Pontifícia Universidade Católica
do Rio de Janeiro

Yuneisy Esthela García Guzmán

Direction Finding Techniques Based on Compressive Sensing and Multiple Candidates

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Engenharia Elétrica da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica.

Advisor: Prof. Rodrigo Caiado de Lamare

Rio de Janeiro
March 2018

Pontifícia Universidade Católica
do Rio de Janeiro

Yuneisy Esthela García Guzmán

Direction Finding Techniques Based on Compressive Sensing and Multiple Candidates

Dissertation presented to the Programa de Pós–graduação em Engenharia Elétrica da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica. Approved by the undersigned Examination Committee.

Prof. Rodrigo Caiado de Lamare
Advisor
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Rosângela Fernandes Coelho
Instituto Militar de Engenharia – IME

Prof. João Terêncio Dias
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – CEFET/RJ

Prof. Márcio da Silveira Carvalho
Vice Dean of Graduate Studies
Centro Técnico Científico – PUC-Rio

Rio de Janeiro, March the 21st, 2018

Yuneisy Esthela García Guzmán

The author graduated in Telecommunication and Electronic Engineering from the Technology University of Havana "Jose A Echeverria", Cuba, 2014

Bibliographic data

García Guzmán, Yuneisy Esthela

Direction Finding Techniques Based on Compressive Sensing and Multiple Candidates / Yuneisy Esthela García Guzmán; advisor: Rodrigo Caiado de Lamare. – Rio de janeiro: PUC-Rio, Departamento de Engenharia Elétrica, 2018.

v., 69 f: il. ; 30 cm

Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Engenharia Elétrica – Teses. Processamento de Arranjos de Sinais; Compressed Sensing (CS); Estimação de Direção de Chegada (DoA); recuperação esparsa; Iterative Hard Thresholding (IHT) algorithm.I. Caiado de Lamare, Rodrigo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 620.11

Acknowledgments

This work would not be possible without the continuous help and guidance of my supervisor, Prof. Rodrigo C. de Lamare. I thank him for trusting in my abilities and giving me the opportunity to work in his research group. I would like to thanks also Prof. Martin Haardt for his comments about my work and his collaboration.

My sincere thanks to the CNPq, FAPERJ, and PUC-Rio, for the financial support.

I also want to say thanks to my colleagues from CETUC for all the moments that we shared together and especially to my friend Daylis.

I also would like to thank to Bruno and his parents: Marilia and Rogelio for opening the door of their house for me.

Thanks to my friends: Odette, Daylis, Clau, Laura, Thais, Piri, Zarza, Yoe, Manu, Hansel, Asiel, Lisi, Ale and Alex for being part of my live all these years.

I thank Lukas for being my friend, my partner and my support. I am very grateful for all the excellent moments that we lived together in these almost two years.

I also would like to thank my family for their infinite support. My grandparents who are my treasure and my aunts and cousins. Especially my mom and my sister for always believing in me. My mom for being a brave and wonderful woman and always encourages me to follow my dreams and my sister for being the kindest and extraordinary person that I have ever known.

Finally, to the most important man of my life, my father, who will always be in my heart.

Abstract

García Guzmán, Yuneisy Esthela; Caiado de Lamare, Rodrigo (Advisor). **Direction Finding Techniques Based on Compressive Sensing and Multiple Candidates**. Rio de Janeiro, 2018. 69p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Direction of arrival (DoA) estimation is a key area of sensor array processing which is encountered in a broad range of important engineering applications. This fact together with the development of the Compressed Sensing (CS) area in the last years are the principal motivation of this thesis. In this dissertation, a formulation of the source localization problem as a sparse signal representation problem is presented and several sparse recovery algorithms are derived and investigated for solving the current problem. The proposed algorithms are based on the incorporation of the prior information about the sparse signal in the estimation process. In the first part, we focus on the development of two Bayesian greedy algorithms which are principally based on the iterative hard thresholding (IHT) algorithm. Due to the inferior performance of the conventional DoA estimation algorithm in scenarios with correlated sources, we pay special attention to the performance of the proposed algorithms under this condition. In the second part, the optimization problem using a ℓ_1 penalty is introduced and a Bayesian algorithm for solving the basis pursuit denoising problem is presented.Simulation results shows that Bayesian estimators which take into account the prior knowledge of the signal distribution outperform and improve substantially the performance of the non-Bayesian estimators.

Keywords

Sensor array signal processing; Compressed Sensing (CS); Direction of Arrival estimation (DoA); sparse recovery; Iterative Hard Thresholding (IHT) algorithm.

Resumo

García Guzmán, Yuneisy Esthela; Caiado de Lamare, Rodrigo. **Técnicas de Estimação de Direção Baseadas em Sensoriamento Compressivo e Múltiplos Candidatos**. Rio de Janeiro, 2018. 69p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A estimação de direção de chegada (DoA) é uma importante área de processamento de arranjos de sensores que é encontrada em uma ampla gama de aplicações de engenharia. Este fato, juntamente com o desenvolvimento da área de *Compressed Sensing* (CS) nos últimos anos, são a principal motivação desta dissertação. Nesta dissertação, é apresentada uma formulação do problema de estimação de direção de chegada como um problema de representação esparsa da sinal e vários algoritmos de recuperação esparsa são derivados e investigados para resolver o problema atual. Os algoritmos propostos são baseados na incorporação da informação prévia sobre o sinal esparso no processo de estimativa. Na primeira parte, nos concentramos no desenvolvimento de dois algoritmos Bayesianos , que se baseiam principalmente no algoritmo *iterative hard thresholding* (IHT). Devido ao desempenho inferior dos algoritmos convencionais de estimação de chegada em cenários com fontes correlacionadas, nós prestamos atenção especial ao desempenho dos algoritmos propostos nesta condição. Na segunda parte, o problema de otimização baseados na minimização da norma ℓ_1 é apresentado e um algoritmo bayesiano é proposto para resolver o problema chamado *basis pursuit denoising* (BPDN). Os resultados da simulação mostram que os estimadores Bayesianos superam os estimadores não Bayesianos e que a incorporação do conhecimento prévio da distribuição do sinal melhorou substancialmente o desempenho dos algoritmos.

Palavras-chave

Processamento de Arranjos de Sinais; Compressed Sensing (CS); Estimação de Direção de Chegada (DoA); recuperação esparsa; Iterative Hard Thresholding (IHT) algorithm.

Table of contents

List of figures

List of tables

List of Abreviations

BHT – Bayesian Hard Thresholding
BPDN – Basis Pursuit Denoising
B-BPDN–Bayesian Basis Pursuit Denoising
CMV – Constrained Minimum Variance
CS – Compressed Sensing
CRB – Cramer- Rao Bound
DoA – Direction of Arrival
ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques
EVD- Eigen Value Decomposition
FBSS – Forward-Backward Spatial Smoothing
FPC – Fixed Point Continuation
FISTA – Fast Iterative Soft-thresholding
HTP – Hard Thresholding Pursuit
IHT – Iterative Hard Thresholding
LASSO-Least Absolute Shrinkage and Selection Operator
MAP – Maximum a posteriori estimator
MCN – Matrix Variate Complex Normal
ML – Maximum Likelihood
MMV – Multiple Measurement Vector
MUSIC – Multiple Signal Classification
MVDR – Minimum Variance Distortionless Response
NESTA – Nesterov's Smoothing Techniques
NP-hard – Non-deterministic Polynomial time
OMP – Orthogonal Matching Pursuit
RIP – Restricted Isometry Property
RMC-IHT – Randomized Multiple Candidate Iterative Hard Thresholding
RMSE – Root Mean Square Error
SMV – Single Measurement Vector
SNR – Signal-to-noise ratio
SPGL1 – Spectral Projected Gradient
SOC – Second Order Cone programming
SS – Spatial Smoothing
SVD – Singular Value Decomposition

The only limit to the height of your achievements is the reach of your dreams and your willingness to work hard for them

Michelle Obama

1
Introduction

Sensor array signal processing is an active area of research in the broad field of signal processing and focuses on the problem of estimating signal parameters from data collected by an array of sensors. It is a key element in many applications such as sonar, radar, exploration seismology, radio astronomy, seismology, machine condition monitoring and wireless communications.

One of the most relevant topics within array signal processing is direction of arrival estimation (DoA) or direction finding, which has as objective to determine the direction of a given signal that propagates over space and impinges on an antenna array. To this end, the spatial separation of multiple sensor elements is exploited to obtain the location of the energy-radiating source.

Another area that has also recently gained significant attraction in the signal processing community is Compressed Sensing (CS) which goes against the common knowledge in data acquisition and allows the representation of the compressible signals at a rate significantly below to Nyquist rate.

This fact has motivated the development of several approaches which casting the source localization problem as a sparse representation problem based on the principles of CS. In this thesis we analyze the problem of direction of arrival estimation from this perspective and proposed some algorithms for solving the sparse recovery problem.

1.1
Motivation

Many practical applications require that the estimates of the locations be not only accurate under ideal conditions, but also robust to factors such as measurement noise, limitations in the amount of data, correlation of the sources, and modeling errors. It is also desired that the spectra have narrow peaks, low sidelobes, and the ability to achieve superresolution [1].

Conventional DoA estimation techniques show lack of resolution under some conditions as short data records, low signal-to-noise (SNR) scenarios, correlated sources and others. Motivated by this fact, in the last years a new

approach to DoA estimation based on sparse signal representation has been introduced which achieves superresolution by exploiting sparsity.

As result of the development of this new framework based on sparse representation, the development of signal reconstruction algorithms has also evolved rapidly in the last years. Several research works have been concentrated on the design of recovery algorithms that obeys some important properties such as high speed, low storage requirements, ease of implementation, flexibility, and good recovery performance.

1.2 Dissertation Outline

This master thesis is organized as follows.

Chapter 2

In this chapter we formulate the problem of source localization using an array of sensors. We describe several existing source localization methods. We explain some of the limitations of existing techniques thus motivating the need for the source localization framework. We also introduce the basic ideas about the CS theory and present an overview about the most effective recovery algorithms described in the literature.

Chapter3

In this chapter we propose two Bayesian greedy algorithms called randomized multiple candidate iterative hard thresholding (RMC-IHT) and Bayesian iterative hard thresholding (BHT). Different to the deterministic algorithms, the proposed techniques take into account the prior knowledge about the sparse signal in the estimation process. The performance of the developed algorithms is numerically evaluated and compared with some widely-used algorithm.

Chapter4

A novel Bayesian algorithm for solving the sparse recovery problem based on ℓ_1-regularization is presented in this chapter. Bayesian basis pursuit denoising (B-BPDN) uses the prior knowledge of the sparse signal for reducing the complexity of the basis pursuit denoising problem. We numerically evaluate signal recovery performance of the B-BPDN.

Chapter 5

In this chapter, conclusions of this work are presented and future directions for this research topic are discussed.

1.3 Contributions

The main contributions of this thesis are:

- A brief review about the signal model for DoA estimation and its formulation as a sparse representation problem based on CS are presented in Chapter 2.
- Several existing theoretical results regarding the measurement system and the recovery algorithms in compressed sensing are reviewed in Chapter 2.
- A general Bayesian framework for sparse recovery algorithms, that exploits the a priori knowledge of the sparse signal have been proposed. As consequence two Bayesian algorithms applied to the direction finding problem, RMC-IHT and BHT, are developed. The derivation of the maximum a posteriori (MAP) estimator of the sparse signal and the Bayesian hard thresholding operator is explained in details. A version of the RMC-IHT algorithm for the case of correlated sources is also presented. Simulations results show the effectiveness of the proposed techniques fundamentally in scenarios with low values of SNR and a limited number of snapshots.
- A Bayesian approach to the basis pursuit denoising problem is presented in Chapter 4. The proposed B-BPDN algorithm uses the prior knowledge of the sparse signal for reducing the space of search in the ℓ_1 minimization problem. Some analysis of the computational complexity is also presented. Simulation results show that the proposed algorithm achieves a better performance than the standard recovery reconstruction algorithms and also reduce the computational complexity.
- The benefits of using a sparse regularization framework for directional of arrival estimation are illustrated. These include robustness to SNR, to limited number of samples and to correlated sources.

1.4 List of Publications

Some of the results in this dissertation have been published.
Conference Paper:

- Y. García, R. C. de Lamare and M. Haardt "Randomized Multiple Candidate Iterative Hard Thresholding algorithm for Direction of Arrival Estimation", 22nd International ITG Workshop on Smart Antennas (WSA 2018), Bochum, Germany, 2018.

2
Direction of Arrival (DoA) Estimation and Compressed Sensing (CS)

The problem of estimating the wavenumber or angle of arrival of a planewave (or multiple plane waves) is commonly referred to as the direction finding (DF), source localization or direction of arrival (DOA) estimation problem [2]. It is a topic of broad interest in a variety of fields including wireless communications, radar, sonar and seismic systems, electronic surveillance, medical imaging and treatment, seismology and radio astronomy.

With the field of applications involving DOA estimation constantly expanding, numerous direction finding techniques have been devised over the past few decades. The most well-known parameter estimation strategies can be classified into three main categories, namely conventional, subspace-based, and maximum likelihood (ML) methods. These techniques successively steer the main beam in all possible look directions and measure the output power, which is recorded in the form of a pseudo spectrum over the angle range. The largest peaks in the pseudo spectrum are associated with the DOA estimates [3].

The most prominent approach within the conventional algorithms is Capon's method [4] based on the constrained minimum variance (CMV) criterion. It minimizes the power induced by interfering signals and noise while keeping the gain towards the look direction fixed.

The class of subspace-based methods exploits the spectral decomposition of the covariance matrix for achieving high resolution. Among the most important techniques are the multiple signal classification (MUSIC) [5], its extension Root-MUSIC [6], the estimation of signal parameters via rotational invariance techniques (ESPRIT) [7] and others [8–11].

ML-type methods are based on a parametric approach. They effectively exploit the underlying data model, resulting in sufficiently high accuracy that is superior to that of conventional and subspace-based methods. However, the efficiency is at the expense of the computational intensity as a multidimensional search is required [3].

In the last years a new method for acquisition of sparse signals and reconstruction from compressed measurements has been introduced, namely Compressed Sensing (CS). The discovery and further development of CS has

motivated the introduction of a new approach to DoA in [1] by formulating the source localization problem as a sparse representation problem. Hence, several sparse recovery algorithms have been developed in the literature for solving this problem such as iterative hard thresholding (IHT) [13,14], orthogonal matching pursuit (OMP) [15], basis pursuit denoising (BPDN) [16] and others.

2.1 Notation

The following notation will be used in this thesis. The complex-conjugate transpose will be denoted by $(.)^H$, the transpose by $(.)^T$ and the pseudo-inverse matrix by $(.)^\dagger$. The operator supp(.) is defined as the set index of the non-zero elements of (.).

The ℓ_p norms are define for $p \in [1, \infty]$ as

$$\|\mathbf{x}\|_p = \begin{cases} (\sum_{i-1}^{n} |x_i|^p)^{\frac{1}{p}}, & p \in [0, \infty), \\ \max_{i=1,2,\ldots n} |x_i|, & p = \infty. \end{cases} \tag{2-1}$$

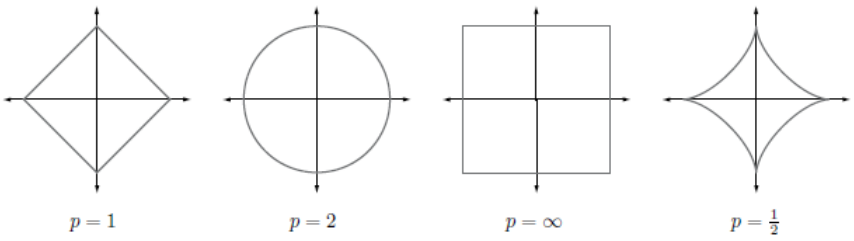


Figure 2.1: Unit spheres in $\mathbb{R}^2$ for the ℓ_p norms with $p = 1, 2, \infty$, and the quasinorm with $p = \frac{1}{2}$ [17].

The Figure 2.1 shows the unit sphere induced by each of these norms in $\mathbb{R}^2$.

The ℓ_0 norm is denoted as $\|\mathbf{x}\|_0 := |\text{supp}(\mathbf{x})|$, where $\text{supp}(\mathbf{x}) = \{i : x_i \neq 0\}$ denotes the support of $\mathbf{x}$ and $|\text{supp}(\mathbf{x})|$, the cardinality of $\text{supp}(\mathbf{x})$ [17].

For the multiple measurement vectors (MMV) model, the row support of the matrix $\mathbf{X}$ is defined to be equal to the index set of the non-zero rows of $\mathbf{X}$. This is equal to the union of the supports of all the columns of $\mathbf{X}$, i.e.,

$$\text{rsupp}(\mathbf{X}) = \bigcup_{j=1}^{q} \text{supp}(x_j), \tag{2-2}$$

where x_{ij} denotes the element in the i-th row and j-th column of $\mathbf{X}$ [18].

The ℓ_0 norm of the rowsparse $\mathbf{X}$ is defined to be equal to the cardinality of its row support:

$$\|\mathbf{X}\|_0 = |\text{rsupp}(\mathbf{X})|. \tag{2-3}$$

The Frobenius norm is defined as [19]:

$$\|\mathbf{X}\|_F = \sqrt{\text{Tr}(\mathbf{X}^H\mathbf{X})} = \|\text{vec}(\mathbf{X})\|,$$

where vec($\mathbf{X}$) is a vector formed by stacking the columns of $\mathbf{X}$ on top of each other and the mixed $\ell_{p,q}$ norm of $\mathbf{S}$ with rows $\mathbf{s}^i, i = 1, 2, \ldots, P$, is defined as [12]:

$$\|\mathbf{S}\|_{p,0} = \left\| \left[\left\|\mathbf{s}^1\right\|_p, \left\|\mathbf{s}^2\right\|_p, \ldots, \left\|\mathbf{s}^P\right\|_p \right] \right\|_q. \tag{2-4}$$

The $spark(\mathbf{A})$ of a given matrix $\mathbf{A}$ is defined as the smallest number of columns of $\mathbf{A}$ that are linearly dependent [17]. The $H_K(.)$ operator sets all but the K largest in magnitude elements of (.) to 0. Moreover, for any matrix $\mathbf{X}$, $\mathbf{X}_{(i)}$ denotes the i-th row of $\mathbf{X}$ and $\mathbf{X_I}$ represents the submatrix of $\mathbf{X}$ obtained by selecting the columns of $\mathbf{X}$ indexed by the set $\mathbf{I}$. Columns vectors and matrices are denoted by bold lowercase and uppercase letters, respectively.

2.2 Signal Model

Before we describe the most important methods for direction finding, it is necessary to present a mathematical model for the problem.

Let us assume that K uncorrelated narrowband zero mean signals $u_k(t), k = 1, 2, \ldots, K$ from far-field sources impinge on a uniform linear array (ULA) of M ($M > K$) sensor elements with inter-element spacing of half a wavelength ($d = \lambda/2$) from directions $\theta_k \in [-90°, 90°)$ corresponding to the spatial frequency $\mu_k = -\pi \sin\theta_k$. At time instant $t, t = 1, 2, \ldots, N$, where N is the total number of available snapshots, the received signal at the mth sensor can be expressed as:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{u}(t) + \mathbf{n}(t), \tag{2-5}$$

where $\mathbf{u}(t) = [u_1(t), \ldots, u_K(t)]^T \in \mathbb{C}^K$ represents the zero-mean source data vector and the entries of noise vector $\mathbf{n}(t) \in \mathbb{C}^M$ are assumed to be i.i.d, zero mean complex Gaussian distribution with variance σ_n^2 . The Vandermonde matrix $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$, known as the array manifold, contains the array steering vectors $\mathbf{a}(\theta_k)$ corresponding to the kth source, which can be expressed as:

$$\mathbf{a}(\theta_k) = [1, e^{j2\pi\frac{d}{\lambda}\sin\theta_k}, \ldots, e^{j2\pi(M-1)\frac{d}{\lambda}\sin\theta_k}]^T. \tag{2-6}$$

Equation (2-5) can be expressed for the case of multiple snapshots or multiple measurement vectors (MMV) as,

$$\mathbf{X} = \mathbf{AU} + \mathbf{N}. \tag{2-7}$$

2.3 The Cramer-Rao Bound (CRB)

In estimation theory, the Cramer-Rao Bound (CRB) [2] provides a bound on the covariance matrix of any unbiased estimate of $\boldsymbol{\theta}$. We denote the covariance matrix of the estimation errors by $\mathbf{C}(\boldsymbol{\theta})$. Then,

$$\mathbf{C}(\boldsymbol{\theta}) \triangleq E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]. \tag{2-8}$$

The multiple-parameter CRB states that

$$\mathbf{C}(\boldsymbol{\theta}) \geq \mathbf{C}_{CR}(\boldsymbol{\theta}) \triangleq \mathbf{J}^{-1}, \tag{2-9}$$

for any unbiased estimate of $\boldsymbol{\theta}$. The matrix inequality means that $\mathbf{C}(\boldsymbol{\theta}) - \mathbf{C}_{CR}(\boldsymbol{\theta})$ is a non-negative definite matrix. The $\mathbf{J}$ matrix is commonly referred as the Fisher's information matrix (or FIM) [2].

