto the subspace $\langle \mathbf{V} \rangle$ as opposed to the subspace $\langle \mathbf{U} \rangle$,

$$
\begin{aligned}
\sum_{i=1}^{n} \mathbf{e}_i^T (\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}})^H (\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}}) \mathbf{e}_i &= \operatorname{tr}\left((\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}})^H (\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}}) \right) \\
&= \left\| \mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}} \right\|_F^2 \\
&= 2d\left(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle\right)^2
\end{aligned}
$$

2.2.2 Order Selection Rule

Let us consider a collection of measured subspaces $\{\langle \mathbf{V}_r \rangle\}_{r=1}^R$ of $\mathbb{C}^n$, each with respective dimension $\dim(\langle \mathbf{V}_r \rangle) = q_r < n$. To simplify the notation, we denote the orthogonal projection matrix onto the rth subspace as $\mathbf{P}_r$. Each subspace $\langle \mathbf{V}_r \rangle$ is a point on the Grassmann manifold $\mathbb{G}(q_r, n)$, and the collection of subspaces lives on a disjoint union of Grasmannians. Without loss of generality, the dimension of the union of all subspaces is assumed to be the ambient space dimension n.

Using the extrinsic distance metric between subspaces, an order estimation criterion for the central subspace that "best approximates" the collection is

$$
(s^*, \mathbf{P}_s^*) = \underset{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}}{\arg\min} \quad \frac{1}{2R} \sum_{r=1}^{R} \left\| \mathbf{P} - \mathbf{P}_r \right\|_F^2, \qquad (2.2)
$$

where $\mathbb{P}(s, n)$ denotes the set of all idempotent projection matrices of rank s. For completeness, we also accept solutions $\mathbf{P} = \mathbf{0}$ with rank $s = 0$, meaning that there is no central "signal subspace" shared by the collection of input subspaces.

Expanding the cost function in (2.2) we obtain the equivalent problem

$$
\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{2} \operatorname{tr}\left(\mathbf{P}(\mathbf{I} - 2\overline{\mathbf{P}}) + \overline{\mathbf{P}} \right), \qquad (2.3)
$$

where $\overline{\mathbf{P}}$ is an average of orthogonal projection matrices

$$\overline{\mathbf{P}} = \frac{1}{R}\sum_{r=1}^{R}\mathbf{P}_r, \tag{2.4}$$

with compact eigendecomposition $\overline{\mathbf{P}} = \mathbf{F}\mathbf{K}\mathbf{F}^H$, where $\mathbf{K} = \operatorname{diag}(k_1, \ldots, k_n)$ with $1 \geq k_1 \geq k_2 \geq \ldots \geq k_n$.

Now, discarding constant terms and writing the projection matrix as $\mathbf{P} = \mathbf{U}\mathbf{U}^H$, where $\mathbf{U}$ is a unitary $n \times s$ matrix, problem (2.3) can be rewritten as

$$\min_{\mathbf{U}\in\mathbb{S}(s,n)} \quad \operatorname{tr}\left(\mathbf{U}^H(\mathbf{I} - 2\overline{\mathbf{P}})\mathbf{U}\right),$$

where $\mathbb{S}(s,n)$ denotes the complex Stiefel manifold of orthonormal s-frames in $\mathbb{C}^n$. Hence, the optimal order s^* is the number of negative eigenvalues of the matrix

$$\mathbf{S} = \mathbf{I} - 2\overline{\mathbf{P}},$$

or, equivalently, the number of eigenvalues of $\overline{\mathbf{P}}$ larger than 1/2, which is the order fitting rule proposed in [Santamaria et al., 2016]. The proposed rule can be written alternatively as

$$s^* = \underset{s\,\in\,\{0,1,\ldots,n\}}{\arg\min} \quad \sum_{i=1}^{s}(1 - k_i) + \sum_{s+1}^{n} k_i.$$

A similar rule was developed in [Hlawatsch and Kozek, 1994] for the problem of designing optimum time-frequency subspaces with a specified time-frequency pass region.

Once the optimal order s^* is known, a basis for the average subspace can be obtained as the solution of the following optimization problem

$$\max_{\mathbf{U}\in\mathbb{S}(s^*,n)} \quad \operatorname{tr}\left(\mathbf{U}^H\mathbf{F}\mathbf{K}\mathbf{F}^H\mathbf{U}\right)$$

whose solution is given by any unitary matrix whose column space is the same as the subspace spanned by the s^* principal eigenvectors of $\mathbf{F}$

$$\mathbf{U}^* = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{s^*}) = \mathbf{F}_{s^*},$$

and $\mathbf{P}^* = \mathbf{U}^*(\mathbf{U}^*)^H$. So the average subspace is constructed by quantizing the eigenvalues of the average projection matrix at 0 or 1.

2.2.3 Properties of the Average Projection Matrix

The average of projection matrices in (2.4) is not a projection matrix itself, and therefore is not idempotent. However, it has the following properties:

1. It is symmetric and positive semidefinite.

2. Its eigenvalues are real and satisfy $0 \leq k_i \leq 1$.

1) is trivially proved by noticing that $\overline{\mathbf{P}}$ is an average of symmetric and positive semidefinite projection matrices. To prove 2) let us take without loss of generality the ith eigenvalue-eigenvector pair $(k_i, \mathbf{f}_i)$, then we have

$$\begin{aligned} k_i = \mathbf{f}_i^H \overline{\mathbf{P}} \mathbf{f}_i &= \frac{1}{R} \sum_{r=1}^{R} \mathbf{f}_i^H \mathbf{P}_r \mathbf{f}_i \\ &\overset{(a)}{=} \frac{1}{R} \sum_{r=1}^{R} \mathbf{f}_i^H \mathbf{P}_r^2 \mathbf{f}_i = \frac{1}{R} \sum_{r=1}^{R} ||\mathbf{P}_r \mathbf{f}_i||^2 \leq 1, \end{aligned}$$

where (a) holds because all $\mathbf{P}_r$ are idempotent, and the inequality follows from the fact that each term $||\mathbf{P}_r \mathbf{f}_i||^2$ is the squared norm of the projection of a unit norm vector, $\mathbf{f}_i$, onto the subspace $\langle \mathbf{V}_r \rangle$ and therefore $||\mathbf{P}_r \mathbf{f}_i||^2 \leq 1$ with equality only if the eigenvector belongs to the subspace.

Assuming that $\mathbf{V}_r = \left[\mathbf{v}_{r1}, \ldots, \mathbf{v}_{rq_r}\right]$ is a unitary basis for the rth subspace, the eigenvalues of the average projection matrix can be further expressed as

$$k_i = \frac{1}{R} \sum_{r=1}^{R} \sum_{j=1}^{q_j} ||\mathbf{v}_{rj}^H \mathbf{f}_i||^2,$$

and hence they can be interpreted as the squared norm of the average projection along the direction $\mathbf{f}_i$. It is important to remark, however, that the eigenvalues of $\overline{\mathbf{P}}$ are invariant to a common change in the basis of all subspaces. That is, we can apply an arbitrary change of basis $\mathbf{V}_r' = \mathbf{V}_r \mathbf{Q}'$ for $r = 1, \ldots, R$ with $\mathbf{Q}'$ unitary, and the eigenvalues k_i do not change.

2.2.4 Robust Version

In some applications there is a need to account for outliers within our collection of measured or extracted subspaces. To this end, we discuss in this section a robust formulation of the proposed order fitting rule based on majorization-minimization (MM) algorithms [Sun et al., 2017].

The simplest robust formulation of Problem (2.2) is

$$\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{R} \sum_{r=1}^{R} \rho\left(d_r^2(\mathbf{P})\right) \tag{2.5}$$

where

$$d_r^2(\mathbf{P}) = \frac{1}{2} \|\mathbf{P} - \mathbf{P}_r\|_F^2$$

and $\rho(\cdot)$ is a smooth concave increasing function that saturates so that outliers or subspaces far away from the average have a limited effect. Examples of robust functions are [Lerman and Maunu, 2018]:

- ℓ_p-norm:
$$\rho(t) = t^{p/2} \tag{2.6}$$
where $0 < p \leq 2$ with the nonrobust ℓ_2-norm formulation recovered for $p = 2$.

- Huber: for $T > 0$,
$$\rho_H(t) = \begin{cases} t/\sqrt{T} & t \leq T, \\ \sqrt{t} & t > T. \end{cases} \tag{2.7}$$
For $T = 0$ we obtain the median estimator $\rho_H(t) = \sqrt{t}$.

- Log-loss:
$$\rho_{LL}(t) = \theta \ln(\theta + t),$$
where $\theta \geq 1$.

- Logistic:
$$\rho_L(t) = \frac{1}{1 + e^{-t}}$$

- Geman-McClure estimator [Geman and Reynolds, 1992, Barron, 2017]: for $\theta > 0$,
$$\rho_{GM}(t) = \frac{t}{\theta + t}.$$

The idea of the MM algorithm [Sun et al., 2017] is, at each iteration, to find a majorizer of the objective function. Since the robust function $\rho(\cdot)$ is a smooth concave function, we can easily majorize at some point simply by linearizing:

$$\rho(t) \leq \rho(t_0) + \rho'(t_0)(t - t_0).$$

In the context of our problem, the problem at iteration k (where a central subspace $\mathbf{P}^{(k)}$ of dimension $s^{(k)}$ is available) is

$$\begin{aligned} \min_{\mathbf{P} \in \mathbb{P}(s,n)} \quad & \frac{1}{R} \sum_{r=1}^{R} \rho\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right) \\ & + \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)\left(d_r^2(\mathbf{P}) - d_r^2\left(\mathbf{P}^{(k)}\right)\right) \end{aligned}$$

or, removing unnecessary constant terms,

$$\min_{\mathbf{P} \in \mathbb{P}(s,n)} \frac{1}{R} \sum_{r=1}^{R} \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right) d_r^2(\mathbf{P}).$$

Let us now define the normalized weights that define a simplex

$$\bar{w}_r^{(k)} = \frac{\rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)}{\sum_{r=1}^{R} \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)}, \quad \bar{w}_r^{(k)} \geq 0, \quad \sum_r \bar{w}_r^{(k)} = 1.$$

With this definition we can obtain the next iterate $\mathbf{P}^{(k+1)}$ as the solution to

$$\min_{\substack{s\in\{0,1,\dots,n\}\\ \mathbf{P}\in\mathbb{P}(s,n)}} \quad \frac{1}{2}\sum_{r=1}^{R} \bar{w}_r^{(k)} \|\mathbf{P}-\mathbf{P}_r\|_F^2 . \tag{2.8}$$

Expanding the cost function (2.8), we obtain

$$\min_{\substack{s\in\{0,1,\dots,n\}\\ \mathbf{P}\in\mathbb{P}(s,n)}} \quad \frac{1}{2}\mathrm{tr}\left(\mathbf{P}(\mathbf{I}-2\overline{\mathbf{P}}_w^{(k)})+\overline{\mathbf{P}}_w^{(k)}\right), \tag{2.9}$$

where we have used the fact that $\sum_{r=1}^{R}\bar{w}_r^{(k)}\mathrm{tr}(\mathbf{P}) = \mathrm{tr}(\mathbf{P})$ and $\overline{\mathbf{P}}_w^{(k)}$ is the weighted average projection matrix given by

$$\overline{\mathbf{P}}_w^{(k)} = \sum_{r=1}^{R} \bar{w}_r^{(k)}\mathbf{P}_r$$

Writing the projection matrix in (2.9) as $\mathbf{P}=\mathbf{U}\mathbf{U}^H$ and discarding constant terms, the optimization problem can be rewritten as

$$\min_{\mathbf{U}\in\mathbb{S}(s,n)} \quad \mathrm{tr}\left(\mathbf{U}^H(\mathbf{I}-2\overline{\mathbf{P}}_w^{(k)})\mathbf{U}\right).$$

Then, the optimal order at iteration $k+1$, $s^{(k+1)}$, is the number of negative eigenvalues of the matrix

$$\mathbf{S}^{(k)} = \mathbf{I}-2\overline{\mathbf{P}}_w^{(k)}. \tag{2.10}$$

While the non-robust average projection matrix in (2.4) equally weights all subspaces in the collection by $\bar{w}_r^{(k)} = 1/R$, the robust version uses different weights at each iteration. It is also clear that $\overline{\mathbf{P}}_w^{(k)}$ is symmetric with real eigenvalues bounded above by one, like its non-robust version $\overline{\mathbf{P}} = \sum_r \mathbf{P}_r/R$. A unitary basis for the central subspace at iteration $k+1$ is given by the $s^{(k+1)}$ largest eigenvectors of $\overline{\mathbf{P}}_w^{(k)}$, where recall that $s^{(k+1)}$ is the number of non-negative eigenvalues of $\mathbf{S}^{(k)}$ in (2.10).

Notice that the objective function (2.5) is bounded below and that the sequence of objective values at each iteration is non-increasing. Then, the convergence of the sequence of robust order estimates $s^{(1)}, s^{(2)}, \dots,$ to a stationary point s^* is guaranteed. For a more detailed study of the convergence of MM algorithms, the reader is referred to [Sun et al., 2017].

2.3 Distributions on the Manifold of Projection Matrices

In many problems it is useful to assume that the measured subspaces $\{\langle\mathbf{V}_r\rangle\}_{r=1}^{R}$ are random samples drawn from an underlying distribution. Uniform and non-uniform

distributions on the Grassmann manifold $\mathbb{G}(k,n)$ or, equivalently, on the manifold of projection matrices of rank $= k$, $\mathbb{P}(k,n)$, have been extensively discussed in [Chikuse, 2003]. For uniform distributions the basic experiment is this: generate $\mathbf{X}$ as a random $n \times k$ matrix with i.i.d $\mathcal{CN}(0,1)$ random variables. Perform a QR decomposition of this random matrix as $\mathbf{X} = \mathbf{QR}_x$, then, $\mathbf{QH}_x$ where $\mathbf{H}_x \in \mathbb{U}(k)$ is any unitary matrix independent of $\mathbf{X}$, is uniformly distributed on $\mathbb{G}(k,n)$, and $\mathbf{QQ}^H$ is uniformly distributed on $\mathbb{P}(k,n)$. Remember that points on $\mathbb{G}(k,n)$ are equivalence classes of $n \times k$ matrices, where $\mathbf{Q}_1 \sim \mathbf{Q}_2$ if $\mathbf{Q}_1 = \mathbf{Q}_2\mathbf{H}_x$ for some $\mathbf{H}_x \in \mathbb{U}(k)$.

If $\mathbf{P}$ is uniformly distributed on $\mathbb{P}(k,n)$, it is immediate to prove that (see [Chikuse, 2003], pp. 29)

$$E[\mathbf{P}] = \frac{k}{n}\mathbf{I}_n,$$

so all eigenvalues of the mean projection matrix when the subspaces are uniformly distributed are identical to $k_i = k/n$, $i = 1,\ldots,n$, indicating no preference for any particular direction. In this way, the proposed order fitting rule, applied to an average of uniformly-distributed subspaces, will tend to return 0 if $k < n/2$, and n otherwise, in both cases suggesting there is no central low-dimensional subspace.

The matrix Langevin (or von Mises-Fisher) has been suggested as a non-uniform distribution on the Stiefel and Grassmann manifolds [Chikuse, 2003, Khatri and Mardia, 1977, Chikuse, 1990]. For real $n \times k$ orthogonal frame $\mathbf{X}$, the matrix Langevin, as defined by Downs in [Downs, 1972], has an exponential distribution of the form $\mathcal{L}(\mathbf{X}) \propto \exp\{\mathrm{tr}(\mathbf{B}^T\mathbf{X})\}$, where $\mathbf{B} = \mathbf{UDV}^T$ is a matrix that parameterizes the distribution with $\mathbf{U}$ an $n \times k$ slice of an $n \times n$ orthogonal matrix, $\mathbf{V}$ a $k \times k$ orthogonal matrix, and $\mathbf{D}$ a $k \times k$ diagonal matrix with positive entries. The matrices $\mathbf{U}$ and $\mathbf{V}$ are interpreted as orientations, while the diagonal elements of $\mathbf{D}$ are concentration parameters along the k directions determined by $\mathbf{U}$ and $\mathbf{V}$. The matrix Langevin $\mathcal{L}(\mathbf{X})$ is unimodal and the density is maximized at $\mathbf{X} = \mathbf{UV}^T$, which is the central k-frame or subspace of the distribution. Note that when $\mathbf{B} = \mathbf{0}$ we recover the uniform distribution. As suggested in [Chikuse, 2003], to generate samples from $\mathcal{L}(\mathbf{X})$ we might use a rejection sampling mechanism with the uniform as proposal density. This rejection sampling, however, can be very inefficient for large n and $k > 1$. More efficient sampling algorithms have been proposed in [Hoff, 2013].

The uniform and the matrix Langevin are continuous distributions on the manifold of projection matrices of fixed rank $= k$. To deal with subspaces or projection matrices that do not live on the same manifold we would need distributions defined over unions of Grassmannians, which, to the best of our knowledge, have not been studied. Nevertheless, it is possible to define the following discrete distribution that will be useful for the application of the proposed subspace averaging technique to array processing in Section 2.4.

Definition 2.2. *Let* $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix} \in \mathbb{U}(n)$ *be an arbitrary unitary basis of the ambient space, and let* $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n)$ *with* $0 \leq \alpha_i \leq 1$*; the* α_i *are ordered from largest to smallest, but they need not sum to* 1*. We define a discrete distribution* $\mathcal{D}$ *on the set of random projection matrices* $\mathbf{P} = \mathbf{VV}^H$ *(or, equivalently, the set of random*

subspaces $\langle \mathbf{V} \rangle$*, or set of frames* $\mathbf{V}$*) with parameter vector* $\boldsymbol{\alpha}$ *and orientation matrix* $\mathbf{U}$*. The distribution of* $\mathbf{P}$ *will be denoted* $\mathbf{P} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ *or* $\mathbf{V} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$.

To shed some light on this distribution, let us explain the experiment that determines $\mathcal{D}$. Draw 1 includes $\mathbf{u}_1$ with probability α_1, and excludes it with probability $(1 - \alpha_1)$; draw 2 includes $\mathbf{u}_2$ with probability α_2, and excludes it with probability $(1 - \alpha_2)$; continue in this way until draw n includes $\mathbf{u}_n$ with probability α_n, and excludes it with probability $(1 - \alpha_n)$. We may call the string $i_1, i_2, \ldots, i_n$, the indicator sequence for the draws; that is, $i_k = 1$, if $\mathbf{u}_k$ is drawn on draw k, and $i_k = 0$ otherwise. In this way Pascal's triangle shows that the probability of drawing the subspace $\langle \mathbf{V} \rangle$ is $p(\langle \mathbf{V} \rangle) = \prod_I \alpha_i \prod_{\bar{I}} (1 - \alpha_j)$, where the index set I is the set of indices k for which $i_k = 1$ in the construction of $\mathbf{V}$. This is also the probability law on frames $\mathbf{V}$ and projections $\mathbf{P}$. For example, the probability of drawing an empty frame is $\prod_1^n (1 - \alpha_i)$, the probability of drawing the dimension-1 frame $\mathbf{u}_i \mathbf{u}_i^H$ is $\alpha_i \prod_{j \neq i} (1 - \alpha_j)$, and so on. It is clear from this pdf on the 2^n frames that all probabilities lie between 0 and 1, and that they sum to 1.

Let $\mathbf{P}_r \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$, $r = 1, \ldots, R$, be a sequence of i.i.d. draws from the distribution $\mathcal{D}$, and let $\overline{\mathbf{P}} = \sum_r \mathbf{P}_r / R$ be its sample mean with eigenvalues $(k_1, \ldots, k_n)$. Then, we have the following properties:

1. $E[\mathbf{P}_r] = \mathbf{U} \mathrm{diag}(\boldsymbol{\alpha}) \mathbf{U}^H$, that is, the mean is not generally a projection.
2. $E\left[\mathrm{tr}(\mathbf{P}_r)\right] = \sum_{i=1}^n \alpha_i$.
3. $E\left[k_i\right] = \alpha_i$.

These properties follow directly from the definition of $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$.

Remark 1. *The* α_i*'s control the concentrations or probabilities in the directions determined by the unitary basis* $\mathbf{U}$*. For instance, if* $\alpha_i = 1$ *all random subspaces contain direction* $\mathbf{u}_i$*, whereas if* $\alpha_i = 0$ *the angle between* $\mathbf{u}_i$ *and all random subspaces drawn from that distribution will be* $\pi/2$*.*

Example 1. *Suppose* $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ *is the standard basis in* $\mathbb{R}^3$ *and let* $\boldsymbol{\alpha} = (3/4, 1/4, 1/4)$*. The discrete distribution* $\mathbf{P} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ *has an alphabet of* $2^3 = 8$ *subspaces with the following probabilities:*

- $\Pr(\mathbf{P} = \mathbf{0}) = 9/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_1 \mathbf{u}_1^H\right) = 27/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_2 \mathbf{u}_2^H\right) = 3/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_3 \mathbf{u}_3^H\right) = 3/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_1 \mathbf{u}_1^H + \mathbf{u}_2 \mathbf{u}_2^H\right) = 9/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_1 \mathbf{u}_1^H + \mathbf{u}_3 \mathbf{u}_3^H\right) = 9/64$

- $\Pr\left(\mathbf{P} = \mathbf{u}_2\mathbf{u}_2^H + \mathbf{u}_3\mathbf{u}_3^H\right) = 1/64$
- $\Pr\left(\mathbf{P} = \mathbf{I}\right) = 3/64$

The distribution is unimodal with mean

$$E[\mathbf{P}] = \mathbf{U}diag(\boldsymbol{\alpha})\mathbf{U}^H.$$

and expected dimension $E[tr(\mathbf{P})] = 5/4$. *Given* R *draws from the distribution* $P \sim D(\mathbf{U}, \alpha)$, *the eigenvalues of the sample average projection* $\overline{\mathbf{P}} = \sum_{r=1}^{R} \mathbf{P}_r/R$ *converge to* $k_i \to \alpha_i$ *as* R *grows, and the proposed order fitting rule will return* $s^* = 1$ *as the dimension of the central subspace for this example. It is easy to check that the probability of drawing a dimension-*1 *subspace for this example is* $33/64$.

2.4 Subspace Averaging for Source Enumeration

In this section we apply the proposed order fitting rule for subspace averaging to the problem of estimating the number of signals received by a sensor array, which is referred to as source enumeration. This is a classic and well-researched problem in radar, sonar, and communications [Scharf, 1991, Trees, 2002], and numerous criteria have been proposed over the last decades to solve this problem, most of which are given by functions of the eigenvalues of the sample covariance matrix [Akaike, 1974, Wax and Kailath, 1985, Rissanen, 1978, Lu and Zoubir, 2013, Kay, 2005, Xu and Kay, 2008, Zhu and Kay, 2018, Quinlan et al., 2007]. These methods tend to underperform when the number of antennas is large and the number of snapshots is relatively small in comparison to the number of antennas, the so-called small-sample regime, which is the situation of interest here. The proposed method is to construct a collection of subspaces based on the array geometry and a random sampling procedure from a specifically designed distribution $\mathcal{D}$, and then use the order fitting rule for averages of projections to enumerate the sources.

We consider K narrowband signals impinging on a large, uniform, half-wavelength linear array with M antennas, in the presence of white noise (Noise Model 1). Therefore, the noise is modeled as $\mathbf{e}[n] \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I})$, and the covariance matrix in (1.2) for this model is

$$\mathbf{R} = \mathbf{A}\boldsymbol{\Psi}\mathbf{A}^H + \sigma^2\mathbf{I}. \tag{2.11}$$

As described in Section 1.1, the objective here is to estimate K from $\mathbf{X}$ or $\hat{\mathbf{R}}$.

2.4.1 Subspace Generation

A key ingredient of the proposed method is the method of generating the collection of subspaces from which to estimate the average projection matrix and its dimension. In some applications, such as image or video processing, the collection of subspaces might be given, but in array processing the signal subspace is an array manifold, to be estimated from array snapshots. For instance, when a uniform linear array (ULA)

is used, we can exploit the shift-invariance property to estimate a subspace from the data acquired by a subarray of the sensors. The subspaces could also be estimated from subsets of $L < N$ snapshots randomly selected from the original dataset, which would be appropriate for cyclostationary snapshots, or using any other bootstrapping scheme.

For the SA method to be effective, it is important that the subspaces to average overlap as much as possible with the true signal subspace. In fact, as long as each extracted subspace contains a large common portion of the signal subspace and (more or less) independent portions of the noise subspace, then, the averaging procedure enhances signal coordinates while averaging out noise coordinates.

In the following, we describe a subspace generation procedure that has proven to be effective for this particular application. It generates random subspaces by randomly sampling from the distribution $\mathcal{D}(\mathbf{U}, \alpha)$, whose orientations $\mathbf{U}$ and concentrations $\boldsymbol{\alpha}$ are determined by the eigenvectors and eigenvalues of the sample covariance matrix. Moreover, it exploits the shift invariance property of ULAs. A preliminary version of this method that only exploited the shift-invariance property was presented in [Santamaria et al., 2018].

2.4.2 Shift Invariance

When uniform linear arrays are used, a property called shift invariance holds, which forms the basis of the ESPRIT method [Paulraj et al., 1986, Roy and Kailath, 1989] and its many variants. Let $\mathbf{A}_d$ be the $L \times K$ matrix with rows $d, \ldots, d+L-1$ extracted from the steering matrix $\mathbf{A}$. This steering matrix for the dth subarray is illustrated in Fig.2.1.

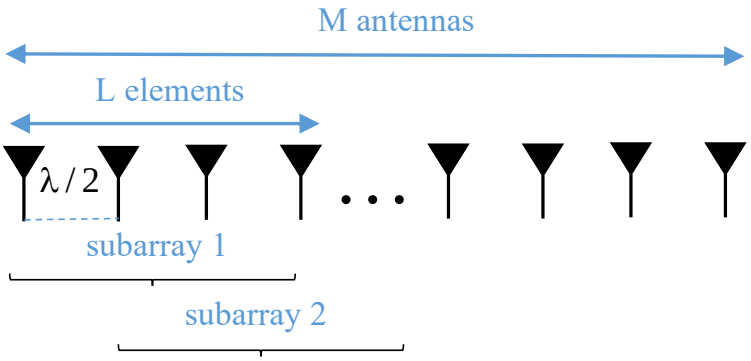


Figure 2.1: L-dimensional subarrays extracted from a uniform linear array with $M > L$ elements.

Then, from (1.1) it is readily verified that

$$\mathbf{A}_d \text{diag}(e^{-j\theta_1}, \ldots, e^{-j\theta_K}) = \mathbf{A}_{d+1}, \; d = 1, \ldots, M-L+1,$$

which is the shift invariance property. In this way, $\mathbf{A}_d$ and $\mathbf{A}_{d+1}$ are related by a nonsingular rotation matrix,

$$\mathbf{Q} = \text{diag}(e^{-j\theta_1}, \ldots, e^{-j\theta_K}),$$

and therefore they span the same subspace. That is, $\langle \mathbf{A}_d \rangle = \langle \mathbf{A}_{d+1} \rangle$, with $\dim(\langle \mathbf{A}_d \rangle) = K < L$. In ESPRIT, two sub-arrays of dimension $L = M - 1$ are considered, and thus we have $\mathbf{A}_1\mathbf{Q} = \mathbf{A}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ select, respectively, the first and the last $M - 1$ rows of $\mathbf{A}$.

There is an interesting characterization of the shift invariance property. Let $\mathbf{x}_d[n]$ be an $L \times 1$ vector containing the noise-free observations acquired by sensors $d, \ldots, d + L - 1$ of $\mathbf{x}[n]$, and let $\mathcal{S}^r$ denote a shift operator, so that $\mathcal{S}^r\mathbf{x}_d[n] = \mathbf{x}_{d+r}[n]$. Then, in the noise-free model $\mathbf{x}_d[n] = \mathbf{A}_d\mathbf{s}[n]$, this shift invariance produces.

$$\mathcal{S}^r\mathbf{x}_d[n] = \mathcal{S}^r\mathbf{A}_d\mathbf{s}[n] = \mathbf{A}_{d+r}\mathbf{s}[n] = \mathbf{A}_d\mathbf{Q}^r\mathbf{s}[n].$$

The source signal $\mathbf{Q}^r\mathbf{s}[n]$ is distributed as $\mathbf{s}[n]$ is distributed, provided $\mathbf{s}[n]$ is complex normal with covariance matrix $\sigma_s^2\mathbf{I}$, making $\mathbf{Q}$ a measure-preserving transformation. So shift on $\mathbf{x}_d[n]$ is measure-preserving on $\mathbf{s}$. This property leaves second-order matrices invariant.

It is also possible to show that, when the signal covariance matrix in (2.11) is $\mathbf{\Psi} = \sigma_s^2\mathbf{I}$, the K principal eigenvectors of $\mathbf{R}$ are also shift-invariant [Roy and Kailath, 1989, Li and Vaccaro, 1991]. When noise is present, however, the shift-invariance property does not hold for the main eigenvectors extracted from the sample covariance matrix. The Optimal Subspace Estimation (OSE) technique proposed by Vaccaro et al. obtains an improved estimate of the signal subspace with the required structure (up to the first order) [Li and Vaccaro, 1991]. The OSE has recently been used with the subspace averaging of [Santamaria et al., 2016] to improve DOA estimation [Vaccaro, 2017, Palka, 1996, Palka and Vaccaro, 2017]. Nevertheless, the OSE technique requires the dimension of the signal subspace to be known in advance and, therefore, does not apply directly to the source enumeration problem.

From the $L \times 1$ ($L > K$) sub-array snapshots $\mathbf{x}_d[n]$ we can estimate an $L \times L$ sample covariance as

$$\hat{\mathbf{R}}_d = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_d[n]\mathbf{x}_d^H[n].$$

Note that each $\hat{\mathbf{R}}_d$ block corresponds to an $L \times L$ submatrix of the full sample covariance $\hat{\mathbf{R}}$ extracted along its diagonal, that is, in Matlab notation $\hat{\mathbf{R}}_d = \hat{\mathbf{R}}(d : d + L - 1, d : d + L - 1)$.

Due to the shift invariance property of uniform linear arrays the noiseless signal subspaces of the theoretical $\mathbf{R}_d$ are identical. Since there are M sensors and we extract L-dimensional subarrays, there are $D = M - L + 1$ different submatrices $\hat{\mathbf{R}}_d$, $d = 1, \ldots, D$. For each $\hat{\mathbf{R}}_d$ we compute its eigendecomposition $\hat{\mathbf{R}}_d = \mathbf{U}_d\mathbf{\Lambda}_d\mathbf{U}_d^H$, where $\mathbf{\Lambda}_d = \text{diag}\,(\lambda_{d,1}, \ldots, \lambda_{d,L})$, $\lambda_{d,1} \geq \ldots \geq \lambda_{d,L}$.

For each $\hat{\mathbf{R}}_d$ we can define a distribution $\mathcal{D}(\mathbf{U}_d, \boldsymbol{\alpha}_d)$ from which to draw random subspaces: $\mathbf{P}_{dk}, k = 1, \ldots, K$. Obviously a key point for the success of the SA method

is to determine a good distribution $\mathcal{D}(\mathbf{U}_d, \boldsymbol{\alpha}_d)$ and a good sampling procedure to draw random subspaces. This is described in the next subsection.

2.4.3 Random Generation of Subspaces

To describe the random sampling procedure for subspace generation, let us take for simplicity the full $M \times M$ sample covariance matrix $\hat{\mathbf{R}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\boldsymbol{\Lambda} = \text{diag}\,(\lambda_1, \ldots, \lambda_M)$, $\lambda_1 \geq \ldots \geq \lambda_M$.

Each random subspace $\langle \mathbf{V} \rangle$ has dimension $\dim(\langle \mathbf{V} \rangle) = k_{max}$, where $k_{max} < \min(M, N)$ is an overestimate of the maximum number of sources that we expect in our problem. The subspace is iteratively constructed as follows:

1. Initialize $\langle \mathbf{V} \rangle = \emptyset$
2. While $\text{rank}(\mathbf{V}) \leq k_{max}$ do
   (a) Generate a random draw $\langle \mathbf{G} \rangle \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$, according to the sampling description in Definition 2.2
   (b) $\langle \mathbf{V} \rangle = \langle \mathbf{V} \rangle \bigcup \langle \mathbf{G} \rangle$

The orientation matrix $\mathbf{U}$ of the distribution $\mathcal{D}$ is given by the eigenvectors of the sample covariance matrix. On the other hand, the concentration parameters should be chosen such that the signal subspace directions are selected more often than the noise subspace directions, and, consequently, they should be a function of the eigenvalues of the sample covariance λ_k. In this work we propose to use the following concentration parameters for $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$

$$\alpha_i = \frac{\Delta\lambda_i}{\sum_i \Delta\lambda_i}, \tag{2.12}$$

where

$$\Delta\lambda_i = \begin{cases} \lambda_i - \lambda_{i+1}, & i = 1, \ldots, M-1, \\ 0, & i = M. \end{cases} \tag{2.13}$$

Here is a motivating example for this choice. Consider the wide-sense stationary time series $\{x[n]\}$ with covariance $r[k] = E[x[n]x^*[n+k]] = \beta \sin(\beta\pi k)/(\beta\pi k)$ for all k, where $0 < \beta < 1$ determines the signal bandwidth. A snapshot $\mathbf{x} = [x[0], \ldots, x[M-1]]$ has symmetric Toeplitz covariance matrix $\mathbf{R}$ with entries $r[k]$ on its kth diagonal. This covariance matrix may be written

$$\mathbf{R} = \int_{-\beta\pi}^{\beta\pi} \frac{d\theta}{2\pi} S(\theta)\boldsymbol{\psi}(\theta)\boldsymbol{\psi}^H(\theta)$$

where $\boldsymbol{\psi} = [1, e^{j\theta}, \ldots, e^{j\theta(M-1)}]^T$, and $S(\theta) = 1$. The trace of this matrix is βM, so the sum of its eigenvalues is this trace. The eigenvalue decomposition of the covariance $\mathbf{R}$ is $\mathbf{R} = \mathbf{F}\mathbf{K}\mathbf{F}^T$, where the columns of $\mathbf{F}$ are the discrete prolate spheroidal wave functions, or Slepian sequences [Slepian and Pollack, 1961]. The first $\lfloor \beta M \rfloor$

Algorithm 1: Generation of a Random Subspace.

Input: $\hat{\mathbf{R}} = \mathbf{U}_0 \boldsymbol{\Lambda} \mathbf{U}_0^H$, k_{max}
Output: Unitary basis for a random subspace $\mathbf{V}$
Initialization: $\mathbf{U} = \mathbf{U}_0$, $\boldsymbol{\lambda} = \text{diag}(\boldsymbol{\Lambda})$, and $\mathbf{V} = \emptyset$
while $rank(\mathbf{V}) \leq k_{max}$ **do**
- `/* Generate concentration parameters` $\boldsymbol{\alpha}$ `*/`
- $M = |\boldsymbol{\lambda}|$
- $\alpha_i = \frac{\Delta\lambda_i}{\sum_i \Delta\lambda_i}$, $i = 1, \ldots, M$ with $\Delta\lambda_i$ given by (2.13)
- `/* Sample from` $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ `*/`
- $\mathbf{g} = (g_i, \ldots, g_M)$ with $g_i \sim \mathcal{U}(0,1)$
- $\mathcal{I} = \{i \mid g_i \leq \alpha_i\}$
- $\mathbf{G} = \mathbf{U}(:, \mathcal{I})$
- `/* Append new subspace */`
- $\mathbf{V} = \begin{bmatrix} \mathbf{V} & \mathbf{G} \end{bmatrix}$
- `/* Eliminate selected directions */`
- $\mathbf{U} = \mathbf{U}(:, \overline{\mathcal{I}})$
- $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\overline{\mathcal{I}})$

eigenvalues λ_i are approximately 1, and the trailing $M - \lfloor \beta M \rfloor$ are approximately 0. Moreover, the matrix $\mathbf{FKF}^T$ is approximately a rank $\lfloor \beta M \rfloor$ projection onto the subspace spanned by the first $\lfloor \beta M \rfloor$ columns of the matrix $\mathbf{FKF}^T$. An estimator of this rank is $\text{argmax}_i(\lambda_1 - \lambda_{i+1})$, which returns the integer part of βM. This suggests that the function $\alpha_i = \lambda_i - \lambda_{i+1}$, after proper normalization, would be a suitable function of the eigenvalues to use in a sequence of stochastic draws of subspaces, as outlined previously.

