Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

GLOBAL OPTIMIZATION METHODS FOR LOCALIZATION IN COMPRESSIVE SENSING

by
Marco Rossi

The dissertation discusses *compressive sensing* and its applications to localization in multiple-input multiple-output (MIMO) radars. Compressive sensing is a paradigm at the intersection between signal processing and optimization. It advocates the sensing of "sparse" signals (i.e., represented using just a few terms from a basis expansion) by using a sampling rate much lower than that required by the Nyquist-Shannon sampling theorem (i.e., twice the highest frequency present in the signal of interest). Low-rate sampling reduces implementation's constraints and translates into cost savings due to fewer measurements required. This is particularly true in localization applications when the number of measurements is commensurate to antenna elements. The theory of compressive sensing provides precise guidance on how the measurements should be acquired, and which optimization algorithm should be used for signal recovery.

The first part of the dissertation addresses the application of compressive sensing for localization in the spatial domain, specifically direction of arrival (DOA), using MIMO radar. A sparse localization framework is proposed for a MIMO array in which transmit and receive elements are placed at random. This allows for a dramatic reduction in the number of elements needed, while still attaining performance comparable to that of a filled (Nyquist) array. By leveraging properties of structured random matrices, a bound on the coherence of the resulting measurement matrix is obtained, and conditions under which the measurement matrix satisfies the so-called *isotropy* property are detailed. The coherence and isotropy concepts are used to establish uniform and non-uniform recovery guarantees within the proposed spatial

compressive sensing framework. In particular, it is shown that non-uniform recovery is guaranteed if the product of the number of transmit and receive elements, MN (which is also the number of degrees of freedom), scales with $K\left(\log G\right)^2$, where K is the number of targets and G is proportional to the array aperture and determines the angle resolution. In contrast with a filled virtual MIMO array where the product MN scales linearly with G, the logarithmic dependence on G in the proposed framework supports the high-resolution provided by the virtual array aperture while using a small number of MIMO radar elements.

The second part of the dissertation focuses on the sparse recovery problem at the heart of compressive sensing. An algorithm, dubbed Multi-Branch Matching Pursuit (MBMP), is presented which combines three different paradigms: being a greedy method, it performs iterative signal support estimation; as a rank-aware method, it is able to exploit signal subspace information when multiple snapshots are available; and, as its name foretells, it possesses a multi-branch structure which allows it to trade-off performance (e.g., measurements) for computational complexity. A sufficient condition under which MBMP can recover a sparse signal is obtained. This condition, named MB-coherence, is met when the columns of the measurement matrix are sufficiently "incoherent" and when the signal-to-noise ratio is sufficiently high. The condition shows that successful recovery with MBMP is guaranteed for dictionaries which do not satisfy previously known conditions (e.g., coherence, cumulative coherence, or the Hanman relaxed coherence).

Finally, by leveraging the MBMP algorithm, a framework for target detection from a set of compressive sensing radar measurements is established. The proposed framework does not require any prior information about the targets' scene, and it is competitive with respect to state-of-the-art detection compressive sensing algorithms.

GLOBAL OPTIMIZATION METHODS FOR LOCALIZATION IN COMPRESSIVE SENSING

by
Marco Rossi

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering, NJIT

May 2014

APPROVAL PAGE

GLOBAL OPTIMIZATION METHODS FOR LOCALIZATION IN COMPRESSIVE SENSING

Marco Rossi

Dr. Alexander M. Haimovich, Dissertation Advisor — Date
Professor, Department of Electrical and Computer Engineering, NJIT

Dr. Yeheskel Bar-Ness, Committee Member — Date
Distinguished Professor, Department of Electrical and Computer Engineering, NJIT

Dr. Osvaldo Simeone, Committee Member — Date
Associate Professor, Department of Electrical and Computer Engineering, NJIT

Dr. Ali Abdi, Committee Member — Date
Associate Professor, Department of Electrical and Computer Engineering, NJIT

Dr. Mark A. Govoni, Committee Member — Date
Chief Scientist, Radar ISR Technology SME, US Army CERDEC I2WD

BIOGRAPHICAL SKETCH

Author: Marco Rossi

Degree: Doctor of Philosophy

Date: May 2014

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2014
- Master of Science in Telecomunication Engineering,
  Politecnico di Milano, Milan, 2007
- Bachelor of Science in Telecomunication Engineering,
  Politecnico di Milano, Milan, 2004

Major: Electrical Engineering

Presentations and Publications:

M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Multi-Branch Marching Pursuit with Applications to MIMO Radar," submitted to IEEE Transactions on Signal Processing, Dec. 2013.