The comparison with the CRB has become an important ingredient of the analysis of any source localization method for array processing. When an estimator meets the CRB, it is called efficient. Many of the existing estimators of source location are biased, but they have the property of asymptotic unbiasedness, as either the number of snapshots or the number of sensors or both approach infinity [20]. For a review of the derivation of the CRB for source localization refer to [21] and Appendix A.

2.4 Classical and Parametric methods for direction finding

In this section, we describe several classical methods for direction finding, which include Capon's method and the maximum likelihood (ML) method.

2.4.1 Optimal beamforming: Capon's method (MVDR)

The Minimum Variance Distortionless Response Estimator (MVDR) [4], also known as Capon, is a spectral estimation method which uses the MVDR beamforming solution to obtain the angles of arrival. It attempts to minimize the variance due to noise, while keeping the gain in the direction of the steering vector equal to unity [20]:

$$\mathbf{w}_{CAP}(\theta) = \arg\min_{\mathbf{w}} E(\mathbf{w}^H \mathbf{x}\mathbf{x}^H \mathbf{w}) \quad \text{s.t.} \quad Re[\mathbf{w}^H \mathbf{a}(\theta)] = 1. \tag{2-10}$$

The solution of this optimization problem can be shown to have the following form:

$$\mathbf{w}_{CAP}(\theta) = \frac{\mathbf{R}_{xx}^{-1}\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{R}_{xx}^{-1}\mathbf{a}(\theta)}, \tag{2-11}$$

where $\mathbf{R}_{xx} = E[\mathbf{x}(t)\mathbf{x}(t)^H]$ is the covariance matrix of the array output $\mathbf{x}(t)$.

Then the Capon method applies the beamformer $\mathbf{w}_{CAP}(\theta)$ to the received data and computes the mean power of the array by:

$$\begin{aligned} P_{CAP}(\theta) &= E[||\mathbf{w}_{CAP}(\theta)^H \mathbf{x}(t)||_2^2] \\ &= \mathbf{w}_{CAP}(\theta)^H \mathbf{R}_{xx} \mathbf{w}_{CAP}(\theta) = \frac{1}{\mathbf{a}(\theta)^H \mathbf{R}_{xx}^{-1}\mathbf{a}(\theta)} \end{aligned} \tag{2-12}$$

and after that it finds the peaks of $P_{CAP}(\theta)$ which ones correspond to the DoAs.

Although the implementation of this technique is simple, it suffers from a lack of angular resolution and requires a large number of sensors to achieve a higher resolution.

2.4.2 Maximum Likelihood techniques (ML)

Maximum Likelihood (ML) methods belong to the class of parametric methods. In contrast to the method described above, the spectrum is not computed, but instead parameters of the model are estimated. A variety of methods resides under the ML header. One notable classification is in the assumed form of the signal. When signals are modeled as deterministic, the method is called Deterministic ML (DML), when the signals are modeled as Gaussian, the method is called Stochastic ML (SML) [20]. Noise is usually modeled as stationary Gaussian. For deterministic maximum likelihood, the objective is to find $\boldsymbol{\theta}, \mathbf{u}(t)$ and σ^2, to maximize the likelihood function:

$$L_{DML}(\boldsymbol{\theta}, \mathbf{u}(t), \sigma^2) = \mathbf{\Pi}_{t=1}^{T} (\pi\sigma^2)^{-M} \exp(- \left\| \mathbf{x}(t) - \mathbf{A}(\boldsymbol{\theta})\mathbf{u}(t) \right\|_2^2 / \sigma^2), \tag{2-13}$$

where $\boldsymbol{\theta}$ is the vector of sources localizations. The log-likelihood is [20]:

$$l_{DML}(\boldsymbol{\theta}, \mathbf{u}(t), \sigma^2) = -2M \log\sigma + \frac{1}{\sigma^2 T} \sum_{t=1}^{T} (- \left\| \mathbf{x}(t) - \mathbf{A}(\boldsymbol{\theta})\mathbf{u}(t) \right\|_2^2), \tag{2-14}$$

Fortunately, it is not necessary to optimize over all the parameters, $\boldsymbol{\theta}$, $\mathbf{u}(t)$ and σ^2 simultaneously, since once $\boldsymbol{\theta}$ is known, we can use $\mathbf{A}(\boldsymbol{\theta})$ to get explicit values for the other parameters:

$$\hat{\sigma}^2 = \frac{1}{M}\mathrm{tr}\left\{\Pi^{\perp}_{\mathbf{A}(\boldsymbol{\theta})}\mathbf{R}_{xx}\right\} \quad \text{and} \quad \hat{\mathbf{u}}(t) = \mathbf{A}(\boldsymbol{\theta})^{\dagger}\mathbf{x}(t), \tag{2-15}$$

where $\Pi^{\perp}_{\mathbf{A}(\boldsymbol{\theta})}$ is the projection matrix onto the orthogonal complement of the range space of $\mathbf{A}(\boldsymbol{\theta})$.

The remaining unknown, the locations of the sources, can be found by minimizing the following cost function [20]:

$$\hat{\boldsymbol{\theta}}_{\mathrm{DML}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left\|\Pi^{\perp}_{\mathbf{A}(\boldsymbol{\theta})}\mathbf{x}(t)\right\|_2^2 = \arg\min_{\boldsymbol{\theta}} \mathrm{tr}\left\{\Pi^{\perp}_{\mathbf{A}(\boldsymbol{\theta})}\mathbf{R}_{xx}\right\} \tag{2-16}$$

The computational complexity of ML is considerably higher than the others DoA estimation methods. The benefits of ML family of methods is the ability to resolve coherent signals, ability to handle single snapshot scenarios, and better statistical properties. A major problem with the ML-family of methods is the need for a very accurate starting point for the optimization procedure; otherwise the solution may converge to a local extremum [20].

2.5 Subspace methods

In this section, we describe several subspaces-based methods for direction finding, which include multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT).

2.5.1 Multiple Signal Classification (MUSIC)

The MUSIC method [5, 21] is the most prominent member of the family of eigen-expansion based direction finding techniques. The underlying idea is to separate the eigenspace of the covariance matrix of sensor outputs into the signal and noise components using the knowledge about the covariance matrix of the noise [20].

MUSIC makes the assumption that the noise in each channel is uncorrelated making the noise correlation matrix diagonal and the sensor output correlation matrix admits the following decomposition:

$$\begin{aligned}\mathbf{R}_{xx} = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_{uu}\mathbf{A}(\boldsymbol{\theta})^H &= \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \\ &= \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_n\boldsymbol{\Lambda}_n\mathbf{U}_n^H \\ &= \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^H + \sigma^2\mathbf{U}_n\mathbf{U}_n^H\end{aligned} \tag{2-17}$$

where $\mathbf{U} \in \mathbb{C}^{M\times M}$ is a unitary matrix containing the eigenvector of the covariance matrix $\mathbf{R}_{xx} \in \mathbb{C}^{M\times M}$, and $\boldsymbol{\Lambda} \in \mathbb{C}^{M\times M}$ is the diagonal matrix containing the eigenvalues of $\mathbf{R}_{xx}$.

$\mathbf{U}_s \in \mathbb{C}^{M\times K}$, $\mathbf{U}_n \in \mathbb{C}^{M\times(M-K)}$, $\boldsymbol{\Lambda}_s \in \mathbb{C}^{K\times K}$, $\boldsymbol{\Lambda}_n = \sigma^2\mathbf{I}_{M-K} \in \mathbb{C}^{(M-K)\times(M-K)}$ are the partitions of the eigenspectrum into signal plus noise and signal subspaces. Provided that $\mathbf{R}_{uu} \in \mathbb{C}^{K\times K}$ is nonsingular, $\mathbf{A}(\boldsymbol{\theta})\mathbf{R}_{uu}\mathbf{A}(\boldsymbol{\theta})^H$ has rank K. The number of sources K has to be strictly less than the number of sensors M, for the method to work. Hence, $\mathbf{R}_{xx}$ has K eigenvalues which are due to the combined signal plus noise subspace, and $M-K$ eigenvalues due to the noise subspace alone. Assuming that the noise has a flat spectrum of σ^2, K eigenvalues corresponding to the signal and noise subspaces are larger than the remaining $M-K$ noise eigenvalues, which are equal to σ^2. This information can be used to separate the two eigensubspaces. Due to the orthogonality of eigensubspaces corresponding to different eigenvalues for Hermitian matrices, the noise subspace is orthogonal to the steering vectors corresponding to the direction of propagation, thus $\mathbf{U}_n^H\mathbf{a}(\theta) = \mathbf{0}$ for all directions from which the signals are impinging [20]. The MUSIC spectrum is obtained by putting the squared norm of this term into the denominator, which leads to very sharp estimates of the positions of the sources:

$$P_{MUSIC}(\theta) = \frac{1}{\mathbf{a}(\theta)^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\theta)}. \tag{2-18}$$

In the noiseless case the peaks of the spectrum approach infinity.

2.5.2 Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT)

Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) exploits an underlying rotational invariance among signal subspaces induced by an array of sensors with a translational invariance structure [7].

To describe mathematically the effect of the translational invariance of the sensor array, ESPRIT describes conveniently the array as being comprised of two subarrays (called doublets), identical in every respect although physically displaced from each other by a know displacement of magnitud $\triangle$ [7]. Hence, the received data vector can be written as follows:

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} \mathbf{u}(t) + \begin{pmatrix} \mathbf{n}_1(t) \\ \mathbf{n}_2(t) \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_1\mathbf{\Phi} \end{pmatrix} \mathbf{u}(t) + \begin{pmatrix} \mathbf{n}_1(t) \\ \mathbf{n}_2(t) \end{pmatrix}, \quad (2\text{-}19)$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the steering matrix of the first and second subarrays respectively, and $\mathbf{\Phi} = \text{diag}\{\phi_k\}_{k=1}^{K} \in \mathbb{C}^{K\times K}$ is a diagonal matrix of the phase delays between the sensor doublets for the K wavefronts and its diagonal elements, the phase factors ϕ_k, are given by

$$\phi_k = e^{-j\omega_0 \triangle \sin(\theta_k)/c} \quad 1 \leq k \leq K, \quad (2\text{-}20)$$

where c is the signal propagation velocity and ω_0 is the common center frequency [22]. The matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ can be expressed for the following relation:

$$\mathbf{A}_1 = \mathbf{J}_1\mathbf{A} \quad \text{and} \quad \mathbf{A}_2 = \mathbf{J}_2\mathbf{A},$$

where $\mathbf{J}_1 = [\mathbf{1}_{M-1}, \mathbf{0}_{(M-1)\times 1}] \in \{0,1\}^{(M-1)\times(M)}$ and $\mathbf{J}_2 = [\mathbf{0}_{(M-1)\times 1}, \mathbf{1}_{M-1}] \in \{0,1\}^{(M-1)\times(M)}$ are selection matrices with the identity matrix $\mathbf{1}_{M-1}$ and the zero vector $\mathbf{0}_{(M-1)\times 1}$. Then the invariance equation is given by

$$\mathbf{J}_1\mathbf{U}_s\mathbf{\Phi} = \mathbf{J}_2\mathbf{U}_s. \quad (2\text{-}21)$$

Solving the equation in the Least Squares (LS) sense and after computing the eigenvalues of $\mathbf{\Phi}$, the estimated DoAs of the K sources can be obtained as [12]

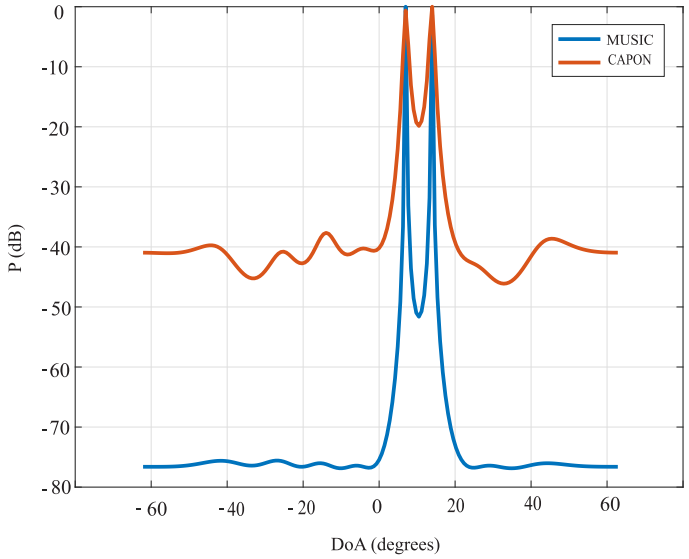
$$\hat{\theta}_k = -\arcsin\left(\frac{\arg(\phi_k)}{\pi}\right). \quad (2\text{-}22)$$

2.5.3 Limitations of current methods

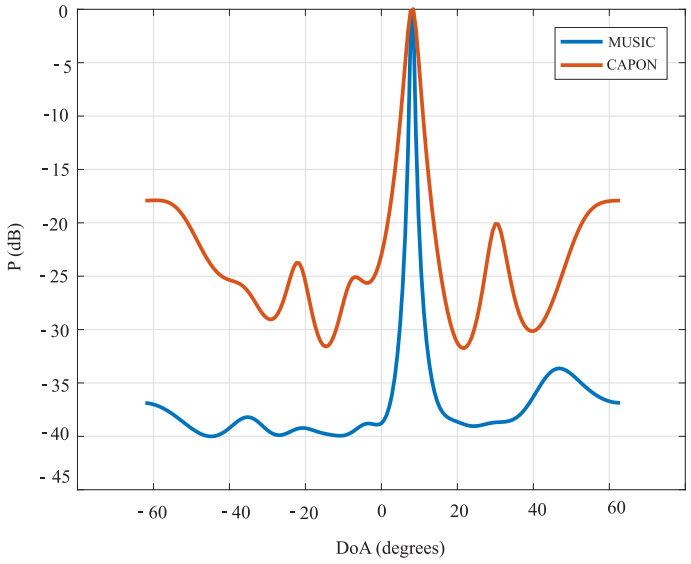
ESPRIT, MUSIC and Capon's methods present some limitations over different conditions; when the sources are close and the SNR is low, they lose resolution and eventually are unable to separate the sources.

Figure 2.2 depicts what happens in a scenario of lower SNR and when the sources are close together. In plot (a) SNR is 10 dB, and separation between the sources is 10°, so both MUSIC and Capon's methods are able to resolve the two sources well. However, plot (b) shows that when SNR is decreased to 0 dB, and source separation is decreased to 5°, neither of the two methods can resolve the two sources, the peaks are merged for both.

Some additional limitations of these two methods include inferior performance for correlated and coherent sources, and for scenarios with limited



(a)



(b)

Figure 2.2: Limitations of MUSIC and Capon's methods (a) SNR=10 dB, separation between the sources is 10°, (b) SNR=0 dB, separation between the sources is 5°.

number of snapshots. These limitations are the principal reason why several research efforts in the last years have concentrated on improving the performance of existing methods.

2.5.4 Techniques for dealing with correlated source signals

As we mentioned before one of the main problems of DoA estimation is to determine the correct DoA in the case when we have a scenario with correlated or coherent sources because the source covariance matrix $\mathbf{R}_{uu} = E[\mathbf{u}(t)\mathbf{u}(t)^H]$ becomes rank deficient and non-diagonal, so that some of its eigenvalues are zero. As a result, the orthogonality of the noise and signal subspaces does not hold . This means that part of the signal subspace is indistinguishable from the noise subspace. There are some techniques with the objective of overcoming this problem. Among the most effective techniques are spatial smoothing (SS) and forward-backward spatial smoothing (FBSS) described in [23, 24] and [25, 26], respectively, which are based on the preprocessing scheme that partitions the total array of sensors into subarrays. We will explain in the next sections the application of these techniques in detail.

2.5.4.1 Spatial Smoothing technique

Spatial Smoothing [26] is a preprocessing scheme that is applied to circumvent the problems encountered in DoA estimation of fully correlated signals [24]. The basic idea is to form covariance matrices from subsets of the array, which is equivalent to partitioning the original covariance matrix. A geometrical interpretation is shown in Figure 2.3.

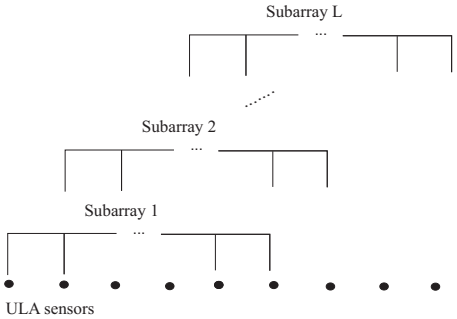


Figure 2.3: Spatial Smoothing subarray formation.

If there are L subarrays, each subarray is of size $M_{sub} = M - L + 1$ and the submatrix for the lth subarray is given by:

$$\mathbf{R}_l = \mathbf{J}_l \mathbf{R}_{xx} \mathbf{J}_l^T \tag{2-23}$$

where the $\mathbf{J}_l \in \mathbb{C}^{M_{sub} \times M}$ matrix is defined by:

$$\mathbf{J}_l = [\mathbf{0}_{M_{sub} \times (l-1)} \ \mathbf{I}_{M_{sub}} \ \mathbf{0}_{M_{sub} \times (L-l)}] \quad 1 \leq l \leq L \tag{2-24}$$

Then the partitioned matrices are used to form a smoothed matrix $\mathbf{R}_{SS} \in \mathbb{C}^{M_{sub} \times M_{sub}}$, which is calculated as follows:

$$\mathbf{R}_{SS} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{J}_l \mathbf{R}_{xx} \mathbf{J}_l^T \tag{2-25}$$

For this technique in order to solve K coherent sources, an array that contains at least $2K$ elements is required [23].

2.5.4.2 Forward-Backward Spatial Smoothing (FBSS) technique

Forward-Backward Spatial Smoothing (FBSS) is an additional technique that can be used to improve performance in the case of scenario with correlated sources. It applies the spatial smoothing step in both directions and form the resulting matrix $\mathbf{R}_{FBSS}$ using the forward and backward covariance matrices as follows:

$$\mathbf{R}_{FB} = \frac{1}{2}(\mathbf{R}_F + \mathbf{R}_B) = \frac{1}{2}(\mathbf{R}_F + \mathbf{\Pi}\mathbf{R}_F^*\mathbf{\Pi}) = \frac{1}{2}(\mathbf{R}_{SS} + \mathbf{\Pi}\mathbf{R}_{SS}^*\mathbf{\Pi}), \tag{2-26}$$

where $\mathbf{\Pi}$ is the exchange matrix with ones on its antidiagonal and zeros elsewhere and $\mathbf{R}_{SS}^*$ denotes the complex conjugation [23].

A geometrical interpretation of this technique is illustrated in Figure 2.4.

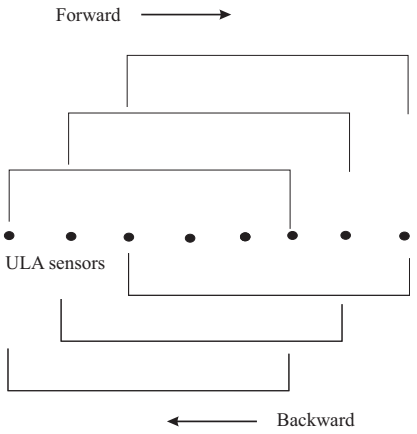


Figure 2.4: Forward-Backward scheme.

2.6 Direction of Arrival (DoA) estimation based on Compressed Sensing (CS)

In this section we introduce the concept of CS and we explain how CS allows one to transform direction finding or source localization problem into a problem of sparse signal representation.

2.6.1 Compressive Sensing (CS) theory

Compressive sensing (CS) is an exciting, rapidly growing, field that has attracted considerable attention in signal processing, statistics, and computer science, as well as the broader scientific community [17].

CS is a method to capture and represent compressible signals at a rate significantly below the Nyquist rate. This method employs non-adaptive linear projections that preserve the structure of the signal; the signal is then reconstructed from these projections using an optimization procedure [27].