With this choice for $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$, the probability of picking the ith direction from $\mathbf{U}$ is proportional to $\lambda_i - \lambda_{i+1}$, thus placing more probability on jumps of the eigenvalue profile. Notice also that whenever $\lambda_i = \lambda_{i+1}$ then $\alpha_i = 0$, which means that $\mathbf{u}_i$ will never be chosen in any random draw. We take the convention that if $\Delta\lambda_i = 0$, $\forall i$, then we do not apply the normalization in Eq. (2.12) and hence the concentration parameters are also all zero: $\alpha_i = 0$, $\forall i$.

A summary of the proposed algorithm is shown in Algorithm 1.

2.4.4 SA Criterion

For each subarray sample covariance matrix we can generate T random subspaces according to the procedure described in the previous section. Since we have D subarray matrices, we get a total of $R = T\,D$ subspaces. The SA approach simply finds the average projection matrix

$$\overline{\mathbf{P}} = \frac{1}{TD} \sum_{d=1}^{D} \sum_{t=1}^{T} \mathbf{P}_{dt},$$

Algorithm 2: Subspace Averaging Criterion.

Input: $\hat{\mathbf{R}}$, $L = M - 5$, T and k_{max};
Output: Order estimate $\hat{k}_{SA}$
for $d = 1, \ldots, D$ **do**
  Extract $\hat{\mathbf{R}}_d$ from $\hat{\mathbf{R}}$ and obtain $\hat{\mathbf{R}}_d = \mathbf{U}_d \boldsymbol{\Sigma}_d \mathbf{U}_d^H$
  Generate T random subspaces from $\hat{\mathbf{R}}_d$ using Algorithm 1
  Compute the projection matrices $\mathbf{P}_{dt} = \mathbf{V}_{dt} \mathbf{V}_{dt}^H$
Compute $\overline{\mathbf{P}}$ and its eigenvalues $(k_1, \ldots, k_L)$
Estimate $\hat{k}_{SA}$ as the number of eigenvalues of $\overline{\mathbf{P}}$ larger than 1/2

to which the order estimation method described in Sec. 2.2 can be applied. Notice that the only parameters in the method are the dimension of the subarrays, L, the dimension of the extracted subspaces, k_{max}, and the number T of random subspaces extracted from each subarray. For large-scale arrays ($M \geq 50$), we have found that $L = M - 5$, $k_{max} = \lfloor M/5 \rfloor$, and $T = 20$ provide in general good performance for many scenarios.

A summary of the proposed algorithm is shown in Algorithm 2.

2.4.5 Consistency of the SA Method

In this section we show that the SA criterion equipped with the proposed subspace generation procedure is consistent in the classic asymptotic regime (fixed M, $N \to \infty$).

Theorem 2.3. *Let $\mathbf{R} = \mathbf{A}\boldsymbol{\Psi}\mathbf{A}^H + \sigma^2\mathbf{I}$ be an $M \times M$ theoretical covariance matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \ldots, \geq \mu_K > \mu_{K+1} = \ldots = \mu_M$, corresponding to an scenario with K uncorrelated sources. Let $\hat{\mathbf{R}}$ be the sample covariance matrix formed by N snapshots and let us denote its eigenvalues as λ_i, $i = 1, \ldots, M$. Then, if $M \geq L > k_{max} \geq K$, $\hat{k}_{SA}$ is a consistent estimator of K as $N \to \infty$.*

Proof. Let us take without loss of generality the case $M = L$, that is, we only generate subspaces from the full sample covariance matrix. In the classic fixed system-size M, large sample-size asymptotic regime $N \to \infty$, the eigenvalues of the sample covariance estimated from independent snapshots converges to the theoretical ones $\lambda_i \to \mu_i$, $i = 1, \ldots, M$. Then, in the first draw to construct each random subspace we sample from $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ where the orientation matrix $\mathbf{U}$ contains the eigenvectors of $\mathbf{R}$ and the concentration parameters are

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K, 0, \ldots, 0)\,. \tag{2.14}$$

In (2.14) the $M - K$ trailing concentration values are zero because they are constructed from (normalized) differences of eigenvalues, whose $M - K$ smallest values are identical in the asymptotic regime: $\mu_{K+1} = \ldots = \mu_M$.

Therefore, the first draw samples exclusively from the signal directions. Since we sample until the dimension of the subspace is k_{max}, or until all concentration parameters are zero, then, as long as $k_{max} \geq K$, all random subspaces will be the true signal subspace. Consequently, for any number of generated subspaces the average projection matrix has exactly K eigenvalues equal to 1 and $M - K$ eigenvalues equal to zero, and the SA criterion returns $\hat{k}_{max} = K$, thus proving the result.

□

Notice that this consistency results also holds as the noise variance $\sigma^2 \to 0$, unlike the MDL criterion which, as shown in [Ding and Kay, 2011], is inconsistent with increasing signal-to-noise ratio.

2.5 Simulation Results

In this section we evaluate the performance of the proposed order fitting rule by means of numerical examples. Firstly, we study the performance of the order fitting rule, as well as its robust version. Secondly, we consider the application of subspace averaging techniques as a method of enumerating sources in large linear arrays, under conditions of low sample support.

2.5.1 Performance of the Order Fitting Rule

Experiment 1: In the first example, we generate a collection of R subspaces, $\langle \mathbf{V}_r \rangle \in \mathbb{G}(k,n)$, $r = 1,\ldots,R$, as follows: we first generate

$$\mathbf{G}_r = \Big[\mathbf{V}_0 \mid \mathbf{0}_{n\times(n-k)}\Big] + \sigma\mathbf{Z}_r, \quad r = 1,\ldots,R \tag{2.15}$$

where $\mathbf{V}_0 \in \mathbf{C}^{n\times k}$ is a matrix whose columns form an orthonormal basis for a central subspace $\langle \mathbf{V}_0 \rangle$, $\mathbf{0}_{n\times(n-k)}$ is an $n \times (n-k)$ zero matrix, and $\mathbf{Z}_r \in \mathbf{C}^{n\times n}$ is a matrix whose entries are independent and identically distributed (iid) complex Gaussian random variables with zero mean and variance $\sigma^2 = 1/n$. The value of σ determines the signal-to-noise-ratio, which determines the spread of the subspaces around its mean and is defined as $\text{SNR} = 10\log_{10}\left(\frac{k}{n\sigma^2}\right)$.

An orthogonal basis for the rth subspace, $\mathbf{V}_r$, is then constructed from the first k orthonormal vectors of the QR decomposition of $\mathbf{G}_r$. For this example all subspaces in the collection have exactly the same dimension.

Fig. 2.2 shows the estimated order as a function of the SNR for different values of (k,n) and a total number of $R = 200$ subspaces. The curves represent averaged results of 500 independent simulations. As we can see, there is phase-transition behavior between $s^* = 0$ (no central subspace) and the right order $s^* = k$.

Experiment 2: In the second experiment we evaluate the robust order fitting rule proposed in 2.2.4. To this end, we create a collection of subspaces contaminated by outliers as follows: we first generate

$$\mathbf{G}_r = \Big[\mathbf{V}_0 \mid \mathbf{0}_{n\times(n-k)}\Big] + \mathbf{Z}_r, \quad r = 1,\ldots,R \tag{2.16}$$

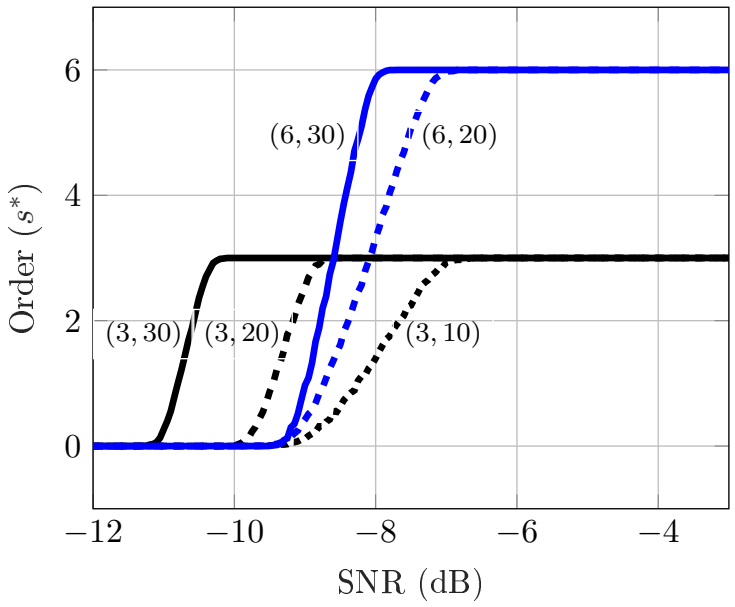


Figure 2.2: Estimated order as a function of the SNR for different values of (k, n). In all examples the number of measured subspaces is $R = 200$.

where now the elements of $\mathbf{Z}_r$ are drawn from a Gaussian mixture $\mathbf{Z}_r(i,j) \sim (1-\epsilon)\mathcal{CN}(0, \sigma_1^2/n) + \epsilon\,\mathcal{CN}(\sigma_2^2/n)$, where $\sigma_2^2 >> \sigma_1^2$. In words, with probability $(1-\epsilon)$ the central subspace is additively perturbed by a random matrix whose entries are i.i.d. zero-mean Gaussians random variables with variance σ_1^2/n, whereas with probability ϵ the entries of the noise matrix are drawn from a Gaussian distribution with variance $\sigma_2^2/n >> \sigma_1^2/n$. Again, an orthogonal basis for the rth subspace, $\mathbf{V}_r$, is constructed from the first k orthonormal vectors of the QR decomposition of $\mathbf{X}$. In this way, we emulate a Gaussian mixture model for this problem. For low values of σ_1^2, with probability $1-\epsilon$ the subspaces are well clustered around $\mathbf{V}_0$. On the other hand, with probability ϵ the subspaces are generated with a much higher variance $\sigma_2^2 >> \sigma_1^2$ and hence they can be interpreted as outliers.

The signal-to-noise-ratio for the normal data (inliers) and the outliers is defined as $\text{SNR}_i = 10\log_{10}\left(\frac{k}{n\sigma_i^2}\right)$ for $i = 1, 2$. For this example, we estimate the order of the central subspace using the extrinsic mean squared error distance, and the robust versions using the ℓ_1 norm (2.6) and the Huber loss function with $T = 0.5$ (2.7). In all simulations, the MM algorithm converged in less than 5 iterations. We consider a set of $R = 100$ subspaces of dimension $k = 3$ in an ambient space of dimension $n = 20$. The proportion of outliers to $\epsilon = 0.5$, and its signal-to-noise ratio is $\text{SNR}_2 = -20$ dB. Fig. 2.3 shows the probability of correct order estimation as the signal-to-noise-ratio, SNR_1, for the inliers varies, where increased robustness of both the ℓ_1 and Huber cost functions are evident.

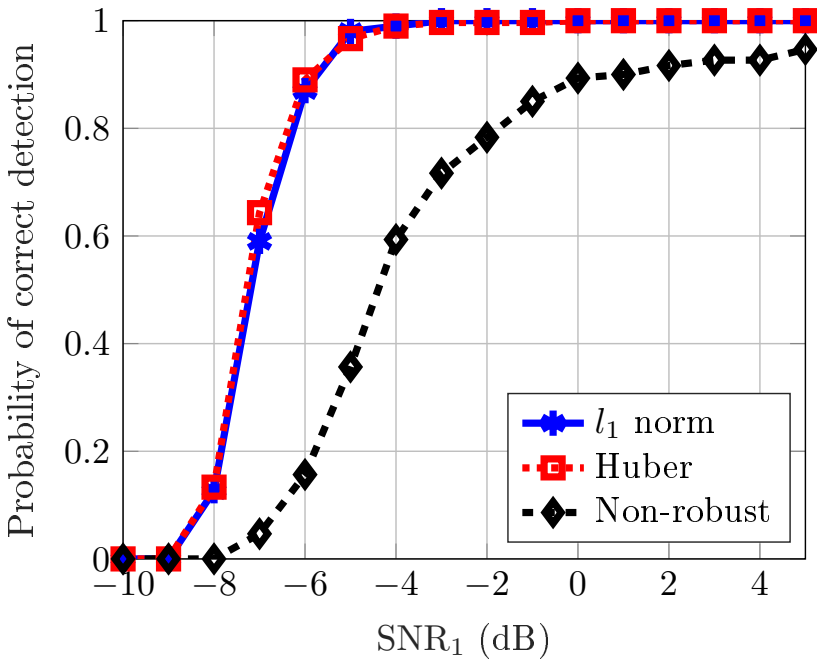


Figure 2.3: Probability of correct order detection for robust and non-robust methods ($M = 100$, $k = 3$, $n = 40$, $\text{SNR}_2 = -20$ dB and $\epsilon = 0.5$).

2.5.2 Application to Source Enumeration

We consider a scenario with K narrowband incoherent unit-power signals and DOAs separated Δ_θ (in electrical angle) impinging on a uniform linear array with M antennas and half-wavelength element separation (cf. Fig. 1.1). The number of snapshots is N. The proposed SA method uses subarrays of size $L = M - 5$, hence for $M = 100$, the total number of subarrays is $D = 6$. From the sample covariance matrix of each subarray we generate $T = 20$ random subspaces of dimension $k_{max} = \lfloor M/5 \rfloor$, which gives us a total of $R = 120$ subspaces (for $M = 100$) on the Grassmann manifold $\mathbb{G}(k_{max}, L)$ to compute the average projection matrix $\overline{\mathbf{P}}$.

Some representative methods for source enumeration with high-dimensional data and few snapshots have been selected for comparison. They exploit random matrix results and are specifically designed to operate in this regime. Further, all of them are functions of the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_M$ of the sample covariance matrix $\hat{\mathbf{R}}$. For comparison, LS-MDL [Huang and So, 2013], NE [Nadakuditi and Edelman, 2008] and BIC [Huang et al., 2016] are considered. These methods have been briefly described in Section 1.2.

Experiment 3: In this example we consider an array with $M = 100$ antennas receiving $K = 3$ sources with electrical angle separation $\Delta_\theta = 10°$, and $N = 60$ snapshots, thus yielding a rank-deficient sample covariance matrix. The Rayleigh limit for this scenario is $2\pi/100 \approx 3.6°$, so in this example the sources are well separated.

Fig. 2.4 shows the probability of correct detection vs. the signal-to-noise-ratio (SNR) for all methods under comparison. Increasing the number of snapshots to $N = 150$ and keeping fixed the rest of the parameters, we obtain the results shown in Fig. 2.5. For this scenario, where source separations are roughly 3 times the Rayleigh limit, the SA method outperforms competing methods.

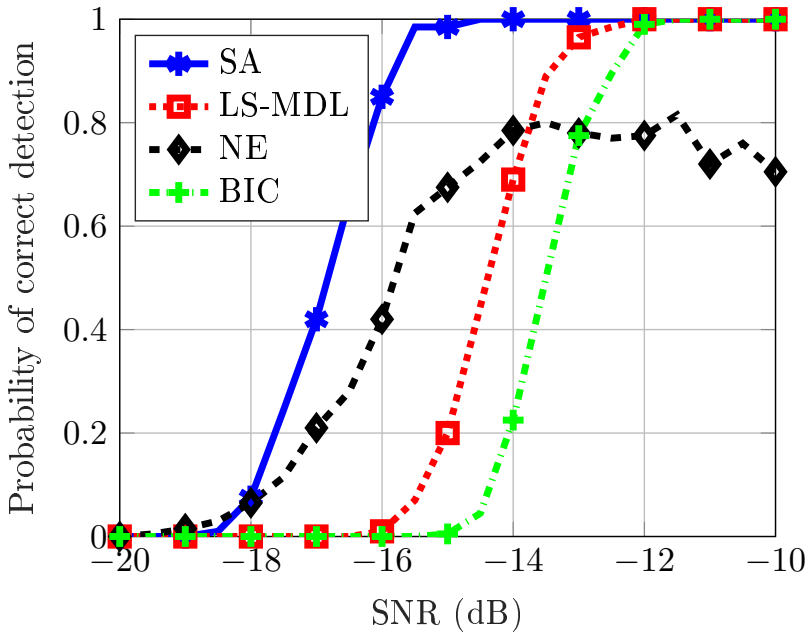


Figure 2.4: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources with electrical angle separation $\Delta_\theta = 10^\circ$, the number of antennas is $M = 100$, the number of snapshots is $N = 60$ and $L = \lfloor M - 5 \rfloor$

Experiment 4: To analyze the impact of the separation between sources, we consider again a scenario with $M = 100$ antennas, and $K = 3$ sources but now separated at angles of $\Delta_\theta = 2^\circ$, which is within the Rayleigh limit of approximately 3.6°. The results for $N = 150$ and $N = 60$ snapshots are shown in Figs. 2.6 and 2.7, respectively. Again, the SA method provides the best performance. Also, the LS-MDL method seems to be more robust than NE and BIC for very low sample support scenarios ($N = 60$).

Experiment 5: In this experiment we compare the performance for an increasing number of snapshots when the number of antennas is fixed to $M = 100$ antennas, the signal-to-noise-ratio is SNR $= -16$ dB, and there are $K = 3$ uncorrelated sources separated $\Delta_\theta = 10^\circ$. As Fig. 2.8 shows, the SA method provides very competitive results with only a few snapshots, while the other methods require a much higher number of snapshots to consistently estimate the right number of sources.

Experiment 6: In the next experiment we consider an array with $M = 120$ antennas, the signal-to-noise-ratio is SNR $= -16$ dB, and there are $K = 6$ uncorrelated sources now with separation of $\Delta_\theta = 12^\circ$. As Fig. 2.9 shows, the SA method starts performing well with very few snapshots.

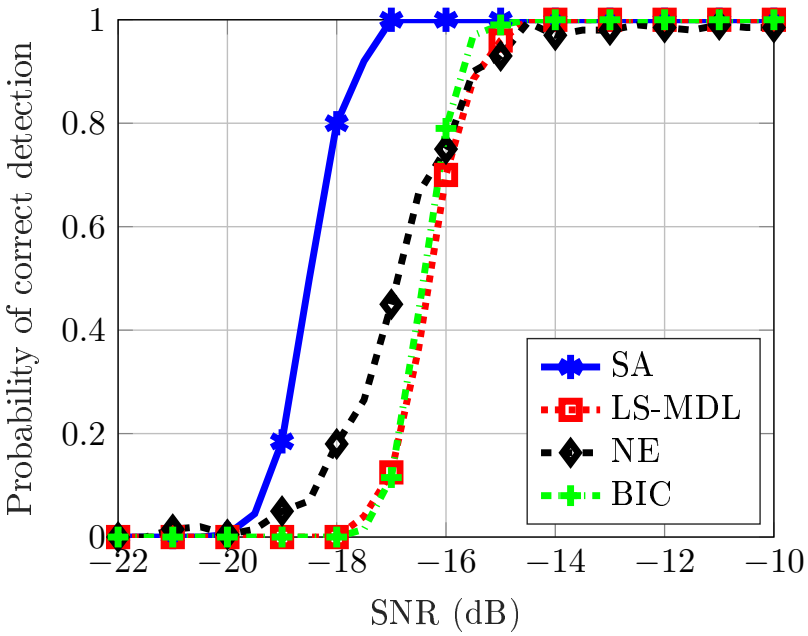


Figure 2.5: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources with electrical angle separation $\Delta_\theta = 10^\circ$, the number of antennas is $M = 100$, the number of snapshots is $N = 150$ and $L = \lfloor M - 5 \rfloor$

Experiment 7: In the next experiment we assume two sources with unequal powers P_1 and P_2 respectively. We assume that initially both sources have equal power as $P_1 = P_2$, and then, we fix the P_1 and decrease the P_2 gradually, and the probability of correct detection vs. the ratio of two powers $\frac{P_2}{P_1}$ is shown in Fig. 2.10. We consider an array with $M = 100$ antennas, the number of samples $N = 60$, the signal-to-noise-ratio is SNR $= -10$ dB, and the source separation $\Delta_\theta = 10^\circ$. As the P_2 decreases, all methods start underestimating the number of sources, however, SA performs well for small values of ratio $\frac{P_2}{P_1}$.

Experiment 8: In the last experiment, we evaluate the impact of having correlated sources. We consider a scenario with $K = 2$ correlated sources when $M = 100$, $N = 150$, $SNR = -10$ dB, and $\Delta_\theta = 10^\circ$. The correlation coefficient between two sources, ρ, varies from 0 to 1, where $\rho = 0$ denotes uncorrelated signals and $\rho = 1$ denotes fully correlated signals. As Fig. 2.11 shows, SA outperforms the rest of methods and provides accurate results even for highly correlated sources $\rho < 0.8$. Nevertheless, the performance of SA under correlated sources needs additional theoretical analysis.

Discussion: It may be said that the method NE uses asymptotic results, based on large random matrix theory, to derive an order fitting rule. The rule is then applied to randomly generated eigenvalues computed from finite samples of finite matrices, as if these eigenvalues behaved as the eigenvalues for a large random matrix. They do, approximately. The methods LS-MDL (based on MDL), and BIC use geometric and arithmetic means of *sub-dominant eigenvalues*, derived from likelihood formulas,

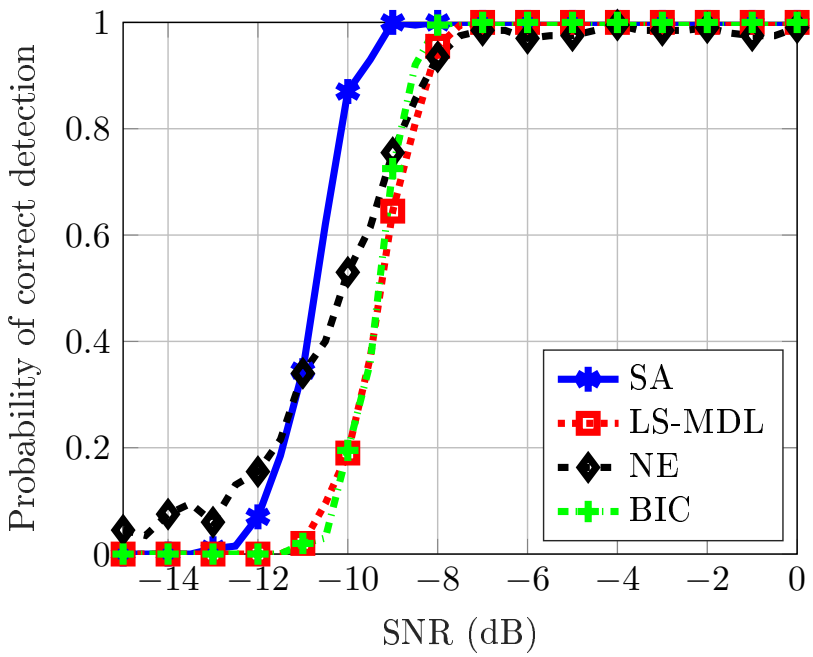


Figure 2.6: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources with electrical angle separation $\Delta_\theta = 2^\circ$, the number of antennas is $M = 100$, the number of snapshots is $N = 150$ and $L = \lfloor M - 5 \rfloor$

to determine the likelihood of a factor model of a fixed order. In fact, in the computation of likelihood, it is the *likelihood of a signal covariance matrix* $\mathbf{F}\mathbf{\Lambda}\mathbf{F}^H$*, of order k, plus a diagonal noise covariance* $\sigma^2\mathbf{I}$ *of unknown variance* σ^2 *that is computed.* So, in a very real sense, all these methods are based on the likelihood of a *full covariance model for multivariate normal data*, and likelihoods of different models are rank- ordered after penalties for large order are applied. This rank ordering depends critically on the scales of the components $\mathbf{F}\mathbf{\Lambda}\mathbf{F}^H$ and σ^2 that are identified in the likelihood computation. The method SA, treats eigenvalues computed from finite samples of finite matrices as variables that only indicate which model *subspace* could have produced these eigenvalues as the eigenvalues of a corresponding covariance matrix. Importantly, all eigenvalues are used in a bootstrap, and not only the sub-dominant eigenvalues. Perhaps more importantly, scale is removed from consideration. That is, the methods LS-MDL and BIC account for scale of the low-rank and diagonal components of covariance in the fitting of a covariance model to the data, whereas the method of SA is scale-invariant, as it computes scale-invariant probabilities from the covariance eigenvalues, without a low-rank-plus diagonal covariance model, and then produces draws of randomly-generated, scale-invariant, subspaces. Subspace modeling seems better matched to the problem of order determination for an array manifold than does covariance modeling. The experiments in this section indicate that this normalization with respect to scale is useful for estimating model order in

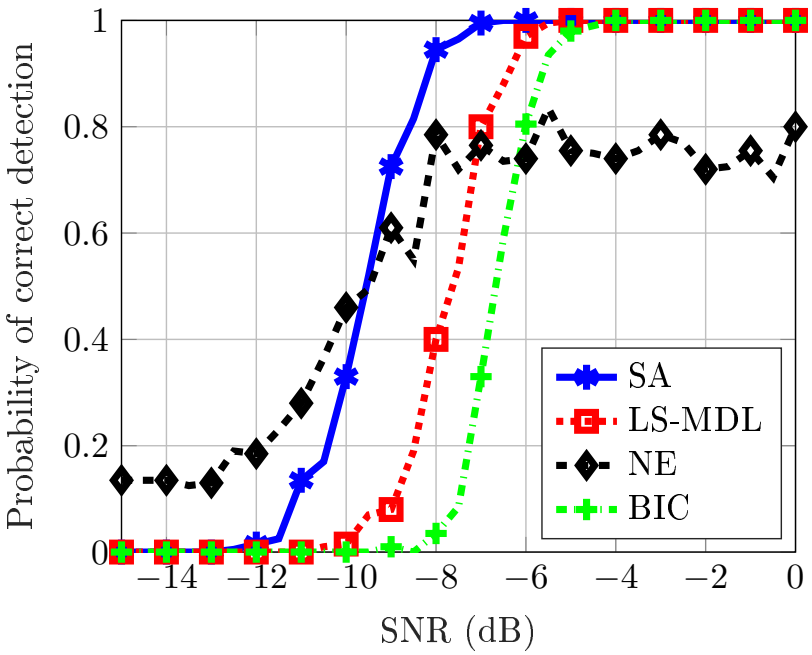


Figure 2.7: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources with electrical angle separation $\Delta_\theta = 2°$, the number of antennas is $M = 100$, the number of snapshots is $N = 60$ and $L = \lfloor M - 5 \rfloor$

experiments where the scales of the signal covariance and the noise covariance are unknown, and the SNR and/or sample support are small.

2.6 Conclusions

In this chapter we have studied the problem of source enumeration from measurements in a uniform linear array. The approach is to extract a subspace from each of several subarrays, and then average these subspaces for a subspace whose dimension is the estimated number of far-field sources. A key element of the method is the automatic order-fitting rule for extracting the dimension of the average subspace that minimizes the mean-squared error between the average and each individual subspace. The net of this procedure is that eigenvalues of a sample covariance matrix determine a distribution on subspaces that could have produced the measured covariance matrix. This procedure normalizes scale by replacing scale-dependent covariance models by scale-invariant subspace models. The method requires no penalty terms for controlling the estimated order.

Simulations indicate performance that is superior to other published methods, over a range of signal-to-noise ratios, sample supports, and source separations. The results suggest that the problem of source enumeration may be viewed as a problem of identifying an approximating subspace, and its dimension, from a set of subspaces estimated from measurements. This point of view stands in contrast to methods that

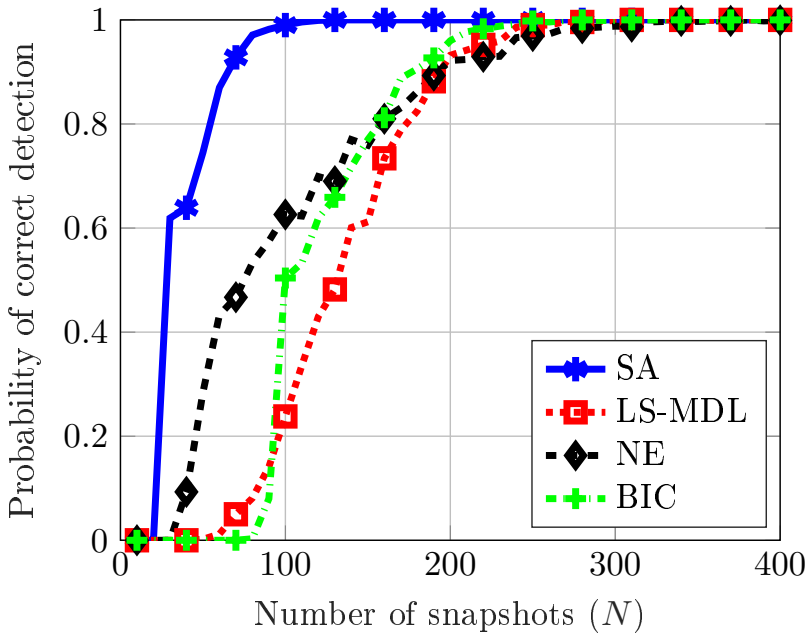


Figure 2.8: Probability of correct detection vs. Number of Snapshots for all methods. In this experiment, there are $K = 3$ sources with electrical angle separation $\Delta_\theta = 10^\circ$, the number of antennas is $M = 100$, the Signal to Noise Ratio is $SNR = -16$ dB and $L = \lfloor M - 5 \rfloor$

compute likelihood for covariance models, where scale is retained, and then penalize these likelihoods for large dimension.

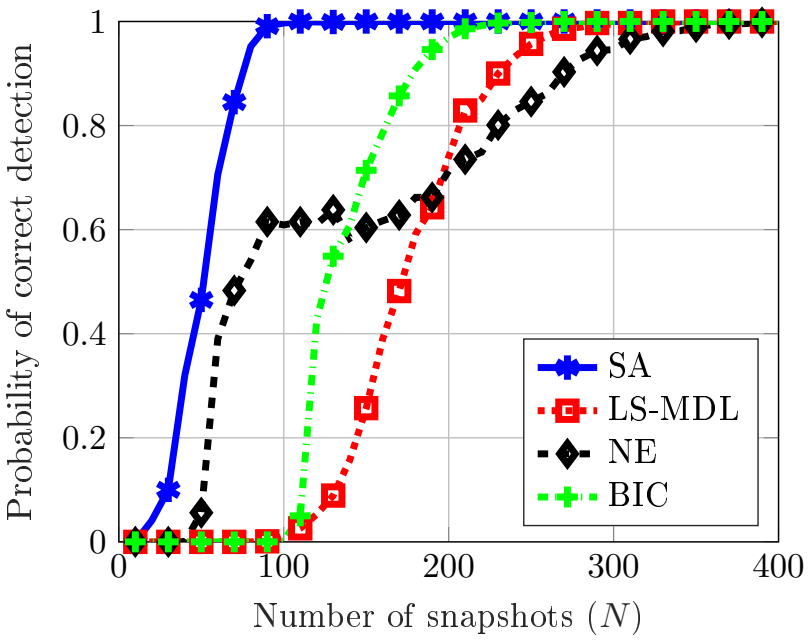


Figure 2.9: Probability of correct detection vs. number of snapshots for all methods. In this experiment, there are $K = 6$ sources with electrical angle separation $\Delta_\theta = 12°$, the number of antennas is $M = 120$, $SNR = -16$ dB and $L = \lfloor M - 5 \rfloor$.

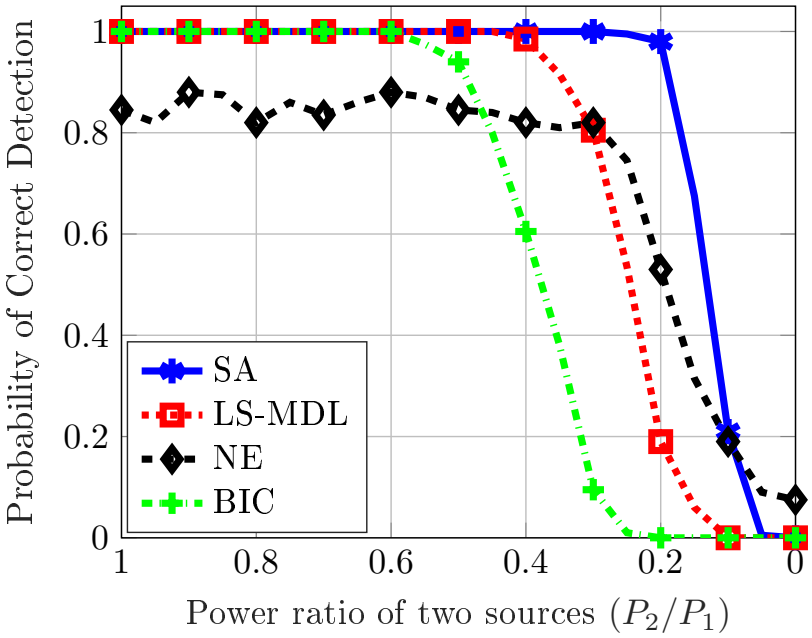


Figure 2.10: Probability of correct detection vs. power ratio of two sources $\frac{P_2}{P_1}$ for all methods. In this experiment, there are $K = 2$ sources with electrical angle separation $\Delta_\theta = 10°$, the number of antennas is $M = 100$, $N = 60$, $SNR = -10$ dB and $L = \lfloor M - 5 \rfloor$.

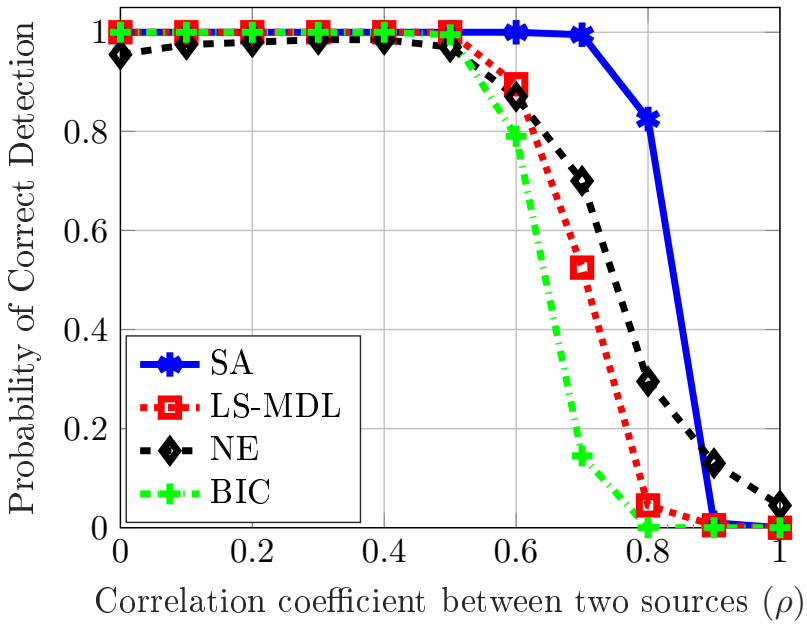


Figure 2.11: Probability of correct detection vs. correlation coefficient between two sources ρ for all methods. In this experiment, there are $K = 2$ sources with electrical angle separation $\Delta_\theta = 10^\circ$, the number of antennas is $M = 100$, $N = 150$, $SNR = -10$ dB and $L = \lfloor M - 5 \rfloor$.

Chapter 3

Extension to Non-White Noises and/or Non-Uniform Linear Arrays

In Chapter 2, we discussed the problem of source enumeration for uniform linear arrays in the presence of white noise, and provided a solution based on subspace averaging. However, in many practical applications, the assumptions of ULA and white noise are too restrictive, and hence, SA method can not be applied. In this chapter, we relax these assumptions and extend SA for: i) non-white noises, and ii) non-uniform linear arrays. For the first extension, we combine SA technique with a majority vote approach. The number of sources is detected for increasing dimensions of the SA technique and then a majority vote is applied to determine the final estimate. To extend SA for arrays with arbitrary geometry, each received snapshot is approximated by a sparse linear combination of rest of the snapshots. Based on the sparse reconstruction (SR) solution, a sampling mechanism is proposed in the second step to generate a collection of subspaces, all of which approximately span the same signal subspace. Finally, the dimension of the average of this collection of subspaces provides a robust estimate for the number of sources.