M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial Compressive Sensing for MIMO Radar," IEEE Transactions on Signal Processing, vol. 62, no. 2, pp. 419-430, Jan. 2014.

M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Conditions for Target Recovery in Spatial Compressive Sensing for MIMO Radar," in Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013), Vancouver, Canada, May 26-31, 2013.

M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Compressive Sensing with Unknown Parameters," in Proc. Asilomar Conference on Signals, Systems and Computers, Monterey, CA, Nov. 4-7, 2012.

M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial Compressive Sensing in MIMO Radar with Random Arrays," in Proc. Conference on Information Sciences and Systems (CISS), Princeton, NJ, Mar. 21-23, 2012.

M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Global Methods for Compressive Sensing in MIMO Radar with Distributed Sensors," (Invited paper) in Proc. Asilomar Conference on Signals, Systems and Computers, Monterey, CA, Nov. 6-9, 2011.

M. Rossi, A. M. Tulino, O. Simeone, and A. M. Haimovich, "Nonconvex Utility Maximization in Gaussian MISO Broadcast and Interference Channels," in Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, May 22-27, 2011.

M. Rossi, S. Sorrentino, U. Spagnolini, and L. Moretti, "Convex Optimization Strategies for Precoding of Broadcast Channels," in Proc. International ITG Workshop on Smart Antennas (WSA 2008), Feb. 2008.

To my Family.

ACKNOWLEDGMENT

I would like to thank my advisor, Prof. Alexander Haimovich, for giving me this opportunity and for enabling the special relationship we built. He lead me through the Ph.D. studies, not only as an advisor, but also as a paternal figure. This had made these four years invaluable.

I will be always thankful to Prof. Osvaldo Simeone for his precious advice, his encouragement and his friendship. He is a very reliable "estimate" of what an ideal researcher looks like. Special thanks go to the committee members. Prof. Yeheskel Bar-Ness for sharing his energy and deep passion toward research and innovation. Prof. Ali Abdi for his thoughtful comments about my work, and Dr. Mark Govoni for the precious remarks and insights he provided during our discussions about my research.

I'm deeply grateful to Prof. Yonina Eldar for the outstanding guidance and support she provided throughout our collaboration. Very special thanks go to Prof. Antonia Tulino for her advice and encouragement during our work together.

I want to thank all the people in the CWCSPR lab, who shared with me the academic experience. Fabio, Nicola, Igor, Yu, Nil, Haley, Palin, Ali, Behazad, Ciprian, Vlad, Eyal, Tariq, and all the others that are not mentioned but walked with me during these years.

I want to thank Ms. Marlene Toeroek and Ms. Angela Retino, for their support and friendship. I also want to thank all the people at NJIT that assisted me during these years, especially Prof. Ehrlich, Dr. Grundy, Dr. Kline, Ms. Gonzales, Mr. McDermott, Prof. Ansari and Prof. Misra.

Finally, this accomplishment wouldn't have been possible without my family. Their never-ending love, help and encouragement are keeping me walking towards higher heights.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The dissertation discusses the theory of *compressive sensing* [1], and its applications to localization in MIMO radars. Detection, localization, and tracking of targets are basic radar functions. Limited data support and low signal-to-noise ratios (SNR) are among the many challenges frequently faced by localization systems. Another challenge is the presence of nearby targets, whether in terms of location or Doppler, since closely spaced targets are more difficult to discriminate. In multiple-input multiple-output (MIMO) radar, targets are probed with multiple, simultaneous waveforms. Relying on the orthogonality of the waveforms, returns from the targets are jointly processed by multiple receive antennas. MIMO radar is typically used in two antenna configurations, namely distributed [2] and colocated [3]. Depending on the mode of operation and system architecture, MIMO radars have been shown to boost target detection, enhance spatial resolution, and improve interference suppression. These advantages are achieved by providing and exploiting a larger number of degrees of freedom than "conventional" radar.