Suppose that $\mathbf{x} \in \mathbb{C}^N$ is a signal vector of size $N \times 1$, which can be written as:

$$\mathbf{x} = \sum_{i-1}^{N} s_i \psi_i \quad \text{or} \quad \mathbf{x} = \mathbf{\Psi s}, \tag{2-27}$$

where $\mathbf{\Psi}$ is the $N \times N$ sparsity basis matrix and $\mathbf{s}$ an $N \times 1$ vector with $K << N$ non-zero (and large enough) entries [28]. Then, the signal $\mathbf{x}$ is K sparse due to the fact that it is a linear combination of only K basis vectors; that is, only K of the s_i coefficients in equation (2-27) are nonzero and $(N - K)$ are zero [27]. The CS theory states that $\mathbf{x}$ can be recovered using $M = K\mathcal{O}(\log N)$ non-adaptive linear projection measurements onto an $M \times N$ basis matrix $\mathbf{\Phi}$ that is incoherent with $\mathbf{\Psi}$ [28]. The measurement vector $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi s} = \mathbf{\Theta s}. \tag{2-28}$$

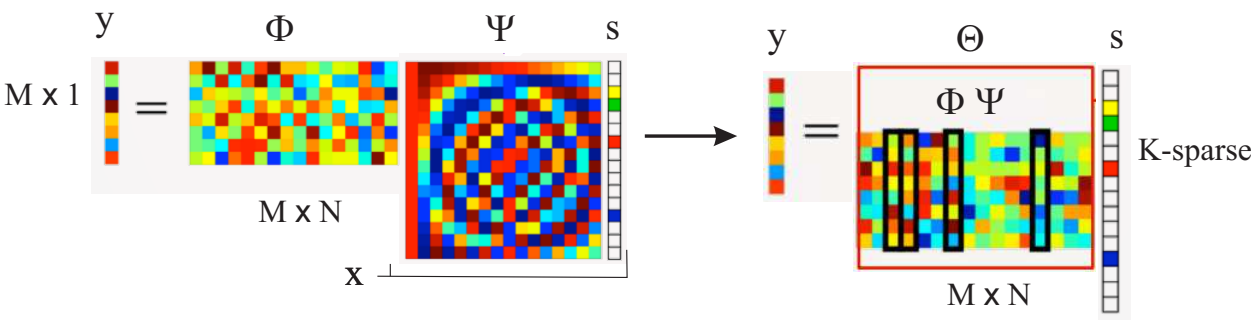


Figure 2.5: The vector of coefficients $\mathbf{s}$ is sparse with $K = 4$ and there are four columns of $\mathbf{\Theta}$ that corresponds to the nonzero s_i coefficients; the measurement vector $\mathbf{y}$ is a linear combination of these columns [27].

The matrix $\mathbf{\Phi}$ has to be a stable measurement matrix such that the salient information of any K-sparse or compressible signal is not damaged by the dimensionality reduction from $\mathbf{x} \in \mathbb{R}^N$ to $\mathbf{y} \in \mathbb{R}^M (M < N)$. A necessary and sufficient condition for this is that, for any vector $\mathbf{v}$ sharing the same K non-zeros entries as $\mathbf{s}$ and for some $\delta > 0$, the matrix $\mathbf{\Theta}$ have to satisfy the restricted isometry property (RIP) [27] given by

$$(1 - \delta) \leq \frac{\| \mathbf{\Theta v} \|_2}{\| \mathbf{v} \|_2} \leq (1 + \delta). \tag{2-29}$$

A related condition referred to the incoherence, requires that the rows $\{\phi_j\}$ of $\mathbf{\Phi}$ cannot sparsely represent the columns $\{\psi_i\}$ of $\mathbf{\Psi}$. However, both the RIP and incoherence can be achieved with high probability simply by selecting $\mathbf{\Phi}$ as a random matrix [27], e.g., Gaussian distribution.

2.6.2 Sparse recovery algorithms

There are a variety of algorithmic approaches to the problem of signal recovery from CS measurements. These algorithms have been used in applications such as sparse approximation, statistics, geophysics, and theoretical computer science that were developed to exploit sparsity in other contexts and can be brought to bear on the CS recovery problem [17].

The sparse recovery algorithms must take the M measurements in the vector $\mathbf{y}$, the random measurement matrix $\mathbf{\Phi}$, the basis $\mathbf{\Psi}$ and reconstruct the length-N signal $\mathbf{x}$ or, equivalently, its sparse coefficient vector $\mathbf{s}$ [27].

There are many $\tilde{\mathbf{s}}$ that satisfy $\mathbf{\Theta}\tilde{\mathbf{s}} = \mathbf{y}$. This is because $\mathbf{\Theta}(\mathbf{s} + \mathbf{r}) = \mathbf{y}$ for any vector $\mathbf{r}$ in the null space $\mathcal{N}(\mathbf{\Theta})$ of $\mathbf{\Theta}$. Therefore, the signal reconstruction algorithm aims to find the signal´s sparse coefficient vector in the $(N - M)$ translated null space $\mathcal{H} = \mathcal{N}(\mathbf{\Theta}) + \mathbf{s}$ [27].

The classical approach to this problem is to find the vector using the smallest ℓ_2 norm by solving [27]:

$$\hat{\mathbf{s}} = \min_{\tilde{\mathbf{s}}} \|\tilde{\mathbf{s}}\|_2 \quad \text{s.t} \quad \mathbf{y} = \mathbf{\Theta}\tilde{\mathbf{s}}. \tag{2-30}$$

This optimization have the convenient closed-form solution $\hat{\mathbf{s}} = \mathbf{\Theta}^T(\mathbf{\Theta}\mathbf{\Theta}^T)^{-1}\mathbf{y}$. Unfortunately, ℓ_2 minimization will almost never find the K-sparse solution [27].

Since the ℓ_2 norm does not guarantee sparsity, it is replaced for the ℓ_0 norm and the modified optimization is given by

$$\min_{\tilde{\mathbf{s}}} \|\tilde{\mathbf{s}}\|_0 \quad \text{s.t} \quad \mathbf{y} = \mathbf{\Theta}\tilde{\mathbf{s}}. \tag{2-31}$$

In this case the K-sparse signal can be recovery with high probability using only $M = K + 1$ i.i.d Gaussian measurements. Unfortunately, solving

(2-31) is both numerically unstable and Non-deterministic Polynomial (NP) hard problem, requiring an exhaustive enumeration of all $\binom{P}{K}$ possible locations of the non-zero entries in $\mathbf{s}$ [27], making it computationally intractable. Then the ℓ_0-norm is frequently substituted by the ℓ_1-norm, which can exactly recovery K-sparse signal and closely approximate compressible signals with high probability using $M = cK\log(P/K)$ [27], as follows:

$$\hat{\mathbf{s}} = \min_{\tilde{\mathbf{s}}} \|\tilde{\mathbf{s}}\|_1 \quad \text{s.t} \quad \mathbf{y} = \boldsymbol{\Theta}\tilde{\mathbf{s}}. \tag{2-32}$$

The above minimization problem is also know as basis pursuit (BP). Although the ℓ_1-norm is weaker than ℓ_0-norm in ensuring sparsity, ℓ_1-regularized optimization is a convex problem and admits efficient solution via linear programming techniques.

Figure 2.6 shows the geometry representation of the compressed sensing problem in $\mathcal{R}^3$. (In practice $N, M, K >> 3$.)

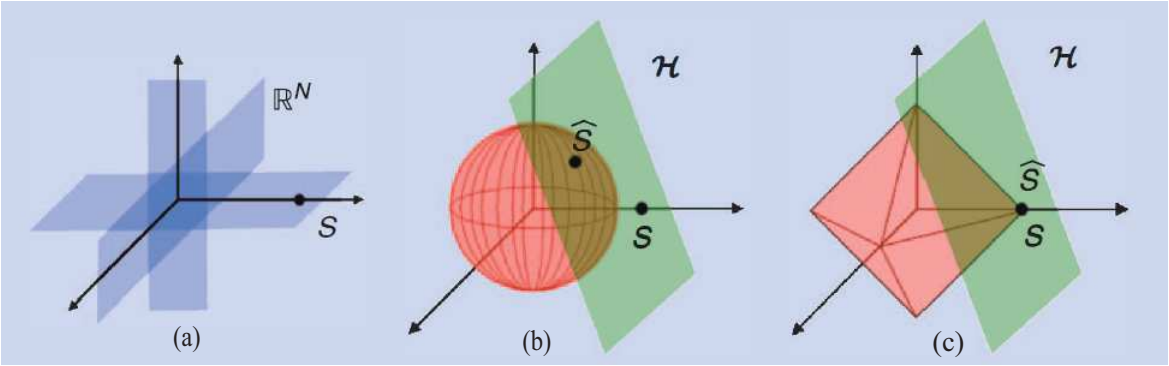


Figure 2.6: (a) The subspaces containing two sparse vectors in $\mathcal{R}^3$ lie close to the coordinate axes. (b) Visualization of the ℓ_2 minimization that finds the non-sparse point-of-contact $\hat{\mathbf{s}}$ between the ℓ_2 ball and the translated measurement matrix null space. (c) Visualization of the ℓ_1 minimization that finds the sparse point-of-contact $\hat{\mathbf{s}}$ with high probability thanks to the pointiness of the ℓ_1 ball [27].

In many applications, it is desirable to trade off exact congruence of $\boldsymbol{\Theta}\mathbf{s}$ and $\mathbf{y}$ in exchange for a sparser $\mathbf{s}$. In these cases, a more appropriate formulation is basis pursuit denoising (BPDN) [29], this involves solving the following problem:

$$\min_{\tilde{\mathbf{s}}} \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Theta}\tilde{\mathbf{s}}\|_2^2 + \lambda\|\tilde{\mathbf{s}}\|_1 \tag{2-33}$$

BPDN, closely related to least absolute shrinkage and selection operator (LASSO) regression is simply least-squares minimization with an ℓ_1-regularized to penalize complex solutions. The regularization parameter establishes the cost of complexity relative to the least-squares error $\frac{1}{2}\|\mathbf{y} - \boldsymbol{\Theta}\tilde{\mathbf{s}}\|_2^2$ [29].

However, CS is often synonymous with ℓ_1-based optimization [17] there are other approaches to the sparse recovery problem such as the Greedy algorithms. The concept of "greedy" implies strategies that, at each step, make a “hard” decision usually based upon some locally optimal optimization criterion [17].

2.7 Signal Model for direction finding using CS approach

DoA estimation using CS consists of formulating the source localization problem as a sparse representation problem through the introduction of an overcomplete representation matrix $\mathbf{A}$ in terms of all possible angles of interest [1].

Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n], n = 1, 2, \ldots, P$, be a sampling grid of all source locations of interest where P is equal to the number of potential sources which must be much greater than the number of sources K or even the number of sensors M, then the signal field can be represented by a vector $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T \in \mathbb{C}^P$ whose nth element is equal to $u_k(t)$ if the kth source comes from direction θ_n and zero otherwise [1], as described by

$$s_n(t) = \begin{cases} u_k(t) & \text{if } \theta_n = \theta_k, \\ 0 & \text{otherwise.} \end{cases} \tag{2-34}$$

The measurement model for one snapshot or a single measurement vector (SMV) $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T \in \mathbb{C}^M$ at the time t can be written as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{2-35}$$

where the matrix $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_P)] \in \mathbb{C}^{M \times P}$ contains the steering vectors corresponding to each potential source localization as its columns.

Note that the matrices $\mathbf{A}(\boldsymbol{\theta})$ in equation (2-5) and $\mathbf{A}$ in (2-35) have different dimensions. The first one correspond to the steering matrix and the second is the measurement matrix using in the sparse representation.

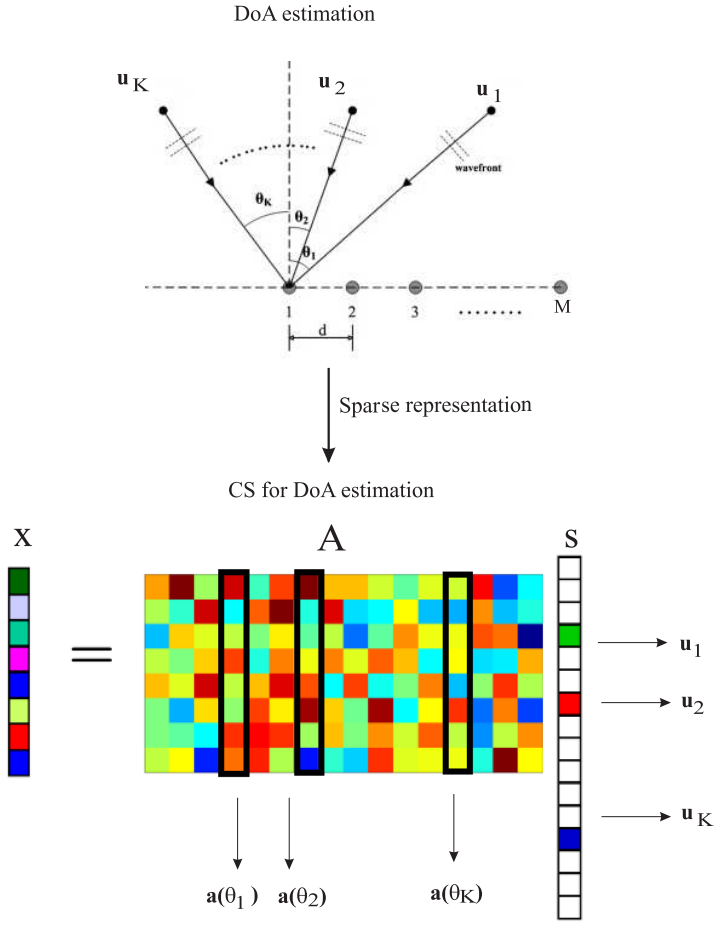


Figure 2.7: Formulation of DoA estimation as a sparse signal recovery problem.

This representation allows to exchange the problem of parameter estimation of $\boldsymbol{\theta}$ for the sparse recovery problem of obtaining a K-sparse estimate $\hat{\mathbf{s}}(t) = [\hat{s}_1(t), \ldots, \hat{s}_P(t)]^T \in \mathbb{C}^P$ for the true K-sparse vector $\mathbf{s}(t)$ from the measurements $\mathbf{x}(t)$ with the array steering matrix $\mathbf{A}$ as the measurement matrix where finally the non-zero indices of $\hat{\mathbf{s}}(t)$ determine the estimates $\hat{\theta}_k$ of the DoAs of the K sources.

The model in 2-35 can be extended to the MMV model as

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}, \tag{2-36}$$

where the matrix of measurements is $\mathbf{X} = [\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N)] \in \mathbb{C}^{M \times N}$, the signal matrix is $\mathbf{S} = [\mathbf{s}(t_1), \ldots, \mathbf{s}(t_N)] \in \mathbb{C}^{P \times N}$ and the noise matrix is $\mathbf{N} = [\mathbf{n}(t_1), ..., \mathbf{n}(t_N)] \in \mathbb{C}^{M \times N}$. In this case the signal matrix $\mathbf{S}$ is row K-sparse and its non-zero rows indices correspond to the DoAs of the K sources.

2.7.1 Recovery guarantees under the MMV model

We assume that a row K-sparse matrix $\mathbf{S}$ is measured with a measurement matrix $\mathbf{A}$ to produce the observed matrix $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{AS}$. Then it is

guaranteed that one can recover $\mathbf{S}$ from $\mathbf{X}$ exactly if and only if [18]

$$spark(\mathbf{A}) > 2K + 1 - rank(\mathbf{S}). \tag{2-37}$$

For a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, every set of $M+1$ columns of $\mathbf{A}$ is guaranteed to have linearly dependent columns. On the other hand, one needs to take at least two column vectors to form a linearly dependent set. Therefore, $2 \leq spark(\mathbf{A}) \leq M + 1$. The definition of the spark of a matrix draws some parallels to the definition of the rank of a matrix, i.e., spark is the cardinality of the smallest set of linearly dependent columns whereas rank is the cardinality of the largest set of linearly independent columns. While rank of a matrix can be computed efficiently, computing spark of a matrix is NP-Hard [18].

Since $\mathbf{S}$ has only K non-zero rows, we have $1 \leq rank(\mathbf{S}) \leq K$. Then from signal recovery perspective it is desired that $\mathbf{A}$ has maximum possible spark. Replacing $spark(\mathbf{A})$ in (2-37) with its maximum possible value, we get

$$M > 2K - rank(\mathbf{S}). \tag{2-38}$$

When $rank(\mathbf{S}) = 1$ this would result in the condition $M \geq 2K + 1$ and in the best case, when $rank(\mathbf{S}) = K$, we have

$$M \geq K + 1,$$

i.e., the minimum number of measurements required for the MMV model is $K + 1$.

2.8 Sparse recovery algorithms applied to DoA estimation

We briefly review in this section some of sparse recovery algorithms applied to the source localization problem.

2.8.1 Greedy Algorithms

The sparse recovery algorithms based on ℓ_1-minimization are an effective methodology which, under certain conditions, can result in an exact signal recovery. In addition, ℓ_1-minimization also has very good performance guarantees which make it a reliable tool for sparse signal recovery but one drawback of the methods based on ℓ_1-minimization is their higher computational cost in large-scale problems. Therefore, algorithms that scale up better and are similar in performance in comparison to the convex optimization methods are needed, this fact motivated the development of the greedy algorithms. These algorithms

make significant savings in computation by performing locally optimal greedy iterations. Some of these methods also have certain performance guarantees that are somewhat similar to the guarantees for ℓ_1-minimization [18].

Greedy algorithms rely on iterative approximations of the signal coefficients and support, either by iteratively identifying the support of the signal until a convergence criterion is met, or alternatively by obtaining an improved estimate of the sparse signal at each iteration that attempts to account for the mismatch to the measured data [17]. The greedy algorithms reviewed in this chapter recover the unknown sparse vector in a non-Bayesian framework, i.e., the sparse vector is treated as a fixed unknown and no prior assumption is made about the probability distribution of the sparse vector.

2.8.1.1 Iterative Hard Thresholding (IHT)

The iterative hard thresholding (IHT) algorithm [13] solves the sparse recovery problem presented above as an optimization problem which is formulated as

$$\hat{\mathbf{S}} = \arg \min_{\tilde{\mathbf{S}} \in \mathbb{C}^{P \times N}} \| \mathbf{A}\tilde{\mathbf{S}} - \mathbf{X} \|_F^2 \text{ s.t. } \| \tilde{\mathbf{S}} \|_{p,0} \leq K, \tag{2-39}$$

where the constraint ensures that the estimate $\hat{\mathbf{S}}$ is row K-sparse.

Each iteration of IHT consists of two steps, a gradient descent step and a hard thresholding step [12]. In the first step the consistency-enforcing objective is reduced by computing the gradient descent step described by

$$\check{\mathbf{S}}^{i+1} = \hat{\mathbf{S}}^i + \mu \mathbf{A}^H (\mathbf{X} - \mathbf{A}\hat{\mathbf{S}}^i), \tag{2-40}$$

where μ is the step size. The second step is to apply the hard thresholding operator $H_K(.)$ to the resulting matrix $\check{\mathbf{S}}^{i+1}$, which sets all but the K rows with the largest l_2 norm to $\mathbf{0}^T$ to get the new estimate $\hat{\mathbf{S}}^{i+1} = H_K(\check{\mathbf{S}}^{i+1})$. This step ensures that the constraint is fulfilled [12]. Then the indices of the non-zero rows of $\hat{\mathbf{S}}^i$, i.e., $\mathbf{\Gamma} = \text{rsupp}(\hat{\mathbf{S}}^i) \subseteq \mathbb{R}^K$ finally determine the estimated angles $\hat{\theta}_k$ for the DoAs of the K sources.

IHT also has a normalized variant called normalized iterative hard thresholding (NIHT) where the fixed step size μ is replaced by a near-optimal step size μ^i that maximally reduces the error at each iteration [30] described by,

$$\mu^{i+1} = \frac{\mathbf{G}_{\mathbf{\Gamma}^i}^H \mathbf{G}_{\mathbf{\Gamma}^i}}{\mathbf{G}_{\mathbf{\Gamma}^i}^H \mathbf{A}_{\mathbf{\Gamma}^i}^H \mathbf{A}_{\mathbf{\Gamma}^i} \mathbf{G}_{\mathbf{\Gamma}^i}} = \frac{\|\mathbf{G}_{\mathbf{\Gamma}^i}\|_F^2}{\|\mathbf{A}_{\mathbf{\Gamma}^i} \mathbf{G}_{\mathbf{\Gamma}^i}\|_F^2}. \tag{2-41}$$

NIHT determines the step size μ adaptively, so then it becomes scale-independent and as a consequence guarantees convergence to a local minimum

of the cost function, whatever the operator norm of $\mathbf{A}$, while still retaining its performance guarantees if a restricted isometry property holds [30].