3.1 Introduction

The assumption of white noise is too restrictive in many scenarios. For example, when the receivers are uncalibrated or due to hardware nonidealities, the noise covariance matrix might be diagonal with unknown elements. In addition to that, for underwater sonar systems, the noise is generally spatially correlated [Urick, 1983, Wenz, 1972, Cron and Sherman, 1962]. Also, close sensors generally have a strong correlation among each-other [Stoica and Cedervall, 1997]. As discussed in Chapter 1, source enumeration in these non-white noise cases with high-dimensional data (i.e., large arrays) and few snapshots is still a challenging problem for which not many methods have been found in the literature that work robustly under different models for the noise covariance matrix.

On the other hand, SA exploits the shift invariance property of ULAs to produce submatrices spanning the same signal subspace. If a non-uniform linear array is used,

the shift invariance property does not hold, and therefore SA in its original form can not be applied.

In this chapter, we first extend SA for non-white noises and ULAs. To tackle this problem, SA is combined with a majority vote approach, and an improved version of the SA technique for order estimation is presented. Under non-white noise scenarios, the original SA criterion is very sensitive to the chosen dimension of extracted subspaces k_{max}. To increase the robustness of the SA method, we apply the SA order estimation rule for increasing values of k_{max} and then apply a majority vote to determine the final estimate of the number of sources. This simple modification, makes SA a very robust method of enumerating sources in large linear arrays under conditions of low sample support and under different models for the noise covariance matrix.

Furthermore, we extend the SA enumeration technique to arrays with arbitrary geometry, and Gaussian noises with arbitrary covariance matrix. The method first approximates each snapshot by a sparse linear combination of rest of the received snapshots. In a noiseless situation, the number of nonzero coefficients of the sparse expansion would directly reveal the signal subspace dimension. In a noisy situation, the sparse expansions can be used to generate a collection of subspaces that, when averaged, will provide the order estimate using the SA method. This method, which combines sparse reconstruction with subspace averaging, is termed sparse subspace averaging (SSA).

The sparse reconstruction (SR) problem is solved by generalizing to complex-valued signals the well-known logarithm-based surrogate of the ℓ_0-norm [Candès et al., 2008]. Using the majorization-minimization (MM) framework [Sun et al., 2017], we show that the SR problem amounts to solving a reweighted regularized least squares (LS) problem. Based on the SR solution, a sampling mechanism is proposed to generate a collection of random subspaces, all of which approximately span the same signal subspace. Finally, an average or central subspace is estimated as in Chapter 2, and the dimension of this average provides a robust estimate for the number of sources impinging on the array. Sparse reconstruction approaches have been used before in array signal processing, but mainly for direction-of-arrival (DOA) estimation problems [Li et al., 2014, Yin and Chen, 2011, Malioutov et al., 2005]. Differently from these works, here we use sparse reconstruction methods as a means to generate multiple approximations to the signal subspace that, when averaged, provide accurate estimates of the underlying signal subspace rank.

The structure of the chapter is as follows. Section 3.2 extends the SA method for non-white noises for ULAs, and a majority vote approach is proposed in Section 3.2.1. Section 3.3 discusses the source enumeration problem for non-uniform linear arrays. A sparse representation method is discussed in Section 3.3.1, and then a method to generate subspaces randomly is proposed in Section 3.3.2. Further, the SSA criterion is proposed in Section 3.3.3. Section 3.4 evaluates the performance of the proposed methods. Finally, the conclusions are summarized in Section 3.5.

3.2 Extension to Non-White Noises: SA with Majority Vote

This section extends the subspace averaging technique for non-white noises. It is noted here that for this section, we are still working with ULAs. We assume that K narrowband signals are impinging on a large, uniform, half-wavelength linear array with M antennas. However, differently to Chapter 2, we do not assume any particular structure for the noise covariance matrix. Therefore, the noise is modeled here as $\mathbf{e}[n] \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{R}_n)$, and the covariance matrix for this scenario is

$$\mathbf{R} = \mathbf{A}\boldsymbol{\Psi}\mathbf{A}^H + \mathbf{R}_n, \tag{3.1}$$

where $\mathbf{R}_n$ is an $M \times M$ non-white noise covariance matrix. Similar to Chapter 2, we assume here as well that the array is composed of a large number of antenna elements, and that the number of snapshots is possibly smaller than the number of antennas, that is, $N < M$.

As discussed earlier, the SA method provides competitive results when the noises are i.i.d., but its performance degrades under spatially correlated noises. In the next subsection we present a simple modification to SA to accurately detect the number of sources in non-white noises and with a small sample size. This modification suggests to apply SA technique for varying values of dimension of extracted subspace k_{max}. It is important to note here that preliminary SA method proposed in [Santamaria et al., 2018] is used here, which does not exploit the generation of random subspace discussed in Section 2.4.3. For the sake of clarity and completeness, the SA method proposed in [Santamaria et al., 2018] is reviewed here, which is as follows:

Let $\mathbf{x}_d[n]$ be the nth snapshot of the dth subarray (Fig. 2.1), and let $\hat{\mathbf{R}}_d$ be the $L \times L$ sample covariance, which corresponds to a submatrix of the full sample covariance. Since there are M sensors and we extract L-dimensional subarrays, there are $D = M - L + 1$ different submatrices $\hat{\mathbf{R}}_d$, $d = 1, \ldots, D$. Due to the shift invariance property of uniform linear arrays the noiseless signal subspaces of the theoretical $\mathbf{R}_d$ are identical. The SA method extracts for each subarray a subspace formed by the k_{max} largest eigenvectors of $\mathbf{R}_d$. A unitary basis for this subspace is denoted as $\mathbf{V}_d \in \mathbb{C}^{L \times k_{max}}$, and the orthogonal projection matrix onto the subspace is $\mathbf{P}_d = \mathbf{V}_d\mathbf{V}_d^H$. The SA order determination rule first computes the average projection matrix

$$\overline{\mathbf{P}} = \frac{1}{D}\sum_{d=1}^{D} \mathbf{P}_d, \tag{3.2}$$

and its eigenvalues $1 \geq k_1 \geq k_2 \geq \ldots \geq k_L$. Finally, the number of sources is determined as the number of eigenvalues larger than 1/2.

3.2.1 The Majority Vote Approach

A limitation of the SA method is that its performance is sensitive to the dimension of extracted subspaces k_{max}. This problem becomes more important under spatially

correlated noises because the directions of the extracted subspaces that correspond to the noise subspace tend to be correlated for consecutive subarrays.

Fig. 3.1 illustrates the influence of k_{max} in the order estimate provided by SA when the noise is spatially correlated with arbitrary invariance matrix $\mathbf{R}_n \succ 0$. For this example we assume $M = 100$ antennas, $N = 50$ snapshots and different number of sources with electrical angle separation $\Delta_\theta = 10^\circ$. Fig. 3.1 clearly shows that $\hat{K}_{SA}$ first increases with the value of k_{max}, and then becomes constant for $\hat{K}_{SA} = K$. However, if we further increase the value of k_{max}, the $\hat{K}_{SA}$ increases again. This example suggests a simple procedure for estimating K. First, the SA order estimation rule is applied for a sequence of increasing values of k_{max}, $1 \leq k_{max} \leq d_{max}$, where d_{max} is an overestimate of the maximum number of sources that we expect in our problem. The final estimate is obtained by majority vote. This simple modification makes the SA method a robust source enumeration technique suitable for noises with different spatial correlation models and with very few snapshots. In comparison to the original SA method for white noise (as discussed in Section 3.2), the computational cost is increased by a factor of d_{max}. A summary of the final algorithm is shown in Algorithm 3.

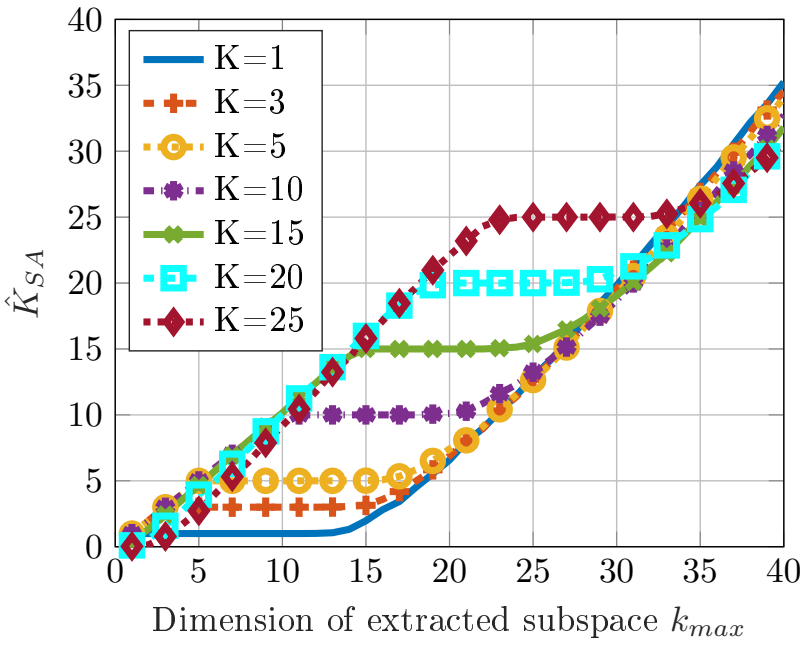


Figure 3.1: Estimated number of sources for SA method vs. k_{max} (dimension of extracted subspaces) in non-white noise (arbitrary noise with $\mathbf{R}_n \succ 0$) for $M = 100$, $N = 50$, $\Delta_\theta = 10^\circ$ and SNR $= 0$ dB.

Algorithm 3: Subspace Averaging with Majority Vote Approach.

Input: $\hat{\mathbf{R}}$, L;
Output: Order estimate $\hat{k}_{SA-MajVote}$
Initialization: $\boldsymbol{K} = \mathbf{0}_{d_{max}\times 1}$
for $t = 1 \ldots d_{max}$ **do**
  for $d = 1, \ldots, D$ **do**
    Extract $\hat{\mathbf{R}}_d$ from $\hat{\mathbf{R}}$ and obtain $\hat{\mathbf{R}}_d = \mathbf{U}_d \boldsymbol{\Sigma}_d \mathbf{U}_d^H$
    Compute the projection matrices $\mathbf{P}_{dt} = \mathbf{V}_{dt}\mathbf{V}_{dt}^H$, where $\mathbf{V}_{dt} = [\mathbf{u}_{d,1}, \ldots, \mathbf{u}_{d,t}]$
  Compute $\overline{\mathbf{P}}$ and its eigenvalues $(k_1, \ldots, k_L)$
  Estimate $\hat{k}$ as the number of eigenvalues of $\overline{\mathbf{P}}$ larger than 1/2
  $\boldsymbol{K}(t) = \hat{k}$
Select $\hat{k}_{SA-MajVote}$ as the majority decision from the collection of all estimated $\hat{k}$ in $\boldsymbol{K}$

3.3 Extension to Non-Uniform Linear Arrays: Sparse Subspace Averaging

The method proposed in previous section is designed to perform in the presence of non-white noises. However, this method still exploits the shift invariance property, and hence, it is not suitable for non-uniform linear arrays. In this section, we extend the SA method for arrays with arbitrary geometries, which is referred to as Sparse Subspace Averaging (SSA). We assume K narrowband signals impinging on an array of arbitrary geometry composed of M antennas. The received signal is

$$\mathbf{x}[n] = \mathbf{H}_s \mathbf{s}[n] + \mathbf{e}[n], \tag{3.3}$$

where $\mathbf{H}_s$ is the $M \times K$ unknown multiple input multiple output (MIMO) channel. The signals are assumed to be uncorrelated. Here as well, we do not assume any particular structure (scale identity, diagonal) for the noise covariance matrix $\mathbf{R}_n$.

Similar to the previous section and from (3.3), the covariance matrix of the received signal is

$$\mathbf{R} = \mathbf{H}_s \boldsymbol{\Psi} \mathbf{H}_s^H + \mathbf{R}_n$$

Note that since the MIMO channel $\mathbf{H}_s$ is unknown, we can assume without loss of generality that $\boldsymbol{\Psi} = \mathbf{I}$.

The key idea of subspace averaging for order estimation is to average subspaces that contain correlated (or ideally identical) versions of the signal subspace, but uncorrelated portions of the noise subspace. For ULAs, these subspaces are obtained from consecutive subarrays by exploiting the shift invariance property, as discussed in Chapter 2. When the MIMO channel $\mathbf{H}_s$ has no particular structure, however, a different approach to generate these subspaces is needed. The SSA method generates the subspaces from the sparse reconstructions of each snapshot, as we describe in the following subsections.

3.3.1 Sparse Representation

In a noiseless situation a basis for the K-dimensional signal subspace can be constructed from K randomly chosen columns of the data matrix $\mathbf{X}$. This means that each snapshot can be represented as linear combination of other K snapshots that lie in the same subspace. Note that the K snapshots will be linearly independent with probability one. In a noisy situation, such an expansion will not exist in general, but still we can find a sparse representation of each snapshot in terms of the rest of snapshots. The snapshots selected by the sparse reconstruction algorithm will allow us to build (or generate) approximate basis for the signal subspace. Note also that the number of sources is typically much smaller than the number of snapshots $K \ll N$. This is the idea behind the sparse reconstruction stage in SSA. Similar ideas have successfully been applied to sparse subspace clustering problems in [Elhamifar and Vidal, 2009].

The sparse expansion for each snapshot is obtained by solving

$$\min_{\mathbf{c}_n} \|\mathbf{c}_n\|_0 \quad \text{subject to} \quad \mathbf{x}[n] = \mathbf{X}_n \mathbf{c}_n \tag{3.4}$$

where $\mathbf{X}_n$ is the data matrix $\mathbf{X}$ after removing the nth column/snapshot and $\mathbf{c}_n = [c_{n,1}, \ldots, c_{n,N-1}]^T$ is the sparse representation for nth snapshot that provides the sparsest linear combination. However, the optimization problem in (3.4) is NP-hard. One well-known solution to this problem is to replace the ℓ_0-norm with the ℓ_1-norm [Donoho, 2006], which can be solved by the least absolute shrinkage and selection operator (LASSO) optimization algorithm [Tibshirani, 1996]. Nevertheless, the ℓ_1-norm surrogate function can sometimes perform poorly and other proxies for the ℓ_0-norm are preferred, such as the logarithm of the absolute value [Candès et al., 2008]. Moreover, the equality constraint in (3.4) is not adequate for low signal-to-noise-ratio (SNR) scenarios. Hence, here the SR will be obtained as the solution to

$$\min_{\mathbf{c}_n} \|\mathbf{x}[n] - \mathbf{X}_n \mathbf{c}_n\|_2^2 + \mu_s \sum_{i=1}^{N-1} \log(|c_{n,i}| + \delta) \tag{3.5}$$

which is the Lagrangian form of (3.4) where we have substituted the ℓ_0-norm by the log-surrogate and the equality by an inequality. Here, μ_s is a regularization parameter and δ is a small constant to avoid numerical issues. Since (3.5) is not convex, we propose to use an MM-based approach [Sun et al., 2017]. The majorizer of the cost function in (3.5) is based on a first-order Taylor series of the logarithm, that is,

$$\begin{aligned}\sum_{i=1}^{N-1} \log(|c_{n,i}| + \delta) = \sum_{i=1}^{N-1} \log\left(\sqrt{|c_{n,i}|^2} + \delta\right) \\ \leq \sum_{i=1}^{N-1} \frac{1}{2|c_{n,i}^{(t)}|^2 + 2\delta|c_{n,i}^{(t)}|} \left(|c_{n,i}|^2 - |c_{n,i}^{(t)}|^2\right)\end{aligned}$$

where $c_{n,i}^{(t)}$ is the ith component of the solution at the tth iteration. Hence, the solution at the $(t+1)$th iteration is computed from

$$\min_{\mathbf{c}_n} \|\mathbf{x}[n] - \mathbf{X}_n \mathbf{c}_n\|_2^2 + \mu_s \mathbf{c}_n^H \mathbf{D}_s \mathbf{c}_n, \tag{3.6}$$

where $\mathbf{D}_s = \text{diag}\,(d_1, \ldots, d_N)$, with

$$d_i = \frac{1}{2\left|c_i^{(t)}\right|^2 + 2\delta\left|c_i^{(t)}\right| + \delta}$$

where a small value δ has been added to the denominator to avoid divisions by zero. Now, we can find a closed-form solution to (3.6) by taking the derivative and setting it to zero, which yields

$$\mathbf{c}_n^{(t+1)} = \left(\mathbf{X}_n^H \mathbf{X}_n + \mu_s \mathbf{D}_s\right)^{-1} \mathbf{X}_n^H \mathbf{x}[n]. \tag{3.7}$$

3.3.2 Random Generation of Subspaces

In this subsection we develop a method to randomly generate subspaces based on the SR obtained in Section 3.3.1, which follows the lines of the method presented in Section 2.4.3. In noisy conditions, it is important for the SA method that the extracted subspaces contain a large common portion of the signal subspace and independent portions of the noise subspace. This translates into an averaging procedure that enhances signal coordinates while averaging out noise coordinates. Concretely, the subspace associated to the nth snapshot is constructed by randomly selecting columns from $\mathbf{X}_n$, with probabilities proportional to the sparse coefficients $\mathbf{c}_n$ as

$$\gamma_{n,i} = \frac{|c_{n,i}|}{\sum_{j=1}^{N-1} |c_{n,j}|}. \tag{3.8}$$

This process can be interpreted as sampling from the discrete distribution $\mathcal{D}(\mathbf{X}_n, \boldsymbol{\gamma}_n)$, where $\boldsymbol{\gamma}_n = [\gamma_{n,1}, \ldots, \gamma_{n,N-1}]^T$ in (3.8) are the concentration parameters of the distribution.

Finally, the subspaces are the column spaces of the matrices iteratively constructed as shown in Algorithm 4. This algorithm considers two thresholds: 1) b_{max}, which is the maximum dimension of the signal subspace (in practice an overestimation of K), and is selected as the number of $|c_{i,n}|$ greater than $\mu_c \max_i(|c_{i,n}|)$, with $0 < \mu_c < 1$ (we used in the simulations $\mu_c = 0.1$); and 2) T_γ, which is a minimum value for the probabilities in γ, and is chosen as $T_\gamma = \frac{1}{2}\sum_{i=1}^{N-1} |c_{n,i}|$. Note that, if b_{max} is high enough, the method is not sensitive to its value.

3.3.3 Sparse Subspace Averaging Criterion

The random generation procedure can be repeated T times to generate T subspaces for each value of $\mathbf{c}_n, n = 1, \ldots, N$. Therefore, we get a total of NT orthogonal projection matrices to be used in the SA procedure. The average projection matrix is

$$\overline{\mathbf{P}} = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \mathbf{P}_{nt}, \tag{3.9}$$

Algorithm 4: Generation of a Random Subspace for SSA

Input: $\mathbf{X}_n$, $\mathbf{c}_n$, b_{max}, T_γ
Initialization: $\tilde{\mathbf{X}}_n = \mathbf{X}_n$, $\mathbf{A}_{nt} = \emptyset$, compute $\boldsymbol{\gamma}_n$ from (3.8), and $M = N - 1$
while $\text{rank}(\mathbf{A}_{nt}) \leq b_{max}$ *or* $\sum_i \gamma_{n,i} \geq T_\gamma$ **do**
  `/* Sample from` $\mathcal{D}(\tilde{\mathbf{X}}_n, \boldsymbol{\gamma}_n)$ `*/`
  $\mathbf{g} = [g_i, \ldots, g_M]^T$, with $g_i \sim \mathcal{U}(0,1)$
  $\mathcal{I} = \{i \mid g_i \leq \gamma_{n,i}\}$ and $\overline{\mathcal{I}} = \{i \mid g_i > \gamma_{n,i}\}$
  $\mathbf{G} = \tilde{\mathbf{X}}_n(:, \mathcal{I})$
  `/* Append new submatrix */`
  $\mathbf{A}_{nt} = \begin{bmatrix} \mathbf{A}_{nt} & \mathbf{G} \end{bmatrix}$
  `/* Eliminate selected directions */`
  $\tilde{\mathbf{X}}_n = \tilde{\mathbf{X}}_n(:, \overline{\mathcal{I}})$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\overline{\mathcal{I}})$, and $M = |\overline{\mathcal{I}}|$
Output: Projection matrix onto the subspace $\mathbf{P}_{nt} = \mathbf{A}_{nt}(\mathbf{A}_{nt}^H \mathbf{A}_{nt})^{-1} \mathbf{A}_{nt}^H$

Algorithm 5: Sparse Subspace Averaging

Input: $\mathbf{X}$, μ_s, d_{max};
Output: Order estimate $\hat{k}_{SSA}$
for $n = 1, \ldots, N$ **do**
  Compute $\hat{\mathbf{c}}_n$ by solving (3.7)
  Select b_{max} and T_γ as suggested in Section 3.3.2
  Generate T random submatrices from $\mathbf{X}_n$ using Algorithm 4
Compute $\overline{\mathbf{P}}$ using (3.9)
Obtain $\hat{k}_{SSA}$ using (3.10)

Once we get $\overline{\mathbf{P}}$, its eigenvalues $1 \geq k_1 \geq k_2 \geq \cdots \geq k_{d_{max}}$ are used to determine number of sources as

$$\hat{k}_{SSA} = \underset{1 \leq k \leq d_{max}}{\text{argmax}}\, k_k - k_{k+1}. \tag{3.10}$$

Basically, the criterion in (3.10) detects the gap between the signal and noise eigenvalues. It is noted here that to avoid numerical issues and to save computational cost, we are considering only the largest d_{max} eigenvalues of $\overline{\mathbf{P}}$, where $d_{max} \gg K$ is an overestimation of K. A summary of the proposed method is shown in Algorithm 5.

3.4 Simulation Results

In this chapter, we proposed two extensions of the SA method. The first extension is designed to work with ULAs in the presence of non-white noises, and the second extension considers arrays with arbitrary geometries. In this section, the performance of both methods is compared with some representative techniques that exist in the literature.

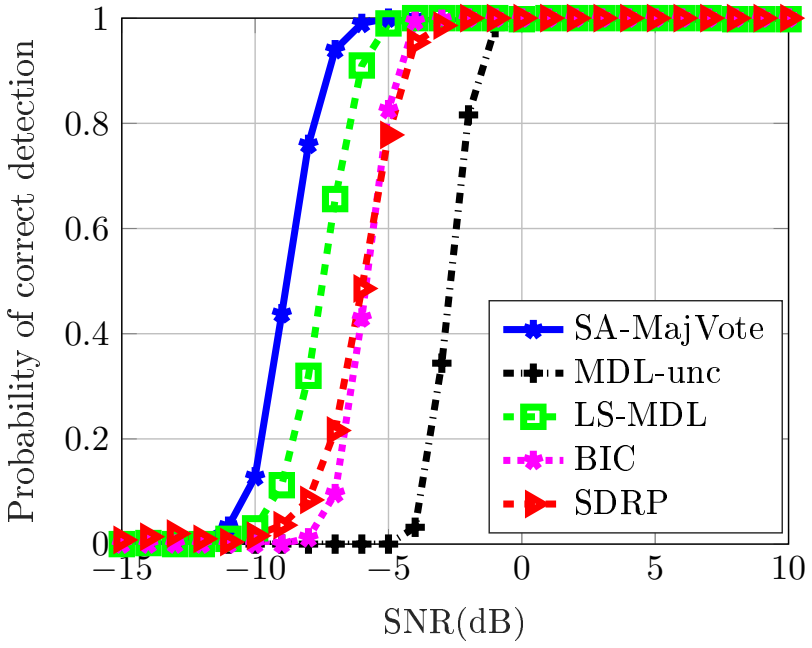


Figure 3.2: Probability of correct detection vs. SNR for ULA and noise with diagonal covariance matrix when $\epsilon_d = 0.4$, $M = 100$, $N = 50$, $K = 3$ and $\Delta_\theta = 2^\circ$.

3.4.1 SA with Majority Vote

This subsection studies the performance of SA method combined with the majority vote approach, and hence denoted as SA-MajVote here. For comparison we use the following methods: the LS-MDL [Huang and So, 2013], the BIC for large-scale arrays [Huang et al., 2016], the MDL for diagonal noise (for noise model with diagonal noise covariance matrix, denoted here as MDL-unc), the SDRP [Eguizabal et al., 2019], the CCA method in [Song et al., 2016], and the VTRS method in [Jiang and Ingram, 2004]. These methods have been briefly described in Section 1.2

We consider a scenario with K narrowband incoherent unit-power signals and DOAs separated Δ_θ impinging on a uniform linear array with M antennas and half-wavelength element separation. For the SA method, we use $L = M - 5$ as the subarray size, and $d_{max} = \frac{M}{5}$. The signal-to-noise-ratio is $\text{SNR} = 10\log\frac{\text{tr}(\mathbf{A}\boldsymbol{\Psi}\mathbf{A}^H)}{\text{tr}(\mathbf{R}_n)}$, where $\text{tr}()$ denotes trace. Average of 500 independent simulations is used to represent each curves.

For the first three experiments we assume uncorrelated noises across antennas so that the noise covariance matrix is diagonal with unknown elements, which is described as Noise Model 2 in Section 1.1.

Experiment 1: In the first experiment we consider an array with $M = 100$ antennas, $K = 3$ sources with electrical angle separation $\Delta_\theta = 2^\circ$, and $N = 50$ snapshots. The noise is drawn according to noise model 2 with $\epsilon_d = 0.4$. Fig. 3.2 shows the

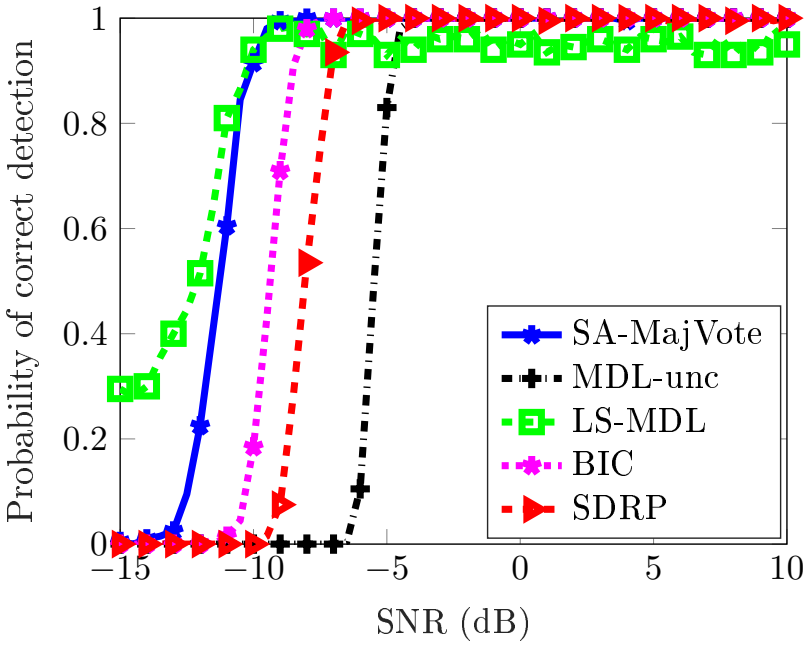


Figure 3.3: Probability of correct detection vs. SNR for ULA and noise with diagonal covariance matrix when $\epsilon_d = 0.4$, $M = 100$, $N = 150$, $K = 3$ and $\Delta_\theta = 2^\circ$.

probability of correct detection vs SNR. Although the LS-MDL and the BIC methods assume i.i.d. noises, in this scenario with very few snapshots their performance is rather robust against a mismatched model. The MDL-unc and the SDRP perform competitively as well. Nevertheless, the SA-MajVote outperforms the other methods.

Increasing the number of snapshots to $N = 150$ and keeping fixed the rest of the parameters, we obtain the results shown in Fig. 3.3. As N increases, the performance of LS-MDL improves, however, it shows an inconsistent behaviour at high SNR due to non-whiteness of the noise. SA-MajVote outperforms competing methods here as well.

Experiment 2: Fig. 3.4 shows the results when ϵ_d varies (recall that for $\epsilon_d = 0$ noise is spatially white). For large values of ϵ_d, the performance of both the LS-MDL and the BIC criteria degrade, whereas the MDL-unc specifically designed for this model performs better. The SDRP does not perform consistently for this scenario and its performance further degrades for higher values of ϵ_d. Interestingly, the SA-MajVote provides accurate estimates across the whole range of ϵ_d.

Experiment 3: In the next experiment, we assume two sources with unequal powers P_1 and P_2 respectively. We assume that initially sources have equal power ($P_1 = P_2$), and then, we fix the P_1 and decrease the P_2 gradually, and the probability of correct detection vs. the ratio of two powers $\frac{P_2}{P_1}$ is shown in Fig. 3.5. We consider an array with $M = 100$ antennas, the number of samples $N = 60$, the signal-to-noise-ratio is SNR $= -10$ dB, the electrical angle separation $\Delta_\theta = 10^\circ$, and $\epsilon_d = 0.6$. LS-MDL is not consistent in this very example. In addition, as the P_2 decreases, all

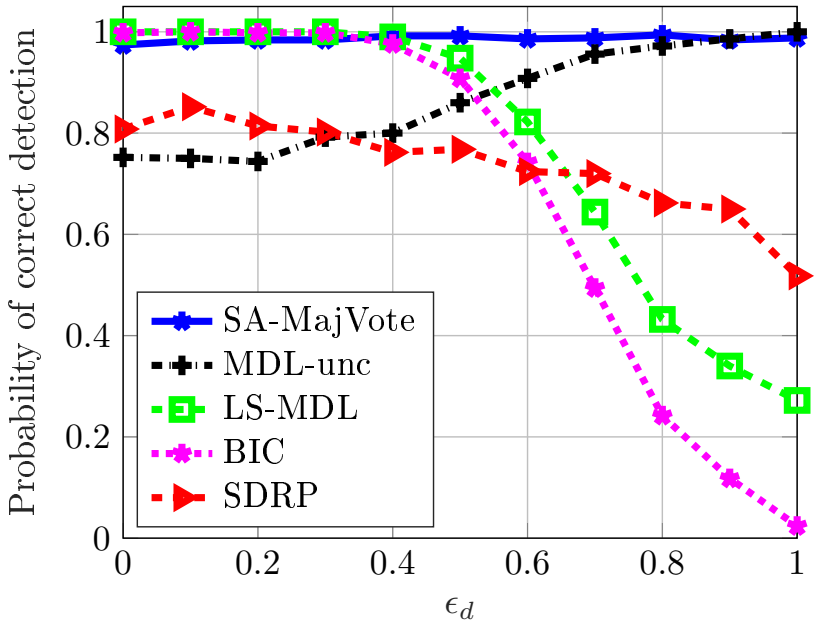


Figure 3.4: Probability of correct detection vs. ϵ_d for ULA and noise with diagonal covariance matrix when $M = 50$, $N = 80$, $K = 3$, SNR $= -9$ dB and $\Delta_\theta = 10°$.

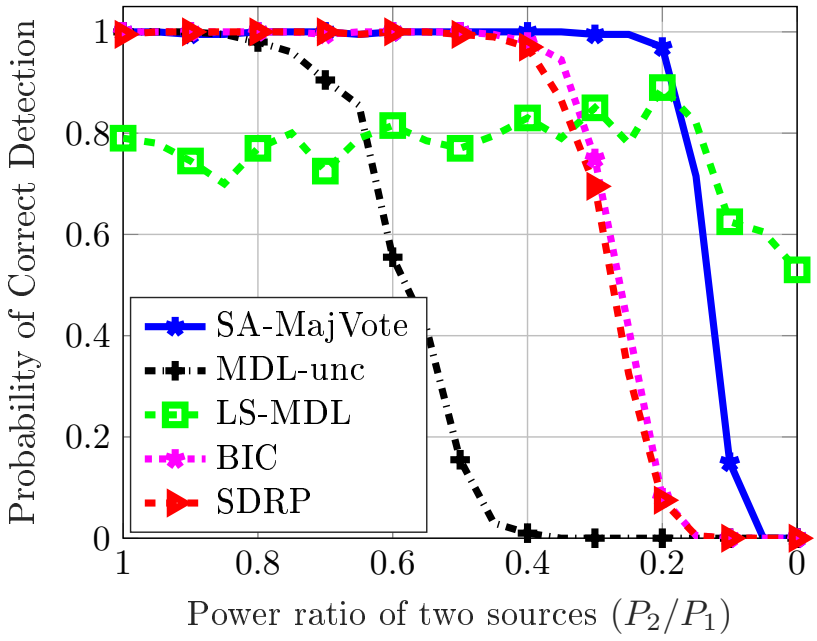


Figure 3.5: Probability of correct detection vs. power ratio of two sources $\frac{P_2}{P_1}$ for ULA and noise with diagonal covariance matrix when $\epsilon_d = 0.6$, $M = 100$, $N = 80$, $K = 2$, SNR $= -10$ dB and $\Delta_\theta = 10°$.

other methods start underestimating the number of sources, however, SA-MajVote performs well for small values of ratio $\frac{P_2}{P_1}$.

For the experiments 4 and 5, we assume that the noise has an arbitrary psd covariance matrix $\mathbf{R}_n \succ 0$ (Noise model 4 in Section 1.1). The performance of the proposed method for this noise model is compared with the CCA, the VTRS, and the SDRP methods.

Experiment 4: For this experiment we consider an array with $M = 80$ sensors, $K = 3$ sources separated $\Delta_\theta = 5°$, and the number of snapshots is $N = 150$. We observe in Fig. 3.6 that the CCA method does not perform consistently in this scenario, and the SDRP method requires high SNR to perform satisfactorily. For the CCA method we use two equal-sized subarrays of 40 antennas each. The assumed noise model for the CCA criterion is mismatched to noise model 4, so, not surprisingly, its estimates are not always good. The proposed SA-MajVote method starts detecting the correct number of sources from very low SNR, and therefore it is a clear winner here. Now, if we decrease the number of snapshots to $N = 50$, the performances of the CCA and the SDRP degrade further, as it can be observed in Fig. 3.7. In the small sample regime, only SA-MajVote and VTRS perform well at low SNR, however, SA-MajVote outperforms the competing methods for this scenario as well.

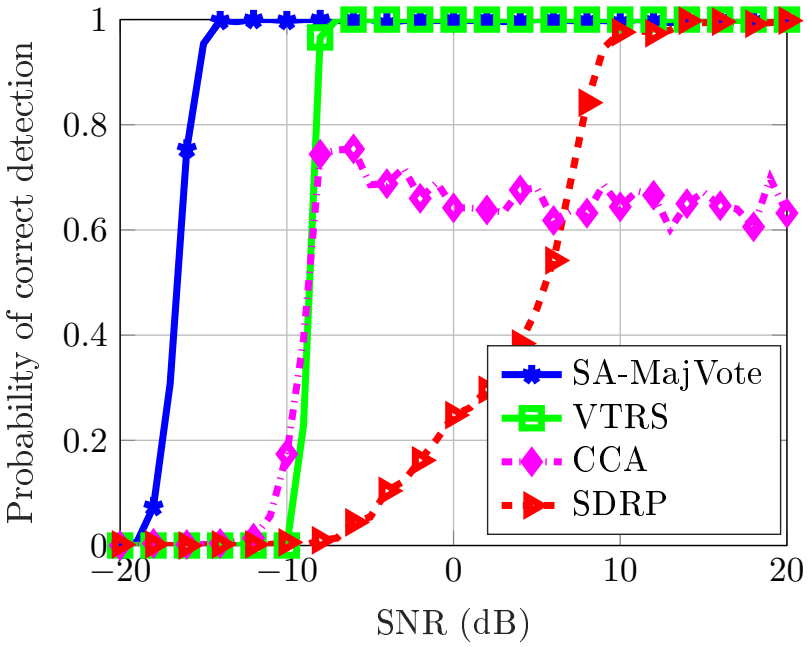


Figure 3.6: Probability of correct detection vs. SNR for ULA and an arbitrary noise model when $M = 80$, $N = 150$, $K = 3$, and $\Delta_\theta = 5°$.