In general, target localization with radar consists of two stages: detection and estimation [4]. The detection process establishes the presence of a target in a prescribed resolution cell. This process is characterized by two parameters [5]: probability of false alarm (P_{FA}) and probability of detection (P_D). The goal is to maximize the probability of detection for a fixed level of false alarms. Classical detection is a process that inherently relies on a single target point of view. Detection performance is usually represented by receiver operating curves (ROC). Estimation builds on detection by seeking to improve the accuracy of localization for detected targets. In principle, estimation adopts a multi-target viewpoint. For example,

maximum likelihood estimation accounts for the interaction between closely spaced targets, hence localization performance is generally improved compared to detection.

1.1 Motivation of the Dissertation

The dissertation focuses on the detection and estimation of targets from direction-of-arrival (DOA) measurements using colocated MIMO radar. It is well known in array signal processing [4] that DOA resolution improves by increasing the array aperture. However, increasing the aperture without increasing the number of sensors may lead to ambiguities, i.e., measurements explained by erroneous sets of locations. A non-ambiguous uniform linear array (ULA) must have its elements spaced at intervals no larger than $\lambda/2$, where λ is the signal wavelength. For MIMO radar, unambiguous direction finding of targets is possible if N receive elements are spaced $\lambda/2$, and M transmit elements are spaced $N\lambda/2$, a configuration known as *virtual ULA* [3]. In sampling parlance, the $\lambda/2$-spaced array and the MIMO virtual ULA perform spatial sampling at the Nyquist rate. The main disadvantage of this Nyquist setup is that the product of the number of transmit and receive elements, MN, needs to scale linearly with the array aperture, and thus with resolution.

The dissertation advocates the use of a sparse, random array architecture in which a low number of transmit/receive elements are placed at random over a large aperture. Random array theory can be traced back to the 1960's. In [6], it is shown that as the number of sensors is increased, the random array pattern, a well known quantity to radar practitioners, converges to its average. This is because the array pattern's variance decreases linearly with the number of elements. This work was extended to MIMO radar in [7]. The main conclusion of the classical random array literature was that the random array pattern can be controlled by using a sufficient number of sensors. However, two fundamental questions were left pending: How

many sensors are needed for localization as a function of the number of targets, and which method should be used for localization?

It is suggested that the theory and algorithms of compressed sensing may be used to address these questions. Indeed, localizing targets from undersampled array data links random arrays to compressed sensing. In particular, it is an example of spatial compressive sensing since spatial sampling is applied at sub-Nyquist rates. The goal of spatial compressive sensing is to achieve similar resolution as a filled array, but with significantly fewer elements.

Undersampling, inherent to compressive sensing, causes ambiguities, i.e., interaction between targets may give rise to false peaks. Classical detection, in which resolution bins are tested one-by-one for the presence of a target, is then not suitable for compressive sensing scenarios. In contrast, classical estimation algorithms can handle compressive sensing scenarios, but are not equipped to handle unknown number of targets.

1.2 State of The Art

Compressive sensing considers linear inverse problems. These problems can be found throughout engineering and the mathematical sciences. Usually these problems are ill-conditioned or underdetermined, so that regularization must be introduced in order to obtain meaningful solutions. Sparsity constraints have emerged as a fundamental type of regularizer, and in the last decade, an enormous body of work has been generated around the theory of compressed sensing [1]. Radar has been among the many areas where compressive sensing has found application, and in particular, sparse recovery has been effectively applied to multiple input multiple output (MIMO) radar [8, 9, 10]. To fit sparse recovery in localization applications, one generates a grid of possible targets' locations and an associated unknown vector of responses, such that only locations associated with targets are non-zero. Therefore, the localization

problem aims to recover the support of such unknown vector (non-zero elements of the vector).