The IHT and NIHT algorithms are described in Algorithm 1.

Algorithm 1: IHT /NIHT

Input: $\mathbf{X}, \mathbf{A}, K, \mu, \mathbf{R}^0 = \mathbf{X}, \boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]$
Initialization: $\hat{\mathbf{S}}^0 = \mathbf{0}, i = 0$
while stopping criterion is not met **do**
  if IHT **then**
    $\mu^{i+1} = \mu$
    else if NIHT **then**
      $\mu^{i+1} = \frac{\left\|\mathbf{A}_{\mathbf{\Gamma}^i}^H \mathbf{R}_{\mathbf{\Gamma}^i}^i\right\|_F^2}{\left\|\mathbf{A}_{\mathbf{\Gamma}^i}(\mathbf{A}_{\mathbf{\Gamma}^i}^H \mathbf{R}_{\mathbf{\Gamma}^i}^i)\right\|_F^2}$
    end
  end
  $\hat{\mathbf{S}}^{i+1} = H_K(\hat{\mathbf{S}}^i + \mu^{i+1}\mathbf{A}^H\mathbf{R}^i)$
  $\mathbf{R}^{i+1} = \mathbf{X} - \mathbf{A}\hat{\mathbf{S}}^{i+1}$
  $i = i + 1$
end
$\mathbf{\Gamma}$ =rsupp($\hat{\mathbf{S}}^i$)
Output: DoAs: $[\hat{\theta}_1, \ldots, \hat{\theta}_K] = \boldsymbol{\theta}(\mathbf{\Gamma})$

2.8.1.2 Hard Thresholding Pursuit (HTP)

The hard thresholding pursuit (HTP) similar to IHT update the previous approximation $\hat{\mathbf{S}}^i$ by taking a step of predefined, fixed length μ in the steepest descent direction $\mathbf{A}^H\mathbf{R}^i$. A new proxy for the support set, $\mathbf{\Gamma}^i$, is then obtained by selecting the rows of with greatest row-ℓ_2-norms. The two algorithms differ in how the support proxy is utilized: IHT employs a hard thresholding operator which restricts the approximation to the rows indexed by while HTP projects the measurements onto the support set as follows [14]:

$$\hat{\mathbf{S}}^{i+1} = \arg\min \left\{ \|\mathbf{X} - \mathbf{AZ}\|_F : \text{rsupp}(\mathbf{Z}) \subseteq \mathbf{\Gamma}^{i+1} \right\}. \tag{2-42}$$

The HTP algorithm and its normalized version are described in Algorithm 2, where DetectSupport$(\mathbf{Z}, k)$ is a subroutine identifying the index set of the rows of $\mathbf{Z}$ with the k largest row-ℓ_2-norms [14].

Algorithm 2: HTP /NHTP

Input: $\mathbf{X}, \mathbf{A}, K, \mu, \mathbf{R}^0 = \mathbf{X}, \boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]$
Initialization: $\hat{\mathbf{S}}^0 = \mathbf{0}, i = 0$
while stopping criterion is not met **do**
  if HTP **then**
    $\mu^{i+1} = \mu$
    else if NHTP **then**
      $\mu^{i+1} = \frac{\left\|\mathbf{A}^H_{\mathbf{\Gamma}^i}\mathbf{R}^i_{\mathbf{\Gamma}^i}\right\|_F^2}{\left\|\mathbf{A}_{\mathbf{\Gamma}^i}(\mathbf{A}^H_{\mathbf{\Gamma}^i}\mathbf{R}^i_{\mathbf{\Gamma}^i})\right\|_F^2}$
    end
  end
  $\hat{\mathbf{S}}^{i+1} = \hat{\mathbf{S}}^i + \mu^{i+1}\mathbf{A}^H\mathbf{R}^i$
  $\mathbf{\Gamma}^{i+1} = \text{DetectSupport}(\hat{\mathbf{S}}^{i+1}, k)$
  $\hat{\mathbf{S}}^{i+1} = \arg\min \{\|\mathbf{X} - \mathbf{AZ}\|_F : \text{rsupp}(\mathbf{Z}) \subseteq \mathbf{\Gamma}^{i+1}\}$
  $\mathbf{R}^{i+1} = \mathbf{X} - \mathbf{A}\hat{\mathbf{S}}^{i+1}$
  $i = i + 1$
end
$\mathbf{\Gamma}$ =rsupp$(\hat{\mathbf{S}}^i)$
Output: DoAs: $[\hat{\theta}_1, \ldots, \hat{\theta}_K] = \boldsymbol{\theta}(\mathbf{\Gamma})$

2.8.1.3 Orthogonal Matching Pursuit (OMP)

Orthogonal Matching Pursuit (OMP) is a commonly used algorithm for recovery signals due to its low complexity and simple implementation [31]. OMP chooses at each iteration i a new atom index j by solving the following optimization problem [32]:

$$j = \arg \min_{j \in \mathbf{\Omega} - \mathbf{\Gamma}^i} \left\| \langle \mathbf{A}_j, \mathbf{R}^i \rangle \right\|_2 . \tag{2-43}$$

where $\mathbf{R} = \mathbf{X} - \mathbf{AS}$.

After that it generates an optimal atom set $\mathbf{\Gamma}^{i+1} = \mathbf{\Gamma}^i \cup \{j\}$ with $|\mathbf{\Gamma}^{i+1}| = i$ and obtains the best $(i+1)$-term approximation $\hat{\mathbf{S}}^{i+1}_{(\mathbf{\Gamma}^{i+1})}$ over $\mathbf{\Gamma}^{i+1}$ by the least-squares (LS) minimization as follows [32]:

$$\hat{\mathbf{S}}^{i+1}_{(\mathbf{\Gamma}^{i+1})} = \arg \min_{\text{rsupp}(\mathbf{Z}) \subseteq \mathbf{\Gamma}^{i+1}} \left\| \mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}^{i+1}} \mathbf{Z} \right\|_F . \tag{2-44}$$

This minimization can simply be performed by standard LS techniques, i.e. $\hat{\mathbf{S}}^{i+1}_{(\mathbf{\Gamma}^{i+1})} = \mathbf{A}^{\dagger}_{\mathbf{\Gamma}^{i+1}} \mathbf{X}$ [31]. The pseudocode of OMP is described in Algorithm 3

Algorithm 3: OMP

Input: $\mathbf{X}, \mathbf{A}, K, \mathbf{R}^0 = \mathbf{X}, \boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]$
Initialization: $\hat{\mathbf{S}}^0 = \mathbf{0}, \mathbf{\Gamma}^0 = \emptyset, i = 0$
while $i \leq K$ **do**
  $j = \arg \min_{j \in [1,\ldots,P]} \left\| \langle \mathbf{A}_j, \mathbf{R}^i \rangle \right\|_2$
  $\mathbf{\Gamma}^{i+1} = \mathbf{\Gamma}^i \cup \{j\}$
  $\hat{\mathbf{S}}^{i+1}_{(\mathbf{\Gamma}^{i+1})} = \arg \min_{\text{rsupp}(\mathbf{Z}) \subseteq \mathbf{\Gamma}^{i+1}} \left\| \mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}^{i+1}} \mathbf{Z} \right\|_F$
  $\hat{\mathbf{S}}^{i+1}_{(\bar{\mathbf{\Gamma}}^{i+1})} = \mathbf{0}$
  $\mathbf{R}^{i+1} = \mathbf{X} - \mathbf{A}\hat{\mathbf{S}}^{i+1}$
  $i = i + 1$
end
$\mathbf{\Gamma}$ =rsupp($\hat{\mathbf{S}}^i$)
Output: DoAs: $[\hat{\theta}_1, \ldots, \hat{\theta}_K] = \boldsymbol{\theta}(\mathbf{\Gamma})$

2.8.2 Signal recovery via ℓ_1 minimization

Since the ℓ_0 minimization problem is a NP hard problem, several recovery algorithms have introduced the convex relaxation to the ℓ_1 norm for translating the optimization problem into a more tractable problem. Moreover, with the development of fast methods of linear programming in the eighties, the idea of convex relaxation became truly promising.

2.8.2.1 ℓ_1-Singular value decomposition algorithm (ℓ_1-SVD)

The ℓ_1-singular value decomposition (ℓ_1-SVD) algorithm, introduced by Malioutov, Çetin and Willsky [1], is a tractable approach to use a large number of time samples coherently. To reduce the computational complexity and the sensetivity to noise, ℓ_1-SVD use the SVD of the observed data matrix $\mathbf{X}$. The idea is to decompose the data matrix into signal and noise subspaces and keep the signal subspace [1].

Mathematically, this translates into the following representation. Take the SVD:

$$\mathbf{X} = \mathbf{ULV}. \tag{2-45}$$

Keep a reduced $M \times K$ dimensional matrix $\mathbf{X}_{SV}$, which contains most of the signal power $\mathbf{X}_{SV} = \mathbf{ULD}_K = \mathbf{XVD}_K$ where $\mathbf{D}_k = [\mathbf{I}_K \mathbf{0}]$. Here, $\mathbf{I}_K$ identity matrix, and $\mathbf{0}$ is a $K \times (N-K)$ matrix of zeros. In addition, let $\mathbf{S}_{SV} = \mathbf{SVD}_K$ and $\mathbf{N}_{SV} = \mathbf{NVD}_K$, to obtain which yields [1],

$$\mathbf{X}_{SV} = \mathbf{AS}_{SV} + \mathbf{N}_{SV}. \tag{2-46}$$

For typical situations where the number of sources is small and the number of time samples may be in the order of hundreds, this reduction in complexity is very substantial [1].

The matrix $\mathbf{S}$ is parameterized temporally and spatially, but sparsity only has to be enforced in space since the signal $\mathbf{s}(t)$ in not generally sparse in time [1]. To accommodate this issue, ℓ_1-SVD first computes the ℓ_2-norm of all time-samples of a particular spatial index of $\mathbf{s}$, i.e., $s_i^{(l_2)} = \|[s_i(t_1), s_i(t_2), \ldots, s_i(t_N)]\|_2$, and penalizing the ℓ_1-norm of $\mathbf{s}^{(l_2)} = [s_1^{(l_2)}, \ldots, s_P^{(l_2)}]$. Then , the minimization problem becomes

$$\min \|\mathbf{X}_{SV} - \mathbf{AS}_{SV}\|_F^2 + \lambda \left\|\mathbf{s}^{(l_2)}\right\|_1, \tag{2-47}$$

where the optimization is performed over $\mathbf{S}$; $\mathbf{s}^{(l_2)}$ is a function of $\mathbf{S}$ and the optimization problem is solved using second-order cone (SOC) programming

framework by an interior point implementation. The block diagram of the steps of the ℓ_1-SVD algorithm is illustrated in Figure 2.8.

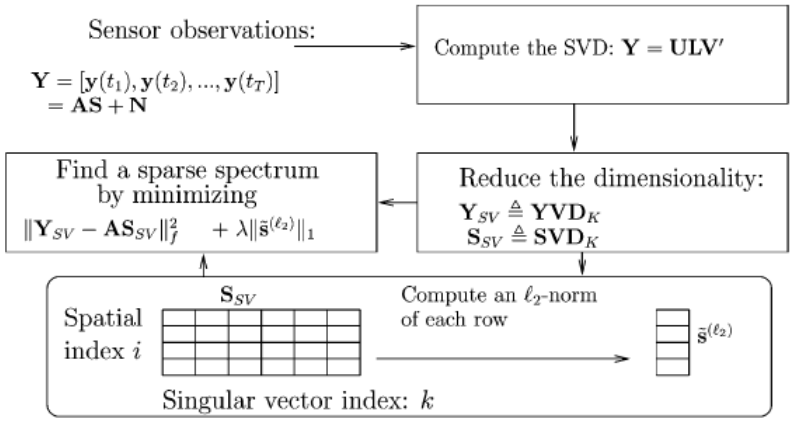


Figure 2.8: Block diagram of the steps of the ℓ_1-SVD [1].

2.8.2.2 Basis Pursuit Denoising (BPDN)

The ℓ_1-norm optimization problem is a convex problem that can be conveniently reduced to a linear problem known as basis pursuit denoising (BPDN) [27]. In the BPDN problem, the row K-sparse signal matrix $\mathbf{S}$ can be recovered by

$$\hat{\mathbf{S}} = \arg \min_{\tilde{\mathbf{S}} \in \mathbb{C}^{P \times N}} \left\|\tilde{\mathbf{S}}\right\|_{2,1} s.t \left\|\mathbf{A}\tilde{\mathbf{S}} - \mathbf{X}\right\|_F^2 \leq \beta, \tag{2-48}$$

where the objective of the optimization problem ensures that the estimate $\hat{\mathbf{S}}$ is row sparse while the constraint forces it to be consistent with the measurements of $\mathbf{X}$ [12]. The parameter β has to be chosen according to the statistics of the noise. When these are not known the choice of this parameter is a difficult question which is the main drawback of this algorithm.

The BPDN problem can be solved using several methods as spectral projected gradient (SPGL1) [33], fixed point continuation method (FPC) [34], fast iterative soft-thresholding algorithm (FISTA) [35], ℓ_1 regularized least squares [36], in-crowd algorithms [29], Nesterov's smoothing techniques with continuation (NESTA) [37] and others.

2.9 Computational Complexity

In practical applications, not only the DoA estimation performance of the methods is important but also their computational complexity. Therefore, the computational complexity of the subspace-based methods MUSIC and

ESPRIT as well as the one of the CS-based methods BPDN and IHT is analyzed in this section.

If the matrix of measurements is compressed by the ℓ_1-SVD, the DoAs can be estimated using BPDN with a computational complexity of $\mathcal{O}(K^3P^3)$ [1]. This is larger than the costs of subspace-based methods MUSIC and ESPRIT, which are $\mathcal{O}(M^2P + M^2N)$ and $\mathcal{O}(M^3 + M^2N)$, respectively. MUSIC is computationally more complex than ESPRIT due to the spectral search [12]. The computational complexity of ESPRIT is dominated by the subspace estimation using the EVD of the covariance matrix. Each iteration of IHT consists only of matrix additions and multiplications as well as the thresholding operation, whose cost is $\mathcal{O}(MNP)$.

If the number of snapshots N is small and the number of antenna elements M becomes very large, the subspace estimation using the EVD of the covariance matrix with a cost of $\mathcal{O}(M^3)$ for the subspace-based methods and the spectral search of MUSIC with a cost of $\mathcal{O}(M^2P)$ become computationally intractable such that the computational complexity of IHT is smaller than the one of the subspace-based methods [12]. Table 2.1 lists the computational complexity of all considered methods.

Table 2.1: Computational Complexity [12]

Algorithm	
BPDN	$\mathcal{O}(K^3P^3)$
IHT	$\mathcal{O}(MNP)$ (per iteration)
MUSIC	$\mathcal{O}(M^2P + M^2N)$
ESPRIT	$\mathcal{O}(M^3 + M^2N)$

3
Bayesian and Iterative Hard Thresholding Methods for Direction of Arrival Estimation

In the case when some prior knowledge about the distribution of the sparse matrix is available, it would make sense to incorporate that prior knowledge into the estimation process. Bayesian methods, which view the unknown sparse matrix as random, provide a systematic framework for doing that. By making use of Bayes' rule, these methods update the prior knowledge about the sparse matrix in accordance with the new evidence or observations [18]. This chapter proposes two greedy algorithms called randomized multiple candidate iterative hard thresholding (RMC-IHT) [38] and Bayesian hard thresholding (BHT) which estimate the sparse matrix $\mathbf{S}$ in a Bayesian framework.

The RMC-IHT algorithm generates a set of potential candidates using the iterative hard thresholding algorithm and selects the best candidate based on the prior knowledge of the distribution of the signal and noise matrices. It is also formulated a version of RMC-IHT for the case of correlated sources based on the computation of an approximation of the covariance matrix at each iteration [38].

The BHT algorithm works similar to IHT, except that instead of the deterministic hard thresholding operator H_K, it uses a Bayesian hard thresholding operator H_{p_j} which select the K atoms with highest probability than belong to the support $\mathbf{\Gamma}$.

3.1
Derivation of the maximum a posteriori (MAP) estimator

In order to derive the maximum a posteriori (MAP) estimator, we assume that the elements of the noise matrix $\mathbf{N}$ in equation (2-36) are assumed to be independent and identically distributed (i.i.d) complex normal random variables with zero mean and variance σ_n^2, so $\mathbf{N}$ has a matrix variate complex normal (MCN) distribution, denoted as $\mathbf{N} \sim \mathcal{MCN}_{M,N}(\mathbf{0}, \sigma_n^2\mathbf{I}_M, \mathbf{I}_N)$, where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix [19].

Let $\mathbf{\Gamma}$ denote the row support of $\mathbf{S}$ with known fixed cardinality K and the total number of different possible row supports of $\mathbf{S}$ is $\binom{P}{K}$ and let Ω

denote the set of all such row supports, with cardinality $|\Omega| = \binom{P}{K}$. It is further assumed that, within the set Ω, $\mathbf{\Gamma}$ has uniform prior distribution, i.e.,

$$p(\mathbf{\Gamma}) = \begin{cases} \frac{1}{|\Omega|} & \text{if } \mathbf{\Gamma} \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{3-1}$$

Furthermore, let $\mathbf{S}_{(\mathbf{\Gamma})}$ denote the $K \times N$ matrix restricted to those K rows of $\mathbf{S}$ that are indexed by the support set $\mathbf{\Gamma}$. Then $\mathbf{S}_{(\bar{\mathbf{\Gamma}})} = \mathbf{0}$ by definition, where $\bar{\mathbf{\Gamma}}$ is the complement of $\mathbf{\Gamma}$. The restriction $\mathbf{A}_{\mathbf{\Gamma}}$ represents the sub-matrix of $\mathbf{A}$ obtained by selecting the columns of $\mathbf{A}$ indexed by $\mathbf{\Gamma}$. n Then the elements of $\mathbf{S}_{(\mathbf{\Gamma})}$ are assumed to be i.i.d complex normal random variables with zero mean and known variance σ_s^2, so $\mathbf{S}_{(\mathbf{\Gamma})} \sim \mathcal{MCN}_{M,N}(\mathbf{0}, \sigma_s^2\mathbf{I}_K, \mathbf{I}_N)$, with probability density function (pdf) given by [19]

$$p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{\Gamma}) = \frac{1}{(\pi\sigma_s^2)^{KN}} \exp\left(-\frac{1}{\sigma_s^2} \parallel \mathbf{S}_{(\mathbf{\Gamma})} \parallel_F^2\right) \tag{3-2}$$

Then we compute $p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{X}, \mathbf{\Gamma})$ using the Bayes's rule as follows:

$$p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{X}, \mathbf{\Gamma}) = \frac{p(\mathbf{X}|\mathbf{S}_{(\mathbf{\Gamma})}, \mathbf{\Gamma})p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{\Gamma})}{p(\mathbf{X}|\mathbf{\Gamma})}, \tag{3-3}$$

where $p(\mathbf{X}|\mathbf{\Gamma})$ is a normalization constant for fixed $\mathbf{X}$ and $\mathbf{\Gamma}$. Then, $\mathbf{X}|\mathbf{S}_{(\mathbf{\Gamma})}, \mathbf{\Gamma} = \mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})} + \mathbf{N}$ for a given $\mathbf{S}_{(\mathbf{\Gamma})}$ is a Gaussian matrix of mean $\mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})}$ and matrix covariance $\sigma_n^2\mathbf{I}_M$, so we obtain

$$p(\mathbf{X}|\mathbf{S}_{(\mathbf{\Gamma})}, \mathbf{\Gamma}) = \frac{1}{(\pi\sigma_n^2)^{MN}} \exp\left(-\frac{1}{\sigma_n^2} \parallel \mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})} \parallel_F^2\right) \tag{3-4}$$

Ignoring the normalization constant $p(\mathbf{X}|\mathbf{\Gamma})$ and using the expressions (3-2) and (3-4) we can rewrite (3-3) as [19]:

$$p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{X}, \mathbf{\Gamma}) \propto \exp\left(-\frac{\parallel \mathbf{S}_{(\mathbf{\Gamma})} \parallel_F^2}{\sigma_s^2} - \frac{\parallel \mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})} \parallel_F^2}{\sigma_n^2}\right) \tag{3-5}$$

Therefore, the MAP estimator of $\mathbf{S}_{(\mathbf{\Gamma})}$ for a given support $\mathbf{\Gamma}$ will be equal to [2]:

$$\begin{aligned} \hat{\mathbf{S}}_{(\mathbf{\Gamma})_{\text{MAP}}} &= \arg \max_{\mathbf{S}_{(\mathbf{\Gamma})} \in \mathbb{C}^{K \times N}} p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{X}, \mathbf{\Gamma}) \\ &= \arg \max_{\mathbf{S}_{(\mathbf{\Gamma})} \in \mathbb{C}^{K \times N}} \log p(\mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{X}, \mathbf{\Gamma}) \\ &= \arg \max_{\mathbf{S}_{(\mathbf{\Gamma})} \in \mathbb{C}^{K \times N}} \left(-\frac{\parallel \mathbf{S}_{(\mathbf{\Gamma})} \parallel_F^2}{\sigma_s^2} - \frac{\parallel \mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})} \parallel_F^2}{\sigma_n^2}\right) \end{aligned} \tag{3-6}$$

The above convex optimization problem can be solved by setting the gradient of the objective function to a zero matrix and solving the resulting set of equations which yields:

$$\hat{\mathbf{S}}_{(\mathbf{\Gamma})_{\mathrm{MAP}}} = \left(\mathbf{A}_{\mathbf{\Gamma}}^H \mathbf{A}_{\mathbf{\Gamma}} + \frac{1}{\gamma}\mathbf{I}_K\right)^{-1} \mathbf{A}_{\mathbf{\Gamma}}^H \mathbf{X} \tag{3-7}$$

where $\gamma = \sigma_s^2/\sigma_n^2$.