Experiment 5: In the last experiment we consider an array with $M = 100$ sensors, $K = 4$ sources separated $\Delta_\theta = 10°$, the SNR is fixed, and the results is shown in Fig. 3.8. The number of snapshots varies between $N = 5$ and $N = 400$ for SA-MajVote and VTRS, and since CCA and SDRP do not perform well for small N, it is

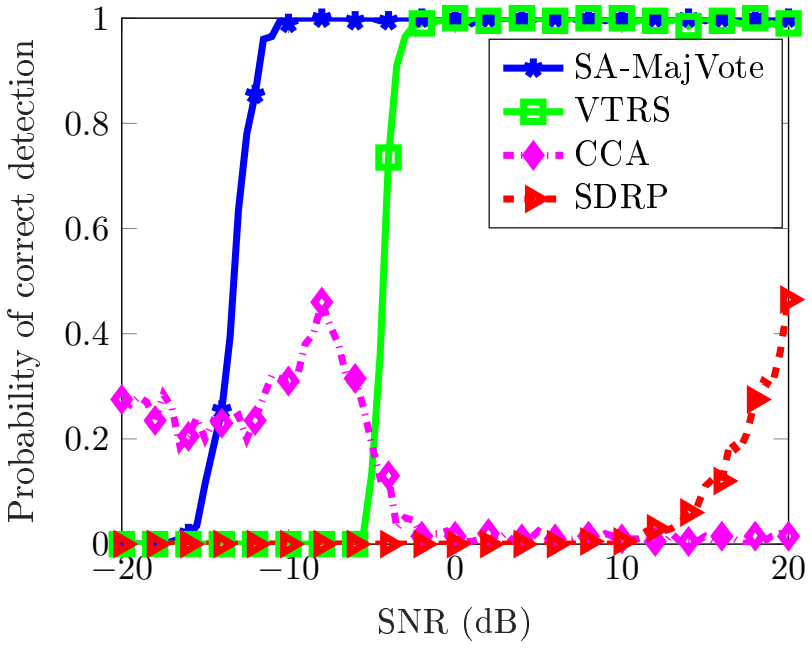


Figure 3.7: Probability of correct detection vs. SNR for ULA and an arbitrary noise model when $M = 80$, $N = 50$, $K = 3$, and $\Delta_\theta = 5^\circ$.

varied between $N = 20$ and $N = 400$ for these two methods. We observe that the VTRS method requires at least $N = 300$ snapshots to provide accurate estimates, and the SDRP requires even more snapshots to perform well at low SNR. For the CCA method we use two equal-sized subarrays of 50 antennas each and for this scenario as well, its estimates are not always good. Finally, the proposed SA method with majority vote (SA-MajVote) detects the correct number of sources with high accuracy with only $N = 50$ snapshots.

3.4.2 Sparse Subspace Averaging (SSA)

In this subsection, the performance of the SSA method is compared with some representative methods for different noise models. For all simulations we assume that K uncorrelated narrowband equal-power signals are impinging on an array with arbitrary geometry with M antennas. The signal-to-noise-ratio is defined as $\text{SNR} = 10 \log \frac{\text{tr}(\mathbf{H}_s \mathbf{\Psi} \mathbf{H}_s^H)}{\text{tr}(\mathbf{R}_n)}$, and 2000 Monte Carlo simulations are averaged for all the results. The parameters of the proposed method are selected as follows: $\mu_s = \frac{M}{2N}$, $T = 20$, $d_{max} = \frac{M}{5}$ and $\delta = 10^{-15}$.

Experiment 1: In the first experiment, we assume that the noise is uncorrelated across antennas with different variances at each sensor, namely noise model 2 in Section 1.1. As methods for comparison, we select the LS-MDL, the BIC, the MDL-unc, and the SDRP methods. Fig. 3.9 shows the probability of correct detection vs. ϵ_d for $M = 70$, $N = 100$, $K = 5$, and $\text{SNR} = -2$ dB. In this example SSA

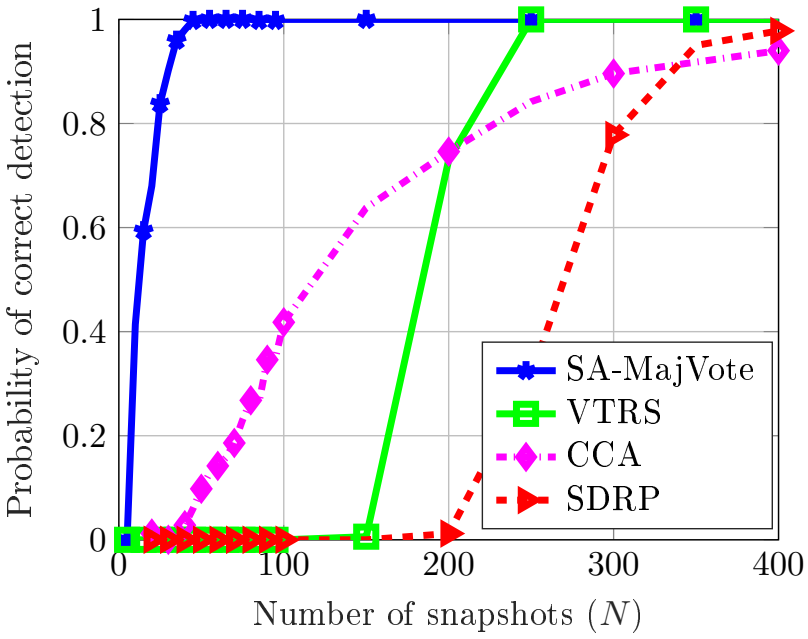


Figure 3.8: Probability of correct detection vs. N for ULA and an arbitrary noise model when $M = 100$, $K = 4$, SNR $= -10$ dB and $\Delta_\theta = 10^\circ$.

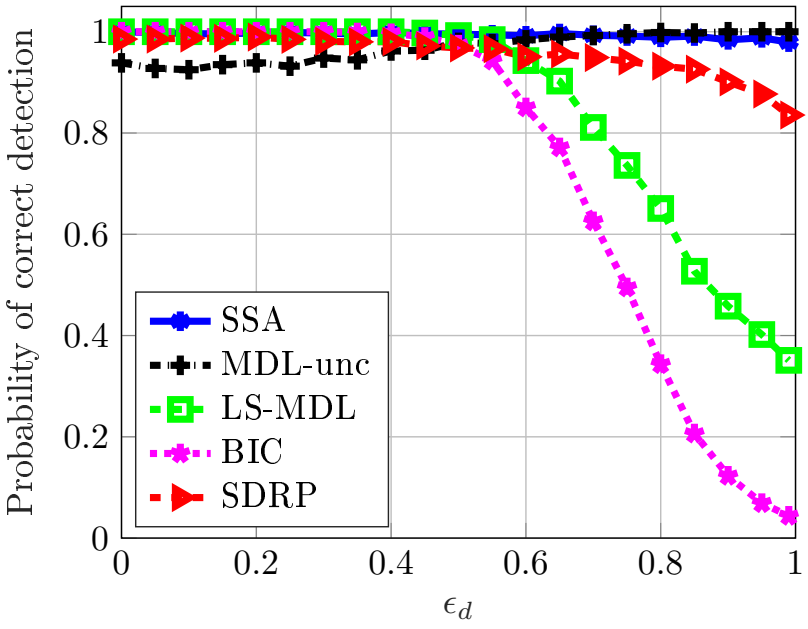


Figure 3.9: Probability of correct detection vs. ϵ_d for the array with arbitrary geometry in the presence of noise with diagonal covariance matrix in a scenario with $M = 70$, $N = 100$, $K = 5$, and SNR $= -2$ dB.

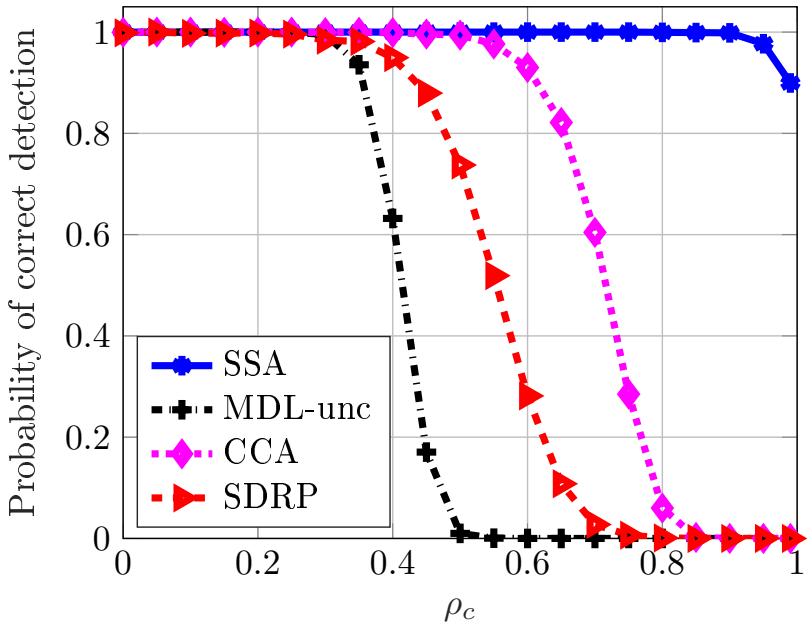


Figure 3.10: Probability of correct detection vs. ρ_c for the array with arbitrary geometry in the presence of exponentially correlated noise in a scenario with $M = 20$, $N = 80$, $K = 3$, and SNR $= 10$ dB.

provides robust and accurate detection results over the entire range of variation of ϵ_d. Since LS-MDL and BIC are originally designed for white noise, their performances degrade for higher values of ϵ_d. The performance of the SDRP also degrades for higher values of ϵ_d. Finally, MDL-unc provides good results for higher values of ϵ_d, but its performance degrades when the noise is nearly spatially white.

Experiment 2: In the second experiment, we consider an exponentially correlated noise model, namely noise model 3 in Section 1.1. Fig. 3.10 shows probability of correct detection vs. ρ_c for an array with $M = 20$ antennas, $N = 80$ snapshots, $K = 3$ sources, and SNR$= 10$ dB. For this noise model, we are comparing the results of MDL-unc, the SDRP and the CCA. As Fig. 3.10 suggests, the proposed SSA method again provides a robust solution for the whole range of ρ_c. All other methods fail when ρ_c increases.

Experiment 3: In the last experiment, we consider Gaussian noise with arbitrary unknown covariance matrix, namely noise model 4 in Section 1.1. The scenario has $M = 100$ antennas, $K = 4$ sources, SNR $= 15$ dB, and the number of snapshots N varies up to 100, so that we are in the small sample regime. Since CCA and SDRP do not usually provide good results with few snapshots, we are considering $N \geq 15$ for these methods, however, lower values of N are considered for SSA. As it is shown in Fig. 3.11, CCA does not perform well for this noise model in the small sample regime, while the SDRP provides good results for sufficiently large N. Finally, SSA performs satisfactorily even when N/M is around 0.1.

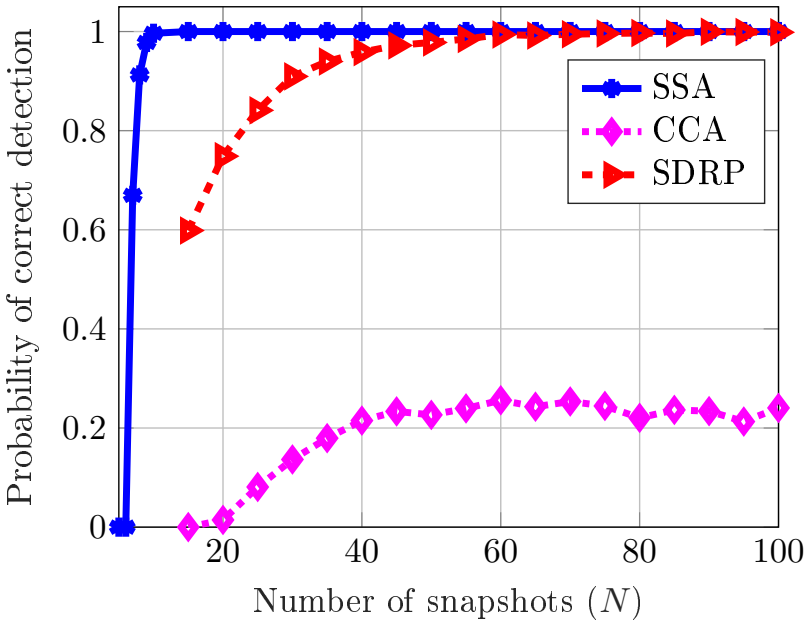


Figure 3.11: Probability of correct detection vs. N for the array with arbitrary geometry in the presence of an arbitrary noise model in a scenario with $M = 100$, $K = 4$, and SNR $= 15$ dB.

3.5 Conclusions

This chapter extended the SA method for different scenarios proposing two different methods. The first method addressed the problem of source enumeration for ULAs in the presence of non-white noises under the condition of low sample support. The method applies subspace averaging (SA) technique for increasing dimensions of the extracted subspaces. For each dimension, we get an estimate of the number of sources, and the final estimate is obtained by a majority vote rule. This method performs robustly under different noise models ranging from uncorrelated noises with different variances to arbitrarily correlated noises.

The second method presented an order estimation technique for arbitrary geometry arrays and noise with unknown spatial correlation. The key idea to generate a collection of subspaces to be averaged is obtaining a sparse representation of each snapshot as a linear combination of the others. To perform this sparse recovery (SR) problem, we proposed a generalization of the log-surrogate of the ℓ_0-norm, which is solved using the majorization-minimization approach. Then, based on the sparse coefficients of the reconstruction, a sampling mechanism is presented to obtain projection matrices that share a large common portion of the signal subspace. Finally, the eigenvalues of an average projection matrix are then used to estimate the number of sources. It is illustrated by some simulation examples that the proposed sparse subspace averaging (SSA) method performs robustly for a wide range of noise models in the small sample regime. Nevertheless, with the help of two proposed methods

in this chapter, the source enumeration problem can be solved for variety of different scenarios from ULA to arrays with arbitrary geometry in the presence of different noise models.

Part II

Order Estimation via Matrix Completion

Chapter 4

Order Estimation via Toeplitz Matrix Completion

In the part I of this dissertation, we discussed the subspace averaging approach to solve the problem of source enumeration for both uniform linear arrays and arrays with arbitrary geometries for different noise models. In part II, a completely different approach to source enumeration is discussed. Here we use the matrix completion (MC) techniques to develop order estimation criteria for: i) noises with diagonal covariance and ii) missing data scenarios. We assume a uniform linear array (ULA) in this part.

This chapter addresses the problem of source enumeration by an array of sensors in the presence of noise whose spatial covariance structure is a diagonal matrix with possibly different variances, when the sources are uncorrelated. The diagonal terms of the sample covariance matrix are removed and, after applying Toeplitz rectification as a denoising step, the signal covariance matrix is reconstructed by using a low-rank matrix completion method adapted to enforce the Toeplitz structure of the sought solution. The proposed source enumeration criterion is based on the Frobenius norm of the reconstructed signal covariance matrix obtained for increasing rank values. The proposed method performs robustly for both small and large-scale arrays with few snapshots, i.e. small-sample regime.

4.1 Introduction

Matrix completion refers to the problem of recovering missing entries of a partially observed low-rank matrix. In last few years, it has proved its usefulness in many applications of various fields such as compressed sensing [Candes and Wakin, 2008], recommendation system [Koren et al., 2009], medical resonance imaging [hu et al., 2018], large-scale network monitoring [Xie et al., 2015], traffic sensing [Du et al., 2015] and multi-task learning [Argyriou et al., 2008]. Many approaches of matrix completion [Keshavan et al., 2009, Ramlatchan et al., 2018, Li et al., 2019] have been proposed for various applications and scenarios. In this dissertation, the MC technique is modified to exploit the properties of ULA in the reconstruction process, which

results in finding a better estimate of the signal subspace, and therefore provides a consistent solution for order estimation in small sample regime and low signal-to-noise ratio scenarios.

Low-rank matrix completion methods have been used before in array signal processing problems. In [Pal and Vaidyanathan, 2014], for instance, MC is used for direction-of-arrival (DOA) estimation, when the number of sources exceeds the number of sensors. An iterative reweighted nuclear norm minimization method is used in [Tan and Feng, 2019] for DOA estimation with nested arrays. The case of non-iid noises is considered in [Liao et al., 2016], where matrix completion algorithms are used to reconstruct the zeroed entries of the sample covariance matrix along its diagonal. All these methods, however, address the DOA estimation problem and assume that the number of sources is known. Differently from these works, here we exploit the MC algorithms to devise novel order estimation criteria.

This chapter presents a method for enumerating sources for ULAs in the presence of Gaussian noise with diagonal covariance matrix (Noise Model 2), that performs satisfactorily in a wide number of scenarios, including the small-sample regime both with large and short arrays, and low signal-to-noise-ratio environments. The method obtains an estimate of the signal covariance matrix that is i) Toeplitz, and ii) low-rank. The idea behind proposed method is that since noise covariance matrix is diagonal, only diagonal entries of sample covariance matrix would get affected by noise. Therefore, we can remove noisy diagonal entries of SCM, and recover them from non-noisy off-diagonal entries by using a matrix completion method. To this end, we use a low-rank matrix completion method that includes an additional regularization term to enforce a Toeplitz structure in the solution. The MC algorithm takes as input the denoised version of the SCM where the elements along the main diagonal have been removed. An additional denoising process known as Toeplitz rectification [Vallet and Loubaton, 2017, Forster, 2001] is used to find a better estimate of the signal covariance matrix. Finally, we propose an order estimation criterion which is based on the Frobenius norm of the matrices reconstructed for increasing rank values.

The structure of this chapter is as follows. Section 4.2 defines the problem statement for this chapter. The Toeplitz matrix completion (TMC) algorithm is proposed in Section 4.3, and then a novel order estimation criterion via TMC is proposed in Section 4.4. Section 4.5 first studies the performance of the TMC method and shows that a better estimated signal subspace is achieved by using TMC. Further, this section compares the performance of the proposed order estimation criterion with some representative methods. Finally the conclusions are summarized in Section 4.6.

4.2 Problem Statement

Let us consider K narrowband signals impinging on a uniform linear array with M antennas. For this chapter we consider that the noise is uncorrelated across antennas with diagonal noise covariance matrix. As mentioned in Chapter 1, the noise covariance matrix for a diagonal noise model can be defined as $\mathbf{R}_n = \text{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2\right)$,

where $\sigma_m^2 \sim \mathcal{U}[\sigma^2(1-\epsilon_d), \sigma^2(1+\epsilon_d)]$ is the noise variance at the mth sensor, and $0 \leq \epsilon_d \leq 1$ allows us to control the spatial non-whiteness of the noise. $\epsilon_d = 0$ would provide the white noise with covariance matrix $\mathbf{R}_n = \sigma^2\mathbf{I}$. Therefore, the covariance matrix for the observations is

$$\mathbf{R} = \mathbf{R}_s + \mathbf{R}_n, \tag{4.1}$$

where $\mathbf{R}_s = \mathbf{A}\mathbf{\Psi}\mathbf{A}^H$ is the signal covariance matrix.

The objective of this chapter is to design an order estimation criterion based on MC technique, which can estimate K from $\mathbf{X}$ or $\hat{\mathbf{R}}$. The proposed order determination criterion builds upon the reconstruction of the signal covariance matrix $\mathbf{R}_s$ in (4.1) for increasing values of its rank. To this end, we apply a matrix completion (MC) approach, which is described in the following section. Then, we introduce the proposed order fitting rule, which is based on the Frobenius norm of the reconstructed matrices.

4.3 Toeplitz Matrix Completion

The signal covariance matrix $\mathbf{R}_s$ is Toeplitz and low-rank, which can be verified as follows. Since signals are uncorrelated, $\mathbf{\Psi}$ is diagonal and can be defined as $\mathbf{\Psi} = \text{diag}(\Psi_1, \ldots, \Psi_K)$, where Ψ_k denotes the power of the kth source. Also, as mentioned in Chapter 1, the steering matrix for ULAs is $\mathbf{A} = [\mathbf{a}(\theta_1) \quad \cdots \quad \mathbf{a}(\theta_K)]$, where $\mathbf{a}(\theta_k) = \begin{bmatrix} 1 & e^{-j\theta_k} & e^{-j\theta_k(M-1)} \end{bmatrix}^T$. Therefore, the entry in the mth row and nth column of $\mathbf{R}_s$ is

$$\mathbf{R}_s(m,n) = \sum_{k=1}^{K} \Psi_k \mathbf{A}(k,m)\mathbf{A}^H(k,n) = \sum_{k=1}^{K} \Psi_k e^{-j(m-n)\theta_k}, \tag{4.2}$$

and it can easily be verified from (4.2) that

$$\begin{aligned} &\mathbf{R}_s(m,n) = \mathbf{R}_s(m+1,n+1), \text{ and} \\ &\mathbf{R}_s(m,n) = \mathbf{R}_s^*(n,m) \end{aligned}$$

which are the conditions for a Toeplitz matrix. Here, $(\cdot)^*$ denotes the complex conjugate. In addition, since $\mathbf{A} \in \mathbb{C}^{M\times K}$ and $\mathbf{\Psi}$ is a $K \times K$ positive semidefinite matrix, $\mathbf{R}_s = \mathbf{A}\mathbf{\Psi}\mathbf{A}^H$ is a rank-K $M \times M$ matrix. However, these two properties are not fulfilled by the sample covariance matrix $\hat{\mathbf{R}}$, which is symmetric (but non-Toeplitz) and full-rank. In addition, since $\mathbf{R}_n$ in (4.1) is diagonal, the off-diagonal terms of $\mathbf{R}$ are unaffected by the noise covariance matrix. Therefore, the diagonal entries of $\hat{\mathbf{R}}$ will be more affected by the noise. These noiser diagonal entries of $\hat{\mathbf{R}}$ can be removed and matrix completion algorithms can then be used to reconstruct the low-rank signal covariance matrix from the off-diagonal terms of $\hat{\mathbf{R}}$. This process is explained step by step in the following subsections.

4.3.1 Matrix Completion for Signal Covariance Matrix Estimation

Let $\Omega = \{(i,j) : i \neq j, i = 1,\dots,M, j = 1,\dots,M\}$ be the set of indices for the off-diagonal entries of $\hat{\mathbf{R}}$. According to [Candes and Plan, 2010], we can recover $\mathbf{R}_s$ by solving

$$\begin{aligned} \min_{\mathbf{R}_s \in \mathbb{C}^{M\times M}} \quad & ||\mathbf{R}_s||_* \\ \text{subject to} \quad & ||P_\Omega(\mathbf{R}_s - \hat{\mathbf{R}})||_F \leq \eta \end{aligned} \tag{4.3}$$

where P_Ω denotes the projection operator that sets to zero the entries with indices not belonging to Ω and leaves the rest unchanged, $||\mathbf{R}_s||_*$ denotes the nuclear norm of $\mathbf{R}_s$, and $\eta > 0$ is a tolerance parameter that limits the fitting error.

Matrix completion can be performed here by means of matrix factorization [Srebro et al., 2005]. Since the signal covariance matrix is Hermitian, we can factorize it as $\mathbf{R}_s = \mathbf{W}\mathbf{W}^H$, where $\mathbf{W} \in \mathbb{C}^{M\times p}$ and p is a fixed value that limits the rank of the reconstructed matrix. Then, using the identity [Srebro et al., 2005]

$$||\mathbf{R}_s||_* = \min_{\mathbf{R}_s = \mathbf{W}\mathbf{W}^H} ||\mathbf{W}||_F^2,$$

$\mathbf{R}_s$ can be estimated by solving the following optimization problem

$$\hat{\mathbf{W}} = \underset{\mathbf{W}\in\mathbb{C}^{M\times p}}{\operatorname{argmin}} \left\| P_\Omega\left(\hat{\mathbf{R}} - \mathbf{W}\mathbf{W}^H\right)\right\|_F^2 + \mu\left\|\mathbf{W}\right\|_F^2 \tag{4.4}$$

where μ is a regularization parameter. Note that the term weighted by μ regulates the nuclear norm of the solution, hence leading to a sparse eigenvalue distribution as μ grows.

4.3.2 Toeplitz Rectification

Due to the fact that we are dealing with a limited number of snapshots, the nonzero entries in $P_\Omega(\hat{\mathbf{R}})$ are still noisy [Vallet and Loubaton, 2017] and thus recovering the full $\mathbf{R}_s$ via (4.4) might yield unreliable estimates in the low-sample regime. Therefore, we propose to use a denoising step called Toeplitz rectification [Vallet and Loubaton, 2017, Forster, 2001, Vinogradova et al., 2019] before applying matrix completion, and enforce the Toeplitz structure in the reconstruction process.

An unbiased estimator of sample covariance with Toeplitz structure is obtained by averaging its entries along each subdiagonal as [Vinogradova et al., 2019]

$$\hat{\mathbf{R}}_{toep}(i,j) = \frac{1}{M - |i-j|} \sum_{l-m=i-j} \hat{\mathbf{R}}(l,m). \tag{4.5}$$

This process reduces the noise in the off-diagonal terms and provides a better reconstruction of the signal covariance matrix.

4.3.3 Find Optimization Problem

After Toeplitz correction, $\hat{\mathbf{R}}$ in (4.4) is replaced by the less noisy $\hat{\mathbf{R}}_{toep}$. Also, since $\mathbf{R}_s$ is Toeplitz, the solution $\hat{\mathbf{R}}_{mc} = \hat{\mathbf{W}}\hat{\mathbf{W}}^H$ should also be Toeplitz. To this end, a Toeplitz constraint is enforced in (4.4). Therefore, (4.4) can be written as:

$$\begin{aligned}\hat{\mathbf{W}} = \underset{\mathbf{W}\in\mathbb{C}^{M\times p}}{\operatorname{argmin}} &\left\| P_\Omega\left(\hat{\mathbf{R}}_{toep} - \mathbf{W}\mathbf{W}^H\right)\right\|_F^2 + \mu\|\mathbf{W}\|_F^2 \\ &\text{s.t.} \quad \mathbf{W}\mathbf{W}^H \in \mathcal{T}\end{aligned} \tag{4.6}$$

where the constraint restricts the solution to a set $\mathcal{T}$, which we define as the set of Toeplitz matrices in $\mathbb{C}^{M\times M}$. An approximate solution to (4.6) can be obtained by introducing the Toeplitz constraint in the form of an additional regularization term as follows

$$\begin{aligned}\hat{\mathbf{W}} = \underset{\mathbf{W}\in\mathbb{C}^{M\times p}}{\operatorname{argmin}} &\sum_{(i,j)\in\Omega} \left|\hat{\mathbf{R}}_{toep}(i,j) - \mathbf{w}_i^H\mathbf{w}_j\right|^2 + \mu\sum_{i=1}^{M}\|\mathbf{w}_i\|_2^2 \\ &+ \alpha\sum_{m=0}^{M-1}\sum_{i=1}^{M-m-1}\left|\mathbf{w}_i^H\mathbf{w}_{i+m} - \mathbf{w}_{i+1}^H\mathbf{w}_{i+1+m}\right|^2\end{aligned} \tag{4.7}$$

where $\mathbf{w}_i^H$ is the ith row vector of $\mathbf{W}$, and α is a regularization scalar. The solution $\hat{\mathbf{R}}_{mc} = \hat{\mathbf{W}}\hat{\mathbf{W}}^H$ can be obtained by iteratively optimizing over each $\mathbf{w}_i$.

4.4 Order Estimation Criterion

The main insight for the proposed order estimation criterion is that, due to the eigenvalue sparsity enforced by the MC algorithm, as long as p (the rank used in the factorization) is larger than K, the reconstructed signal covariance matrix $\hat{\mathbf{R}}_{mc}$ should not change significantly. This intuition can be understood by an example which is corroborated in Fig. 4.1a which shows how the Frobenius norm of $\hat{\mathbf{R}}_{mc}$ changes with p. For this example we assume an array of $M = 100$ sensors with $N = 150$ snapshots, $\epsilon_d = 0.5$ and SNR $= -10$dB. The sources are separated by $\Delta_\theta = 10^\circ$. It can be observed that the norm grows until $p = K$, but once p exceeds K it is almost constant. This suggests to use the difference function $D(p)$ defined as

$$D(p) = \|\hat{\mathbf{R}}_{mc}(p)\|_F^2 - \|\hat{\mathbf{R}}_{mc}(p-1)\|_F^2 \tag{4.8}$$

where $\hat{\mathbf{R}}_{mc}(p)$ denotes the reconstructed signal covariance matrix at a particular value of p. Since $D(p)$ will take very small values for $p > K$, finding the position at which this sharp decline change occurs will yield an order estimate. To this end, we propose the following criterion

$$\hat{k}_{TMC} = \underset{1\le p\le p_{max}}{\operatorname{argmax}} \frac{D(p)}{D(p+1)+\delta} \tag{4.9}$$

Algorithm 6: Order Estimation using TMC

Input: $\hat{\mathbf{R}}$, μ, α, p_{max};
Output: Order estimate $\hat{k}_{TMC}$
Find $\hat{\mathbf{R}}_{toep}$ using (4.5)
for $p = 1, \ldots, p_{max}$ **do**
  Find $\hat{\mathbf{R}}_{mc}(p-1)$ and $\hat{\mathbf{R}}_{mc}(p)$ using (4.7)
  Find $D(p)$ using (4.8)
Estimate number of sources as $\hat{k}_{TMC} = \underset{1 \le p \le p_{max}}{\operatorname{argmax}} \frac{D(p)}{D(p+1)+\delta}$

where p_{max} is an overestimation of K and δ is a small constant to avoid numerical issues. Fig. 4.1b shows $\frac{D(p)}{D(p+1)+\delta}$ for different K for the same example, and each line has been normalized by its maximum value to enhance visibility. We observe that the peaks are positioned at $p = K$, thus evidencing that the chosen criterion is adequate. A summary of the order estimation via Toeplitz matrix completion (TMC) algorithm is shown in Algorithm 6.

4.5 Simulation Results

4.5.1 Performance of the TMC

This subsection discusses the effect of removing diagonal entries from SCM and recovering them back by the TMC method described in Section 4.3. As a figure of merit we use the chordal subspace distance between the true signal subspace, and the estimated signal subspace to assess how different these subspaces are. For comparison, we include the performance of the following methods:

- SCM: The original sample covariance matrix $\hat{\mathbf{R}}$, without eliminating the diagonal terms.
- MC: The diagonal entries of SCM are eliminated and the standard MC algorithm solution given in (4.4) is used to reconstruct $\mathbf{R}_s$.
- TMC: The diagonal entries of SCM are eliminated and TMC algorithm given in (4.7) is used to reconstruct $\mathbf{R}_s$.

For this subsection we assume that K is known and exactly the K largest eigenvectors of the covariance matrices are used to determine the signal subspace in each case. Also, for the MC and the TMC, $p = K$ is considered. Further, for all simulations we assume that K uncorrelated narrowband signals with a separation of Δ_θ are impinging on a uniform linear array with M half-wavelength separated antennas. For TMC, we use $\mu = \frac{M}{2}$ and $\alpha = \frac{M}{10}$. The signal-to-noise-ratio is SNR$= 10 \log \frac{\mathrm{tr}(\mathbf{R}_s)}{\mathrm{tr}(\mathbf{R}_n)}$.

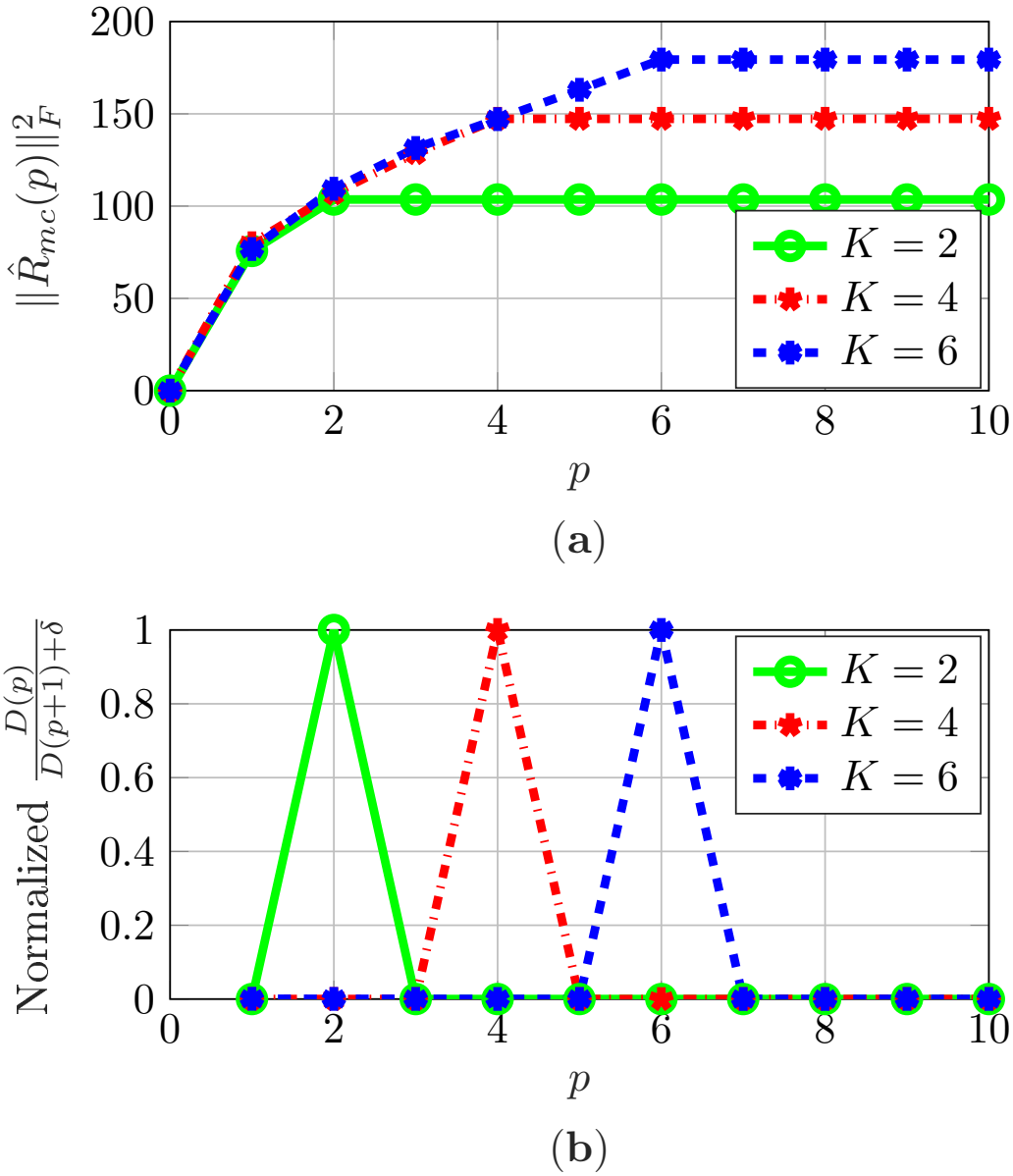


Figure 4.1: (a) $\|\hat{\mathbf{R}}_{mc}(p)\|_F^2$ vs. p, and (b) Normalized $\frac{D(p)}{D(p+1)+\delta}$ vs. p for $M = 100$, $N = 150$, $\epsilon_d = 0.5$, SNR $= -10$ dB and sources are separated by 10°.