1.2.1 Compressive Sensing for Radar

Early works on compressive sensing radar emphasize that the sparse nature of many radar problems supports the reduction of temporal as well as spatial sampling (an overview is given in [11]). Recent work on compressive sensing for single-input single-output radar [12, 13, 14, 15] demonstrates either an increased resolution or a reduction in the temporal sampling rate. Compressive sensing for MIMO radar has been applied both on distributed [16] and colocated [17] setups. Much of the previous literature on compressive sensing for colocated arrays discusses the ULA setup, either within a passive system (with only receive elements) [18] or in a MIMO radar [10, 17] setup. In particular, [10] imposes a MIMO radar virtual ULA and derives bounds on the number of elements to perform range-angle and range-Doppler-angle recovery by using compressive sensing techniques. As discussed above, the (virtual) ULA setup performs Nyquist sampling in the spatial domain. In contrast, *spatial* compressive sensing (i.e., reducing the number of antenna elements while fixing the array aperture) relies on a random array geometry. Links between compressive sensing and random arrays have been explored in [19]. The author shows that spatial compressive sensing can be applied to the passive DOA problem, allowing for a reduction in the number of receiving elements. However, the MIMO radar framework poses a major challenge: contrary to the passive setup, where the rows of the sensing matrix $\mathbf{A}$ are independent, the MIMO radar MN measurements are dependent (they conform to the structure of the MIMO random array steering vector). This lack of independence prevents the application of the vast majority of results in the compressive sensing literature. A MIMO radar random array architecture is studied in [8], but no recovery guarantees are provided.

Low-rate spatial sampling translates into cost savings due to fewer antenna elements involved. It is of practical interest to determine the least amount of elements required to guarantee correct targets recovery. Finding conditions that guarantee recovery has been a main topic of research, and it is one of the underpinnings of compressive sensing theory. Recent work has shown that, for a sufficient number of independent and identically distributed (i.i.d.) compressive sensing measurements, non-uniform recovery can be guaranteed if a specific property of the random sensing matrix, called *isotropy*, holds [20]. While this property plays an important role, this result does not apply to the MIMO radar setup, since the MN rows are not independent. The dependent measurements problem was recently addressed in [21]. There, the authors derived conditions for non-uniform recovery using spatial compressive sensing in a MIMO radar system with N transceivers.

1.2.2 Compressive Sensing Algorithms

Compressive sensing algorithms seek to recover an $G \times P$ matrix $\mathbf{X}$ from a small number of linear observations $\mathbf{Y} = \mathbf{AX}$ (possibly corrupted by noise), where the $m \times G$ matrix $\mathbf{A}$, with $m \ll G$, is commonly referred to as *measurement matrix* or *dictionary*, and its columns are called *atoms*. While the linear system is highly underdetermined ($m \ll G$), the inverse problem has still a unique solution if $\mathbf{X}$ is *sparse*, i.e., it has only K non-zero norm rows out of G (with $K \leq m \ll G$). In this case, the problem of recovering the signal $\mathbf{X}$ from $\mathbf{Y}$ can be cast as a non-convex combinatorial ℓ_0-norm problem, i.e., $\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|_F$ s.t. $\|\mathbf{X}\|_0 \leq K$, where $\|\mathbf{X}\|_0$ counts the number of non-zero norm rows of $\mathbf{X}$. In the following, the rows of $\mathbf{Y}$ are addressed as *measurements*, and the columns of $\mathbf{Y}$ as *snapshots*. The ℓ_0-norm problem is known also under other names, such as sparse approximation or highly nonlinear approximation [22], and it can be related to the Deterministic Maximum Likelihood (DML) estimator [4, 23]. Both problems (ℓ_0-norm minimization and DML) require

a multi-dimensional search with exponential complexity [24], which is infeasible in practical scenarios. A core algorithmic question arises for a given class of dictionaries, how does one design a fast algorithm that provably recovers a K-sparse input signal?