3.2 Description of the proposed RMC-IHT algorithm

The proposed RMC-IHT algorithm is based on the principle of the deterministic IHT algorithm, but instead of IHT which selects a support at each iteration, RMC-IHT generates a set $\mathbf{\Upsilon} = [\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \ldots, \mathbf{\Gamma}_Q]$ of Q potential candidate's supports, where each candidate is computed by the iteration given by

$$\hat{\mathbf{S}}^{(j)} = H_K(\hat{\mathbf{S}}^{(j-1)} + \mathbf{P}\mathbf{A}^H\mathbf{R}^{(j-1)}) \tag{3-8}$$

$$\mathbf{\Gamma}_i = \mathrm{rsupp}(\hat{\mathbf{S}}^{(j)}) \quad i = 1, 2, \ldots, Q \tag{3-9}$$

using the hard thresholding operator H_K.

The criterion for determining the best candidate support is based on the following minimization:

$$\min_{\mathbf{\Gamma}_i \in \mathbf{\Upsilon}} \left\| \mathbf{X} - \mathbf{A}\hat{\mathbf{S}}_{(\mathbf{\Gamma}_i)_{\mathrm{MAP}}} \right\|_F^2 \tag{3-10}$$

Note that for performing this selection it is necessary the computation of the MAP estimator of the signal $\mathbf{S}_{(\mathbf{\Gamma})}$ for each candidate support $\mathbf{\Gamma}_i$ as was determined in the previous section. RMC-IHT based on the result of equation (3-10) finally keeps the signal estimate and the residual corresponding to the best candidate support to initialize the next iteration.

The main steps of RMC-IHT are described in Algorithm 4, where $\mathbf{\Gamma}^*$ is the best support selected among the Q candidates in the current iteration.

Algorithm 4: RMC-IHT

Input: $\mathbf{X}, \mathbf{A}, K, \mu, \boldsymbol{\theta} = [\theta_1, ..., \theta_P], \gamma, Q,$
Initialization:
$\hat{\mathbf{S}}^{(0)} = \mathbf{0}, ii = 0, \boldsymbol{\Gamma}$ =supp$(H_K(\mathbf{A}^H\mathbf{X}))$,
$\mathbf{R}^{(0)} = \mathbf{X}, \mathbf{P}$ =diag$([p_i])$ =diag$([1/\| a_i \|_2^2])$,
$\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}$
while the stopping criterion is not met **do**
  for $j \leftarrow 2$ **to** Q **do**
    $\hat{\mathbf{S}}^{(j)} = H_K(\hat{\mathbf{S}}^{(j-1)} + \mathbf{P}\mathbf{A}^H\mathbf{R}^{(j-1)})$
    $\boldsymbol{\Gamma}_i$ =rsupp$(\hat{\mathbf{S}}^{(j)})$
    $\mathbf{R}^{(j)} = \mathbf{X} - \mathbf{A}^H\hat{\mathbf{S}}^{(j)}$
  end
  for $i \leftarrow 1$ **to** Q **do**
    $\hat{\mathbf{S}}_{(\boldsymbol{\Gamma}_i)} = (\mathbf{A}_{\boldsymbol{\Gamma}_i}^H\mathbf{A}_{\boldsymbol{\Gamma}_i} + \frac{1}{\gamma}\mathbf{I}_K)^{-1}\mathbf{A}_{\boldsymbol{\Gamma}_i}^H\mathbf{X}$
    $\hat{\mathbf{S}}_{(\bar{\boldsymbol{\Gamma}}_i)} = \mathbf{0}$
    $\boldsymbol{\Gamma}^* = \min_{\boldsymbol{\Gamma}_i \in \boldsymbol{\Upsilon}} \left\| \mathbf{X} - \mathbf{A}\hat{\mathbf{S}}_{\boldsymbol{\Gamma}_i} \right\|_F^2$
  end
  keep $\boldsymbol{\Gamma}^*$ as the first candidate for the next iteration
  keep the signal estimate $\hat{\mathbf{S}}_{\boldsymbol{\Gamma}^*}$ and the to corresponding residual to initialize the next iteration
  $ii = ii + 1$
end
Output: DoAs: $[\hat{\theta}_1, \ldots, \hat{\theta}_K] = \boldsymbol{\theta}(\boldsymbol{\Gamma}^*)$

3.3 RMC-IHT for scenarios with correlated sources

As we mentioned before one of the limitations of the conventional DoA estimation algorithms is their inferior performance for scenarios with correlated sources. Due to the illness of the covariance matrix of the signal, the principal techniques based on subspaces methods fail. For this reason, we present in this section a version of RMC-IHT for overcoming this limitation.

In the case of a scenario with correlated sources the covariance matrix of the signal becomes non diagonal. Then we can no longer assume that it is equal to $\sigma_s^2\mathbf{I}_K$. Taking into account this consideration, the MAP estimator of $\mathbf{S}_{(\boldsymbol{\Gamma})}$ is now given by

$$\hat{\mathbf{S}}_{(\boldsymbol{\Gamma})_{\text{MAP}}} = \left(\frac{1}{\sigma_n^2}\mathbf{A}_{\boldsymbol{\Gamma}}^H\mathbf{A}_{\boldsymbol{\Gamma}} + \mathbf{K}_s\right)^{-1}\frac{1}{\sigma_n^2}\mathbf{A}_{\boldsymbol{\Gamma}}^H\mathbf{X} \tag{3-11}$$

when the covariance matrix of the signal is given by

$$\mathbf{K}_s = \mathbf{R}_s = E[\mathbf{S}_{(\boldsymbol{\Gamma})}\mathbf{S}_{(\boldsymbol{\Gamma})}^H], \tag{3-12}$$

and the mean of the signals is equal to zero resulting in equivalent covariance and correlation matrices.

In this case, we need some approximation of the correlation matrix of $\mathbf{S}_{(\mathbf{\Gamma})}$ for computing equation (3-11). Then we propose to obtain the K-term approximation $\mathbf{S}_{(\mathbf{\Gamma})}$ over $\mathbf{\Gamma}$ by a least-squares (LS) minimization as follows:

$$\mathbf{S}_{(\mathbf{\Gamma})} = \arg \min_{\text{rsupp}(\mathbf{Z}) \subseteq \mathbf{\Gamma}} \|\mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}}\mathbf{Z}\|_F^2 . \tag{3-13}$$

This minimization can be simply be performed by standard LS techniques, i.e. $\mathbf{S}_{(\mathbf{\Gamma})} = \mathbf{A}_{\Gamma}^{\dagger}\mathbf{X}$ [31]. Substituting the obtained approximation of $\mathbf{S}_{(\mathbf{\Gamma})}$ in (3-12) we have the following result:

$$\mathbf{R}_s = E[\mathbf{S}_{(\Gamma)}\mathbf{S}_{(\Gamma)}^H] = E[\mathbf{A}_{\Gamma}^{\dagger}\mathbf{X}\mathbf{X}^H(\mathbf{A}_{\Gamma}^{\dagger})^H] = \mathbf{A}_{\Gamma}^{\dagger}\mathbf{R}_{xx}(\mathbf{A}_{\Gamma}^{\dagger})^H. \tag{3-14}$$

We finally obtain an approximation of the correlation matrix of the signal as a function of the measurement matrix $\mathbf{A}_{\mathbf{\Gamma}}$ and the correlation matrix of the received matrix $\mathbf{X}$. Therefore, substituting the equation (3-14) in equation (3-11) we obtain the MAP estimator for the case of the scenario with correlated sources:

$$\hat{\mathbf{S}}_{(\mathbf{\Gamma})_{\text{MAP}}} = \left(\frac{1}{\sigma_n^2}\mathbf{A}_{\mathbf{\Gamma}}^H\mathbf{A}_{\mathbf{\Gamma}} + \mathbf{A}_{\Gamma}^{\dagger}\mathbf{R}_{xx}(\mathbf{A}_{\Gamma}^{\dagger})^H\right)^{-1}\frac{1}{\sigma_n^2}\mathbf{A}_{\mathbf{\Gamma}}^H\mathbf{X}. \tag{3-15}$$

3.4 Derivation of the Bayesian hard thresholding operator

Based on the results presented in the previous section, we have that for a fixed $\mathbf{X}$ and all $\mathbf{\Gamma} \in \mathbf{\Omega}$, both $p(\mathbf{X})$ and $p(\mathbf{\Gamma})$ are constant. Therefore, from Bayes' rule we can conclude that $p(\mathbf{\Gamma}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{\Gamma})$. Moreover, we have [18]

$$\begin{aligned} p(\mathbf{X}|\mathbf{\Gamma}) &= \int_{\mathbf{S}_{(\mathbf{\Gamma})} \in \mathbb{C}^{K\times N}} p(\mathbf{X}, \mathbf{S}_{(\mathbf{\Gamma})}|\mathbf{\Gamma}) \mathrm{d}\mathbf{S}_{(\mathbf{\Gamma})} \propto \\ &\int \exp\left\{-\frac{\left\|\text{vec}(\mathbf{S}_{(\mathbf{\Gamma})})\right\|^2}{\sigma_s^2} - \frac{\left\|\text{vec}(\mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})})\right\|^2}{\sigma_n^2}\right\} \mathrm{d}\mathbf{S}_{(\mathbf{\Gamma})} \end{aligned} \tag{3-16}$$

where the integration is over $\mathbf{S}_{(\mathbf{\Gamma})} \in \mathbb{C}^{K\times N}$. Since $\text{vec}(\mathbf{X} - \mathbf{A}_{\mathbf{\Gamma}}\mathbf{S}_{(\mathbf{\Gamma})}) = \text{vec}(\mathbf{X}) - (\mathbf{I}_N \otimes \mathbf{A}_{\mathbf{\Gamma}})\text{vec}(\mathbf{S}_{(\mathbf{\Gamma})})$, where $\otimes$ denotes the Kronecker product, the integral in (3-16) simplifies to

$$p(\mathbf{X}|\mathbf{\Gamma}) \propto p_{\mathbf{\Gamma}} = \exp\left(\frac{\text{vec}(\mathbf{A}_{\mathbf{\Gamma}}^H\mathbf{X})^H\mathbf{P}_{\mathbf{\Gamma}}^{-1}\text{vec}(\mathbf{A}_{\mathbf{\Gamma}}^H\mathbf{X})}{\sigma_n^4} + \log(\det(\mathbf{P}_{\mathbf{\Gamma}}^{-1}))\right), \tag{3-17}$$

where $\mathbf{P}_{\mathbf{\Gamma}} = \frac{1}{\sigma_n^2}\mathbf{I}_N \otimes \mathbf{A}_{\mathbf{\Gamma}}^H\mathbf{A}_{\mathbf{\Gamma}} + \frac{1}{\sigma_s^2}\mathbf{I}_{KN}$ [19, 39].

Then based on the previous results suppose that the row sparsity of $\mathbf{S}$ is one, i.e., $|\mathbf{\Gamma}| = K = 1$. This implies that $\mathbf{\Omega} = \{1, 2, \dots P\}$, and

$$\mathbf{P}_{\mathbf{\Gamma}=\{j\}} = (c_j/\sigma_s^2)\mathbf{I}_N \quad \text{for} \quad j \in \mathbf{\Omega}, \tag{3-18}$$

where $c_j = 1 + \gamma \left\|\mathbf{A}_j\right\|^2$. Furthermore, $p(\mathbf{\Gamma} = \{j\}|\mathbf{X}) \propto p_{\mathbf{\Gamma}=\{j\}} = p_j$ is given by [19]

$$p_j = \exp\left\{ \frac{\gamma}{\sigma_n^2} \frac{\left\|\mathbf{X}^H \mathbf{A}_j\right\|^2}{c_j} - N \log c_j \right\}. \tag{3-19}$$

Note that p_j corresponds to the conditional probability that given a observed matrix $\mathbf{X}$ the jth atom belong to the support $\mathbf{\Gamma}$.

Then the Bayesian hard operator $H_{p_j}(.)$ is defined as the operator that sets all the values of a given matrix to zero except the K rows with the highest conditional probability p_j.

3.5 Description of the proposed BHT algorithm

The proposed BHT algorithm is similar to IHT. Each iteration of BHT consists of three steps: a gradient descent step, the computation of the conditional probability p_j of each j atom and a Bayesian hard thresholding step.

In the first step the consistency-enforcing objective is reduced by computing the gradient descent step described by

$$\check{\mathbf{S}}^{i+1} = \hat{\mathbf{S}}^i + \mu \mathbf{A}^H (\mathbf{X} - \mathbf{A}\hat{\mathbf{S}}^i), \tag{3-20}$$

where μ is the step size. The second step consists of the computation of the conditional probability p_j of each atoms using the result obtained in equation (3-19) and the last step is to apply the Bayesian hard thresholding operator $H_{p_j}(.)$ to the resulting matrix $\check{\mathbf{S}}^{i+1}$. In this way, BHT exploits the prior knowledge of the distribution of the signal and noise for keeping at each iteration only the K rows of $\mathbf{S}$ that corresponds to the K atoms with the highest probability that belong to the support $\mathbf{\Gamma}$.

The BHT algorithm is summarized in Algorithm 5

Algorithm 5: BHT

Input: $\mathbf{X}, \mathbf{A}, K, \mu, \mathbf{R}^0 = \mathbf{X}, \gamma, \sigma_n^2, \boldsymbol{\theta} = [\theta_1, ..., \theta_P]$
Initialization: $\hat{\mathbf{S}}^0 = \mathbf{0}, i = 0$
while stopping criterion is not met **do**
  $\check{\mathbf{S}}^{i+1} = \hat{\mathbf{S}}^i + \mu \mathbf{A}^H \mathbf{R}^i$
  compute the conditional probability p_j of each atoms as
  for $j \leftarrow 1$ **to** P **do**
    $c_j = 1 + \gamma \left\| \mathbf{A}_j \right\|^2$
    $p_j = \exp\left\{ \frac{\gamma}{\sigma_n^2} \frac{\left\| \mathbf{R}^H \mathbf{A}_j \right\|^2}{c_j} - N \log c_j \right\}$
  end
  $\hat{\mathbf{S}}^{i+1} = H_{p_j}(\check{\mathbf{S}}^{i+1})$
  $\mathbf{R}^{i+1} = \mathbf{X} - \mathbf{A}\hat{\mathbf{S}}^{i+1}$
  $i = i + 1$
end
$\boldsymbol{\Gamma}$ =rsupp$(\hat{\mathbf{S}}^i)$
Output:
DoAs: $[\hat{\theta}_1, \ldots, \hat{\theta}_K] = \boldsymbol{\theta}(\boldsymbol{\Gamma})$

3.6 Computational Complexity

In this section we provide an analysis of the algorithmic complexity of RMC-IHT and BHT algorithms.

In the case of RMC-IHT algorithm, the generation of the Q potential candidate supports is based on IHT algorithm for this reason consists of matrix additions and multiplications whose cost is $\mathcal{O}(MPN)$ at each iteration. The computation of the MAP estimator of the sparse signal for a given candidate consists of matrix multiplications, additions and the inversion of a matrix of size $K \times K$, hence its computational cost is approximately $\mathcal{O}(K^2M + K^3 + K^2N)$.

The BHT algorithm as was described before is a modification of the IHT algorithm that includes one more step that consist of the computation of the conditional probability p_j for each atom whose cost is $\mathcal{O}(PMN)$. Table 3.1 list the computational complexity of the proposed algorithms.

Table 3.1: Computational Complexity of the proposed algorithms

Algorithm	Computational Complexity
RMC-IHT	$\mathcal{O}(QPMN + Q(K^2M + K^3 + K^2N))$
BHT	$\mathcal{O}(2PMN)$ (per iteration)

3.7 Simulations Results

In this section several simulations were carried out to evaluate the proposed algorithms and compare them with conventional directional of arrival estimation algorithms.

Simulation results have been obtained by using a sampling grid of P= 1024 angles and $\theta_i = \arcsin(\frac{2}{P}(i-1)-1)$ corresponding to the equally spaced spatial frequencies. The measurement noise samples are drawn from an i.i.d complex normal random process with zero mean and variance σ_n^2 [12]. The SNR in dB is defined as:

$$\text{SNR} = 10\log_{10}\left(\frac{K\sigma_s^2}{\sigma_n^2}\right) \text{ dB}. \tag{3-21}$$

The DoA estimation performance is measured in terms of the Root Mean Square Error (RMSE) between the DoAs θ_k of the sources $k = 1, 2, \ldots, K$ and their estimates $\hat{\theta}_{r,k}$ in R=100 Monte Carlo runs $r = 1, 2, \ldots, R$.

$$\text{RMSE}_\theta = \sqrt{\left(\frac{1}{RK}\right)\sum_{r=1}^{R}\sum_{k=1}^{K}|\hat{\theta}_{k,r} - \theta_k)|^2}. \tag{3-22}$$

It is important to mention that the simulation results are concentrated in the scenarios where the conventional techniques show a inferior performance, i.e. low values of SNR, a limited number of available snapshots and correlated sources. We also considered antennas array with around 8 to 30 elements and that the noise and the sources are uncorrelated.

Figure **??** represents a scenario with K=2 sources which are assumed uncorrelated complex normal random variables with zero mean and variance σ_s^2 =1 and located at the angles $\theta_1 = 4.93°$ and $\theta_2 = 10.01°$ [12], the RMSE is plotted over the SNR with the number of antenna elements $M = 13$ for $N = 2$ and the size of the set of candidates is set to 4. The number of iteration of IHT algorithm is setting to 10. In addition to the RMSE of the considered DoAs estimation methods, we also show the deterministic Cramér-Rao Lower Bound (CRB) [21].

As it can be noticed the proposed algorithms BHT and RMC-IHT show a better performance than the conventional IHT and the subspaces-based methods MUSIC and ESPRIT. RMC-IHT have a performance similar to OMP algorithm and BHT achieves the best performance among them. Both algorithms are able to estimate the DoAs with a lower RMSE and a small number of snapshots.

Figure 3.2 depicts a scenario with $M = 15$ and $N = 2$ and the rest of the parameters are keeping as before. Note that RMC-IHT can exploit better

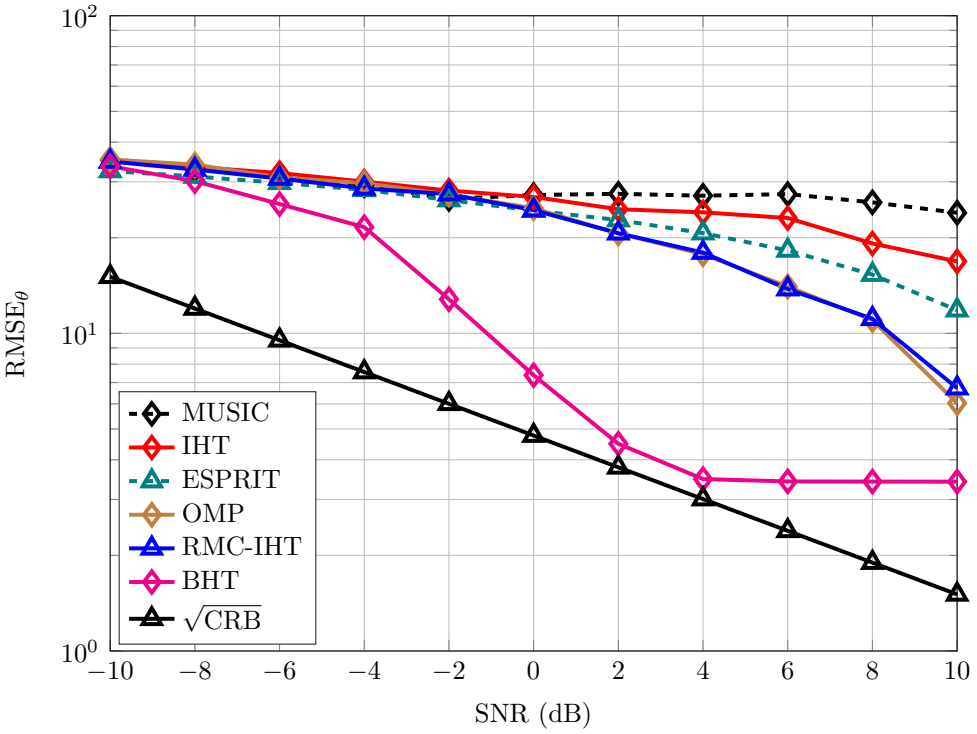


Figure 3.1: RMSE vs. SNR for M=13, N=2

the availability of a larger number of antenna elements than OMP. BHT, as before, shows the best performance.