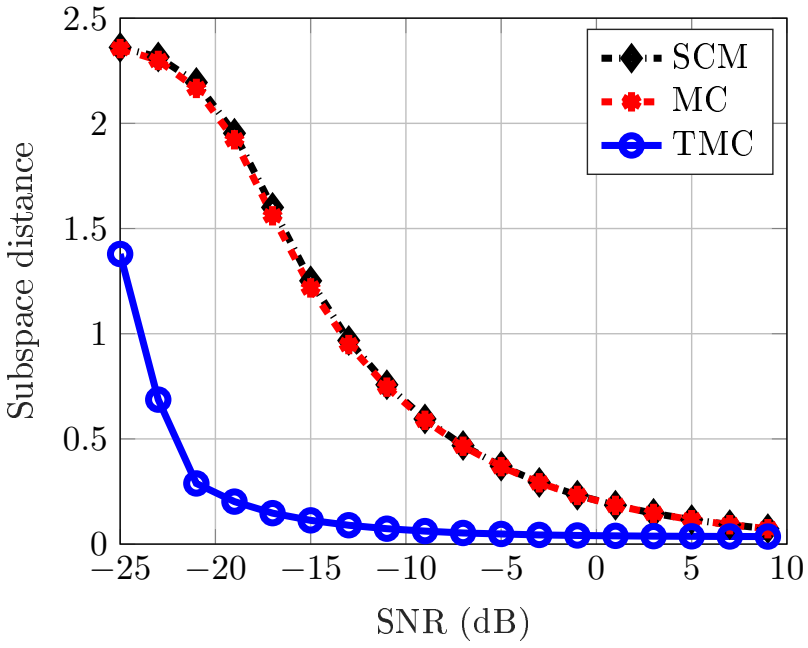


Figure 4.2: Subspace distance vs. SNR for $M = 100$, $N = 150$, $K = 3$, $\epsilon_d = 0.4$, $\Delta_\theta = 5^\circ$ and equal-power sources.

Experiment 1: The first experiment shows subspace distance vs. SNR for an array of $M = 100$ sensors, $N = 150$ snapshots, $K = 3$ sources with the separation of $\Delta_\theta = 5^\circ$ and $\epsilon_d = 0.4$. The sources have equal power. As we can observe from Fig. 4.2, the subspace distances for SCM and MC are almost the same, however, the estimated signal subspace for TMC is much closer to the true signal subspace, thanks to the denoising step of Toeplitz rectification and enforcement of Toeplitz structure into the matrix completion algorithm.

Experiment 2: A similar result is obtained from the next experiment when sources have unequal powers. For this experiment, we considered a scenario with $M = 100$, $N = 80$, $K = 3$, $\epsilon_d = 0.5$, $\Delta_\theta = 10^\circ$, and $\mathbf{\Psi} = \sigma_s^2 \mathrm{diag}(1, 0.8, 0.5)$, where σ_s^2 is the common signal variance. The result is shown in Fig. 4.3 clearly suggesting the usefulness of the TMC technique.

Experiment 3: Fig. 4.4 shows subspace distance vs. number of snapshots for a scenario with $M = 80$, $K = 4$, $\epsilon_d = 0.6$, $\Delta_\theta = 12^\circ$, and SNR $= 0$ dB. The sources are assumed to have equal power here. Here as well the benefit of removing the diagonal entries from the SCM and recover them via TMC is clearly observed. We can conclude that a much better estimate of the signal subspace can be found by removing the more noisy diagonal entries of the SCM and recovering them back from off-diagonal entries by using TMC algorithm.

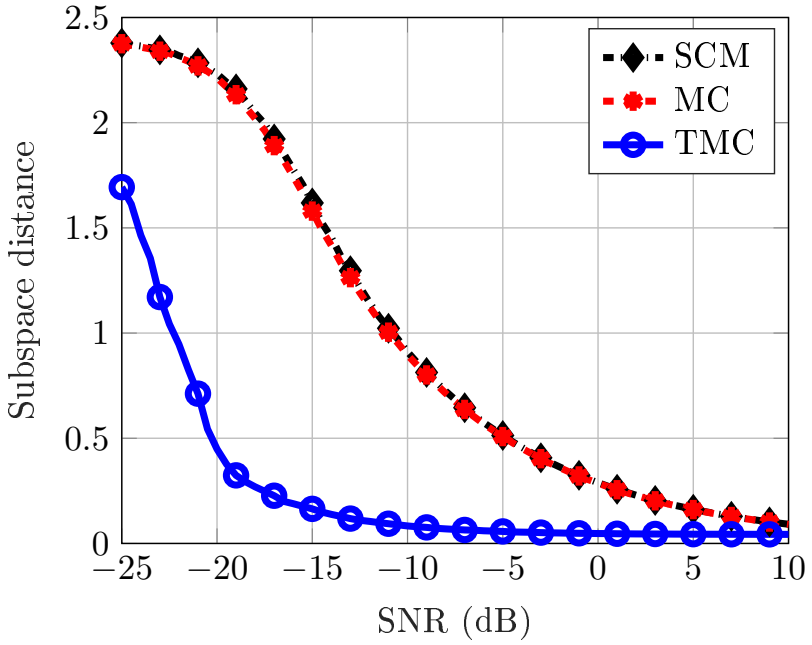


Figure 4.3: Subspace distance vs. SNR for $M = 100$, $N = 80$, $K = 3$, $\epsilon_d = 0.5$, $\Delta_\theta = 10^\circ$ $\mathbf{\Psi} = \sigma_s^2 \text{diag}(1, 0.8, 0.5)$.

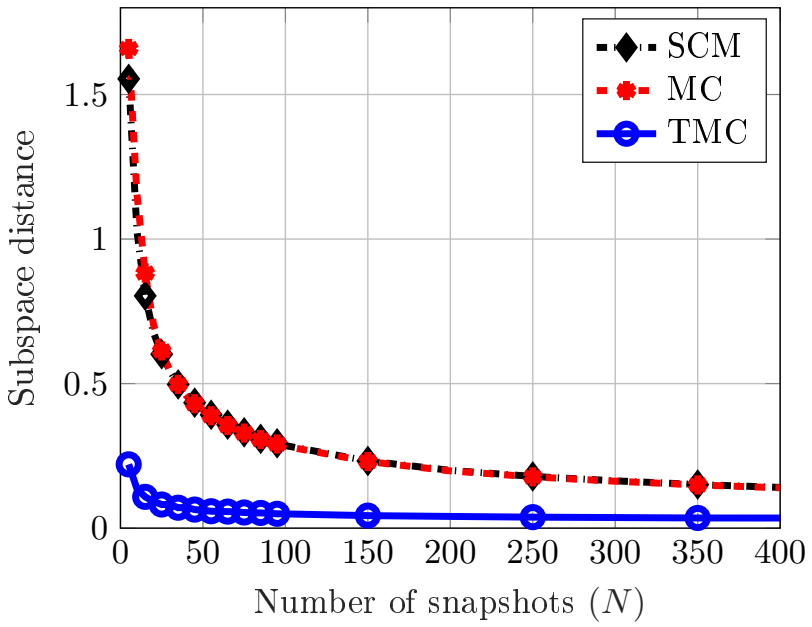


Figure 4.4: Subspace distance vs. number of snapshots (N) for $M = 80$, $K = 4$, $\epsilon_d = 0.6$, $\Delta_\theta = 12^\circ$, SNR $= 0$ dB and equal-power sources.

4.5.2 Application to Source Enumeration

In this subsection we compare the performance of the proposed TMC method for order estimation, referred to as OE-TMC, with some representative methods for order estimation in the presence of the noise with diagonal covariance matrix and/or with low sample support, such as the LS-MDL, the MDL-unc, and the SDRP. These methods have been briefly discussed in Section 1.2. We are also comparing the results with the SA majority vote approach (denoted as SA-MajVote), which was discussed in Chapter 3.

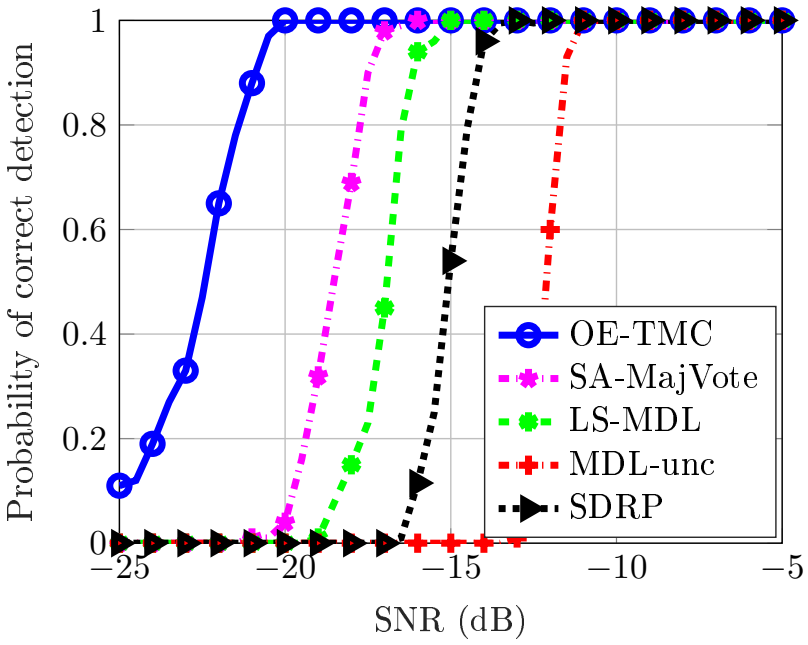


Figure 4.5: Probability of correct detection vs. SNR for $M = 100$, $N = 150$, $K = 3$, $\epsilon_d = 0.3$, $\Delta_\theta = 10^\circ$ and $\mathbf{\Psi} = \sigma_s^2 \text{diag}(1, 0.8, 0.6)$.

Experiment 1: Fig. 4.6 depicts the probability of correct detection with respect to SNR for an array of $M = 100$ antennas, $K = 3$ sources with $\Delta_\theta = 10^\circ$, $N = 150$ snapshots and $\epsilon_d = 0.3$. Sources have unequal powers so that $\mathbf{\Psi} = \sigma_s^2 \text{diag}(1, 0.8, 0.6)$. We observe that OE-TMC provides better performance than competing methods. Thanks to the Toeplitz rectification and the denoising of $\hat{\mathbf{R}}$, the OE-TMC method reliably detects the rank of the signal covariance matrix at lower SNRs than the rest of methods and hence yields a more robust source enumeration method.

Experiment 2: In the next experiment, the probability of correct detection with respect to SNR is shown for $M = 80$, $N = 120$, $K = 4$, $\Delta_\theta = 12^\circ$, and a very high value of ϵ_d is considered as $\epsilon_d = 0.8$. Sources are assumed to have equal power. Since LS-MDL is designed for white noise, it does not performs well in this scenario of very high ϵ_d. The other methods provide consistent results, however, OE-TMC is a clear winner.

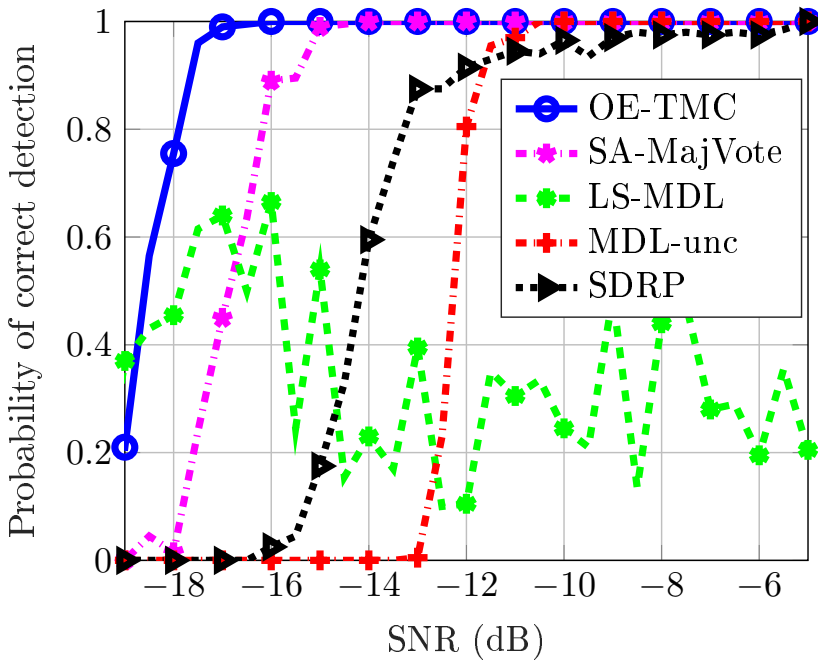


Figure 4.6: Probability of correct detection vs. SNR for $M = 80$, $N = 120$, $K = 4$, $\epsilon_d = 0.8$, $\Delta_\theta = 12^\circ$ and equal-power sources.

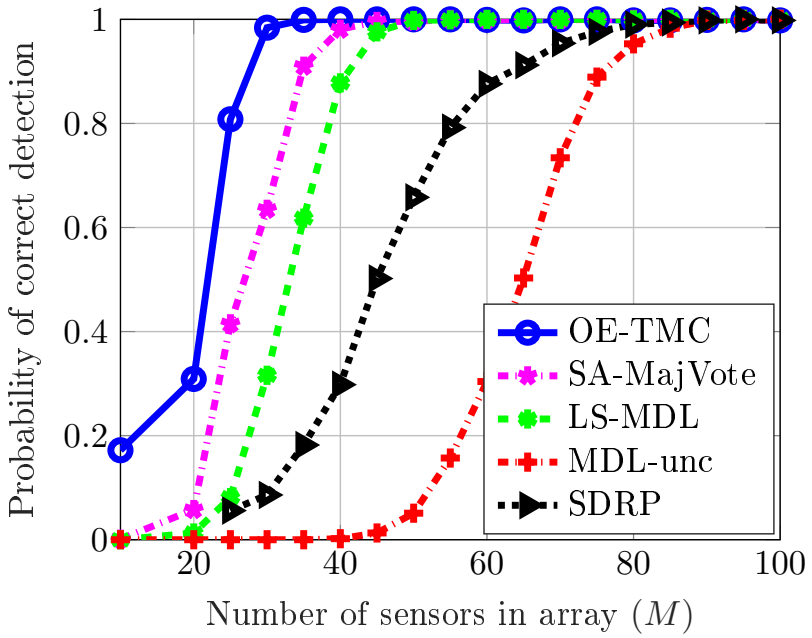


Figure 4.7: Probability of correct detection vs. number of sensors in array (M) for $\frac{M}{N} = 1$, $\Delta_\theta = \frac{2\pi}{M}$, $K = 3$, $\epsilon_d = 0.3$ and SNR = -10 dB and equal-power sources.

Experiment 3: Fig. 4.7 shows the probability of correct detection vs. the number of sensors when $K = 3$, $\epsilon_d = 0.3$, SNR is fixed to -10 dB and all sources have equal power. The ratio between the number of sensors and the number of snapshots is $\frac{M}{N} = 1$ and the separation between sources is $\Delta_\theta = \frac{2\pi}{M}$. Since SDRP does not work properly for small values of N (or M), we consider $M > 20$ for this method. This example shows that OE-TMC provides accurate results in the small sample regime for arrays of varying number of sensors.

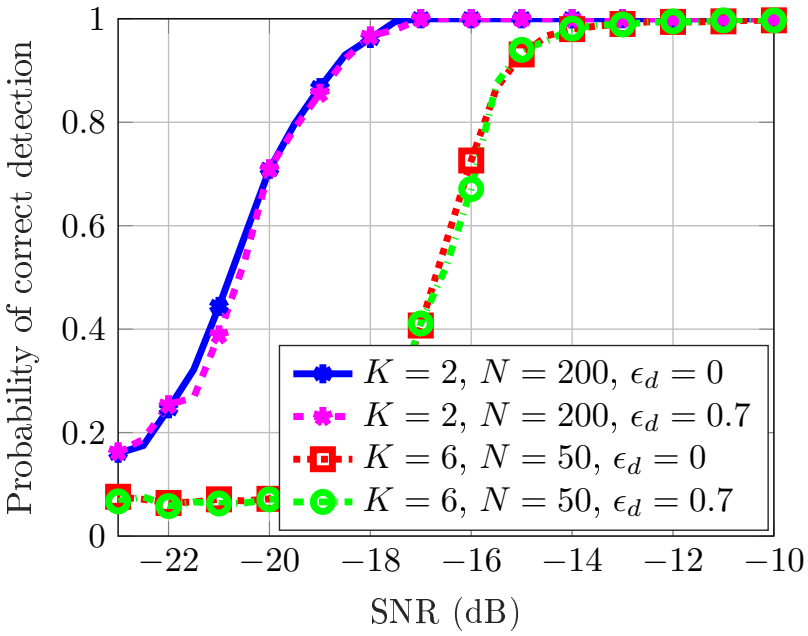


Figure 4.8: Probability of correct detection vs. SNR for OE-TMC when $M = 50$, $\Delta_\theta = 10^\circ$ and equal-power sources for different values of K, N and ϵ_d.

Experiment 4: In the next experiment the robustness of the OE-TMC method against the non-whiteness parameter, ϵ_d, is examined in different scenarios. Since the diagonal terms of $\hat{\mathbf{R}}$ are eliminated as a pre-processing step, it is expected that the OE-TMC results will not be affected by changes in ϵ_d. This behavior can be observed for two different scenarios: i) $K = 2$ and $N = 200$ and ii) $K = 6$ and $N = 50$, in Fig. 4.8, which shows the probability of correct detection with respect to SNR for $M = 50$, $\Delta_\theta = 10^\circ$ and equal power sources. It can be noticed that for both scenarios, results are almost unaffected by ϵ_d. Let us recall that $\epsilon_d = 0$ represents the white noise case; therefore, TMC is robust under both uncorrelated non-iid and iid noises.

Experiment 5: However, TMC is designed for uncorrelated sources, we evaluate the impact of having correlated sources in the last experiment. We consider a scenario with $K = 2$ correlated sources when $M = 100$, $N = 50$, $\epsilon_d = 0.4$, SNR $= -7$ dB, and $\Delta_\theta = 5^\circ$. The correlation coefficient between two sources, ρ, varies from 0 (uncorrelated signals) to 1 (fully correlated signals). As Fig. 4.9 shows, OE-TMC is

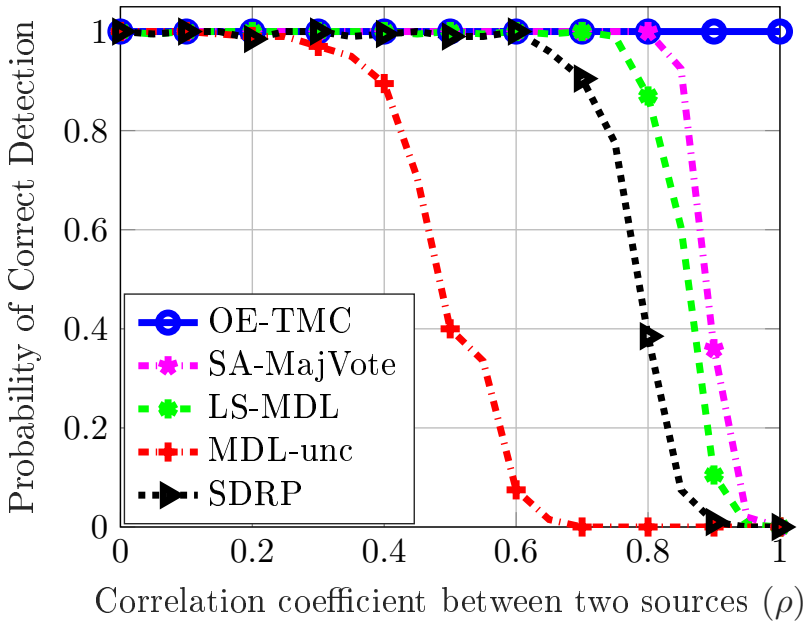


Figure 4.9: Probability of correct detection vs. correlation coefficient between two sources ρ when $M = 100$, $N = 50$, $K = 2$, $\epsilon_d = 0.4$, SNR $= -7$ dB, and $\Delta_\theta = 5°$.

completely unaffected with the ρ. Nevertheless, the performance of OE-TMC under correlated sources needs additional theoretical analysis.

4.6 Conclusions

This chapter addressed the problem of source enumeration for ULAs when the noise covariance matrix is diagonal with unknown entries and with relatively few snapshots. A novel matrix completion based approach is proposed to this problem. The diagonal terms of the sample covariance matrix, which are most affected by this noise model, are eliminated and the off-diagonal terms are further denoised by Toeplitz rectification. The low-rank and Toeplitz signal covariance matrix is then reconstructed with matrix completion techniques. We have shown that the reconstructed matrix can provide a better estimation of the signal subspace, i.e. TMC can be used for denoising the sample covariance matrix for the given noise model. We also have shown that the Frobenius norm of the signal covariance matrix reconstructed by the proposed denoising+MC technique provides an accurate order determination criterion. The method performs robustly for iid and non-iid noises, as well as for small and large arrays, in the small-sample regime.

Chapter 5
Order Estimation with Missing Data

In this chapter, we extend the MC-based order estimation approach for missing data scenario by applying MC directly to the data matrix. We assume that at every time instant only the signals received by a randomly selected subset of antennas are down-converted to baseband and sampled. Low-rank matrix completion techniques are then used to reconstruct the missing entries of the signal data matrix to keep the angular resolution of the original large-scale array. The proposed MC algorithm exploits not only the low- rank structure of the signal subspace, but also the shift-invariance property of ULAs, which results in a better estimation of the signal subspace. Further, this approach is extended to develop an order estimation criterion for a data matrix with missing entries.

5.1 Introduction

The need of large bandwidths in modern 5G networks requires to operate with large-scale antenna arrays. However, a large-scale antenna array introduces new challenges such as hardware and computational complexities and high energy consumption requirements. A possible solution to this problem is to reduce the number of radio frequency (RF) transceiver chains by performing antenna selection at the receiving array (cf. Fig. 5.1). At every time instant a random switch selects a subset of antennas whose RF signals are downconverted and further processed. Since the number of targets or sources is typically much smaller than the number of antennas, it is feasible to reconstruct (or at least to approximate) the low-rank signal data matrix using MC algorithms as if it had been received by the full array, as long as we sample a sufficiently large fraction of the sensors [Candes and Plan, 2010]. However, to achieve the optimum solution from the MC algorithm, the true rank of the noiseless data matrix is required, which is the same as the number of sources, and not always known. This opens up a requirement of an order estimation criterion. However, the performance of existing order estimation methods degrade if some data is missing. A data matrix with missing data can also be found if one or more sensors are damaged. This further increases the importance of developing a source enumeration method for missing data scenario.

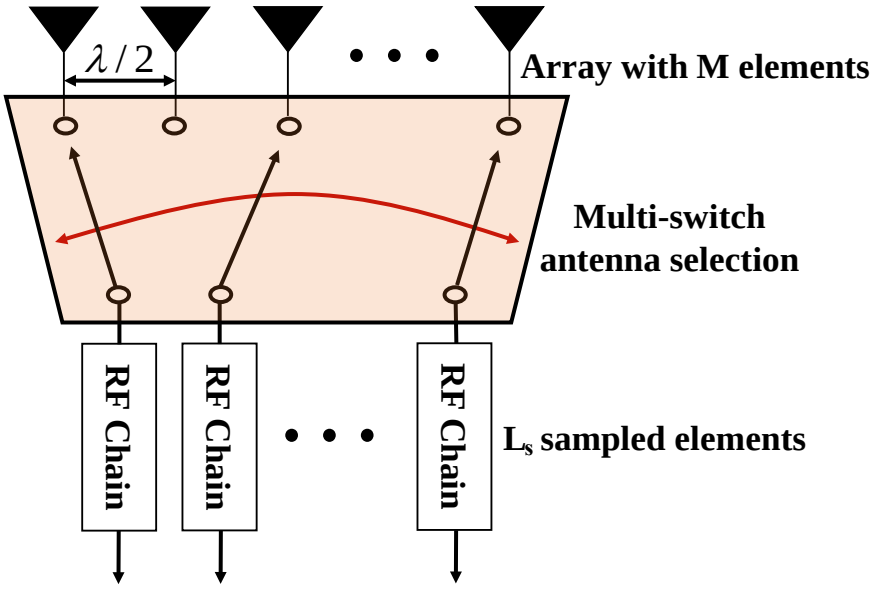


Figure 5.1: Simplified large-scale multi-switch array architecture where L_s out of M sensors are randomly selected and sampled at each time instant.

This chapter first discusses the feasibility of the process of reducing the number of RF chains, and recovering the missing entries by using a matrix completion technique. We show here that the signal subspace of the reconstructed matrix is not much affected by this process and the angular resolution of the original large scale array is kept. In our approach, the matrix completion problem is tailored to enforce the shift-invariance property of ULAs by including an additional regularization term in the MC cost function. Then, the Optimal Subspace Estimation (OSE) technique proposed by Vaccaro and Ding in [Vaccaro and Ding, 1993] is used to estimate the signal subspace. With the help of the simulation results, we show that the signal subspace generated from the reconstructed matrix is close to the original signal subspace generated by complete data matrix, and therefore, the number of RF chains can be largely reduced without significant performance loss.

Furthermore, this approach is extended to account for the order estimation in missing data scenario by exploiting the shift invariance property of ULAs. The proposed source enumeration criterion is based on the chordal subspace distance between two submatrices extracted from the reconstructed matrix after using MC for increasing rank values. By means of simulation results, we show that the proposed order estimation criterion performs consistently in missing data scenario.

The structure of this chapter is as follows. Section 5.2 explains the problem statement for order estimation in missing data scenario and the used system model is explained. Section 5.3 reviews the matrix completion problem, and then, the shift-invariance matrix completion is explained in Section 5.4. A direct application of the Davis–Kahan theorem [Davis and Kahan, 1970] allows us to analyze in Section 5.5

the chordal distance between the true signal subspace and the signal subspace of the sparse and reconstructed matrices. Section 5.6 proposes the order estimation criterion, and then Section 5.7 discusses the corresponding results. Section 5.7 first discusses the feasibility of working with a reduced number of RF chains, and further, it shows the performance of the proposed order estimation criterion. Finally, conclusive remarks are provided in Section 5.8.

5.2 Signal Subspace Estimation with Missing Data

For a fully digital receiver with M RF-branches, the system model is defined in Section 1.1. Let us assume that after collecting N snapshots, the full data matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{X}_s + \mathbf{E} \tag{5.1}$$

where $\mathbf{E} = \big[\mathbf{e}[1] \dots \mathbf{e}[N]\big]$, and $\mathbf{X}_s = \mathbf{AS}$ is the noiseless signal component with $\mathbf{S} = \big[\mathbf{s}[1] \dots \mathbf{s}[N]\big]$. A simplified receiver architecture composed of an $M \times L_s$ RF switching network is considered such that, at each snapshot, it randomly selects L_s out of the M antennas to be downconverted and sampled at baseband (see Fig. 5.1). Multi-switch antenna selection techniques for massive MIMO have been studied and experimentally validated in [Gao et al., 2015]. After downconversion and sampling, the $L_s \times N$ samples are arranged in a $\mathbf{X}_m \in \mathbb{C}^{M \times N}$ matrix so that the missing entries are replaced with zeros. The sampling process can be compactly expressed as

$$\mathbf{X}_m = P_\Omega(\mathbf{X}), \tag{5.2}$$

where $\Omega \subseteq \{1, \dots, M\} \times \{1, \dots, N\}$ is the set of observed (antenna, time) indexes in $\mathbf{X}_m$.

The problem addressed in this chapter is as follows. Given the observed data matrix $\mathbf{X}_m$ in (5.2), to estimate the rank-K noiseless signal subspace by assuming that K in known, which can be used further for direction-of-arrival (DOA) estimation. Further, to design an order estimation method to estimate K from $\mathbf{X}_m$. It is assumed that K satisfies $K \ll L_s < M$.

5.3 Matrix Completion for Data Matrix Estimation

We discussed the matrix completion for a Hermitian covariance matrix in Section 4.3.1. Here, we want to estimate the low-rank signal matrix $\mathbf{X}_s$ from $\mathbf{X}_m \in \mathbb{C}^{M \times N}$, which is not Hermitian, and hence (4.4) can not be applied. Therefore, we redefine the MC problem for data matrix in this section. We can recover $\mathbf{X}_s$ by solving [Candes and Plan, 2010]

$$\begin{aligned} \min_{\mathbf{X}_s \in \mathbb{C}^{M \times N}} \quad & ||\mathbf{X}_s||_* \\ \text{subject to} \quad & ||P_\Omega(\mathbf{X}_s - \mathbf{X}_m)||_F \leq \eta \end{aligned} \tag{5.3}$$

where $||\mathbf{X}_s||_*$ denotes the nuclear norm of $\mathbf{X}_s$.

The main assumption for a successful recovery in low-rank MC is that of incoherence, which means that each singular vector of the unknown matrix must be evenly spread across its coordinates instead of having a few entries with large value i.e., singular vectors are not too sparse. Intuitively, this implies that there is no underlying matrix structure and that all entries have similar importance. Formally, the coherence of the column space of a rank-K matrix $\mathbf{Y} \in \mathbb{C}^{M\times N}$ is defined as

$$\tau(\mathbf{Y}) = \frac{M}{K} \max_{1\leq i\leq M} ||\mathbf{P}_{\mathbf{Y}}\mathbf{e}_i||_2$$

where $\mathbf{P}_{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H$ is the orthogonal projection matrix onto the column space of $\mathbf{Y}$, and $\mathbf{e}_i$ is the ith vector of the Euclidean basis.

As shown in [Weng and Wang, 2012], in array processing $\tau(\mathbf{X}_s)$ and $\tau(\mathbf{X}_s^H)$ are small enough that the complete matrix $\mathbf{X}_s$ can be recovered via (5.3). Indeed, in the noiseless case an exact recovery is possible with high probability provided that we observe $|\Omega| \geq D_m K N^{\frac{6}{5}} \log N$ for a constant D_m assuming a random uniform sampling distribution and $N > M$ [Candes and Plan, 2010]. In our problem we have $|\Omega| = NL_s$, therefore $L_s \geq D_m K N^{\frac{1}{5}} \log N$ antenna elements need to be sampled for successful recovery. In the noisy case, $\mathbf{X}_s$ is recovered with an error proportional to η as long as $||P_\Omega(\mathbf{E})||_F \leq \eta$ [Candes and Plan, 2010].

While standard MC assumes uniform random sampling, this scheme does not exactly match the multi-switch array architecture in Fig. 5.1. In the proposed architecture, exactly L_s sensors, chosen at random, are sampled per snapshot, which is termed as uniform spatial sampling in [Weng and Wang, 2012] and does not correspond to uniform random sampling across $\mathbf{X}$. Nevertheless, as it is proved in [Weng and Wang, 2012], the uniform spatial sampling scheme satisfies the coherence conditions for matrix recovery and hence it can be used in array processing problems.

$\mathbf{X}_s$ can be factored as $\mathbf{X}_s = \mathbf{W}\mathbf{H}^H$, where $\mathbf{W} \in \mathbb{C}^{M\times p}$ and $\mathbf{H} \in \mathbb{C}^{N\times p}$, where $p = K$ limits the rank of the reconstructed matrix. Then, using the identity [Srebro et al., 2005]

$$||\mathbf{X}_s||_* = \min_{\mathbf{X}_s=\mathbf{W}\mathbf{H}^H} \frac{1}{2}\left(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2\right),$$

rank-p $\mathbf{X}_s$ can be estimated by solving the optimization problem

$$\{\hat{\mathbf{W}}, \hat{\mathbf{H}}\} = \underset{\substack{\mathbf{W}\in\mathbb{C}^{M\times p}\\ \mathbf{H}\in\mathbb{C}^{N\times p}}}{\operatorname{argmin}} \left\|P_\Omega\left(\mathbf{X}_m - \mathbf{W}\mathbf{H}^H\right)\right\|_F^2 + \mu\left(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2\right) \tag{5.4}$$

In the next section, we modify (5.4) to exploit the shift-invariance property of the steering matrix $\mathbf{A}$.

5.4 Shift-invariant Matrix Completion (SIMC)

In addition to being a low-rank matrix, $\mathbf{X}_s$ has additional structure inherited from the array geometry that can be exploited by the MC method. Specifically, when ULAs are employed, the shift-invariance property holds, which has been briefly discussed in Chapter 2. According to this property, each row of the steering matrix $\mathbf{A}$ is related to the previous one as follows

$$\mathbf{a}_i^H \mathbf{Q} = \mathbf{a}_{i-1}^H \quad i = 2, \ldots, M \tag{5.5}$$

where $\mathbf{a}_i^H$ is the ith row of $\mathbf{A}$ and $\mathbf{Q} = \text{diag}(e^{j\theta_1}, \ldots, e^{j\theta_K})$, as it can be readily verified from (1.1). From the shift-invariance property, it follows that the column span of $\mathbf{X}_s^{\uparrow}$, formed by the first $M-1$ rows of $\mathbf{X}_s$, and the column span of $\mathbf{X}_s^{\downarrow}$, formed by the last $M-1$ rows of $\mathbf{X}_s$, are identical. In other words, the K-dimensional signal subspaces of $\mathbf{X}_s^{\uparrow}$ and $\mathbf{X}_s^{\downarrow}$ are the same.

It is then clear that the factor $\mathbf{W}$ in

$$\mathbf{X}_s = \mathbf{W}\mathbf{H}^H \tag{5.6}$$

should satisfy the shift-invariance property as well. Since the factorization (5.6) is not unique, we use a relaxed version of (5.5) to enforce the following relation between the rows of $\mathbf{W}$, i.e.,

$$\mathbf{w}_i^H \mathbf{T} = \mathbf{w}_{i-1}^H \quad i = 2, \ldots, M \tag{5.7}$$

where $\mathbf{w}_i^H$ is the ith row of $\mathbf{W}$ and $\mathbf{T} \in \mathfrak{D}$ where $\mathfrak{D}$ is the set of $p \times p$ diagonal complex matrices not necessarily unitary.

To enforce (5.7), the shift-invariant matrix completion (SIMC) problem (5.4) includes an additional regularization term:

$$\begin{aligned}\{\hat{\mathbf{W}}, \hat{\mathbf{H}}, \hat{\mathbf{T}}\} = &\underset{\substack{\mathbf{W}\in\mathbb{C}^{M\times p}\\ \mathbf{H}\in\mathbb{C}^{N\times p}\\ \mathbf{T}\in\mathfrak{D}}}{\text{argmin}} \sum_{(i,j)\in\Omega} \left|\mathbf{X}_m(i,j) - \mathbf{w}_i^H \mathbf{h}_j\right|^2 + \mu\Big(\sum_{i=1}^{M} \|\mathbf{w}_i\|_2^2 + \sum_{j=1}^{N} \|\mathbf{h}_j\|_2^2\Big) \\ &+ \alpha \sum_{i=2}^{M} \|\mathbf{w}_i^H \mathbf{T} - \mathbf{w}_{i-1}^H\|_2^2 \end{aligned} \tag{5.8}$$

where $\mathbf{h}_j^H$ is the jth row of $\mathbf{H}$ and α is an additional regularization parameter.