Finding conditions that guarantee correct recovery with practical algorithms has been a main topic of research and one of the underpinnings of compressive sensing theory. Compressive sensing theory [1] shows that it is possible to recover any K-sparse signal $\mathbf{X}$ using practical algorithms, if the measurement matrix $\mathbf{A}$ satisfies specific properties. For instance a correct solution is guaranteed, if the matrix is sufficiently incoherent (as measured by the cumulative coherence [25]) or if it satisfies the restricted isometry property (RIP). Such properties are satisfied with high probability for a wide class of measurement matrices (e.g., Gaussian, Bernoulli, partial Fourier), as long as a sufficient number of measurements is available (e.g., $m > \beta K \log G$ for some constant β) [1].

While these guarantees are mainly consider the Single Measurement Vector (SMV) scenario (when $P = 1$), they can be extended to the general Multiple Measurement Vector (MMV) setup (when $P \geq 1$). Under mild conditions, usually satisfied by random matrices, a lower bound on the number of measurement m needed for identifiability of the sparse recovery problem is $m > 2K - \text{rank}\,(\mathbf{X})$ [26, 27]. Beside the (unpractical) ℓ_0-norm problem and DML estimator, it is well known in array processing that the so-called "super-resolution" techniques (e.g., subspace methods, such as MUSIC [28]) as well as "rank aware" methods (e.g., RA-ORMP [26]), are able to attain this bound with equality in a noiseless setting, whenever the signal has full-rank ($\text{rank}\,(\mathbf{X}) = K$). This behavior is traced to the ability to reliably estimate the signal subspace (i.e., a vector basis of $\mathbf{Y}$) which orthogonalizes the contribution of non-zero rows of $\mathbf{X}$ to the received signal $\mathbf{Y}$. If the signal subspace cannot be accurately estimated (due to noise or if $\text{rank}\,(\mathbf{X}) < K$), the contribution of non-zero

rows of $\mathbf{X}$ cannot be orthogonalized, and a trade-off emerges between an algorithm's complexity and performance.

1.2.3 Compressive Sensing Detection

The sparsity assumption, which permeates compressive sensing, blurs the distinctions between detection and estimation. In compressive sensing, the main unknown is the signal support, which must be recovered. The recovery of the support is essentially an estimation problem, but it requires decisions on zero and non-zero elements, which is a detection problem. In the MMV setting, the aim is to recover a sparse signal matrix $\mathbf{X}$ from noisy compressive measurements $\mathbf{Y} = \mathbf{AX} + \mathbf{E}$, where $\mathbf{A}$ is a measurement matrix and $\mathbf{E}$ represents the noise. It has been shown in [26] that, under certain conditions on the matrix $\mathbf{A}$ and the sparsity K, the optimal method to recover $\mathbf{X}$ is by solving the nonconvex noisy l_0-norm problem, i.e., $\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \nu \|\mathbf{X}\|_0$, where $\|\mathbf{X}\|_0$ counts the number of non-zero norm *rows* of $\mathbf{X}$. The *regularization parameter* ν governs the trade-off between fitting the data ($\|\mathbf{Y} - \mathbf{AX}\|_F^2$) and reducing the solution cardinality ($\|\mathbf{X}\|_0$), hence it can be set based on prior information, for example, the number of targets K or the noise level σ^2. In a radar detection problem, where there is no *a priori* information on the number of targets or on the noise level, setting the parameter ν is non-trivial. The focus is on compressive sensing methods that bridge between detection and estimation: they support the detection of multiple targets, while accounting for mutual effects between them. In other words, the aim is to recover information about the scene by compressive sensing methods without a priori information about the number of targets.

In [29], the authors addressed target detection in the so-called single measurement vector (SMV) setting (when a single snapshot is available, i.e., the matrix $\mathbf{Y}$ reduces to a column vector) using the Complex Approximate Message Passing (CAMP) algorithm, which aims to solve the l_1-norm convex approximation of the l_0-norm

problem. Building on an asymptotic analysis of CAMP, the authors proposed a detection framework, where K and σ^2 are unknown. In addition to [29], several other authors proposed compressed sensing algorithms to address this type of problem, but either assumed the noise level and/or the number of targets to be known, or required as input the regularization parameter ν.