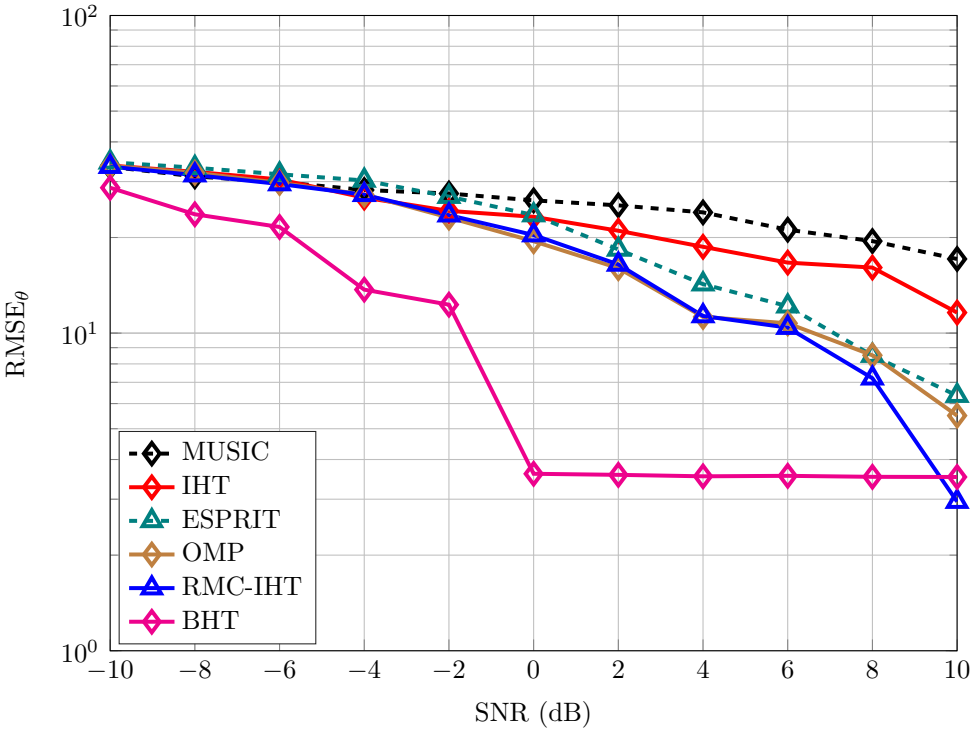


Figure 3.2: RMSE vs. SNR for M=15, N=2

Figure 3.3 represents a scenario with $M = 8$ antenna elements, $N = 5$ and the other parameters are keeping as before, can be notice that with the increase of the number of available snapshots the performance of ESPRIT improves significantly such as it is able to outperform RMC-IHT and OMP algorithms. While the RMSE of BHT decreases to 2° approximately.

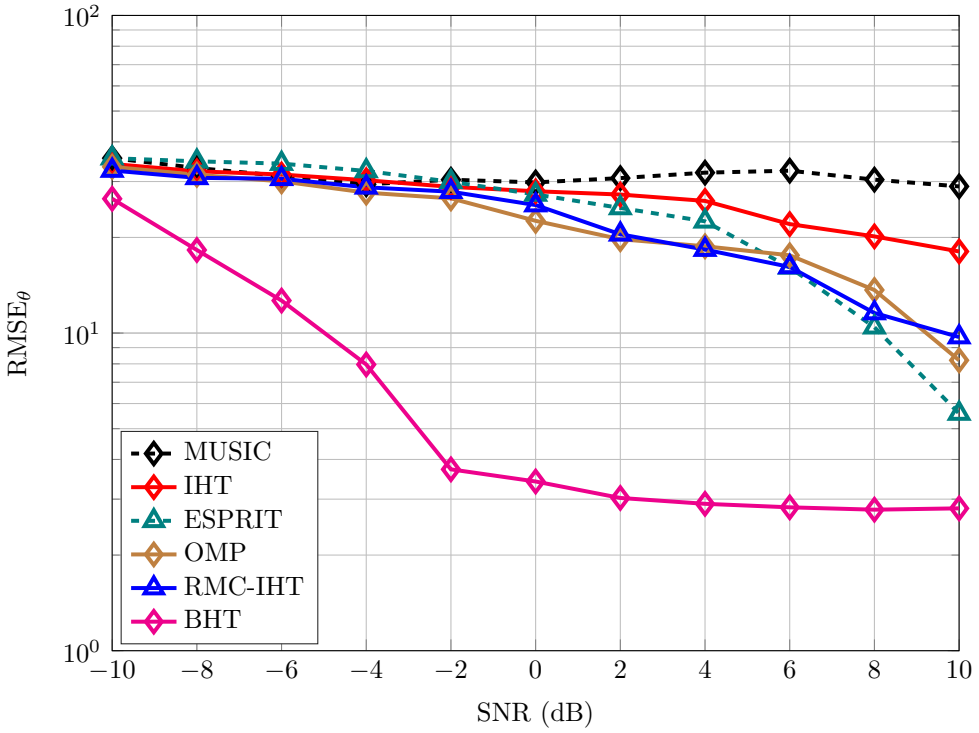


Figure 3.3: RMSE vs. SNR for M=8, N=5

Figure 3.4 depicts the same scenario with $M = 13$ and $N = 2$ but in this case with three sources which are located at the angles $\theta_1 = 4.93°$, $\theta_2 = 10.01°$ and $\theta_3 = 14.9403°$ and we can appreciate that the performance gap between RMC-IHT and OMP increases when the number of sources to be identified increases.

In order to estimate the three sources it is necessary to increase the number of snapshots. Figure 3.5 represents the same scenario but in this case with $M = 25$ antenna elements. Note that RMC-IHT achieves a RMSE of approximately 1° with estimated angles: $\hat{\theta_1} = 4.3686°$, $\hat{\theta_2} = 10.0105°$ and $\hat{\theta_3} = 15.6364°$. In the case of BHT, the increase of the number of antenna elements do not have a significant impact on the performance, because after BHT achieves a RMSE of approximately 3° the behavior stays constant. The principle reason of this saturation is that even with the increase of the antenna elements the difference among the select atoms with highest probability is not considerable.

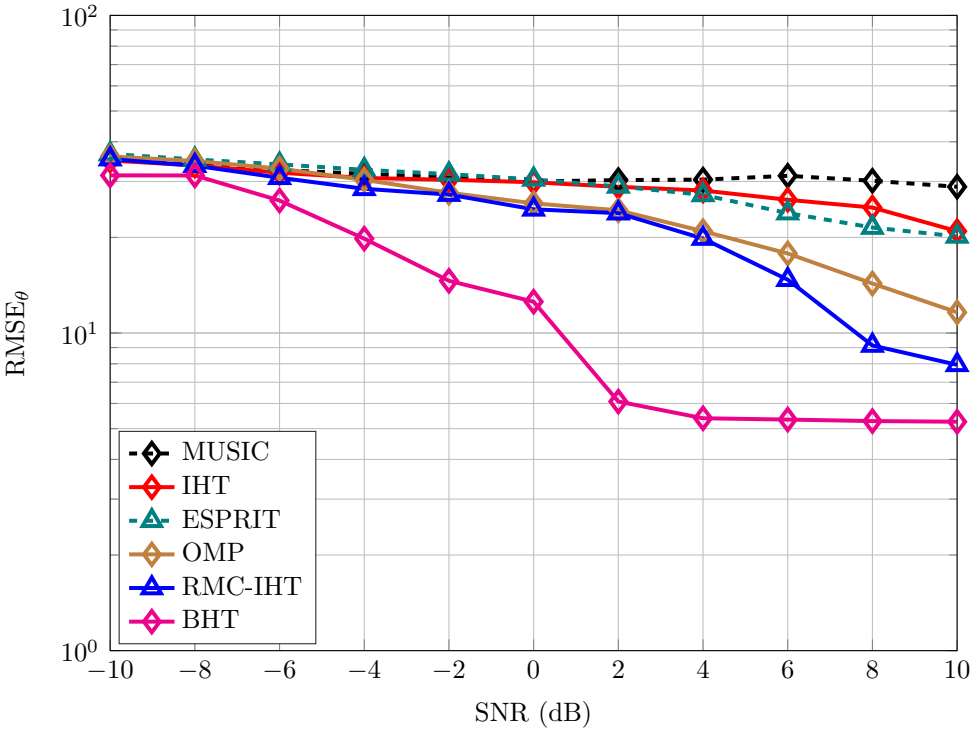


Figure 3.4: RMSE vs. SNR for M=13, N=2 for 3 sources

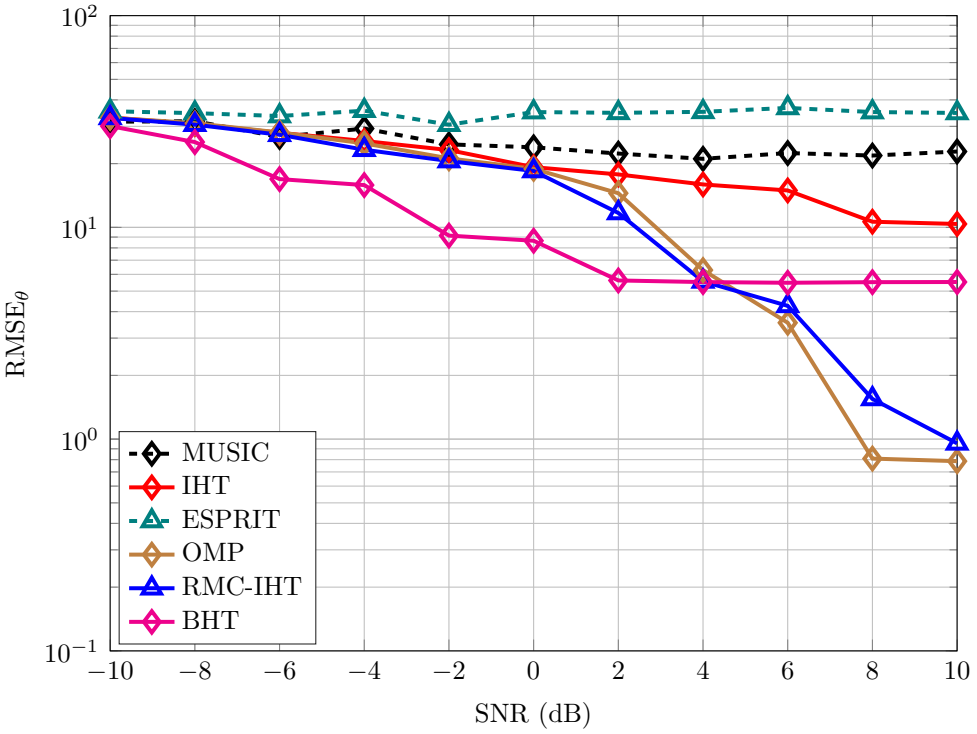


Figure 3.5: RMSE vs. SNR for M=25, N=2

In Figure 3.6 the RMSE is plotted over the number of antenna elements M for $N = 2$ snapshots and the other parameters are kept as before. We can appreciate that the performance of ESPRIT is damaged due to the small number of snapshots and increasing the number of antenna elements cannot compensate this effect. The large performance gap between the two subspace-based methods MUSIC and ESPRIT is due to the fact that MUSIC works on the noise subspace whereas ESPRIT works on the signal subspace. The dimension of the noise subspace is $M - K$ and thus grows with increasing number of antenna elements M. The dimension of the signal subspace, however, is K and thus independent of the number of antenna elements M. If the number of antenna elements M is very large compared to the number of sources K, the dimension of the noise subspace is very large such that the estimation of the noise subspace from a small number of snapshots is more robust than the one of the signal subspace with the relatively small dimension K. This explains why MUSIC is able to perform much better than ESPRIT in this scenario. The performance of IHT, RMC-IHT and MUSIC is comparable and both improve significantly with increasing the number of antenna elements and for the case of a small number of antenna elements RMC-IHT performs better than the rest of the algorithms. Note that the increase of the number of antennas elements do not have a significant impact on the performance of BHT, as it is well suited for scenarios with a moderate number of sensors and a few snapshots.

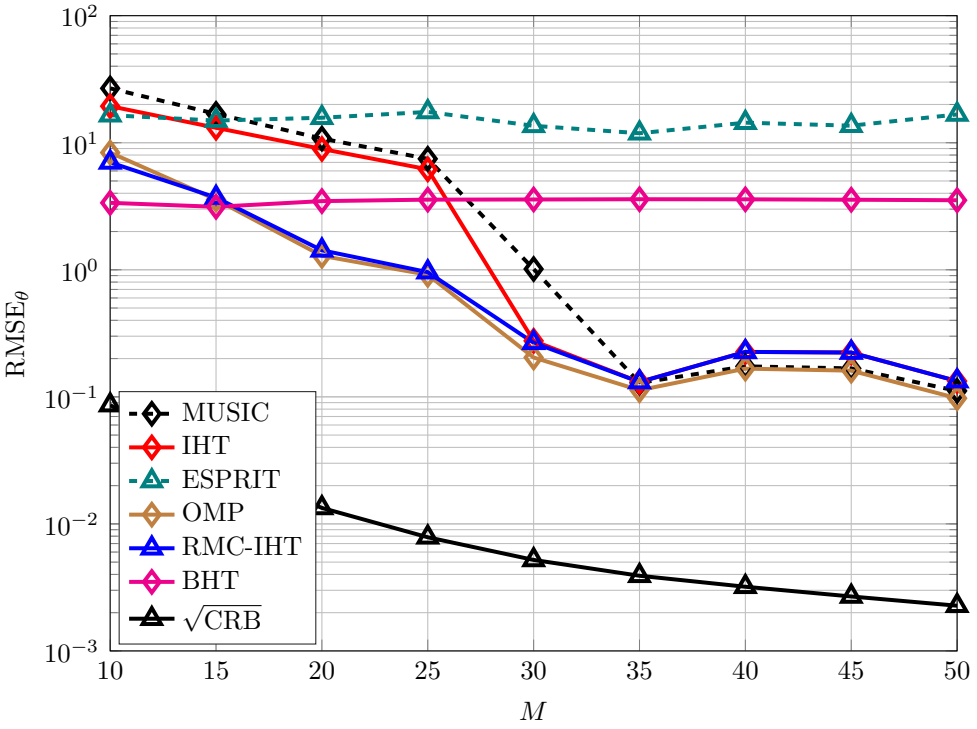


Figure 3.6: RMSE vs. M for $N = 2$

Figure 3.7 depicts a scenario with correlated sources where the correlation coefficient $\rho = 0.8$, $M = 10$, $N = 50$ and the sources are assumed correlated complex normal random variables with zero mean and variance σ_s^2 =1 and are located at the angles $\theta_1 = 4.93°, \theta_2 = 9.8969°, \theta_3 = 14.9403°$. The correlation matrix of the signal in this case can be represented by

$$\mathbf{R}_s = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix} \tag{3-23}$$

Then for confirming the accuracy of our proposed approximation in the case of correlated sources, we plot the RMC-IHT algorithm using the exact value of the covariance matrix given by equation (3-23) and the approximation that we obtain using the equation (3-14).

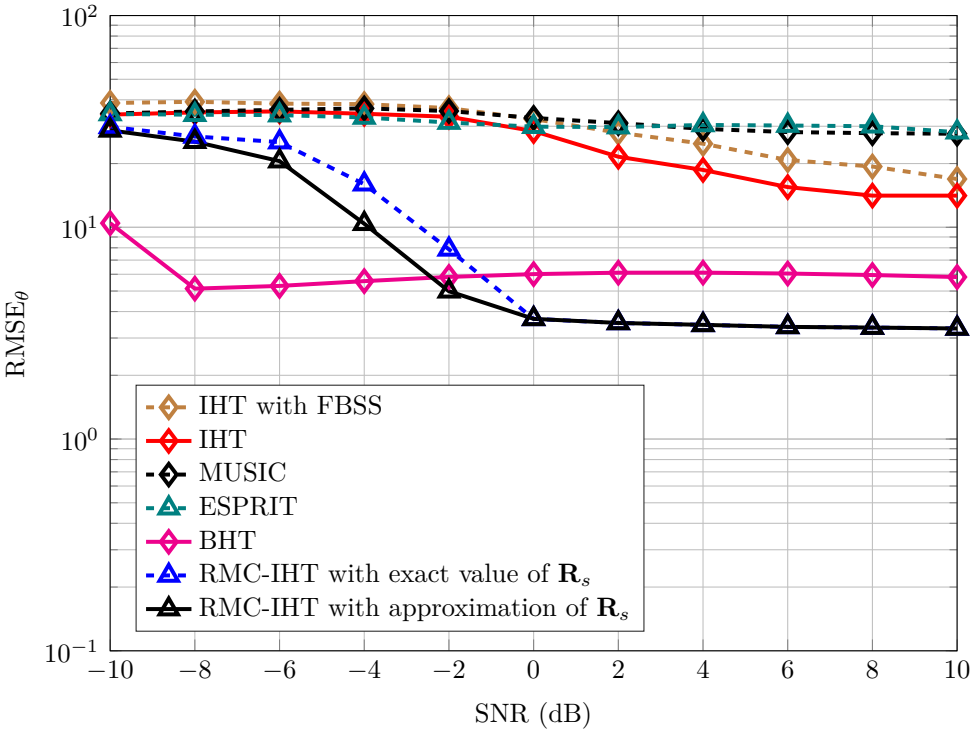


Figure 3.7: RMSE vs. SNR for correlated sources with $\rho = 0.8$

Note that the approximation used by RMC-IHT for estimating the covariance matrix is pretty close to the real value of the matrix, for this reason it shows a good performance in this kind of scenarios. In the case of the BHT algorithm, it suffers from degradation in performance due to the prior assumptions about the signal distribution do not hold, because the derivation obtained of the conditional probability p_j is based on the assumption that the covariance matrix of the signal is equal to $\sigma_s^2\mathbf{I}_K$. Despite that it shows a better behavior than ESPRIT, MUSIC and IHT.

The simulation results reveal that the proposed CS-based algorithms perform better than the subspaces-based methods when a small number of snapshots is available and a scenarios of low values of SNR. Therefore, it can be concluded that BHT algorithm is very well suited for scenarios with a moderate number of antenna elements and RMC-IHT is advantageous for scenarios with a large-scale antenna array. Moreover, both techniques do not require a large number of snapshots for outperforming the conventional algorithms. For the case of scenarios with correlated sources the version of RMC-IHT is a potential candidate as it achieves the best performance under these conditions.

4
Bayesian-BPDN (B-BPDN) Algorithm for Direction of Arrival Estimation

In this chapter an algorithmic approach to the joint sparse recovery problem based on its formulation as a basis pursuit denoising problem is introduced. As it was mentioned before the principal disadvantages of the algorithms based on ℓ_1 penalty is their high computational complexity and slow convergence. As a consequence the Bayesian basis pursuit denoising (B-BPDN) algorithm is fundamentally aimed to overcome these shortcomings.

B-BPDN is an iterative method for solving the BPDN problem that is effective especially for large-scale sparse problems. It exploits the a priori knowledge of the distribution of signal and noise for reducing the BPDN problem to the subspace determined for the atoms with the highest probability of belonging to the row support of the row K-sparse signal $\mathbf{S}$.

4.1
Bayesian Basis Pursuit Denoising (B-BPDN) algorithm

The B-BPDN algorithm is fundamentally based on the idea of the In-Crowd algorithm in [29], which is a numerical method for solving basis pursuit denoising quickly and faster than any other algorithm for large sparse problems. Since every time the In-Crowd algorithm reduced the optimization problem to a subproblem in the space defined for the active set, we decided in similar way to realize this reduction but in our case we apply a Bayesian selection for determining the atoms that belonging to the active set at each iteration. This selection is based on the computation of the conditional probability p_j of the atom belonging to the support of the sparse signal given the observed vector $\mathbf{X}$. The methodology for obtaining p_j is the same that was derived before in Chapter 3, see Section 3.4.