The solution $\hat{\mathbf{X}}_s = \hat{\mathbf{W}}\hat{\mathbf{H}}^H$ can be obtained by iteratively optimizing (5.8) over each $\mathbf{w}_i^H$, $\mathbf{h}_j^H$ and $\mathbf{T}$ until convergence. To optimize (5.8) for $\mathbf{w}_i^H$, we take the derivative with respect to $\mathbf{w}_i^H$, assuming $\mathbf{H}$ and $\mathbf{T}$ fixed, and equate it to zero, which provides the following solution

$$\mathbf{w}_i^H = \begin{cases} \Big(\mathbf{g}_1^H + \mathbf{g}_2^H\Big)\Big(\mathbf{Y}_1 + \alpha\mathbf{I}\Big)^{-1} & \text{if } i = 1 \\ \Big(\mathbf{g}_1^H + \mathbf{g}_2^H + \mathbf{g}_3^H\Big)\Big(\mathbf{Y}_1 + \mathbf{Y}_2 + \alpha\mathbf{I}\Big)^{-1} & \text{if } 1 < i < M \\ \Big(\mathbf{g}_1^H + \mathbf{g}_3^H\Big)\Big(\mathbf{Y}_1 + \mathbf{Y}_2\Big)^{-1} & \text{if } i = M \end{cases} \tag{5.9}$$

where

$$\begin{aligned}
\mathbf{g}_1^H &= \sum_{j\in\mathcal{J}_i} \mathbf{X}_m(i,j)\mathbf{h}_j^H \\
\mathbf{g}_2^H &= \alpha \mathbf{w}_{i+1}^H \mathbf{T} \\
\mathbf{g}_3^H &= \alpha \mathbf{w}_{i-1}^H \mathbf{T}^H \\
\mathbf{Y}_1 &= \sum_{j\in\mathcal{J}_i} \mathbf{h}_j \mathbf{h}_j^H + \mu \mathbf{I} \\
\mathbf{Y}_2 &= \alpha \mathbf{T}\mathbf{T}^H
\end{aligned}$$

and $\mathcal{J}_i$ is the set of observed indices of the ith row of $\mathbf{X}_m$. Similarly, (5.8) can be optimized for $\mathbf{h}_j^H$ to find the solution as

$$\mathbf{h}_j^H = \Big(\sum_{i\in\mathcal{I}_j} \mathbf{X}_m(i,j)^* \mathbf{w}_i^H \Big) \Big(\sum_{i\in\mathcal{I}_j} \mathbf{w}_i \mathbf{w}_i^H + \mu \mathbf{I} \Big)^{-1} \tag{5.10}$$

where $\mathcal{I}_j$ is the set of observed indices of the jth column of $\mathbf{X}_m$. Since $\mathbf{T} = \mathrm{diag}(t_1, \ldots, t_p)$ is a diagonal matrix, (5.8) can be optimized for each diagonal element t_k individually. To this end, the third term in the right hand side of (5.8) can be rewritten in terms of t_k as

$$\sum_{i=2}^{M} \|\mathbf{w}_i^H \mathbf{T} - \mathbf{w}_{i-1}^H\|_2^2 = \sum_{i=2}^{M} \sum_{k=1}^{K} \big| t_k \mathbf{W}(i,k) - \mathbf{W}(i-1,k) \big|^2, \tag{5.11}$$

which can be optimized with respect to t_k to get

$$t_k = \frac{\sum_{i=2}^{M} \mathbf{W}(i-1,k)\mathbf{W}^*(i,k)}{\sum_{i=2}^{M} \big|\mathbf{W}(i,k)\big|^2} \tag{5.12}$$

5.4.1 Post-processing via Optimal Subspace Estimation (OSE)

As the shift-invariance property is enforced through a regularization term, the solution of (5.8) provides a low-rank data matrix, $\hat{\mathbf{X}}_s$, which has the required structure only in an approximate fashion. This motivates applying the Optimal Subspace Estimation (OSE) technique as a final post-processing step of our algorithm. The OSE algorithm takes $\hat{\mathbf{X}}_s$ as input and provides an estimate of the underlying noise-free signal subspace with the required shift-invariant structure. Let $\mathbf{U}_{ose} \in \mathbb{C}^{M\times K}$ be a basis for this subspace, and let $\mathbf{P}_{ose} = \mathbf{U}_{ose}\mathbf{U}_{ose}^H$ be its orthogonal projection matrix. Then, the output of the OSE algorithm is

$$\hat{\mathbf{R}}_{ose} = \mathbf{P}_{ose}\hat{\mathbf{X}}_s \tag{5.13}$$

For a full account of the OSE method the reader is referred to [Vaccaro and Ding, 1993, Vaccaro, 2017, Grayson, 2016].

5.4.2 Selection of Regularization Parameters

The values of α and μ in (5.8) control the trade-off among the fulfillment of the shift-invariance property, the fitting to the observed data and the nuclear norm of the solution. Since α enforces the shift-invariance property into $\hat{\mathbf{X}}_s$, its value should depend on some measure that quantifies the compliance of the shift-invariance property by the original sparse matrix $\mathbf{X}_m$. As we know, given a rank-p matrix for which the shift-invariance property holds, the subspaces spanned by the first and the last $M-1$ rows are identical. Thus, the regularization parameter α is chosen to be a function of the chordal subspace distance [Srivastava and Klassen, 2002] between $\mathbf{X}_m^{\uparrow}$ and $\mathbf{X}_m^{\downarrow}$, which are formed by the first and the last $M-1$ rows of the sparse $\mathbf{X}_m$, respectively.

Specifically, let $\mathbf{U}_1 \in \mathbb{C}^{(M-1)\times p}$ and $\mathbf{U}_2 \in \mathbb{C}^{(M-1)\times p}$ be the p largest left singular vectors (that is, those associated to the p largest singular values) of $\mathbf{X}_m^{\uparrow}$ and $\mathbf{X}_m^{\downarrow}$, respectively. Then, the chordal subspace distance between $\mathbf{X}_m^{\uparrow}$ and $\mathbf{X}_m^{\downarrow}$ is

$$d_{cs} = \|\mathbf{U}_1\mathbf{U}_1^H - \mathbf{U}_2\mathbf{U}_2^H\|_F. \tag{5.14}$$

A large value of d_{cs} implies that the p-dimensional subspaces extracted from $\mathbf{X}_m^{\uparrow}$ and $\mathbf{X}_m^{\downarrow}$ are far apart from each other and, consequently, the shift-invariance property does not hold. This in turn implies that a large α must be used in the reconstruction process. According to our simulations, a value that provides good performance for a wide range of scenarios is

$$\alpha = d_{cs}\mu, \tag{5.15}$$

where $\mu = \frac{M}{20}$.

A summary of the shift-invariant matrix completion method, denoted as SIMC, is provided in Algorithm 7. Once $\hat{\mathbf{R}}_{ose}$ is obtained, any subspace-based method can be used further to estimate the DOAs.

Algorithm 7: Shift-Invariant Matrix Completion (SIMC)

Input: $\mathbf{X}_m$, μ, p, $itr_{\max}$
Output: $\hat{\mathbf{R}}_{ose}$
Initialization: $\hat{\mathbf{T}} = \mathbf{I}$, $itr = 1$
Compute the SVD of $\mathbf{X}_m = \mathbf{F}\boldsymbol{\Lambda}\mathbf{G}^H$ and initialize $\hat{\mathbf{W}} = \mathbf{F}_p\boldsymbol{\Lambda}_p^{1/2}$ and $\hat{\mathbf{H}} = \mathbf{G}_p\boldsymbol{\Lambda}_p^{1/2}$, using the p largest singular vectors and singular values of $\mathbf{X}_m$ (best p-rank approximation of $\mathbf{X}_m$)
Set α as in (5.15)
REPEAT
Compute $\hat{\mathbf{W}}$, $\hat{\mathbf{H}}$ and $\hat{\mathbf{T}}$ using (5.9), (5.10) and (5.12), respectively
$itr = itr + 1$
Until Convergence = true or $itr = itr_{\max}$
Compute $\hat{\mathbf{X}}_s = \hat{\mathbf{W}}\hat{\mathbf{H}}^H$
Apply OSE algorithm to estimate $\hat{\mathbf{R}}_{ose} = \mathbf{P}_{ose}\hat{\mathbf{X}}_s$

The SIMC algorithm has a computational cost of $\mathcal{O}((M+N)K^3)$ per iteration, which is basically the cost of standard MC algorithms based on alternating least

squares, since the extra cost due to (5.12) is negligible. The OSE post-processing step, has a computational complexity of $\big(\mathcal{O}(M^2N) + 2\mathcal{O}((MK)^3)\big)$. Finally, the proposed initialization step performs a compact SVD with cost $\mathcal{O}(MK^2)$. Note that for this problem $K \ll M$.

5.5 Perturbation Analysis

The main factor impacting the performance of the random multi-switch sampling scheme is how well the signal subspace is preserved. The SIMC algorithm aims at estimating an improved signal subspace by leveraging its shift-invariant low-rank structure. Here we analyze how DOA estimation is impacted when performed after MC.

Since the DOA estimates are essentially determined by the singular vectors of the signal subspace, we want to assess how much the principal directions change after each processing step of the original sparse data matrix. To do so, we will analyze the problem from a matrix perturbation standpoint. A perturbed matrix is a matrix which has its singular values and vectors altered after an addition with a second matrix. Thus, $\mathbf{X}_m$ in (5.2) is a perturbed version of $\mathbf{X}_s$, with the perturbation being caused by the missing entries and noise. The Davis-Kahan theorem is a useful tool to measure the angular difference between the singular vectors of two matrices. We show below Theorem 5.1 in [Yu et al., 2015] adapted to our use-case.

Theorem 5.1. Davis-Kahan sin theorem. *[Yu et al., 2015] Let $\mathbf{U}_X$ and $\mathbf{U}_{\tilde{X}}$ denote the first K left singular vectors of $\mathbf{X}_s$ and the perturbed $\tilde{\mathbf{X}}_s$, respectively, and $\Theta(\mathbf{U}_X, \mathbf{U}_{\tilde{X}})$ be the $K \times K$ diagonal matrix containing the principal angles $cos^{-1}(\xi_i)_{i=1}^{K}$, where ξ_i is the ith singular value of $\mathbf{U}_X^H \mathbf{U}_{\tilde{X}}$. Then,*

$$||\sin\Theta(\mathbf{U}_X, \mathbf{U}_{\tilde{X}})||_F \leq \frac{2\sqrt{K}(2||\mathbf{X}_s||_2 + ||\mathbf{X}_s - \tilde{\mathbf{X}}_s||_2)\min(||\mathbf{X}_s - \tilde{\mathbf{X}}_s||_2, \frac{1}{\sqrt{K}}||\mathbf{X}_s - \tilde{\mathbf{X}}_s||_F)}{\lambda_K(\mathbf{X}_s)} \tag{5.16}$$

Theorem 5.1 shows that the subspace distance between the singular vectors $\mathbf{U}_X$ and $\mathbf{U}_{\tilde{X}}$ scales with the norm difference between $\mathbf{X}_s$ and $\tilde{\mathbf{X}}_s$. Interestingly, it also shows that the larger the Kth singular value is, the smaller the subspace distance will be. Below, we leverage the Davis-Kahan theorem to compare the signal space of $\mathbf{X}_s$ firstly with that of the sparse matrix $\mathbf{X}_m$, and secondly with the recovered estimate $\hat{\mathbf{X}}_{MC}$ in (5.4) obtained through MC.

Clearly, due to the missing data, $\mathbf{X}_m$ is a poor approximation to $\mathbf{X}_s$. Nevertheless, the K first singular vectors of the sampled matrix are often used as a crude estimate or initialization point for iterative algorithms [Giménez-Febrer and Pagès-Zamora, 2017]. Let $P_K(\mathbf{X}_m)$ denote the projection of $\mathbf{X}_m$ onto the subspace spanned by its first K left singular vectors, which is obtained by

setting $\lambda_k(\mathbf{X}_m) = 0,\ \forall\, k > K$. Moreover, let us assume a uniform random sampling scheme where each entry in $\mathbf{X}$ is sampled with probability $q_p = L_s/M$. From [Keshavan et al., 2009], we have the bound

$$||\mathbf{X}_s - \frac{1}{q_p}P_K(\mathbf{X}_m)||_2 \leq C||\mathbf{X}_s||_\infty \frac{N^{\frac{3}{4}}}{M^{\frac{1}{4}}\sqrt{q_p}} + C\sigma\sqrt{\frac{N\log M}{q_p}} \tag{5.17}$$

which is satisfied with probability greater than $1 - \frac{1}{M^3}$ for some constant C. Note the scaling $\frac{1}{q_p}$ of $P_K(\mathbf{X}_m)$ in (5.17), which compensates for the norm loss due to the missing entries. Thus, since $\frac{1}{q_p}P_K(\mathbf{X}_m)$ and $\mathbf{X}_m$ share the first K left singular vectors, then

$$||\sin\Theta(\mathbf{U}_X, \mathbf{U}_{X_m})||_F = ||\sin\Theta(\mathbf{U}_X, \mathbf{U}_{\frac{1}{q}P_K(\mathbf{X}_m)})||_F$$

and we can use (5.17) in conjunction with Theorem 5.1 to bound the subspace distance.

With regard to $\hat{\mathbf{X}}_{MC}$, assuming that $N \geq M$ the recently developed bounds in [Chen et al., 2019] show that

$$||\mathbf{X}_s - \hat{\mathbf{X}}_{MC}||_2 \leq ||\mathbf{X}_s||_2 \frac{\sigma}{\lambda_K(\mathbf{X}_s)}\sqrt{\frac{N}{q_p}}. \tag{5.18}$$

with probability exceeding $1 - \frac{1}{N^3}$.

Assuming constant $q_p = L_s/M$ and M, and bounded $||\mathbf{X}_s||_\infty$, we have that the bound for $P_K(\mathbf{X}_m)$ in (5.17) grows as $\mathcal{O}(N^{\frac{3}{4}})$. Therefore, comparing it to that of $\hat{\mathbf{X}}_{MC}$ in (5.18), we observe that the bound for $\hat{\mathbf{X}}_{MC}$ grows as $\mathcal{O}(\sqrt{N})$. Therefore, we can conclude that MC will improve the DOA estimates.

5.6 Order Estimation via SIMC

When K is known, SIMC algorithm can reconstruct the $\hat{\mathbf{R}}_{ose}$ by setting $p = K$, however when K in unknown, an order estimation method is required. The standard order estimation methods do not provide consistent results in the missing data scenario. One possible solution to this problem could be to reconstruct $\hat{\mathbf{X}}_s$ or $\hat{\mathbf{R}}_{ose}$ with an overestimated value of K, such as $K < p < min(M, N)$, and apply any standard source enumeration technique. However, since the process of matrix completion enforces the eigenvalue sparsity (also mentioned in Section 4.4) thus changing the eigenvalue profile of the covariance matrix, we found that the standard source enumeration methods such as LS-MDL and BIC do not provide consistent solutions even after MC. In this section, we propose an order estimation method based on SIMC.

As discussed in Section 5.4.2, in a noiseless condition d_{cs} (see (5.14)) should be a small value, which is achieved for $p = K$ when a complete data matrix is available. On the one hand, the matrix completion algorithm does not provide an optimum solution for $p < K$, and on the other hand, $p > K$ includes the noise directions which

Algorithm 8: Order Estimation using SIMC

Input: $\mathbf{X}_m$, μ, p_{max}, itr_{max};
Output: Order estimate $\hat{k}_{SIMC}$
for $p = 1, \ldots, p_{max}$ **do**
  Find $\hat{\mathbf{R}}_{ose}$ using Algorithm 7
  Find $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ from $\hat{\mathbf{R}}_{ose}$
  Find $\mathbf{U}_{R1}$ and $\mathbf{U}_{R2}$ as the p largest left singular vectors of $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$
Estimate number of sources as $\hat{k}_{SIMC} = \underset{1 \leq p \leq p_{max}}{\operatorname{argmin}} \frac{\|\mathbf{U}_{R1}\mathbf{U}_{R1}^H - \mathbf{U}_{R2}\mathbf{U}_{R2}^H\|_F}{p}$

eventually increases the value of d_{cs}. Based on this, an order estimation criterion is proposed here, which first reconstructs $\hat{\mathbf{R}}_{ose}$ for increasing values of p, and then estimates the chordal subspace distance between $\hat{\mathbf{R}}_1 := \hat{\mathbf{R}}_{ose}(1 : M-1, 1 : M-1)$ and $\hat{\mathbf{R}}_2 := \hat{\mathbf{R}}_{ose}(2 : M, 2 : M)$, by selecting p largest left singular vectors of $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$. The subspace distance is the smallest for $p = K$.

If $\mathbf{U}_{R1} \in \mathbb{C}^{(M-1)\times p}$ and $\mathbf{U}_{R2} \in \mathbb{C}^{(M-1)\times p}$ be the p largest left singular vectors of $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ respectively, the proposed order estimation criterion can be given as

$$\hat{k}_{SIMC} = \underset{1 \leq p \leq p_{max}}{\operatorname{argmin}} \frac{\|\mathbf{U}_{R1}\mathbf{U}_{R1}^H - \mathbf{U}_{R2}\mathbf{U}_{R2}^H\|_F}{p} \tag{5.19}$$

where p_{max} is an overestimation of K. Note that we are normalizing (5.19) with the number of left singular vector used in each case. A summary of proposed method is shown in Algorithm 8.

5.7 Simulation Results

5.7.1 Performance of the SIMC

In this subsection, we illustrate the feasibility of using a reduced number of RF chains by sampling only a random subset of sensors. For this purpose, the performance of SIMC algorithm is evaluated by means of Monte Carlo simulations. As figures of merit we use: i) the chordal subspace distance between the true signal subspace and the estimated signal subspace, and ii) the Root Mean Squared Error (RMSE) for the DOA estimates in radians. The chordal distance between the true signal subspace or column span of $\mathbf{X_s}$, and the estimated signal subspace or column span of $\hat{\mathbf{X}}_s$ is shown to assess how different these subspaces are. Note that this distance is different from the chordal distance in (5.14) used to select the regularization parameter α. The estimation of signal parameters via rotational invariance technique (ESPRIT) method [Roy and Kailath, 1989] is used to compute the DOAs for all competing methods. We assume for this study that K is known and figures of merit are determined for $p = K$. For comparison, we include the performance of the following methods:

- SCM: The sample covariance matrix without MC is estimated as $\hat{\mathbf{R}}_m = \frac{1}{N}\mathbf{X}_m\mathbf{X}_m^H$.

- OSE: The shift-invariance property is enforced by applying OSE to $\mathbf{X}_m$ (without MC).
- MC: The standard MC algorithm solution given by (5.4) is used to reconstruct $\mathbf{X}_s$ from $\mathbf{X}_m$.
- MC-OSE: OSE is applied as a post-processing step to the previous method.
- SIMC: The proposed method.

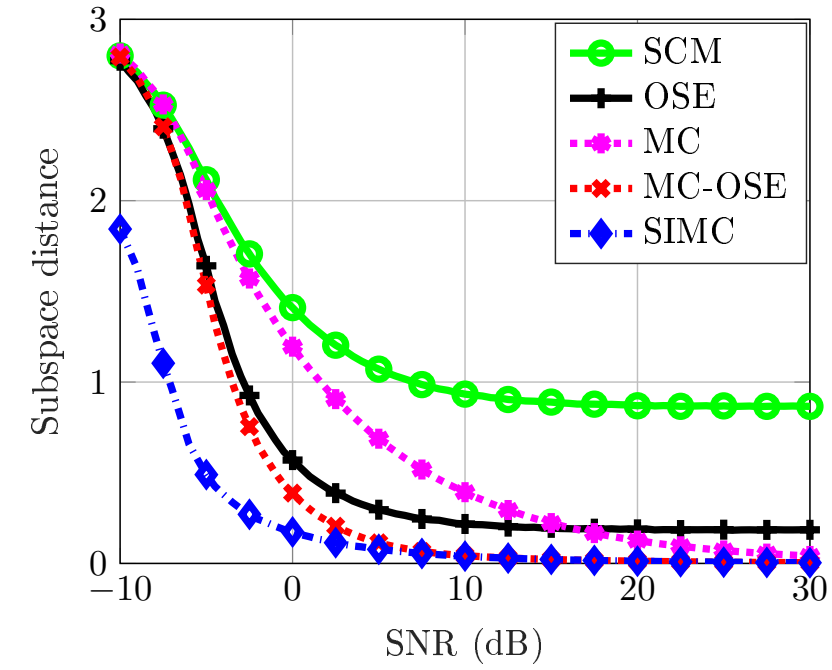
For all simulations we assume that K uncorrelated narrowband signals with a separation of Δ_θ are impinging on a ULA with M half-wavelength separated antennas. Unless stated otherwise, sources have equal power. For both SIMC and MC, we use $\mu = M/20$ and $itr_{max} = 200$. SNR $= 10\log\frac{\mathrm{tr}(\mathbf{R}_s)}{M\sigma^2}$, where $\mathbf{R}_s$ is the signal covariance matrix and σ^2 is the noise variance. L_s denotes the number of randomly sampled sensors per snapshot. The Cramer-Rao lower bound (CRLB) [Stoica and Nehorai, 1990] when the full data matrix $\mathbf{X}$ is available is included as a reference benchmark.

Experiment 1: In the first example, we consider a sample-poor scenario with $M = 100$ antennas, $N = 80$ snapshots, $K = 5$ sources and $\Delta_\theta = 10^\circ$. At each time instant the multi-switch network randomly selects $L_s = 50$ out of the $M = 100$ antennas. Fig. 5.2 shows the subspace distance and the RMSE vs. the SNR. The performance of SCM and OSE without MC saturates at high SNR due to the relatively high fraction of missing entries. The benefits of using MC techniques in combination with enforcing the shift-invariance property are evident, specially at low or moderate SNRs. In fact, even with 50 % of missing data and SNR≈ 0, we observe that SIMC is close to the CRLB (which gives us the performance limit when all data are available). At high SNRs MC-OSE and SIMC have identical performances, which suggest that the post-processing OSE step is sufficient to enforce in the solution the required invariance to displacements.

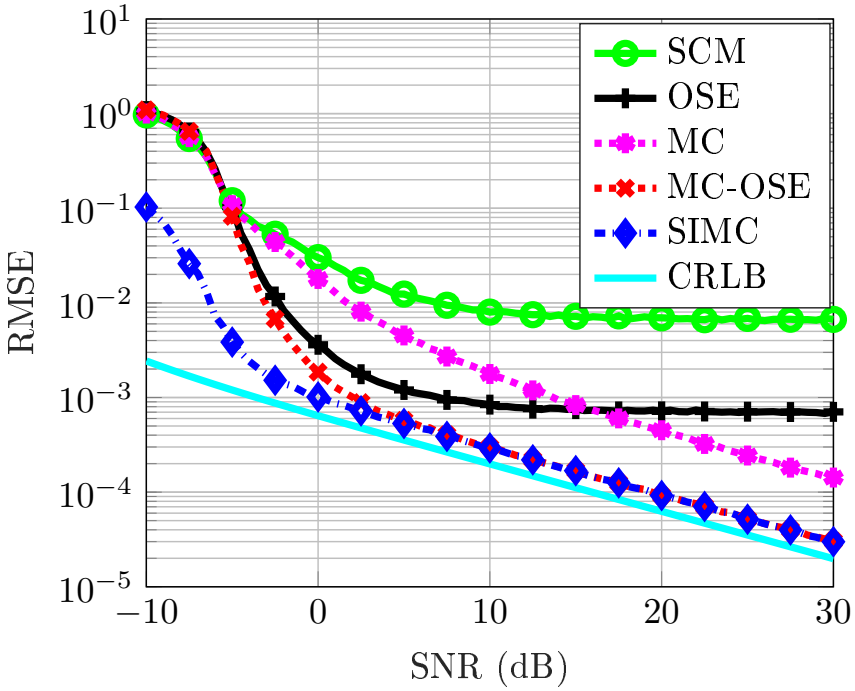
Experiment 2: The second example considers a scenario with $M = 50$ antennas, $K = 4$ sources with $\Delta_\theta = 10^\circ$, $N = 50$ snapshots and $L_s = 25$ (i.e., 50% of missing entries in $\mathbf{X}_m$). The sources in this example have unequal powers with signal covariance matrix $\mathbf{\Psi} = \sigma_s^2\mathrm{diag}(1, 0.8, 0.6, 0.5)$, where σ_s^2 is the common signal variance. A similar behavior to the previous example is observed in Fig. 5.3, with SIMC providing satisfactory performance over a large range of SNR values.

Experiment 3: The third example compares the performance of the methods with respect to N for $M = 200$, $K = 4$, SNR$= -5$ dB, $\Delta_\theta = 5^\circ$, and $L_s = 100$. We can observe in Fig. 5.4 that if N is large enough, SIMC, MC-OSE and OSE provide very similar results. However, SIMC outperforms the rest of methods when N is small. This example demonstrates a clear advantage of the proposed method in the small-sample regime where $N \leq M$.

Experiment 4: The next example compares the performances of arrays of different number of antennas when the number of sampled sensors L_s is fixed. Therefore, the spatial sampling ratio L_s/M decreases as M increases. We consider ULAs with $M = 25$, $M = 50$ and $M = 100$ antennas using a fixed value of $L_s = 25$ so that at

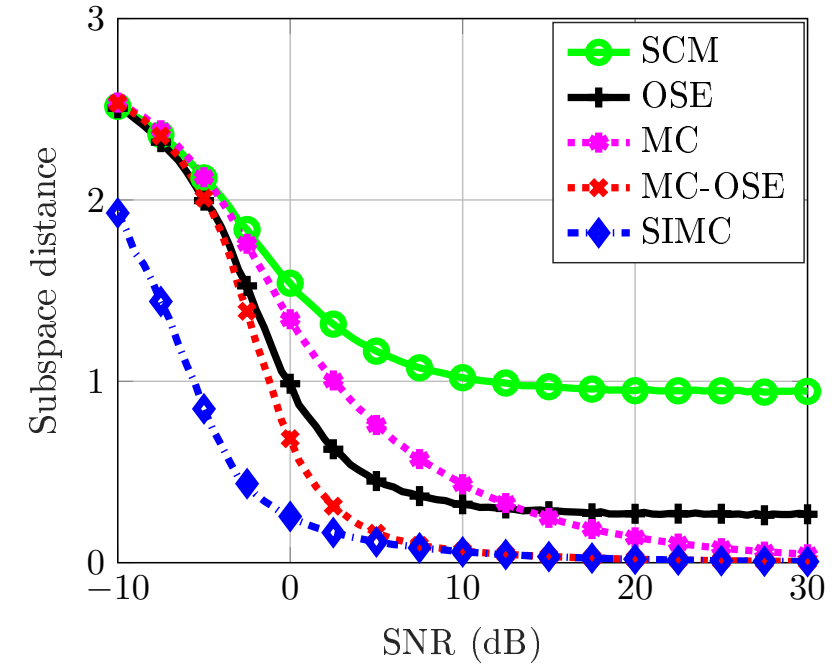


(a)

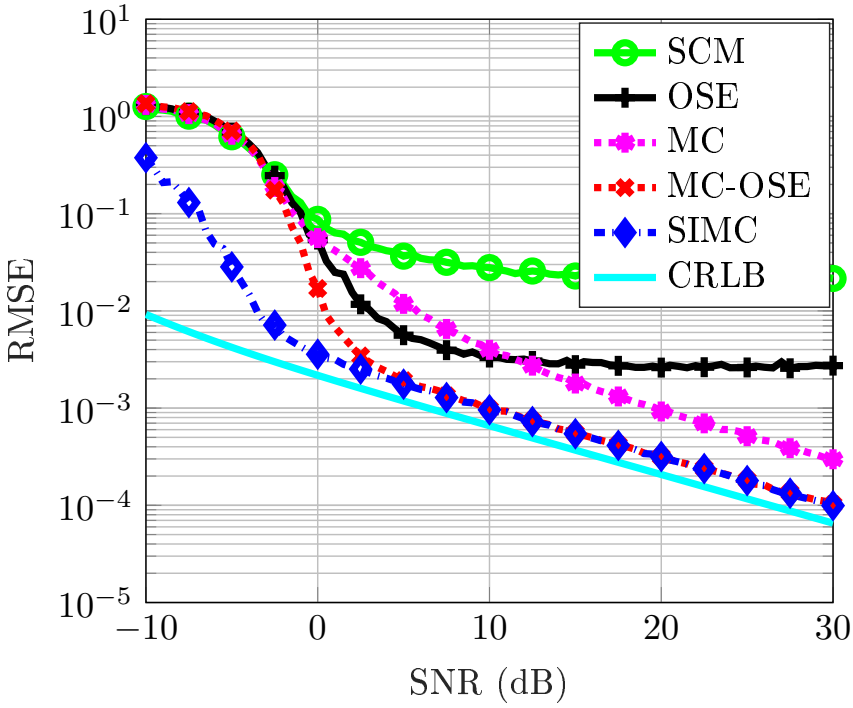


(b)

Figure 5.2: Subspace distance (a) and RMSE (b) vs. SNR for $M = 100$, $N = 80$, $K = 5$, $\Delta_\theta = 10^\circ$ and $L_s = 50$.

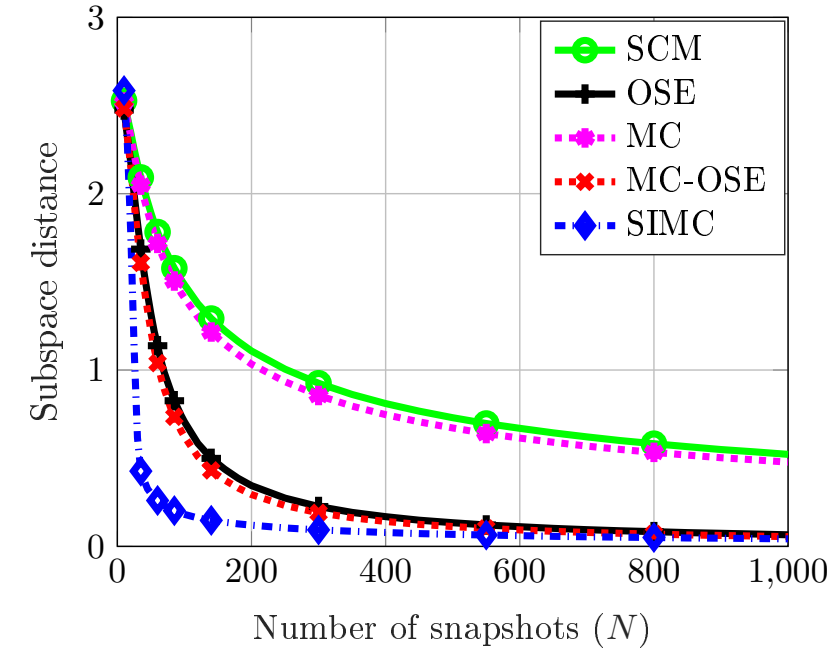


(a)

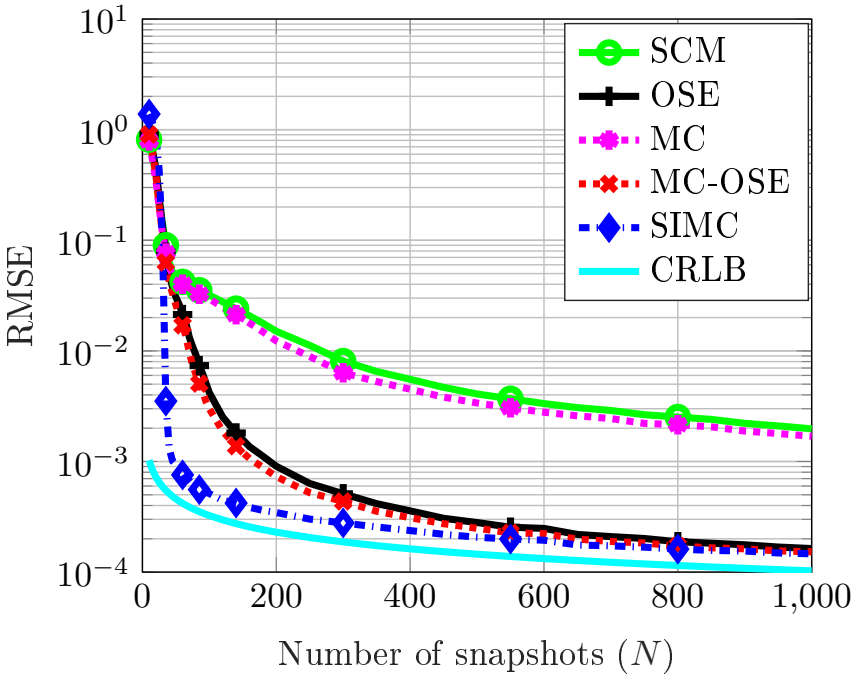


(b)

Figure 5.3: Subspace distance (a) and RMSE (b) vs. SNR for $M = 50$, $N = 50$, $K = 4$, $\Delta_\theta = 10°$, $\mathbf{\Psi} = \sigma_s^2 \mathrm{diag}(1, 0.8, 0.6, 0.5)$ and $L_s = 25$.

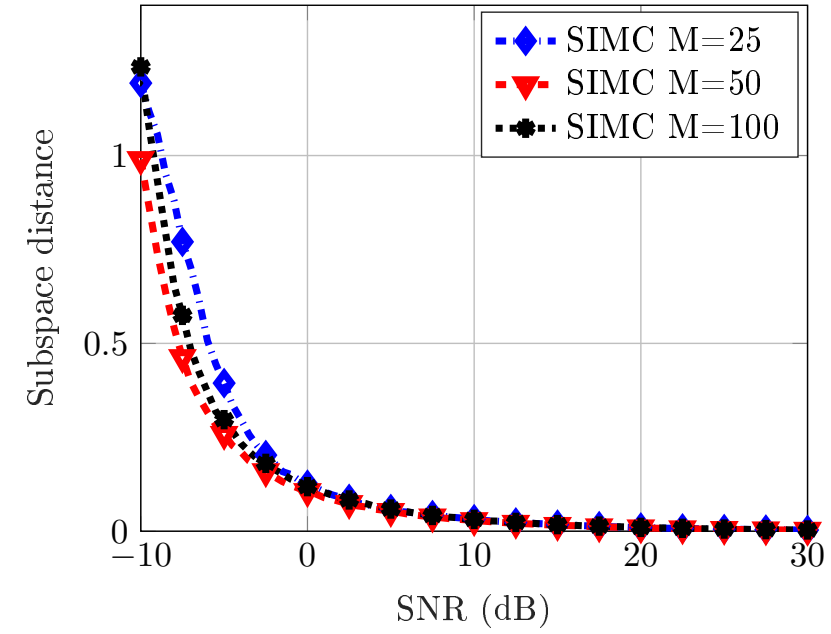


(a)

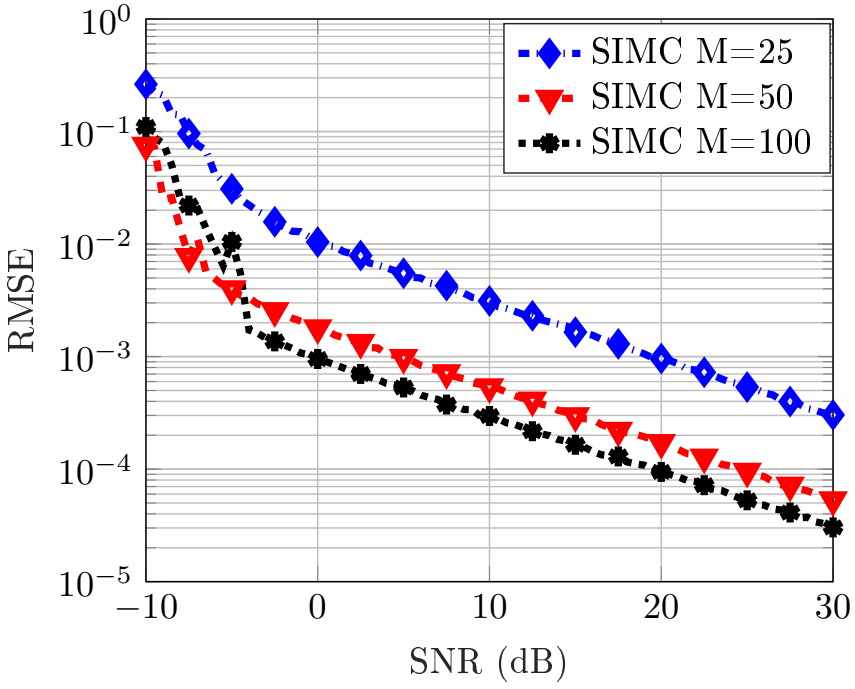


(b)

Figure 5.4: Subspace distance (a) and RMSE (b) vs. number of snapshots (N) for $M = 200$, $K = 4$, SNR$= -5$ dB, $\Delta_\theta = 5°$, and $L_s = 100$.



(a)



(b)

Figure 5.5: Subspace distance (a) and RMSE (b) vs. SNR for ULAs with different number of antennas when $K = 3$, $\Delta_\theta = 10^\circ$, $N = 100$ and $L_s = 25$.

every snapshot the percentages of sampled sensors are 100%, 50% and 25%, respectively. For all three cases, the number of snapshots is $N = 100$ and $K = 3$ sources with $\Delta_\theta = 10^\circ$ of separation impinge on each array. Since L_s and N are fixed, the energy consumption will be roughly the same for all array architectures. However, the effective spatial resolution is improved as M increases, as it is observed in Fig. 5.5. In this way, the proposed SIMC algorithm allows us to increase the spatial resolution of an array with a fixed number of RF chains. In other words, we can trade-off spatial resolution by energy consumption.

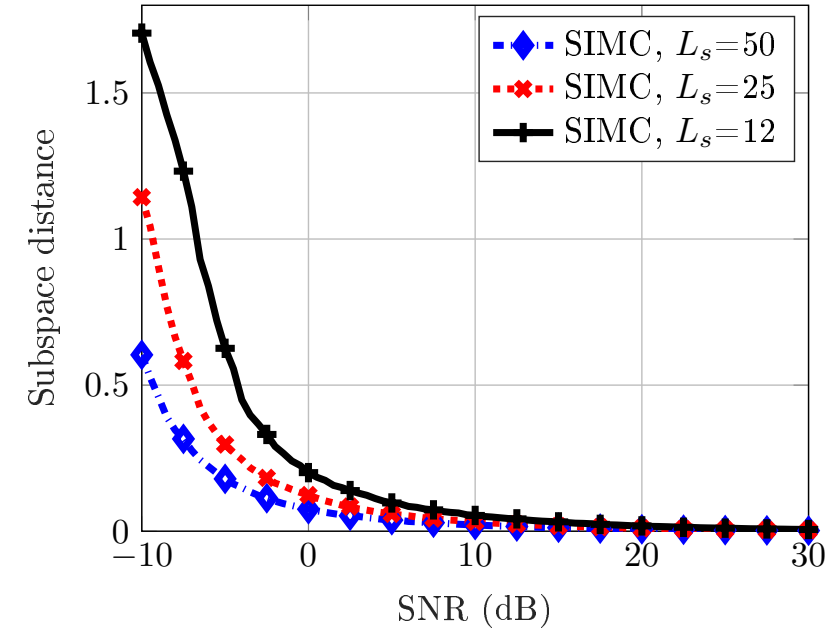
Experiment 5: The following experiments analyze the impact of the percentage of missing data on the methods under comparison. We consider a scenario with $M = 50$ antennas, $N = 80$, snapshots and $K = 3$ sources with a separation of $\Delta_\theta = 10^\circ$. Fig. 5.6 shows the subspace distance and the RMSE curves vs. the SNR when the number of sampled antennas is $L_s = 50$, $L_s = 25$, or $L_s = 12$. Obviously, the best performance is achieved when all sensors are sampled. Nevertheless, performance degrades smoothly with L_s and hence both the hardware costs and energy consumption can be substantially reduced with only a minor performance degradation. As we increase L_s, we observe more entries of the data matrix and the MSE of the SIMC method approaches the CRLB.

Experiment 6: Fig. 5.7 shows results for the same scenario when the number of sampled sensors is $L_s = \lfloor \frac{M(100-P_s)}{100} \rfloor$, where P_s is the percentage of missing data and $\lfloor \cdot \rfloor$ is the floor function. It can be observed in Fig. 5.7 that the performance of SIMC is robust against missing data, providing satisfactory performance for $P_s < 70\%$. The results of Fig. 5.7 allow us to conclude that to obtain accurate signal subspace and DOA estimates it is important to exploit in the reconstruction of the data matrix both its low-rank structure and its shift-invariant structure. When exploited independently, the shift-invariant structure (OSE) provides more benefits than the low-rank structure (MC) for $P_s < 50\%$.

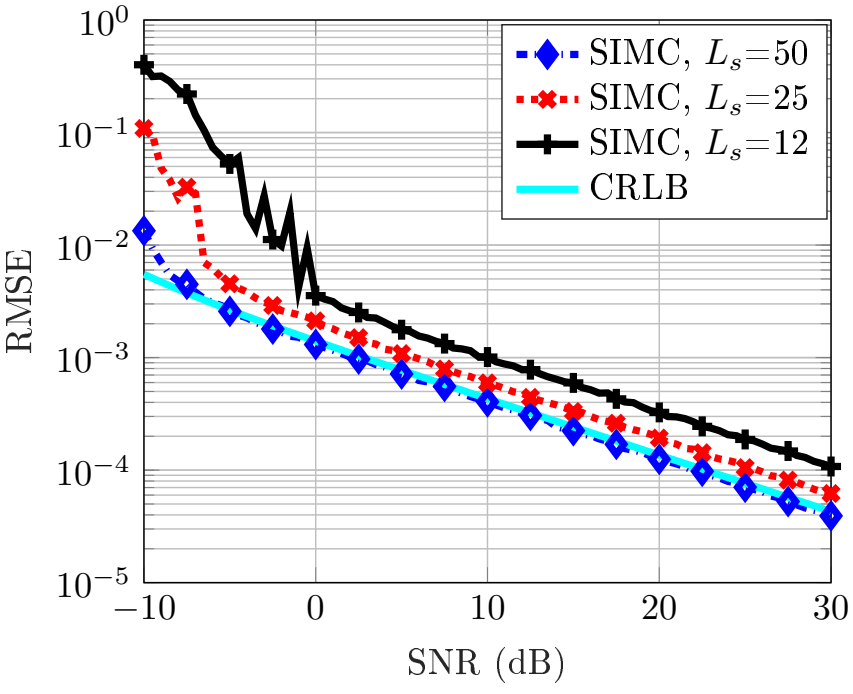
Experiment 7: In the last experiment, we evaluate the impact of having correlated sources. We consider a scenario with $K = 2$ correlated sources when $M = 100$, $N = 80$, SNR$= 0$ dB, $\Delta_\theta = 5^\circ$, and $L = 25$. The correlation coefficient between the two sources, ρ, varies from 0 (uncorrelated) to 1 (fully correlated). As Fig. 5.8 shows, SIMC outperforms the rest of methods and provides accurate DOA estimates even for highly correlated sources $\rho < 0.8$. Nevertheless, the performance of SIMC under correlated sources needs additional theoretical analysis.

5.7.2 Application to Source Enumeration

This subsection demonstrates the performance of the proposed order estimation criterion with missing data, which is referred as OE-SIMC here. As discussed in Section 5.6, no standard method provides a consistent solution with missing data, so we use the SA method, which is described in Algorithm 2 for comparison. For SIMC, we use $p_{max} = \frac{M}{5}$ and the rest of parameters are the same as in the previous subsection. It is noted here that $\mathbf{X}_m$ or $\hat{\mathbf{R}}_m$, without using any matrix completion method is used as input for both methods.

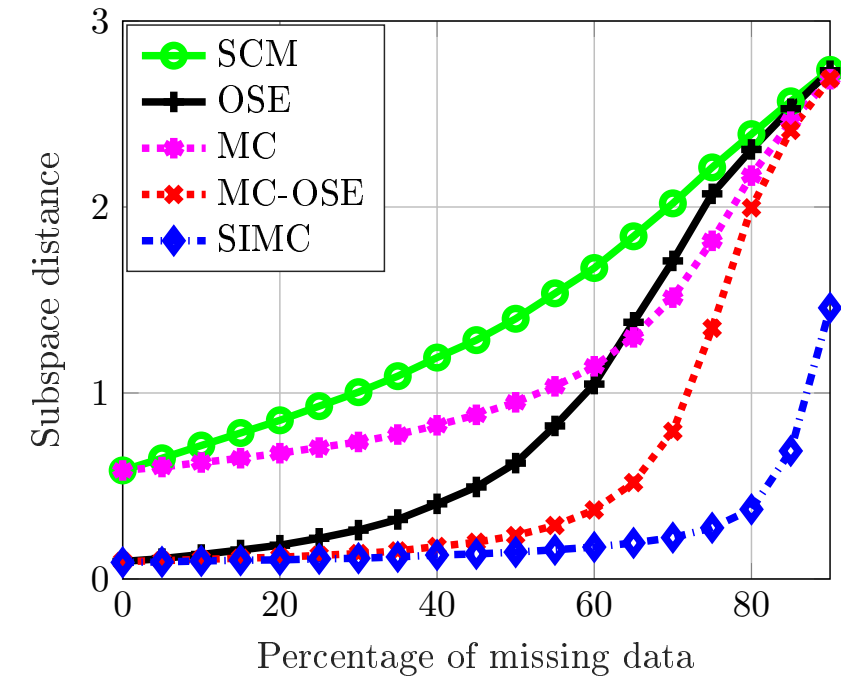


(a)

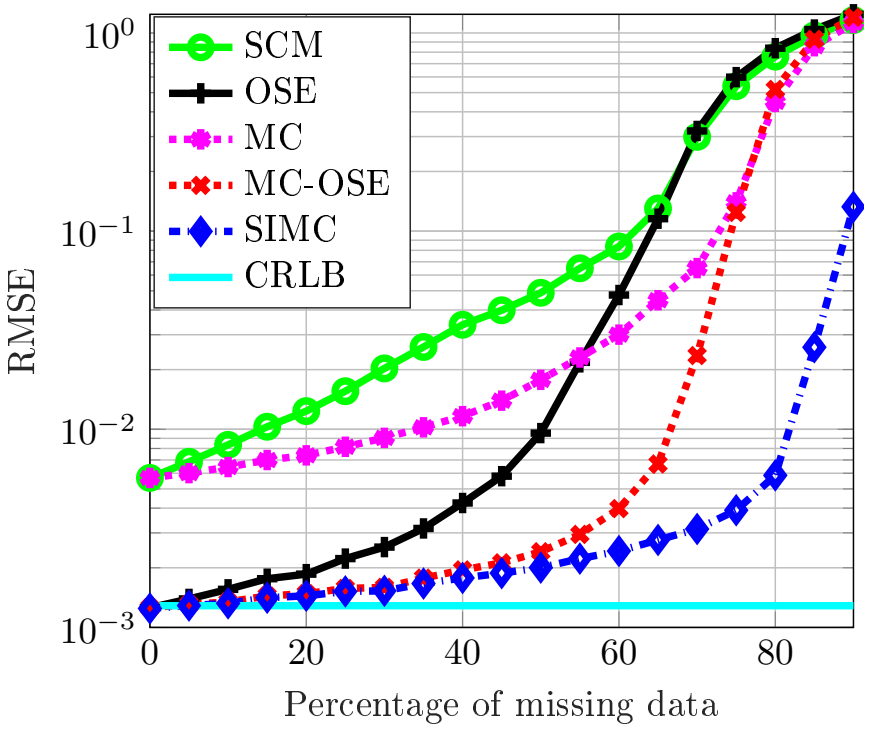


(b)

Figure 5.6: Subspace distance (a) and RMSE (b) vs. SNR when $M = 50$, $N = 80$, $K = 3$ and $\Delta_\theta = 10^\circ$ for $L_s = (50, 25, 12)$.

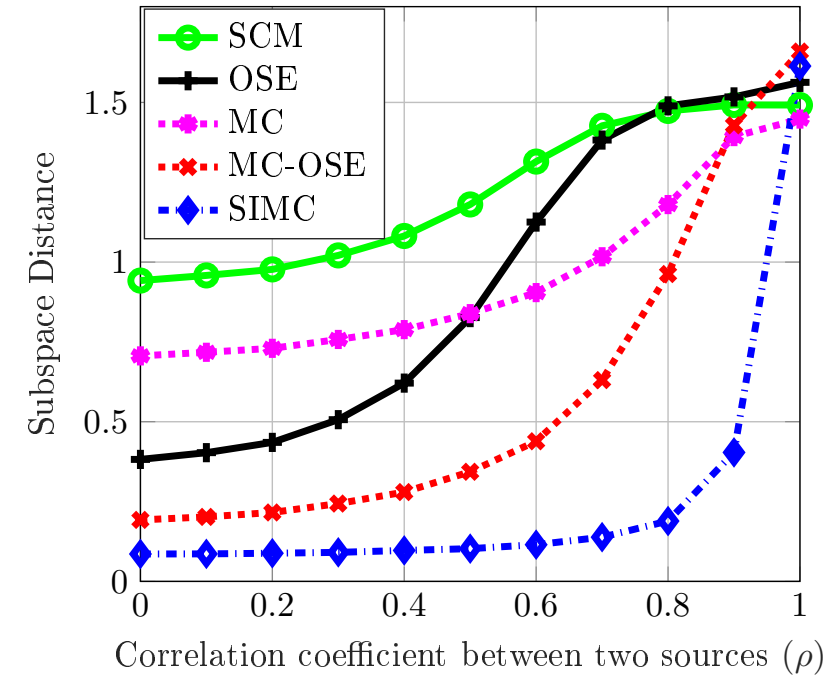


(a)

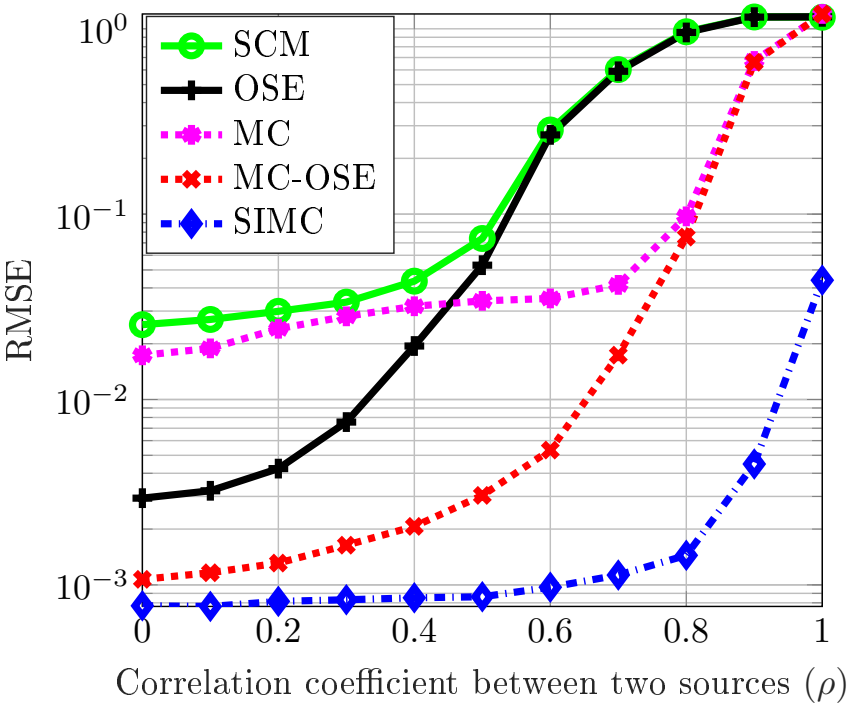


(b)

Figure 5.7: Subspace distance (a) and RMSE (b) vs. Percentage of missing data when $M = 50$, $N = 50$, $K = 5$, $\Delta_\theta = 10°$, $SNR = 5$ dB and equal-power sources.



(a)



(b)

Figure 5.8: Subspace distance (a) and RMSE (b) vs. correlation coefficient between two sources ρ for $M = 100$, $N = 80$, $K = 2$, SNR$= 0$ dB, $\Delta_\theta = 5°$, $L = 25$.

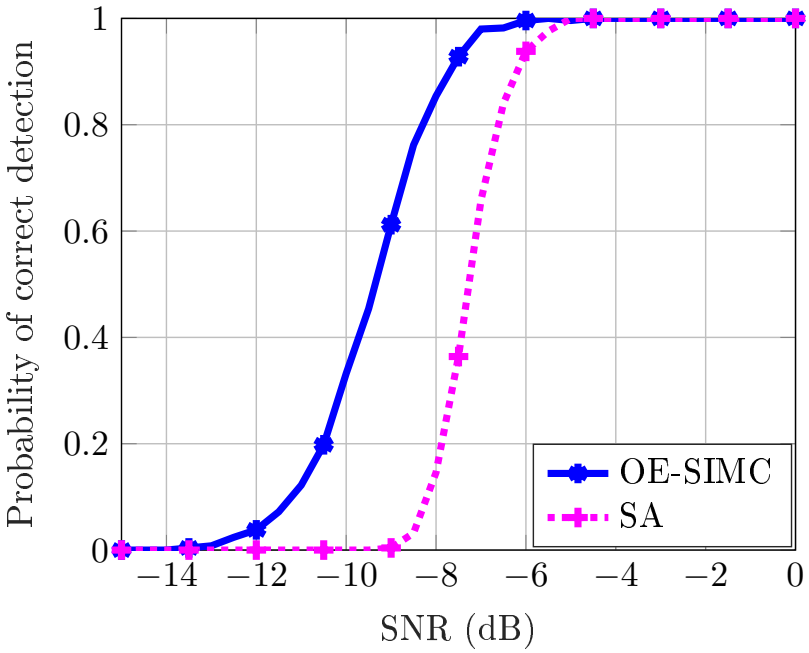


Figure 5.9: Probability of correct detection vs. SNR for $M = 100$, $N = 80$, $K = 4$, $\Delta_\theta = 10^\circ$, $L_s = 50$ and equal-power sources.

Experiment 1: First experiment shows the probability of correct detection vs. SNR for a scenario of $M = 100$, $N = 80$, $K = 4$, $\Delta_\theta = 10^\circ$ and $L_s = 50$, and the result is shown in Fig. 5.9. To our surprise, SA performs well in this scenario when only 50% entries in the data matrix are available. The possible reason behind this behaviour is that the missing entries affect both the signal subspace and noise subspace and hence the averaging mechanism of SA is able to distinguish the two subspaces. OE-SIMC also provides a consistent result, thanks to the enforcement of the shift-invariance property into MC method.

Experiment 2: Fig. 5.10 shows the probability of correct detection vs. SNR for $M = 100$, $N = 50$, $K = 3$, $\Delta_\theta = 8^\circ$ and $L_s = 25$, i.e. only 25% sensors are sampled for each snapshot. A very similar result is obtained in this scenario. SA in general performs well but becomes slightly inconsistent at higher SNRs. Nevertheless, OE-SIMC is a clear winner here as well.

Making the signal powers unequal as $\mathbf{\Psi} = \sigma_s^2(1, 0.8, 0.5)$ and keeping fixed the rest of the parameters, we obtain the results shown in Fig. 5.11. For this scenario, where sources have unequal powers, the performances of both methods are almost the same. This shows the robustness of both methods with unequal power sources.

Experiment 3: In the next experiment, probability of correct detection is evaluated with respect to the percentage of missing data. Like previous subsection, here as well the number of sampled sensors is $L_s = \lfloor \frac{M(100-P_s)}{100} \rfloor$, where P_s is the percentage of missing data. We consider a scenario with $M = 100$, $N = 150$, $K = 5$, $\Delta_\theta = 5^\circ$, and SNR$= 0$ dB. The percentage of missing data is varied between 0% and 95%. As

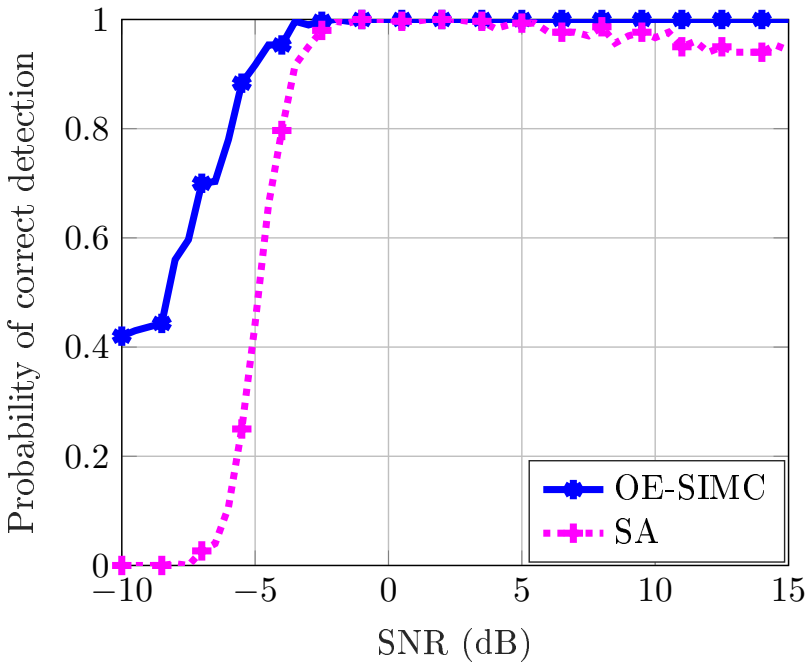


Figure 5.10: Probability of correct detection vs. SNR for $M = 100$, $N = 50$, $K = 3$, $\Delta_\theta = 8^\circ$, $L_s = 25$ and equal-power sources.

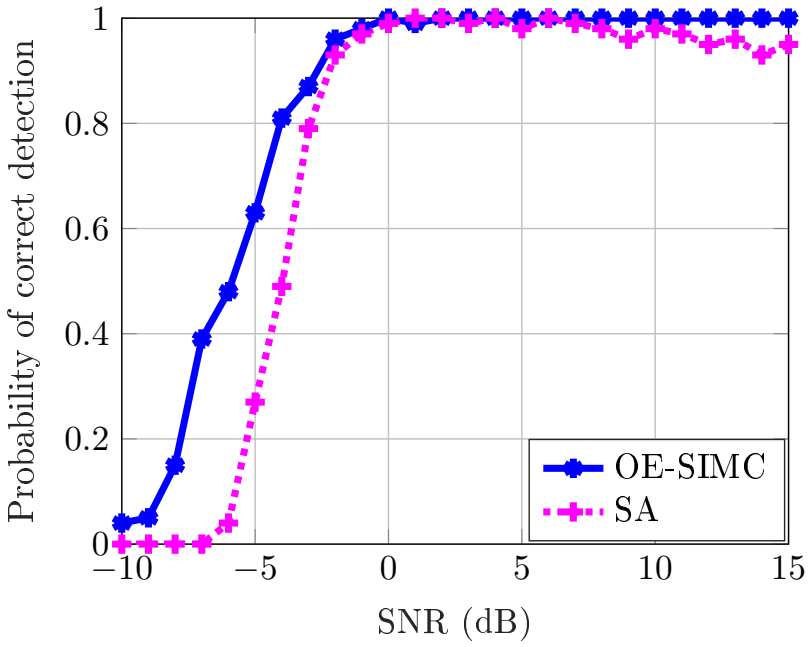


Figure 5.11: Probability of correct detection vs. SNR for $M = 100$, $N = 50$, $K = 3$, $\Delta_\theta = 8^\circ$, $L_s = 25$ and $\mathbf{\Psi} = \sigma_s^2(1, 0.8, 0.5)$.

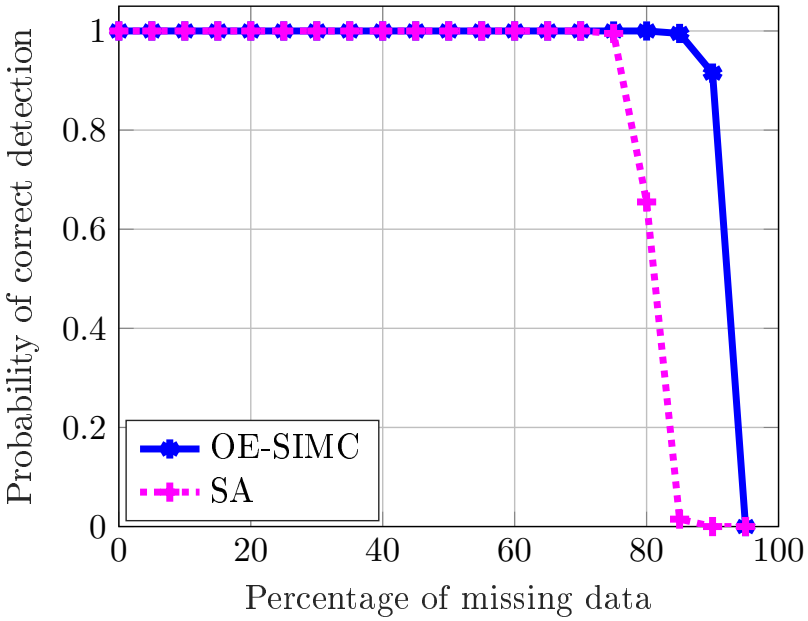


Figure 5.12: Probability of correct detection vs. percentage of missing data for $M = 100$, $N = 150$, $K = 5$, $\Delta_\theta = 5°$, $SNR = 0$ dB and equal-power sources.

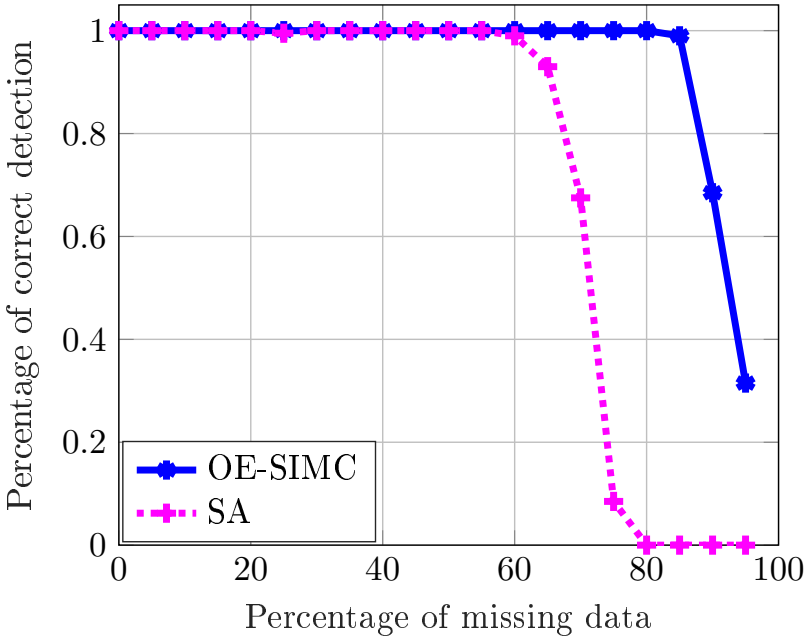


Figure 5.13: Probability of correct detection vs. percentage of missing data for $M = 50$, $N = 50$, $K = 5$, $\Delta_\theta = 10°$, $SNR = 20$ dB and equal-power sources.

it can be observed in Fig. 5.12, the SA performs generally well for a wide range of percentage of missing data, but OE-SIMC outperforms SA and provides good results for very few available data. It is noted here that if more than 95% data is missing, both methods fails.

Now the parameters are changed to $M = 50$, $N = 50$, $K = 5$, $\Delta_\theta = 10^\circ$ and SNR= 20 dB, and the results are shown in Fig. 5.13. In this experiment where a small array is used, both methods provide a robust result in the small sample regime. SA performs robustly till the 60% data are missing and OE-SIMC performs well even if more than 80% data are missing. These two examples shows a clear improvement of OE-SIMC over SA.

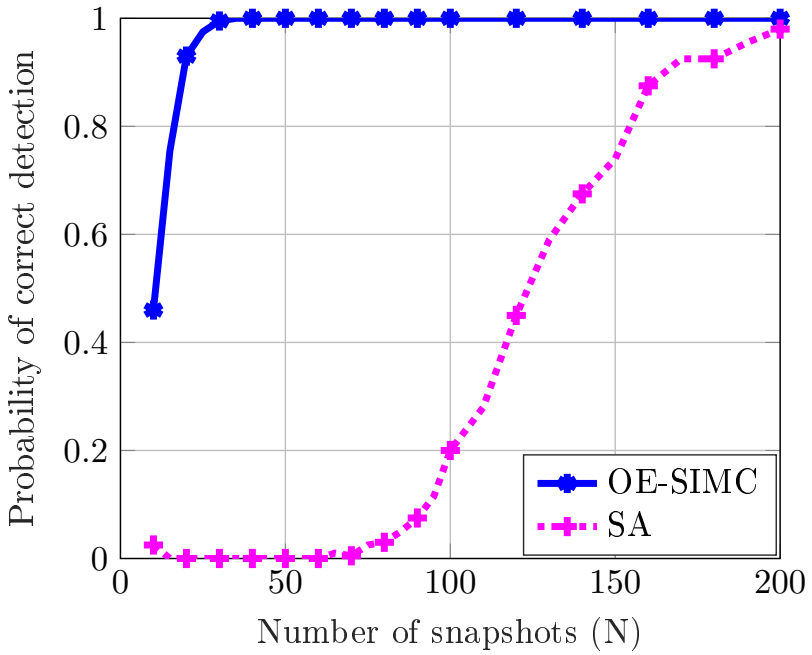


Figure 5.14: Probability of correct detection vs. number of snapshots (N) for $M = 50$, $K = 6$, $\Delta_\theta = 10^\circ$, $L_s = 13$, $SNR = 20$ dB and equal-power sources.

Experiment 4: Now we study the effect of the number of snapshots on the probability of correct detection. We assume that $K = 6$, $\Delta_\theta = 10^\circ$ and SNR $= 20$ dB. We consider two different scenarios with $M = 50$ and $M = 100$, and in both cases only 25% data is available, therefore the number of sampled sensors for each snapshot is $L_s = 13$ and $L_s = 25$, respectively. N is varied from 10 to 200. As Fig. 5.14 suggests, for $M = 50$ and $L_s = 13$, the SA does not perform well in small sample regime, however OE-SIMC performs well with very few snapshots. When we increase the number of sensors in the array to $M = 100$, SA starts performing well in small sample regime for the large-scale array. OE-SIMC is a clear winner here as well. This experiment clearly shows that OE-SIMC performs well in small sample regime with very small value of L_s.

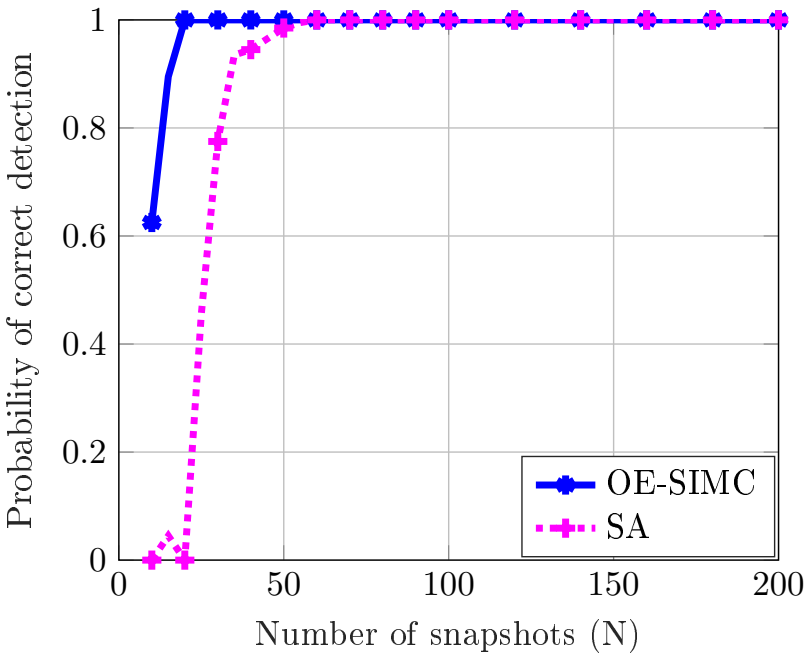


Figure 5.15: Probability of correct detection vs. number of snapshots (N) for $M = 100$, $K = 6$, $\Delta_\theta = 10^\circ$, $L_s = 25$, $SNR = 20$ dB and equal-power sources.

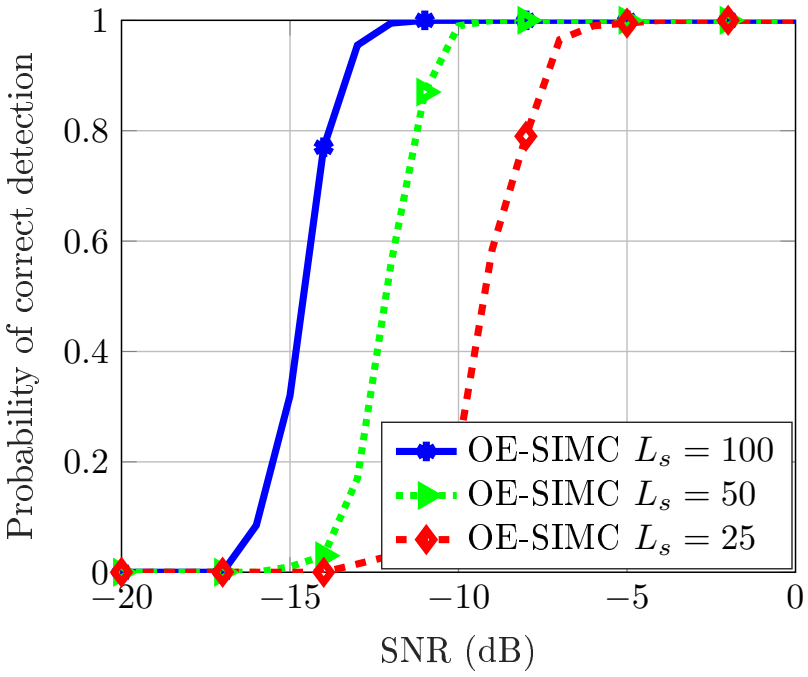


Figure 5.16: Probability of correct detection vs. SNR for $M = 100$, $N = 120$, $K = 3$, and $\Delta_\theta = 10^\circ$, $L_s = (100, 50, 25)$ and equal-power sources.

Experiment 5: The last experiment analyzes the impact of the missing data on the proposed OE-SIMC method. Here, probability of correct detection is shown with respect to SNR for $M = 100$, $N = 120$, $K = 3$, and $\Delta_\theta = 10^\circ$ for different values of L_s for OE-SIMC. As it can be observed from Fig. 5.16, the best performance is achieved when all sensors are sampled, as expected. As L_s decreases, the OE-SIMC method starts performing poorly at lower SNRs. Clearly, L_s affects the performance of the proposed method, nevertheless, for high enough SNR, it provides a consistent solution for the missing data scenario.

5.8 Conclusion

The high hardware complexity and energy consumption of massive MIMO systems is a challenge for its fully-digital implementation. A solution is to reduce the number of RF chains by performing random antenna selection techniques, which result in a data matrix with multiple missing entries. In this chapter we have proposed a matrix completion technique tailored to this array processing architecture. The reconstruction algorithm exploits both the low-rank structure of the partially observed matrix and the shift-invariance property of uniform linear arrays. We showed with the help of simulation results that as long as the number of RF chains is sufficiently larger than the number of sources, the proposed shift-invariant matrix completion (SIMC) method provides a substantial reduction of hardware costs and energy consumption without significant performance loss in resolution or DOA estimation accuracy. Furthermore, SIMC is extended to design a novel order estimation criterion for missing data scenario. The SIMC is used for increasing values of dimension and order is estimated by evaluating chordal subspace distances between two submatrices generated from the reconstructed matrix. The simulation results show that proposed method is consistent and provides good results even if the percentage of available data is very low.

Chapter 6
Conclusions and Further Lines

This dissertation addressed the problem of order estimation for large-scale arrays in low sample support scenarios. To tackle this problem we considered two approaches: i) order estimation via subspace averaging, and ii) order estimation via matrix completion. First approach cast the source enumeration problem as a problem of averaging subspaces, and provided order estimation criteria based on the eigenvalues of an average projection matrix. Second approach used the matrix completion method and exploited the properties of uniform linear array to first design an order estimation criterion for uniform linear array in the presence of a noise with diagonal covariance matrix, and then proposed a method of source enumeration with missing data.

6.1 Conclusions

In the foregoing lines, we briefly summarize the contributions and conclusions corresponding to each of the studies that have been conducted in the dissertation:

- Chapter 2 proposed a subspace averaging based order estimation method for uniform linear array in the presence of the white noise. The approach was to extract a subspace from each of several subarrays, and then average these subspaces for a subspace whose dimension is the estimated number of far-field sources. This procedure normalizes scale by replacing scale-dependent covariance models by scale-invariant subspace models. The method requires no penalty terms for controlling the estimated order. The results suggest that the problem of source enumeration may be viewed as a problem of identifying an approximating subspace, and its dimension, from a set of subspaces estimated from measurements.

- Chapter 3 extended the SA order estimation criterion for: i) non-white noises, and ii) non-uniform linear arrays. A majority vote approach was combined with the SA in order to estimate number of sources for uniform linear array in the presence of non-white noises in small sample regime. The method applies SA technique for increasing dimensions of the extracted subspaces. For each dimension, we get an estimate of the number of sources, and the final estimate is obtained by a majority vote rule.