Another way of formulating the BPDN problem in equation (2-48) is through the introduction of a ℓ_1-norm penalty term in the cost function as follows:

$$\min_{\tilde{\mathbf{S}}\in\mathbb{C}^{P\times N}} \frac{1}{2}\left\|\mathbf{A}\tilde{\mathbf{S}}-\mathbf{X}\right\|_F^2+\lambda\left\|\tilde{\mathbf{S}}\right\|_{2,1} \tag{4-1}$$

where $\lambda > 0$ is the regularization parameter which establishes the cost of complexity relative to the error $\frac{1}{2}\left\|\mathbf{A}\tilde{\mathbf{S}} - \mathbf{X}\right\|_F^2$.

The computational complexity of solving (4-1) with a sufficiently large P is often dominated by searching the dictionary of possible atoms for appropriate additions to the active set $\mathbf{\Gamma}$ [29]. Therefore, the B-BPDN algorithm is partially insulated from the size of the global problem by consulting the dictionary only rarely. Instead of performing a full search over the P possible atoms at each iteration, the B-BPDN algorithm selects the K atoms with the highest conditional probability p_j for adding to the active set $\mathbf{\Gamma}$ and then solves the optimization problem in (4-1) as a subproblem only on the subspace spanned by the K selected atoms.

Note that solving the subproblem in B-BPDN corresponds to solving the following optimization:

$$\min_{\tilde{\mathbf{S}}\in\mathbb{C}^{K\times N}} \frac{1}{2}\left\|\mathbf{A}_{\mathbf{\Gamma}}\tilde{\mathbf{S}} - \mathbf{X}\right\|_F^2 + \lambda\left\|\tilde{\mathbf{S}}\right\|_{2,1} \tag{4-2}$$

We used Matlab's built-in **quadprog** function for the subproblem solver in (4-2), due to the cardinality of $\mathbf{\Gamma} < M$ [29]. Prior to using this function some conditions have to be fulfilled. First the MMV problem has to be expressed as a SMV problem and second it is mandatory that all the values must be real.

For attending the first condition the vectorization operator is applied. Since $\text{vec}(\mathbf{A}_{\mathbf{\Gamma}}\tilde{\mathbf{S}} - \mathbf{X}) = (\mathbf{I}_N \otimes \mathbf{A}_{\mathbf{\Gamma}})\text{vec}(\tilde{\mathbf{S}}) - \text{vec}(\mathbf{X})$, where $\otimes$ denotes the Kronecker product, then equation (4-2) can be expressed as:

$$\min_{\tilde{\mathbf{s}}\in\mathbb{C}^{KN}} \frac{1}{2}\left\|\tilde{\mathbf{A}}\mathbf{s} - \mathbf{x}\right\|_2^2 + \lambda\left\|\mathbf{s}\right\|_1, \tag{4-3}$$

where $\tilde{\mathbf{A}} \in \mathbb{C}^{(MN)\times(KN)} = \mathbf{I}_N \otimes \mathbf{A}_{\mathbf{\Gamma}}$, $\mathbf{s} \in \mathbb{C}^{KN} = \text{vec}(\tilde{\mathbf{S}})$ and $\mathbf{x} \in \mathbb{C}^{MN} = \text{vec}(\mathbf{X})$.

After that for converting all the values to real, the expression $\tilde{\mathbf{A}}\mathbf{s} - \mathbf{x}$ can be split into real and imaginary components (given by $\mathcal{R}(.)$ and $\mathcal{I}(.)$) as follows [40]:

$$\underbrace{\begin{bmatrix} \mathcal{R}(\tilde{\mathbf{A}}) & -\mathcal{I}(\tilde{\mathbf{A}}) \\ \mathcal{I}(\tilde{\mathbf{A}}) & \mathcal{R}(\tilde{\mathbf{A}}) \end{bmatrix}}_{\hat{\mathbf{A}}} \underbrace{\begin{bmatrix} \mathcal{R}(\mathbf{s}) \\ \mathcal{I}(\mathbf{s}) \end{bmatrix}}_{\hat{\mathbf{s}}} - \underbrace{\begin{bmatrix} \mathcal{R}(\mathbf{x}) \\ \mathcal{I}(\mathbf{x}) \end{bmatrix}}_{\hat{\mathbf{x}}}, \tag{4-4}$$

where $\hat{\mathbf{A}} \in \mathbb{R}^{2MN\times 2KN}$, $\hat{\mathbf{s}} \in \mathbb{R}^{2KN}$ and $\hat{\mathbf{x}} \in \mathbb{R}^{2MN}$. Note that as a result of this transformation the dimensions are increased.

Finally, the optimization problem in (4-3) can be formulated as follows:

$$\min_{\hat{\mathbf{s}}\in\mathbb{R}^{2KN}} \frac{1}{2}\left\|\hat{\mathbf{A}}\hat{\mathbf{s}} - \hat{\mathbf{x}}\right\|_2^2 + \lambda\left\|\hat{\mathbf{s}}\right\|_1. \tag{4-5}$$

The **quadprog** routine in Matlab finds a minimum for a problem specified by

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{f}^T\mathbf{x} \quad \text{s.t} \quad \mathbf{Q}\mathbf{x} \leq \mathbf{b}, \tag{4-6}$$

where $\mathbf{H}, \mathbf{Q}$ are matrices and $\mathbf{f}, \mathbf{b}$ are vectors. Therefore, for solving the least squares problem in (4-5) with the help of this quadratic programming is necessary to define first the input parameters. Hence, we rewrite the objective function in quadratic form as:

$$\begin{aligned}\min_{\hat{\mathbf{s}}} \frac{1}{2}\left\|\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{x}}\right\|_2^2+\lambda\left\|\hat{\mathbf{s}}\right\|_1 &= \min_{\hat{\mathbf{s}}} \frac{1}{2}(\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{x}})^T(\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{x}})+\lambda\left\|\hat{\mathbf{s}}\right\|_1 \\ &= \min_{\hat{\mathbf{s}}} \frac{1}{2}(\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{x}}-\hat{\mathbf{x}}^T\hat{\mathbf{A}}\hat{\mathbf{s}}+\hat{\mathbf{x}}^T\hat{\mathbf{x}})+\lambda\left\|\hat{\mathbf{s}}\right\|_1 \\ &= \min_{\hat{\mathbf{s}}} \frac{1}{2}(\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\hat{\mathbf{s}}-2\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{x}})+\lambda\left\|\hat{\mathbf{s}}\right\|_1 \\ &= \min_{\hat{\mathbf{s}}} \frac{1}{2}\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{x}}+\lambda\left\|\hat{\mathbf{s}}\right\|_1 .\end{aligned}$$

Note that the term $\hat{\mathbf{x}}^T\hat{\mathbf{x}}$ can be eliminated because it does not depend on $\hat{\mathbf{s}}$ and since $\|\hat{\mathbf{s}}\|_1$ is simply the sum of the absolute values of the elements of the vector $\hat{\mathbf{s}}$, we can write:

$$\min_{\hat{\mathbf{s}}} \frac{1}{2}\left\|\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{x}}\right\|_2^2+\lambda\left\|\hat{\mathbf{s}}\right\|_1 = \min_{\hat{\mathbf{s}}} \frac{1}{2}\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\hat{\mathbf{s}}-\hat{\mathbf{s}}^T\hat{\mathbf{A}}^T\hat{\mathbf{x}}+\lambda\sum_{i=1}^{K}|s_i|. \tag{4-7}$$

For combining the linear terms in (4-7) it is necessary to remove the absolute operator, but take into account that the vector $\hat{\mathbf{s}}$ can have negative components, the following transformation is introduced:

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}^+ - \hat{\mathbf{s}}^-, \quad \text{with} \quad \hat{\mathbf{s}}^+ \leq \mathbf{0}, \hat{\mathbf{s}}^- \leq \mathbf{0} \tag{4-8}$$

where the vector $\hat{\mathbf{s}}$ is split into the positive and negative components $\hat{\mathbf{s}}^+$ and $\hat{\mathbf{s}}^-$ respectively.

Based on (4-8), it is defined a column vector $\hat{\mathbf{z}} = \begin{bmatrix} \hat{\mathbf{s}}^+ \\ \hat{\mathbf{s}}^- \end{bmatrix}$ and a new matrix $\bar{\mathbf{A}} = [\hat{\mathbf{A}} \quad -\hat{\mathbf{A}}]$. Substituting these elements in equation (4-7), the function to optimize turns into:

$$\min_{\hat{\mathbf{z}}} \frac{1}{2}\hat{\mathbf{z}}^T\bar{\mathbf{A}}^T\bar{\mathbf{A}}\hat{\mathbf{z}}-\hat{\mathbf{z}}^T\bar{\mathbf{A}}^T\hat{\mathbf{x}}+\lambda\mathbb{1}^T\hat{\mathbf{z}} \quad \text{s.t} \quad \mathbf{Q}\hat{\mathbf{z}} \leq \mathbf{0} \tag{4-9}$$

where $\mathbb{1} \in \mathbb{R}^{4KN}$ is a column-vector of ones, $\mathbf{Q} = -\mathbf{I}_{4KN}$ and $\mathbf{0}$ is the zero vector. The constraint ensures that $\hat{\mathbf{s}}^+$ and $\hat{\mathbf{s}}^-$ take non-negative values. Then, equation (4-9) can be written as:

$$\min_{\hat{\mathbf{z}}} \frac{1}{2}\hat{\mathbf{z}}^T \underbrace{\bar{\mathbf{A}}^T\bar{\mathbf{A}}}_{\mathbf{H}}\hat{\mathbf{z}} + \underbrace{(\lambda\mathbb{1}-\bar{\mathbf{A}}^T\hat{\mathbf{x}})^T}_{\mathbf{f}^T}\hat{\mathbf{z}} \quad \text{s.t} \quad \mathbf{Q}\hat{\mathbf{z}} \leq \mathbf{0} \tag{4-10}$$

Finally, equation (4-10) is written in quadratic form and the input parameters of the quadratic programming can be defined as :

$$\mathbf{H} = \bar{\mathbf{A}}^T \bar{\mathbf{A}}$$
$$\mathbf{f} = \lambda \mathbb{1} - \bar{\mathbf{A}}^T \hat{\mathbf{x}}$$
$$\mathbf{Q} = -\mathbf{I}_{4KN}$$
$$\mathbf{b} = \mathbf{0}.$$

The proposed B-BPDN algorithm is described in Algorithm 6.

Algorithm 6: B-BPDN

Input: $\mathbf{X}, \mathbf{A}, \boldsymbol{\theta} = [\theta_1, ..., \theta_P], K, \mu, , \gamma, \sigma_n^2, \lambda, N_{\max}$
Initialization:
$\hat{\mathbf{S}}^{(0)} = \mathbf{0}, \mathbf{R}^{(0)} = \mathbf{X}, i = 0$
while the stopping criterion is not met **do**
  compute the conditional probability p_j of each atoms as
  for $j \leftarrow 1$ **to** P **do**
    $c_j = 1 + \gamma \left\| \mathbf{A}_j \right\|^2$
    $p_j = \exp\left\{ \frac{\gamma}{\sigma_n^2} \frac{\left\| \mathbf{R}^H \mathbf{A}_j \right\|^2}{c_j} - N \log c_j \right\}$
  end
  Determine the active support $\boldsymbol{\Gamma}$ composed for the K atoms which have the highest p_j
  Solve (4-1) on the subspace spanned by all the components in $\boldsymbol{\Gamma}$ using the SolveQuadratic algorithm
  Set all the rows of $\hat{\mathbf{S}}$ to $\mathbf{0}$ except for the K rows in $\boldsymbol{\Gamma}$, set these to the solution found in the previous step($\hat{\mathbf{S}}_{(\boldsymbol{\Gamma})} = \tilde{\mathbf{S}}$ and $\hat{\mathbf{S}}_{(\bar{\boldsymbol{\Gamma}})} = \mathbf{0}$)
  $\mathbf{R} = \mathbf{X} - \mathbf{A}_{\boldsymbol{\Gamma}} \hat{\mathbf{S}}_{(\boldsymbol{\Gamma})}$
  $i = i + 1$
end
Output:
$\hat{\mathbf{S}} = H_K(\hat{\mathbf{S}}^i)$
$\boldsymbol{\Gamma} = \text{rsupp}(\hat{\mathbf{S}})$
DoAs: $[\hat{\theta}_1, \ldots, \hat{\theta}_K] = \boldsymbol{\theta}(\boldsymbol{\Gamma})$

The SolveQuadratic algorithm is described in Algorithm 7.

Algorithm 7: SolveQuadratic

Input: $\mathbf{A_\Gamma}, \mathbf{X}, \lambda$

$\tilde{\mathbf{A}} = \mathbf{I}_N \otimes \mathbf{A_\Gamma}$ % vectorization

$\mathbf{x} = \text{vec}(\mathbf{X})$

Split in real and imaginary part as in equation (4-4)

Split in positive and negative components as in equation (4-8)

Defined the input parameter of the quadratic programming

$\mathbf{H} = \bar{\mathbf{A}}^T\bar{\mathbf{A}}$, $\mathbf{f} = \lambda\mathbb{1} - \hat{\mathbf{A}}^T\hat{\mathbf{x}}$, $\mathbf{Q} = -\mathbf{I}_{4KN}$, $\mathbf{b} = \mathbf{0}$

Using Matlab's **quadprog** function for solving the optimization subproblem

```
\hat{z}=quadprog(H,f,Q,b);
```

Split the found solution $\hat{\mathbf{z}}$ into the positive and negative elements of $\hat{\mathbf{s}}$

Transform to complex form

```
s=[\hat{s}(1:KN) + \hat{s}(KN+1:end)*1i];
```

Revert the vectorization process using Matlab's **vec2mat** function

```
\tilde{S}=vec2mat(s,K);
\tilde{S}=\tilde{S}';
```

Output: $\tilde{\mathbf{S}}$

The B-BPDN algorithm uses a stopping criterion to terminate when the active support $\mathbf{\Gamma}$ does not change from one iteration to the next or when the maximum number of iterations is reached.

4.2 Computational Complexity

In this section we provide an analysis of the algorithmic complexity of B-BPDN. Note that the proposed algorithm has an outer loop which computes the conditional probability p_j of each atom and an inner loop which solves the optimization subproblem using quadratic programming.

– Outer Loop Complexity:

The outer loop requires a matrix multiplication of the residual $\mathbf{R}$ (which is $M \times N$) with the $\mathbf{A}_j$ column (which is $M \times 1$) of the measurement matrix $\mathbf{A}$. This result in MN requiring operations. Considering that the number of runs of the outer loop is P, it finally requires approximately PNM operations.

– Inner Loop Complexity:

Matlab's **quadprog** routine is an interior-point-convex algorithm which attempts to follow a path that is strictly inside the constraints. It uses a presolve module to remove redundancies, and to simplify the problem by

solving for components that are straightforward [41]. It uses an interior point method with complexity $\mathcal{O}(n^3)$ [29], where n is the dimension of the vector that it is minimized.

For this reason, the computational complexity of the subproblem solver in the inner loop is $\mathcal{O}(64K^3N^3)$ due to the dimension of the resultant vector $\hat{\mathbf{z}}$ after the transformations is $4KN$, see equation (4-9).

Finally, the computational complexity of B-BPDN can be approximated to $\mathcal{O}(64K^3N^3 + PMN)$. Note that if we compare this result with the complexity of solving the BPDN problem applied others well known techniques as SPGL1, see Table 2.1, the computational complexity of the B-BPDN algorithm is considerably reduced for the case when the size of the problem is greater than the number of available snapshots, i.e., $P >> N$.

4.3 Simulations Results

In this section we perform several simulation experiments to test the proposed B-BPDN.

Figure 4.1 represents a scenario with K=2 sources which are assumed uncorrelated complex normal random variables with zero mean and variance σ_s^2 =1 and located at the angles $\theta_1 = 4.93°$ and $\theta_2 = 10.01°$, the RMSE is plotted over the SNR with the number of antenna elements $M = 13$ for $N = 2$, $N_{\max} = 20$ and $\lambda = 2\sigma_n^2$. For evaluating the performance of B-BPDN, we compare it with the SPGL1 algorithm whose Matlab code is available in [42], and with others subspaces-based and greedy algorithms.

We can appreciate that the gap between the performance of the algorithms is considerable and the proposed B-BPDN shows the best performance.

In another simulation example in Figure 4.2, we set $M = 8$, $N = 2$ and the number of sources is increasing to three which are located at the angles $\theta_1 = 4.93°$, $\theta_2 = 10.01°$ and $\theta_3 = 14.9403°$. We keep the values of other parameters as before. Note that B-BPDN performs better than SPGL1 even with the decrease of number of antenna elements and the increase of the number of sources. It is also superior to the other techniques.

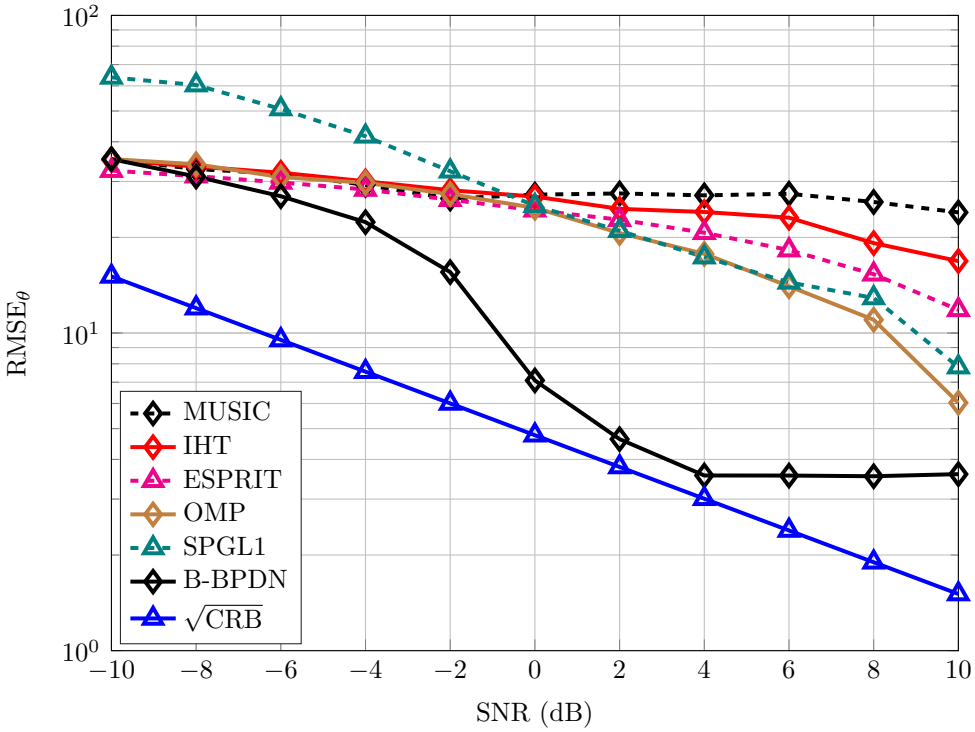


Figure 4.1: RMSE vs. SNR for $M = 13$ and $N = 2$

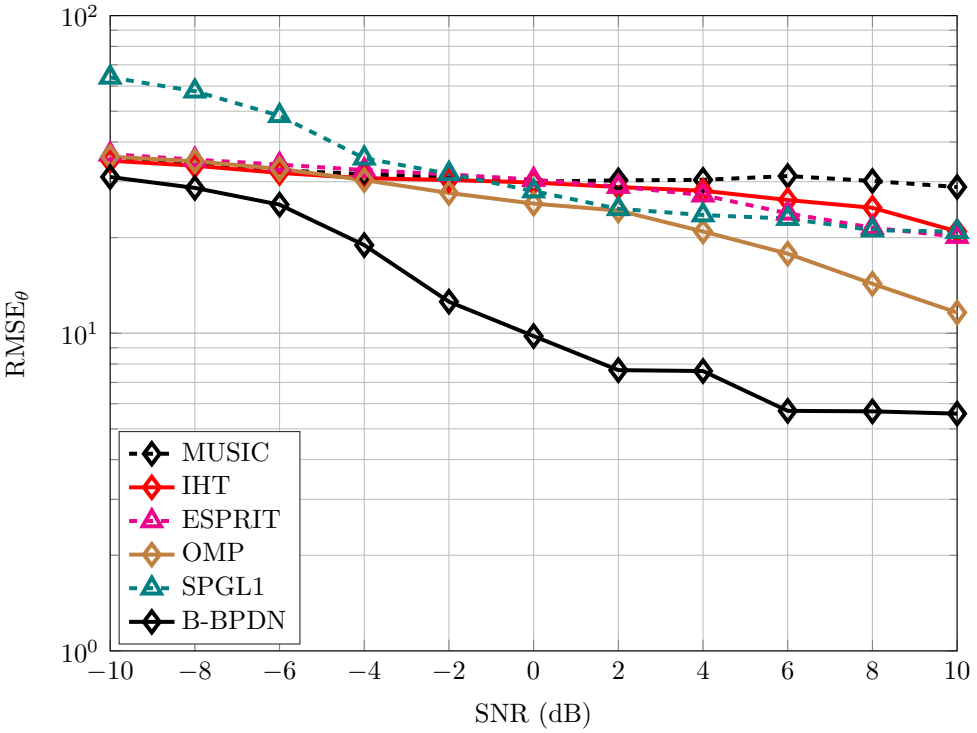


Figure 4.2: RMSE vs. SNR for $M = 8$ and $N = 2$

For evaluating the computational complexity we simulate three different scenarios and compare the simulation times of both algorithms.