- To extend the SA for arrays with arbitrary geometries, SA is combined with a sparse reconstruction (SR) step. The collection of subspaces to be averaged is generated by obtain a sparse representation of each snapshot as a linear combination of the others. A generalization of the log-surrogate of the ℓ_0-norm is proposed to perform the SR, which is solved using the majorization-minimization approach. Then, based on the sparse coefficients of the reconstruction, a sampling mechanism is presented to obtain projection matrices that share a large common portion of the signal subspace. Finally, the eigenvalues of an average projection matrix are then used to estimate the number of sources.

- Chapter 4 proposed a matrix completion based order estimation method for uniform linear array, when the noise covariance is diagonal with unknown entries. The most-noisy diagonal terms of the sample covariance matrix are eliminated and the off-diagonal terms are further denoised by Toeplitz rectification. The low-rank and Toeplitz signal covariance matrix is then reconstructed with matrix completion techniques. The proposed order estimation criterion is based on the Frobenius norm of the reconstructed matrices, which are reconstructed for increasing values of its rank. We have shown that the proposed method outperforms other methods for large and small arrays in small sample regime.

- Finally, Chapter 5 proposed to work with a reduced number of radio-frequency (RF) chains. The hardware complexity and energy consumption of massive MIMO systems is a challenge. A solution is to reduce the number of RF chains by performing random antenna selection techniques, which result in a data matrix with multiple missing entries. A matrix completion technique is then used to recover these missing entries. Matrix completion technique not only exploits the low rank of the noiseless data matrix, but also exploits the shift invariance property of uniform linear arrays. We showed that a better signal subspace was estimated by proposed matrix completion methods, which can be used further for DOA estimation. Then, based on the proposed matrix completion algorithm, an order estimation criterion is proposed for missing data scenario, which is based on the chordal subspace distance of two submatrices extracted from the reconstructed matrix for increasing values of its rank. We showed that the proposed method can estimate the number of sources accurately even when very few data is available.

6.2 Further Lines

This dissertation discussed the order estimation problem and variety of solutions are proposed. However, no research is ever quite complete. Here, we list some research lines that could complement the present work.

- In this dissertation, we reviewed several existing methods for order estimation and proposed some novel approaches which were based on: i) subspace averaging and ii) matrix completion. However, snapshots or subspaces are generally

received sequentially in practical problems, and most methods process the received data in batches. In this case, a large memory is required to store all the data and the computational cost is also high. The solution to this problem is to process the data sequentially as it is received. To this end, an online algorithm is needed so that the proposed methods can work sequentially.

- In this dissertation we worked with linear arrays. For instance, Chapter 2 discussed the order estimation problem for uniform linear array, and shift invariance property was exploited. The shift invariance property would exist for uniform rectangular arrays as well, and subspace averaging technique might be useful to design an order estimation criterion for uniform rectangular 2-D arrays. It would also be interesting to study uniform circular arrays. Conclusively, a thorough study of the performance of subspace averaging technique for planar arrays is required.

- In Chapter 4, a matrix completion approach is used with Toeplitz rectification to denoise a covariance matrix whose noise covariance matrix was diagonal with unknown elements. This approach can be extended to other noise models. For example, if the noise is exponentially correlated, the entries close to the main diagonal of the covariance matrix would be affected more with the noise. Nevertheless, the signal covariance matrix always possesses the Toeplitz structure. This suggests us to extend the Toeplitz matrix completion method to other spatially correlated noise models.

- Chapter 5 proposed an order estimation technique with missing data. The proposed method performs consistently, however, the computational cost of the method is high. In addition, the proposed method is restricted to work with uniform linear array in the presence of white noise. At this point we believe that a more in-depth study is required for missing data scenarios, considering other noise models or other array geometries.

- We observed in Chapter 5 that the subspace averaging method performs robustly with missing data. However, other methods that use the eigenvalues of the sample covariance matrix to estimate the number of sources do not provide consistent results. There is a need to study further the effect of missing entries on the eigenvalues and the eigenvectors of the data matrix or sample covariance matrix.

- In the end, it is important to point it out that we assumed Gaussian noises throughout the dissertation. The performances of the proposed methods for non-Gaussian noises are further required to be studied. It is understandable that some of these methods might not perform well as they are designed for Gaussian noises. To this regard, an order estimation criterion in presence of non-Gaussian noises by using subspace averaging or matrix completion approaches in the challenging scenarios of massive MIMO systems with low sample support is required to be designed.

Conclusiones y Líneas Futuras

En esta tesis, se ha tratado el problema de la estimación de orden para arrays de gran tamaño en escenarios con pocas muestras. Para abordar este problema, se han considerado dos enfoques: i) estimación de orden mediante el promediado de subespacios, y ii) estimación de orden vía matrix completion. Con el primer enfoque, se ha planteado el problema de enumeración de fuentes como un problema de promediado de subespacios y se han propuesto criterios de estimación de orden basados en los autovalores de una matriz de proyección promedio. En el segundo enfoque, se ha utilizado el método matrix completion y se han aprovechado las propiedades de los arrays lineales y uniformes para diseñar primero un criterio de estimación de orden para este tipo de arrays en presencia de un ruido con matriz de covarianza diagonal y después proponer un método de enumeración de fuentes con datos faltantes.

Conclusiones

En las siguientes líneas, se resumen brevemente los aportes y conclusiones correspondientes a cada una de las investigaciones que se han realizado en esta tesis:

- En el capítulo 2, se ha propuesto un método de estimación de orden basado en el promediado de subespacios para arrays lineales y uniformes en presencia de ruido blanco. Esta estrategia consiste en extraer un subespacio de cada uno de los varios subarrrays para luego promediarlos en un subespacio cuya dimensión es el número estimado de fuentes de campo lejano. Este procedimiento lleva a cabo la normalización de escala reemplazando los modelos de covarianza dependientes de ella por modelos de subespacios invariantes a la escala. Este método no requiere términos de penalización para controlar el orden estimado. Los resultados obtenidos sugieren que el problema de la enumeración de fuentes puede verse como un problema de identificación de un subespacio aproximado y su dimensión a partir de un conjunto de subespacios estimados a partir de medidas.

- En el capítulo 3, se ha extendido el criterio de estimación de orden SA a: i) ruidos no blancos y ii) arrays lineales no uniformes. Se ha combinado una estrategia de voto mayoritario con el SA para estimar el número de fuentes para un array lineal y uniforme en presencia de ruidos no blancos en regímenes con pocas muestras. Este método aplica la técnica SA para aumentar las dimensiones de los subespacios extraídos. Para cada dimensión, se obtiene una

estimación del número de fuentes y la estimación final se obtiene mediante una regla de voto mayoritario.

- Para extender el SA a arrays con geometrías arbitrarias, este se ha combinado con una fase de reconstrucción dispersa (SR). La colección de subespacios a promediar se genera obteniendo una representación dispersa de cada snapshot como combinación lineal de las demás. Se ha propuesto una generalización de la sustitución basada en logaritmos de la norma ℓ_0 para realizar la SR, que se resuelve utilizando la técnica de mayorización-minimización. Después, en base a los coeficientes dispersos de la reconstrucción, se ha presentado un mecanismo de muestreo para obtener matrices de proyección que comparten una gran porción común con el subespacio de la señal. Finalmente, los valores propios de la matriz de proyección promedio se utilizan para estimar el número de fuentes.

- En el capítulo 4, se ha propuesto un método de estimación de orden basado en la técnica de matrix-completion para arrays lineales y uniformes cuando la covarianza de ruido es una matriz diagonal con entradas desconocidas. Los términos diagonales más ruidosos de la matriz de covarianza de la muestra son eliminados y a los términos de fuera de la diagonal principal se les reduce el ruido mediante la rectificación de Toeplitz. La matriz de covarianza Toeplitz de bajo rango es reconstruida después con técnicas matrix-completion. El criterio de estimación de orden propuesto está basado en la norma Frobenius de las matrices reconstruidas para valores crecientes de su rango. Se ha demostrado que el método propuesto supera a otros métodos para arrays de pequeño y gran tamaño en regímenes con pocas muestras.

- Finalmente, en el capítulo 5 se ha propuesto trabajar con un número reducido de cadenas de radiofrecuencia (RF). La complejidad del hardware y el consumo de energía de los sistemas MIMO masivos constituyen un auténtico desafío. Una posible solución consiste en reducir el número de cadenas de RF mediante la realización de técnicas de selección aleatoria de antenas, que dan como resultado una matriz de datos con múltiples entradas faltantes. A continuación, se utiliza una técnica matrix-completion para recuperar estas entradas faltantes. La técnica matrix-completion no solo explota el bajo rango de la matriz de datos sin ruido, sino que también aprovecha la propiedad de invariancia de desplazamiento de los arrays lineales y uniformes. Se ha demostrado que se estima un mejor subespacio de señal mediante los métodos matrix-completion propuestos, pudiendo ser utilizados además para la estimación de DOA. A continuación, basado en el algoritmo matrix-completion propuesto, se ha propuesto un criterio de estimación de orden para escenarios con datos faltantes que se basa en la distancia cordal entre dos submatrices extraídas a partir de la matriz reconstruida para valores crecientes de su rango. Se ha demostrado que el método propuesto es capaz de estimar el número de fuentes con precisión incluso cuando hay muy pocos datos disponibles.

Líneas futuras

En esta tesis, se ha tratado el problema de la estimación de orden y se han propuesto varias soluciones al mismo. Sin embargo, ninguna línea de investigación está del todo completa. A continuación, se enumeran algunas líneas de investigación que podrían complementar el presente trabajo.

- En esta tesis, se han revisado varios métodos existentes para la estimación de orden y se han propuesto algunos enfoques novedosos basados en: i) promediado de subespacios y ii) técnicas matrix-completion. Sin embargo, los snapshots o subespacios son recibidos generalmente de forma secuencial en la práctica y la mayoría de los métodos propuestos procesan los datos recibidos por lotes. En estos casos, se requiere una gran capacidad de almacenamiento para guardar todos los datos y el costo computacional también es alto. La solución a este problema consistiría en procesar los datos secuencialmente a medida que se van recibiendo. Para ello, sería necesario un algoritmo online para que los métodos propuestos puedan trabajar de forma secuencial.

- En esta tesis, se ha trabajado con arrays lineales. Por ejemplo, en el Capítulo 2 se ha discutido el problema de la estimación de orden para arrays lineales y uniformes y se ha aprovechado la propiedad de invariancia de desplazamiento. La propiedad de invariancia de desplazamiento también es válida para arrays rectangulares uniformes y la técnica de promediado de subespacios podría ser útil para diseñar un criterio de estimación de orden para arrays rectangulares 2-D uniformes. También sería interesante estudiar arrays circulares uniformes. En conclusión, se requiere un estudio exhaustivo del rendimiento de la técnica de promediado de subespacios para arrays planos.

- En el capítulo 4, se usan técnicas matrix-completion con rectificación de Toeplitz para eliminar el ruido de una matriz de covarianza cuya matriz de covarianza de ruido es diagonal con elementos desconocidos. Este enfoque puede ser extendido a otros modelos de ruido. Por ejemplo, si el ruido está correlado de manera exponencial, las entradas cercanas a la diagonal principal de la matriz de covarianza se verían más afectadas por el ruido. Sin embargo, la matriz de covarianza de la señal siempre posee la estructura de Toeplitz. Esto sugiere extender el método Toeplitz-matrix-completion a otros modelos de ruido espacialmente correlacionados.

- En el capítulo 5, se ha propuesto una técnica de estimación de orden con datos faltantes. El método propuesto funciona de manera consistente, sin embargo, el coste computacional del método es bastante alto. Además, el método propuesto está restringido a trabajar con arrays lineales y uniformes en presencia de ruido blanco. Llegados a este punto, se cree que se requiere un estudio más profundo para escenarios con datos faltantes, considerando otros modelos de ruido u otras geometrías de array.

- En el capítulo 5, también se ha observado que el método de promediado de subespacios funciona de manera robusta cuando faltan algunos datos. Sin embargo, otros métodos que utilizan los autovalores de la matriz de covarianza de la muestra para estimar el número de fuentes no proporcionan resultados consistentes. Sería necesario estudiar más a fondo el efecto de las entradas faltantes sobre los valores y vectores propios de la matriz de datos o de la matriz de covarianza de la muestra.

- Finalmente, es importante señalar que se han asumido ruidos Gaussianos a lo largo de toda la tesis. Resultaría necesario estudiar más a fondo las prestaciones de los métodos propuestos para escenarios con ruidos no gaussianos. Es comprensible que algunos de estos métodos puedan no funcionar del todo bien, ya que están diseñados para ruidos gaussianos. A este respecto, se requeriría diseñar un criterio de estimación de orden en presencia de ruidos no gaussianos mediante el uso del promediado de subespacios o técnicas matrix-completion en los desafiantes escenarios de sistemas MIMO masivos con escasez de muestras.

Publications

Publications derived from this dissertation

The work developed in this thesis has produced the following publications.

Journal Articles

1. [Garg et al., 2021b] V. Garg, A. Pagès-Zamora, and I. Santamaria,“Order estimation with missing data for massive MIMO systems”, Submitted to the IEEE Signal Processing Letters, 2021.

2. [Garg et al., 2021a] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria.“DOA estimation via shift-invariant matrix completion”. Signal Processing, volume 183, 2021.

3. [Garg et al., 2019] V. Garg, I. Santamaria, D. Ramirez, and L. L. Scharf, “Subspace averaging and order determination for source enumeration,” IEEE Trans. Signal Process., vol. 67, pp. 3028–3041, 2019.

Conference Contributions

1. [Garg et al., 2021c] V. Garg, D. Ramirez and I. Santamaria,"Sparse subspace averaging for order estimation", Submitted to IEEE Statistical Signal Processing (SSP) Workshop, July 2021.

2. [Garg et al., 2020] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria, “Source enumeration via Toeplitz matrix completion,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 2020.

3. [Garg and Santamaria, 2019] V. Garg and I. Santamaria, “Source enumeration in non-white noise and small sample size via subspace averaging,” in 27th European Signal Processing Conference (EUSIPCO), A Coruña, Spain, Sep. 2019.

List of Figures

List of Algorithms

Bibliography

[Akaike, 1974] H. Akaike. "A new look at the statistical model identification". *IEEE Trans. Autom. Control*, volume 19, no. 6, pages 716–723, 1974.

[Anderson, 1963] T. W. Anderson. "Asymptotic theory for principal component analysis". *Ann. J. Math. Stat.*, volume 34, pages 122–148, 1963.

[Argyriou et al., 2008] A. Argyriou, T. Evgeniou, and M. Pontil. "Convex multi-task feature learning". *Mach. Learn.*, volume 73, no. 3, pages 243–272, 2008.

[Barron, 2017] J. T. Barron. "A more general robust loss function". *arXiv:1701.03077v5*, 2017.

[Bartlett, 1941] M. S. Bartlett. "The statistical significance of canonical correlations". *Biometrika*, volume 32, pages 29–37, 1941.

[Basri and Jacobs, 2003] R. Basri and D. W. Jacobs. "Lambertian reflectance and linear subspaces". *IEEE Trans. Pattern Anal. Mach. Intell.*, volume 2, no. 25, pages 218–233, 2003.

[Björk and Golub, 1973] A. Björk and G. H. Golub. "Numerical methods for computing angles between linear subspaces". *Mathematics of Computation*, volume 37, no. 123, pages 579–594, 1973.

[Burges, 2009] C. J. C. Burges. "Dimension reduction: a guided tour". *Foundations and trends in machine learning*, volume 2, no. 4, pages 275–365, 2009.

[Candes and Plan, 2010] E. J. Candes and Y. Plan. "Matrix completion with noise". *Proceedings of the IEEE*, volume 98, no. 6, pages 925–936, 2010.

[Candes and Wakin, 2008] E. J. Candes and M. B. Wakin. "An introduction to compressive sampling". *IEEE Signal Processing magazine*, volume 25, pages 21–30, 2008.

[Candès et al., 2008] E. J. Candès, M. B. Wakin, and S. P. Boyd. "Enhancing sparsity by reweighted ℓ_1 minimization". *Fourier*, volume 14, pages 877–905, 2008.

[Chen et al., 2019] Y. Chen, Y. Chi, J. Fan, C. Ma, and Y. Yan. "Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization". 2019.

[Chikuse, 1990] Y. Chikuse. "Distributions of orientations on Stiefel manifolds". *J. Multivariate Analysis*, volume 33, pages 247–264, 1990.

[Chikuse, 2003] Y. Chikuse. "Statistics on special manifolds.". *New York, USA: Springer-Verlag*, 2003.

[Cron and Sherman, 1962] B. F. Cron and C. H. Sherman. "Spatial-correlation functions for various noise models". *The Journal of the Acoustical Society of America*, volume 34, pages 1732–1736, 1962.

[Davis and Kahan, 1970] C. Davis and W. M. Kahan. "The rotation of eigenvectors by a pertubation". *SIAM J. Numer. Anal.*, volume 7, pages 1–46, 1970.

[Ding and Kay, 2011] Q. Ding and S. Kay. "Inconsistency of the MDL: On the performance of model order selection criteria with increasing signal-to-noise ratio". *IEEE Trans. Signal Process.*, volume 59, no. 5, pages 1959–1969, 2011.

[Donoho, 2006] D. L. Donoho. "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution". *Communications on Pure and Applied Mathematics*, volume 59, pages 797–829, 2006.

[Downs, 1972] T. D. Downs. "Orientation statistics". *Biometrika*, volume 59, pages 665–676, 1972.

[Du et al., 2015] R. Du, C. Chen, B. Yang, N. Lu, X. Guan, and X. Shen. "Effective urban traffic monitoring by vehicular sensor networks". *IEEE Trans. Veh. Technol.*, volume 64, no. 1, pages 273–286, 2015.

[Edelman et al., 1998] A. Edelman, T. Arias, and S. T. Smith. "The geometry of algorithms with orthogonality constraints". *SIAM J. Matrix Anal. Appl.*, volume 20, no. 2, pages 303–353, 1998.

[Eguizabal et al., 2019] A. Eguizabal, C. Lameiro, D. Ramírez, and P. J. Schreier. "Source enumeration in the presence of colored noise". *IEEE Signal Process. Letters*, volume 26, pages 475–479, 2019.

[Elhamifar and Vidal, 2009] E. Elhamifar and R. Vidal. "Sparse subspace clustering". In *IEEE Conf. Comp. Vision and Pattern Recog.*, pages 2790–2797. Miami, FL, USA, 2009.

[Forster, 2001] P. Forster. "Generalized rectification of cross spectral matrices for arrays of arbitrary geometry". *IEEE Trans. Signal Processing*, volume 49, pages 972–978, 2001.

[Gao et al., 2015] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson. "Multi-switch for antenna selection in massive MIMO". In *IEEE Global Communications Conference (GLOBECOM)*. San Diego (CA), USA, 2015.

[Garg et al., 2020] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria. "Source enumeration via Toeplitz matrix completion". In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain, 2020.

[Garg et al., 2021a] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria. "DOA estimation via shift-invariant matrix completion". *Signal Processing*, volume 183, 2021.

[Garg et al., 2021b] V. Garg, A. Pagès-Zamora, and I. Santamaria. "Order estimation with missing data for massive mimo systems". *Submitted to the IEEE Signal Processing Letters*, 2021.

[Garg et al., 2021c] V. Garg, D. Ramirez, and I. Santamaria. "Sparse subspace averaging for order estimation". *Submitted to the IEEE Work. Stat. Signal Process. (SSP)*, 2021.

[Garg and Santamaria, 2019] V. Garg and I. Santamaria. "Source enumeration in non-white noise and small sample size via subspace averaging". In *2019 27th European Signal Processing Conference (EUSIPCO)*. A Coruña, Spain, 2019.

[Garg et al., 2019] V. Garg, I. Santamaria, D. Ramirez, and L. L. Scharf. "Subspace averaging and order determination for source enumeration". *IEEE Trans. Signal Process.*, volume 67, pages 3028–3041, 2019.

[Geman and Reynolds, 1992] D. Geman and G. Reynolds. "Constrained restoration and the recovery of discontinuities". *IEEE Trans. Pattern Analysis and Machine Intelligence*, volume 14, no. 3, pages 367–383, 1992.

[Giménez-Febrer and Pagès-Zamora, 2017] P. Giménez-Febrer and A. Pagès-Zamora. "Matrix completion of noisy graph signals via proximal gradient minimization". In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4441–4445. IEEE, 2017.

[Gohary and Davidson, 2009] R. H. Gohary and T. N. Davidson. "Noncoherent MIMO communication: Grassmannian constellations and efficient detection". *IEEE Trans. Inf. Theory*, volume 55, no. 3, pages 1176–1205, 2009.

[Golub and Van Loan, 1996] G. H. Golub and C. F. Van Loan. *Matrix computations*. The John Hopkins University Press, Baltimore, USA, 1996.

[Grayson, 2016] C. Grayson. "Applications of optimal subspace estimation". *Open Access Master's Theses. Paper 959 http://digitalcommons.uri.edu/theses/959*, 2016.

[Hagege and Francos, 2016] R. R. Hagege and J. M. Francos. "Universal manifold embedding for geometric deformed functions". *IEEE Trans. Inf. Theory*, volume 62, no. 6, pages 3676–3684, 2016.

[Hlawatsch and Kozek, 1994] F. Hlawatsch and W. Kozek. "Time-frequency projection filters and time-frequency signal expansions". *IEEE Trans. Signal Process.*, volume 42, no. 12, pages 3321–3334, 1994.

[Hoff, 2013] P. D. Hoff. "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate data". *arXiv:0712.4166*, 2013.

[hu et al., 2018] Y. hu, X. Liu, and M. Jacob. "A generalized structured low-rank matrix completion algorithm for MR image recovery". *IEEE transactions on medical imaging*, volume 38, pages 1841–1851, 2018.

[Huang and So, 2013] L. Huang and H. C. So. "Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix". *IEEE Trans. Signal Process.*, volume 61, no. 19, pages 4806–4821, 2013.

[Huang et al., 2016] L. Huang, Y. Xiao, H. C. So, and J.-K. Zhang. "Bayesian information criterion for source enumeration in large-scale adaptive antenna array". *IEEE Trans. Veh. Technol.*, volume 65, no. 5, pages 3018–3032, 2016.

[Jiang and Ingram, 2004] J. S. Jiang and M. A. Ingram. "Robust detection of number of sources using the transformed rotational matrix". In *IEEE Wireless Comm. and Networking Conf. (WCNC)*. Atlanta, GA, USA, 2004.

[Joreskog, 1967] K. G. Joreskog. "Some contributions to maximum likelihood factor analysis". *psychometrika*, volume 32, pages 443–482, 1967.

[Karcher, 1977] H. Karcher. "Riemannian center of mass and mollifier smoothing". *Comm. Pure Appl. Math.*, volume 5, no. 30, pages 509–541, 1977.

[Kay, 2005] S. Kay. "Exponentially embedded families: New approaches to model order estimation". *IEEE Trans. Aerosp. Electron. Syst.*, volume 41, no. 1, pages 333–345, 2005.

[Keshavan et al., 2009] R. Keshavan, A. Montanari, and S. Oh. "Matrix completion from noisy entries". 2009.

[Khatri and Mardia, 1977] C. G. Khatri and K. V. Mardia. "The von Mises-Fisher matrix distribution in orientation statistics". *J. Roy. Statist. Soc. Ser. B*, volume 39, no. 1, pages 95–106, 1977.

[Koren et al., 2009] Y. Koren, R. Bell, and C. Volinsky. "Matrix factorization techniques for recommender systems". *Computer*, volume 42, pages 30–37, 2009.

[Lawley, 1959] D. N. Lawley. "Tests of significance in canonical correlations". *Biometrika*, volume 46, pages 59–66, 1959.

[Lawley and Maxwell, 1971] D. N. Lawley and A. E. Maxwell. *Factor analysis as a statistical method*. American Elsevier Pub. Co., 1971.

[Lerman and Maunu, 2018] G. Lerman and T. Maunu. "An overview of robust subspace recovery". *arXiv:1803.01013v1*, 2018.

[Li and Vaccaro, 1991] F. Li and R. J. Vaccaro. "Unified analysis for DOA estimation algorithms in array signal processing". *Signal Process.*, volume 25, pages 147–169, 1991.

[Li et al., 2014] S. Li, S. Ma, W. He, and Y. Wang. "An improved sparse representation method for direction-of-arrival estimation in the presence of unknown mutual coupling". In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 676–679. Guilin, China, 2014.

[Li et al., 2019] X. P. Li, L. Huang, H. C. So, and B. Zhao. "A survey on matrix completion: Perspective of signal processing". *arXiv:1901.10885*, 2019.

[Liao et al., 2016] B. Liao, C. Guo, L. Huang, and J. Wen. "Matrix completion based direction-of-arrival estimation in nonuniform noise". In *2016 IEEE International Conference on Digital Signal Processing (DSP)*. Beijing, China, 2016.

[Lu and Zoubir, 2013] Z. Lu and A. M. Zoubir. "Generalized Bayesian information criterion for source enumeration in array processing". *IEEE Trans. Signal Process.*, volume 61, no. 6, pages 1470–1480, 2013.

[Malioutov et al., 2005] D. Malioutov, M. Cetin, and A. Willsky. "A sparse signal reconstruction perspective for source localization with sensor arrays". *IEEE Transactions on Signal Processing*, volume 53, pages 3010–3022, 2005.

[Mardia et al., 1979] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis*. New York: Academic, 1979.

[Marrinan et al., 2014] R. Marrinan, J. R. Beveridge, B. Draper, M. Kirby, and C. Peterson. "Finding the subspace mean or median to fit your need". In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 1082–1089. Columbus, OH, USA, 2014.

[Nadakuditi and Edelman, 2008] R. R. Nadakuditi and A. Edelman. "Sample eigenvalue based detection of high-dimensional signals in white noise with relatively few samples". *IEEE Trans. Signal Process.*, volume 56, no. 7, pages 2625–2638, 2008.

[Pal and Vaidyanathan, 2014] P. Pal and P. P. Vaidyanathan. "A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising". *IEEE Signal Processing Letters*, volume 21, no. 6, pages 737–741, 2014.

[Palka, 1996] T. A. Palka. *Bounds and algorithms for subspace estimation on Riemannian quotient submanifolds*. Ph.D. Dissertation, University of Rhode Island, 1996.

[Palka and Vaccaro, 2017] T. A. Palka and R. J. Vaccaro. "A differential geometry-based framework for constrained subspace estimation". In *Proc. 2017 Underwater Acoust. Signal Process. Work.* University of Rhode Island, 2017.

[Paulraj et al., 1986] A. Paulraj, R. Roy, and T. Kailath. "A subspace rotation approach to signal parameter estimation". *Proc. of the IEEE*, volume 74, pages 1044–1045, 1986.

[Pesavento and Gershman, 2001] M. Pesavento and A. B. Gershman. "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise". *IEEE Trans. Signal Processing*, volume 49, no. 7, pages 1310–1324, 2001.

[Quinlan et al., 2007] A. Quinlan, J. P. Barbot, P. Larzabal, and M. Haardt. "Model order selection for short data: an exponential fitting test (EFT)". *EURASIP J. Adv. Signal Process.*, volume 34, pages 122–148, 2007.

[Ramirez et al., 2011] D. Ramirez, G. Vázquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria. "Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas". *IEEE Trans. Signal Process.*, volume 59, no. 8, pages 3764–3774, 2011.

[Ramlatchan et al., 2018] A. Ramlatchan, M. Yang, Q. Liu, M. Li, J. Wang, and Y. Li. "A survey of matrix completion methods for recommendation systems". *Big Data Mining and Analytics*, volume 1, no. 4, pages 308–323, 2018.

[Rappaport et al., 2015] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock. *Millimeter wave wireless communications*. Prentice Hall, 2015.

[Rappaport et al., 2013] T. S. Rappaport et al. "Millimeter wave mobile communications for 5G cellular: It will work!". *IEEE Access*, volume 1, pages 335–349, 2013.

[Rissanen, 1978] J. Rissanen. "Modeling by shortest data description". *IEEE Trans. Acoust., Speech, Signal Process.*, volume 14, pages 465–471, 1978.

[Roy and Kailath, 1989] R. Roy and T. Kailath. "ESPRIT- estimation of signal parameters via rotational invariance techniques". *IEEE Trans. on Acoust. Speech and Signal Process.*, volume 37, no. 7, pages 984–995, 1989.

[Rusek et al., 2013] F. Rusek et al. "Scaling up MIMO: Opportunities and challenges with very large arrays". *IEEE Signal Processing Magazine*, volume 30, pages 40–60, 2013.

[Santamaria et al., 2018] I. Santamaria, D. Ramirez, and L. L. Scharf. "Subspace averaging for source enumeration in large arrays". In *Proc. IEEE Work. Stat. Signal Process. (SSP)*. Freiburg, Germany, 2018.

[Santamaria et al., 2016] I. Santamaria, L. L. Scharf, M. Kirby, C. Peterson, and J. Francos. "An order fitting rule for optimal subspace averaging". In *Proc. IEEE Work. Stat. Signal Process. (SSP)*. Palma de Mallorca, Spain, 2016.

[Santamaria et al., 2017] I. Santamaria, L. L. Scharf, J. Via, Y. Wang, and H. Wang. "Passive detection of correlated subspace signals in two MIMO channels". *IEEE Trans. Signal Process.*, volume 65, no. 20, pages 5266–5280, 2017.

[Scharf, 1991] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Addison-Wesley, Reading, MA, 1991.

[Scharf and Friedlander, 1994] L. L. Scharf and B. Friedlander. "Matched subspace detectors". *IEEE Trans. Signal Process.*, volume 42, no. 8, pages 2146–2157, 1994.

[Schwarz, 1978] G. Schwarz. "Estimating the dimension of a model". *The Annals of Statistics*, volume 6, no. 2, pages 461–464, 1978.

[Slepian and Pollack, 1961] D. Slepian and H. O. Pollack. "Prolate spheroidal wave functions, fourier analysis and uncertainty, i". *The Bell System Technical Journal*, volume 40, no. 1, pages 43–63, 1961.

[Song et al., 2016] Y. Song, P. J. Schreier, D. Ramirez, and T. Hasija. "Canonical correlation analysis of high-dimensional data with very small sample support". *Signal Processing*, volume 128, pages 449–458, 2016.

[Srebro et al., 2005] N. Srebro, J. Rennie, and T. Jaakkola. "Maximum-margin matrix factorization". In *Advances in Neural Information Processing Systems*, pages 1329–1336. Vancouver, Canada, 2005.

[Srivastava and Klassen, 2002] A. Srivastava and E. Klassen. "Monte Carlo extrinsic estimators of manifold-valued parameters". *IEEE Trans. Signal Process.*, volume 50, no. 2, pages 299–308, 2002.

[Stoica and Cedervall, 1997] P. Stoica and M. Cedervall. "Detection tests for array processing in unknown correlated noise fields". *IEEE Trans. Signal Processing*, volume 45, no. 9, pages 2351–2362, 1997.

[Stoica and Nehorai, 1990] P. Stoica and A. Nehorai. "Performance study of conditional and unconditional direction-of-arrival estimation". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, volume 38, no. 10, pages 1783–1795, 1990.

[Sun et al., 2017] Y. Sun, P. Babu, and P. Palomar. "Majorization-Minimization algorithms in signal processing, communications, and machine learning". *IEEE Trans. Signal Processing*, volume 65, no. 3, pages 794–816, 2017.

[Tan and Feng, 2019] W. Tan and X. Feng. "Covariance matrix reconstruction for direction finding with nested arrays using iterative reweighted nuclear norm minimization". *Int. Journal of Antennas and Propagation*, 2019.

[Tibshirani, 1996] R. Tibshirani. "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society B*, volume 58, pages 267–288, 1996.

[Trees, 2002] H. V. Trees. *Detection, Estimation and Modulation Theory Part IV: Optimum Array Processing*. Wiley, New York, NJ, 2002.

[Turaga et al., 2011] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition". *IEEE Trans. Pattern Anal. Mach. Intell.*, volume 33, no. 11, pages 2273–2286, 2011.

[Urick, 1983] R. J. Urick. *Principles of Underwater Sound, 3rd ed.* McGraw-Hill, New York, 1983.

[Vaccaro, 2017] R. J. Vaccaro. "The role of subspace estimation in sensor array signal processing". In *Proc. 2017 Underwater Acoust. Signal Process. Work.* University of Rhode Island, 2017.

[Vaccaro and Ding, 1993] R. J. Vaccaro and Y. Ding. "Optimal subspace-based parameter estimation". In *Proc. IEEE Int. Conf. Acoustics Speech and Signal Proc. (ICASSP)*. Minneapolis, MN, USA, 1993.

[Vallet and Loubaton, 2017] P. Vallet and P. Loubaton. "On the performance of MUSIC with Toeplitz rectification in the context of large arrays". *IEEE Trans. Signal Processing*, volume 65, pages 5848–5859, 2017.

[Vazquez-Vilar et al., 2011] G. Vazquez-Vilar, D. Ramirez, R. Lopez-Valcarce, J. Via, and I. Santamaria. "Spatial rank estimation in cognitive radio networks with uncalibrated multiple antennas". In *4th International Conference on Cognitive Radio and Advanced Spectrum Management (cogART)*. Barcelona, Spain, 2011.

[Vidal, 2011] R. Vidal. "Subspace clustering". *IEEE Signal Proc. Magazine*, volume 28, no. 2, pages 52–68, 2011.

[Vidal et al., 2008] R. Vidal, R. Tron, and R. Hartley. "Multiframe motion segmentation with missing data using power factorization and GPCA". *Int. J. Comput. Vis.*, volume 79, no. 1, pages 85–105, 2008.

[Vinogradova et al., 2019] J. Vinogradova, R. Couillet, and W. Hachem. "Estimation of toeplitz covariance matrices in large dimensional regime with applications to source detection". *IEEE Trans. Signal Process.*, volume 63, pages 4903–4913, 2019.

[Wax and Kailath, 1985] M. Wax and T. Kailath. "Detection of signals by information theoretic criteria". *IEEE Trans. Acoust., Speech, Signal Process.*, volume 33, no. 2, pages 387–392, 1985.

[Weng and Wang, 2012] Z. Weng and X. Wang. "Low-rank matrix completion for array signal processing". In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2697–2700. IEEE, 2012.

[Wenz, 1972] G. M. Wenz. "Review of underwater acoustics research: Noise". *The Journal of the Acoustical Society of America*, volume 51, pages 1010–1024, 1972.

[Wu and Wong, 1994] Q. Wu and K. M. Wong. "UN-MUSIC and UN-CLE: An application of generalized correlation analysis to the estimation of the directionsl of arrival in unknown correlated noise". *IEEE Trans. Signal Processing*, volume 42, no. 9, pages 2331–2343, 1994.

[Xie et al., 2015] K. Xie, L. Wang, G. Xie, G. Zhang, D. Xie, and J. Wen. "Sequential and adaptive sampling for matrix completion in network monitoring systems". *IEEE Conference on Computer Communications (INFOCOM)*, pages 2443–2451, 2015.

[Xu and Kay, 2008] C. Xu and S. Kay. "Source enumeration via the EEF criterion". *IEEE Signal Process. Letters*, volume 15, pages 569–572, 2008.

[Yin and Chen, 2011] J. Yin and T. Chen. "Direction-of-arrival estimation using a sparse representation of array covariance vectors". *IEEE Transactions on Signal Processing*, volume 59, pages 4489–4493, 2011.

[Yu et al., 2015] Y. Yu, T. Wang, and R. J. Samworth. "A useful variant of the Davis-Kahan theorem for statisticians". *Biometrika*, volume 102, no. 2, pages 315–323, 2015.

[Zhang and Wong, 1993] Q. T. Zhang and K. M. Wong. "Information theoretic criteria for the determination of the number of signals in spatially correlated noise". *IEEE Trans. Signal Processing*, volume 41, no. 4, pages 1652–1663, 1993.

[Zheng and Tse, 2002] L. Zheng and D. N. C. Tse. "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel". *IEEE Trans. Inf. Theory*, volume 48, no. 2, pages 359–383, 2002.

[Zhu and Kay, 2018] Z. Zhu and S. Kay. "On Bayesian exponentially embeded family for model order selection". *IEEE Trans. Signal Process.*, volume 66, no. 4, pages 933–943, 2018.