- Case 1: corresponds to a scenario of $M = 15$, $N = 2$, two sources which are located at the angles $\theta_1 = 4.93°$, $\theta_2 = 10.01°$
- Case 2: corresponds to a scenario of $M = 50$, $N = 10$, two sources which are located at the angles $\theta_1 = 4.93°$, $\theta_2 = 10.01°$
- Case 3: corresponds to a scenario of $M = 100$, $N = 20$, three sources which are located at the angles $\theta_1 = 4.93°$, $\theta_2 = 10.01°$ and $\theta_3 = 14.9403°$.

The SNR is set to 10 dB. Table 4.1 shows the obtained results. It can be noticed that the complexity of B-BPDN is inferior for all the cases which confirm that the performance of a Bayesian selection before solving the BPDN problem reduces the computational complexity of the ℓ_1 regularized problem. We also provide the running times of the RMC-IHT algorithm.

Table 4.1: Running Times in seconds

Case	SPGL1	B-BPDN	RMC-IHT
Case 1	1.5228	0.0592	0.0934
Case 2	4.2892	0.0806	0.1209
Case 3	5.9227	0.1496	0.1848

In order to see how the correlation among the sources affects the performance of the B-BPDN algorithm, we run a simulation for a scenario with three correlated sources which are located at the angles $\theta_1 = 4.93°$, $\theta_2 = 10.01°$ and $\theta_3 = 14.9403°$. The correlation coefficient is $\rho = 0.9$, $M = 10$ and $N = 10$. Figure 4.3 shows that the performance of B-BPDN is also superior to SPGL1 in this kind of scenarios.

The simulation results reveal that the proposed B-BPDN algorithm perform better than the subspaces-based methods and SPGL1 algorithm in scenarios with a limited number of snapshots and low values of SNR. Moreover, the computational complexity of B-BPDN is considerably inferior to SPGL1 for case when the size of the grid is much larger than the number of available snapshots. Therefore, B-BPDN is advantageous for scenarios where a exhaustive search is required, i.e. for large sampling grids.

Finally, Figure 4.4 represents a comparison among the proposed algorithms in this thesis in a scenario with $M = 10$, $N = 2$ and two sources which are located at the angles $\theta_1 = 4.93°$ and $\theta_2 = 10.01°$.

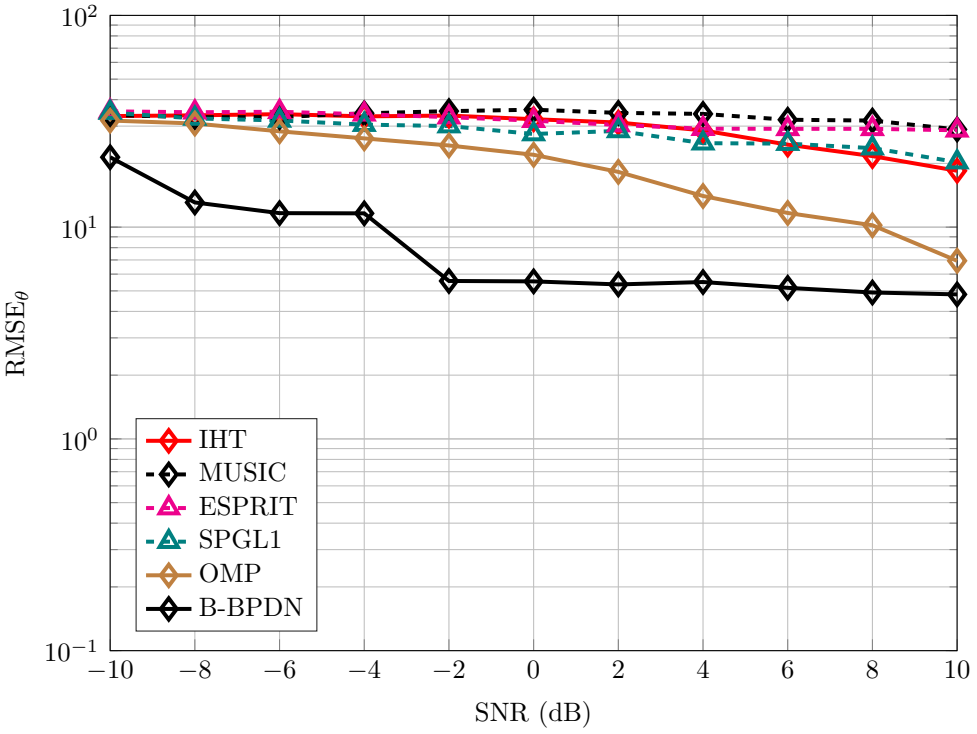


Figure 4.3: RMSE vs. SNR for $M = 10$ and $N = 10$

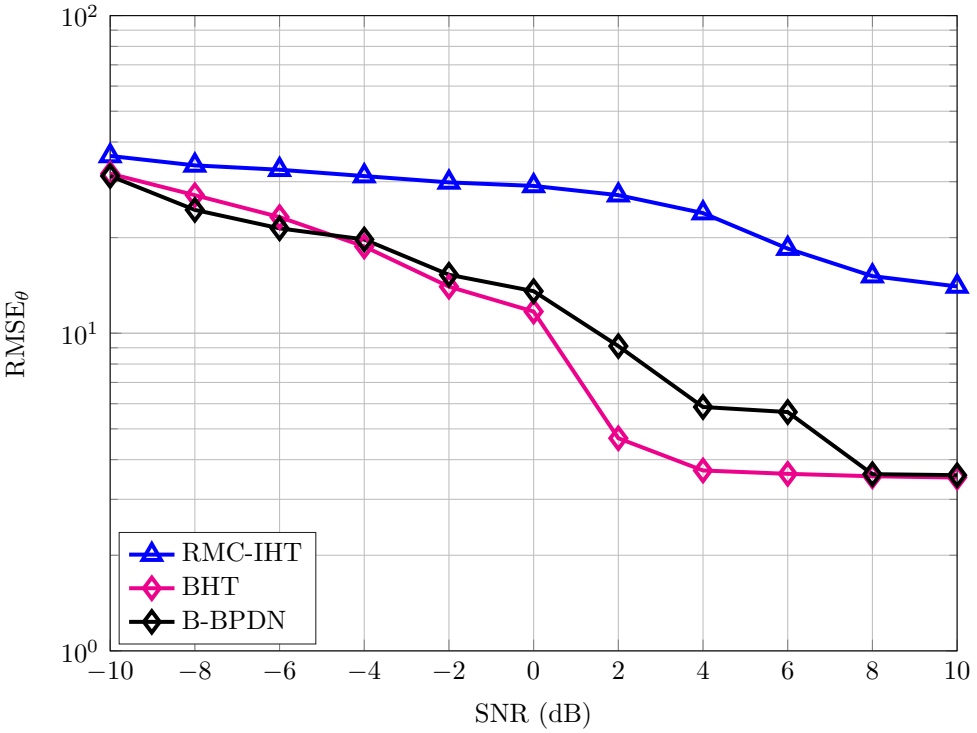


Figure 4.4: RMSE vs. SNR for $M = 10$ and $N = 2$

5
Conclusions and Future Works

In this thesis we have considered the problem of direction finding using sensor array by transforming it into the problem of sparse signal representation. A review of the principle sparse recovery techniques based on ℓ_p penalties with $p \leq 1$ has been also presented. In the case of the ℓ_0 penalty due to the NP-hard nature of the problem, different approaches using greedy algorithms have been described. For optimization techniques involving ℓ_1 penalties which leads to the convex optimization problem, we have used a formulation based on the basis pursuit denoising problem.

In Chapter 3, two Bayesian approaches to the IHT algorithm have been developed. The RMC-IHT algorithm generates a set of potential candidate supports and applies a Bayesian criterion, based on the computation of the MAP estimator of the signal, for selecting the best candidate at each iteration. We have also developed an extension of this algorithm for the case of a scenario with correlated sources.

The BHT which uses the conditional probability associated at each atoms for transforming the deterministic hard thresholding operator in a Bayesian operator. The benefits of using RMC-IHT and BHT are mainly that they are able to estimate the DoAs with a lower RMSE than previously reported techniques in scenarios with only a few available snapshots. RMC-IHT is advantageous for scenarios with large-scale antenna array and only a few snapshots and BHT is well suited for scenarios with a moderate number of sensors and a few snapshots. Moreover, this is also noticed in the case of correlated sources that RMC-IHT performs better than other techniques.

In Chapter 4, a novel way of incorporating prior knowledge for solving the BPDN problem is proposed. Having knowledge of the atoms with highest probabilities of belonging to the support of the sparse signal, the optimization problem is reduced to the subspace determined for these atoms. As a consequence the estimation performance of signals from unknown directions is substantially improved and the computational complexity is reduced.

Simulation results shows that the utilization of the Bayesian algorithms which take into account the a priori knowledge of the signal distribution provide more accurate signal recovery than non-Bayesian algorithms.

The proposed algorithms in this thesis can be considered for future works in other fields of application, for example in image denoising and beamforming problems. Moreover, the proposed Bayesian algorithms can be extended to the cases where the sparse signal and the noise have non-Gaussian prior distributions. This is especially important when dealing with heavy-tailed distributions, since the estimators developed under Gaussian assumption perform poorly in the presence of outliers.

Bibliography

[1] MALIOUTOV, D.; ÇETIN, M. ; WILLSKY, A.. **A sparse signal reconstruction perspective for source localization with sensor arrays**. IEEE Transactions on Signal Processing, 53(8):3010–3022, Aug, 2005.

[2] TREES, H. V.. **Optimum Array Processing. Part IV of Detection, Estimation and Modulation Theory**. J.Wiley and Sons, 2002.

[3] STEINWANDT, J.. **Advanced array signal processing techniques for beamforming and direction finding**. Master thesis, Ilmenau University of Technology, Germany, 2011.

[4] CAPON, J.. **High-resolution frequency-wavenumber spectrum analysis**. Proceedings of the IEEE, 57:1408–1418, Aug, 1969.

[5] SCHMIDT, R.. **Multiple emitter location and signal parameter estimation**. IEEE Transactions Antennas and Propagation, 34:276–280, 2011.

[6] RAO, B. D.; HARI, K.. **Performance analysis of Root-MUSIC**. IEEE Transactions on Acoustics, Speech and Signal Processing, 37(12):1939–1949, December, 1989.

[7] PAULRAJ, A.; ROY, R. ; KAILATH, T.. **Estimation of signal parameter via rotational invariance techniques- ESPRIT**. Circuits Systems and Computers, p. 83–89, 1985.

[8] GROVER, R.; PADOS, D. A. ; MEDLEY, M. J.. **Subspace direction finding with an auxiliary-vector basis**. IEEE Transactions on Signal Processing, 55:758–763, January, 2007.

[9] SEMIRA, H.; BELKACEMI, H. ; MARCOS, S.. **High resolution source localization algorithm based on the conjugate gradients**. EURASIP Journal on Advances in Signal Processing, 2007:1–9, March, 2007.

[10] WANG, L.; DE LAMARE, R. C. ; HAARDT, M.. **Direction finding algorithms based on joint iterative subspace optimization**. IEEE Transactions on Aerospace and Electronic Systems, 50:2541–2553, Oct, 2014.

[11] QIU, L.; CAI, Y.; DE LAMARE, R. C. ; ZHAO, M.. **Reduced-rank doa estimation algorithms based on alternating low-rank decomposition**. IEEE Signal Processing Letters, 2016.

[12] STÖCKLE, C.; MUNIR, J.; MEZGHANI, A. ; NOSSEK, J. A.. **DoA estimation performance and computational complexity of subspace and compressed sensing-based methods**. In: WSA 2015; 19TH INTERNATIONAL ITG WORKSHOP ON SMART ANTENNAS, Ilmenau, Germany, March, 2015.

[13] BLUMENSATH, T.; DAVIES, M. E.. **Iterative hard thresholding for compressed sensing**. Applied and Computational Harmonic Analysis, 27:265–274, 2009.

[14] BLANCHARD, J. D.; CERMAK, M.; HANLE, D. ; JING, Y.. **Greedy algorithms for joint sparse recovery**. IEEE Transactions on Signal Processing, 62(7):1694–1704, April 2014.

[15] TROPP, J. A.; GILBERT, A. C.. **Signal recovery from random measurements via orthogonal matching pursuit**. IEEE Transactions on Information Theory, 53(12):4655–4666, Dec, 2007.

[16] DEN BERG, E.; FRIEDLANDER, M. P.. **Sparse optimization with least squares constraints**. IEEE Transactions on Information Theory, 21(4):1201–1229, 2011.

[17] ELDAR, Y.; KUTYNIOK, G.. **Compressed Sensing: theory and applications**. Cambridge University Press, 2012.

[18] EJAZ, A.. **Algorithms for sparse signal recovery in compressed sensing**. Master thesis, Aalto university school of electrical engineering, Finland, May, 2015.

[19] EJAZ, A.; OLLILA, E. ; KOIVUNEN, V.. **Randomized simultaneous orthogonal matching pursuit**. In: 23RD EUROPEAN SIGNAL PROCESSING CONFERENCE EUSIPCO, Nice, France, Aug, 2015.

[20] MALIOUTOV, D.. **A sparse signal reconstruction perspective for source localization with sensor arrays**. Master thesis, Massachusetts Institute of Technology, EEUU, July, 2003.

[21] STOICA, P.; NEHORAI, A.. **MUSIC, Maximum likelihood, and Cramer-Rao bound: further results and comparisons**. IEEE Transactions on Acoustics, Speech, and Signal Processing, 38(12):2140–2150, Dec 1990.

[22] HAARDT, M.; NOSSEK, J. A.. **Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden**. IEEE Transactions on Signal Processing, 43(5):1232–1242, May, 1995.

[23] J.THOMPSON; GRANT, P. M. ; MULGREW, B.. **Performance of Spatial Smoothing algorithms for correlated sources**. IEEE Transactions on Signal Processing, 44(4):1040–1046, April, 1996.

[24] THAKRE, A.; HAARDT, M. ; GIRIDHAR, K.. **Single snapshot spatial smoothing with improved effective array aperture**. IEEE Signal Processing Letters, 16(6), June, 2009.

[25] S.U.PILLAI; KWON, B. H.. **Foward/backward spatial smoothing techniques for coherent signal identification**. IEEE Transactions on Acoustics, Speech and Signal Processing, 37(1):3489–3499, January, 1989.

[26] SHAN, T.-J.; WAX, M. ; KAILATH, T.. **On spatial smoothing for direction-of-arrival estimation of coherent signals**. IEEE Transactions on Acoustics, Speech, and Signal Processing, 33(4):806–811, Aug 1985.

[27] BARANIUK, R.. **Compressive sensing**. IEEE Signal Processing Magazine, July, 2007.

[28] WANG, Y.; LEUS, G. ; PANDHARIPANDE, A.. **Direction estimation using compressive sampling array processing**. In: 2009 IEEE/SP 15TH WORKSHOP ON STATISTICAL SIGNAL PROCESSING, p. 626–629, Aug 2009.

[29] GILL, P. R.; WANG, A. ; MOLNAR, A.. **The in-crowd algorithm for fast basis pursuit denoising**. IEEE Transactions on Signal Processing, 59(10):4595–4605, Oct,2011.

[30] BLUMENSATH, T.; DAVIES, M. E.. **Normalized iterative hard thresholding : guaranteed stability and performance**. IEEE Journal of Select Topics in Signal Processing, 4(2), April, 2010.

[31] HUANG, S.; ZHU, J.. **Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP**. Inverse Problems, 27(3), 2011.

[32] WANG, Y.; FU, T.; GAO, M. ; DING, S.. **Performance of orthogonal matching pursuit for multiple measurement vectors with noise**.

In: 2013 IEEE CHINA SUMMIT AND INTERNATIONAL CONFERENCE ON SIGNAL AND INFORMATION PROCESSING, p. 67–71, July 2013.

[33] VAN DEN BERG, E.; FRIEDLANDER, M. P.. **Probing the pareto frontier for basis pursuit solutions**. SIAM J. Sci. Comput., 31(2):890–912, Nov, 2008.

[34] ELAINE T. HALE, W. Y.; ZHANG, Y.. **Fixed-point continuation for ℓ_1-minimization: Methodology and convergence**. SIAM J. Sci. Comput., 19(3):1107–1130, 2008.

[35] BECK, A.; TEBOULLE, M.. **A fast iterative shrinkage-thresholding algorithm for linear inverse problems**. SIAM J. Img. Sci., 2(1):183–202, March, 2009.

[36] K.KOH; W.YIN ; S.BOYD. **A fast iterative shrinkage-thresholding algorithm for linear inverse problems**. Stanford University, 2007.

[37] BECKER, S.; BOBIN, J. ; CANDÈS, E. J.. **Nesta: A fast and accurate first-order method for sparse recovery**. SIAM J. Img. Sci., 4(1):1–39, Jan, 2011.

[38] GARCÍA, Y.; DE LAMARE, R. C. ; HAARDT, M.. **Randomized multiple candidate iterative hard thresholding for direction of arrival estimation**. In: WSA 2015; 22TH INTERNATIONAL ITG WORKSHOP ON SMART ANTENNAS, Bochum, Germany, March, 2018.

[39] ELAD, M.. **Sparse and redundant representations**. Springer, 2010.

[40] HAWES, M.; MIHAYLOVA, L.; SEPTIER, F. ; GODSILL, S.. **Bayesian compressive sensing approaches for direction of arrival estimation with mutual coupling effects**. IEEE Transactions on Antennas and Propagation, 65(3):1357–1368, March 2017.

[41] MATHWORKS. **Quadratic Programming: Optimization algorithms and examples (Optimization Toolbox)**, Feb 2011. http://www.mathworks.com/help/toolbox/optim/ug/brnox71.html.

[42] VAN DEN BERG, E.; FRIEDLANDER, M. P.. **SPGL1: A solver for large-scale sparse reconstruction**, June 2007. http://www.cs.ubc.ca/labs/scl/spgl1.

A
The Cramer Rao Bound

The Cramer Rao inequality establishes a lower bound on the variance of any unbiased estimators for nonrandom unknown parameter estimation problems. It is used to evaluate potential estimators. If an unbiased estimator meets the CRB with an equality, that means that the task of search for a good estimator is over, we cannot get the variance any lower (of course there is always a possibility of considering other metrics of merit) [1].

Theorem (Cramer Rao Inequality): *If an estimator* $\hat{\boldsymbol{\theta}}(x)$ *is unbaised* $(E[\hat{\boldsymbol{\theta}}(x)] = \boldsymbol{\theta})$, then

$$Var[\hat{\boldsymbol{\theta}}(x)] \geq \left(\left[\left[\frac{\partial ln p_x(x|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^T \left[\frac{\partial ln p_x(x|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \right] \right)^{-1} = \left(-E \left[\frac{\partial^2 ln p_x(x|\boldsymbol{\theta})}{\partial^2 \boldsymbol{\theta}} \right] \right) \tag{A-1}$$

where the probability $p_x(x|\boldsymbol{\theta})$ *is assumed to be strictly positive and twice continuously differentiable*

The matrix $I_x(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 ln p_x(x|\boldsymbol{\theta})}{\partial^2 \boldsymbol{\theta}}\right]$ is called the expected Fisher information matrix. It is dependent of x but varies with $\boldsymbol{\theta}$, so the bound is a function of the unknown parameter. An estimator meeting the CRB for every value of $\boldsymbol{\theta}$ is labeled an efficient estimator. If an estimator is efficient then it is also the minimum variance unbiased estimator [20].