1.3 Dissertation Outline and Contributions

The dissertation is outlined as follow: Chapter 2 introduces the sparse recovery problem and an outline of compressive sensing. Chapter 3 introduces the proposed spatial compressive sensing framework for MIMO radar DOA localization, analyzes the properties of the measurement matrix $\mathbf{A}$, and derives recovery guarantees. Chapter 4 discusses recovery algorithms to solve the sparse recovery problem in a general compressive sensing scenario. An algorithm, dubbed Multi-Branch Matching Pursuit (MBMP), is presented and its recovery guarantees are derived. Finally, a framework for target detection in compressive sensing using MBMP is proposed in Chapter 5. The dissertation expands the prior literature in several ways, as detailed below.

1.3.1 Spatial Compressive Sensing for MIMO Radar

A sparse localization framework for a MIMO random array is proposed assuming a general setup of M transmitters and N receivers. A bound on the coherence of the measurement matrix is derived, and the conditions under which the isotropy property holds are derived. This allows to develop both uniform and non-uniform recovery guarantees for target localization in MIMO radar systems. The proposed MIMO random array framework is of practical interest to airborne and other radar applications, where the spacing between antenna elements may vary as a function of aspect angle towards the target, or where exact surveying of element locations is

not practical due to natural flexing of the structures involved. The findings of the dissertation show that one can obtain the high-resolution provided by a virtual array aperture while using a reduced number of antenna elements.

1.3.2 Multi-Branch Matching Pursuit

A new approach for sparse recovery, called MBMP, is introduced and its recovery guarantees are formulated: (i) A sufficient condition under which MBMP recovers any sparse signal belonging to a given support; (ii) A sufficient condition under which MBMP can recover any K-sparse signal. Condition (i), named Multi-Branch Exact Recovery Condition (MB-ERC), generalizes the well-known Tropp's ERC [25] to a multi-branch algorithm. Condition (ii), named MB-coherence, is met when the dictionary is sufficiently incoherent. This condition improves the state-of-the-art in the sense that it enables to guarantee the success of the proposed MBMP for dictionaries that do not satisfy previously known conditions (e.g., cumulative coherence [25]). Moreover, MB-coherence provides a guideline to design the multi-branch structure of MBMP. The proposed MBMP method is suited both for MMV and for SMV problems, and in both cases performs better than alternatives. Due to its ability to adjust computational complexity, MBMP is particularly well suited to applications in which measurements are very expensive, such as in MIMO radar applications where the number of measurements is commensurate with the number of antenna elements.

1.3.3 Compressive Sensing Detection via MBMP

A framework for target detection using MBMP is presented. The proposed framework is fully adaptive (i.e., it does not require prior knowledge of the number of targets K or noise level σ^2), it addresses the general MMV setting (rather than the SMV setting addressed in [29]), and it provides an analysis of false alarm and detection

probabilities that holds for finite data records and any measurement matrix $\mathbf{A}$. The proposed algorithm is tested in a spatial compressive sensing-radar setting, where undersampling in space enables considerable savings in the number of array elements, while still enabling high resolution localization.

1.4 Notation

The following notation is used: boldface denotes matrices (uppercase) and vectors (lowercase); for a vector $\mathbf{a}$, the i-th index is $\mathbf{a}_i$, while for a matrix $\mathbf{A}$, $[\mathbf{X}]_{i,j}$ denotes the element at i-th row and j-th column, $\mathbf{A}(i,:)$ denotes the i-th row, and $\mathrm{vec}(\mathbf{X})$ produces a column vector by stacking the columns of $\mathbf{X}$. The complex conjugate operator is $(\cdot)^*$, the transpose operator is $(\cdot)^T$, the complex conjugate-transpose operator is $(\cdot)^H$, and the pseudo-inverse operator is $(\cdot)^\dagger$. For a full rank matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$ with $m > n$, $\mathbf{X}^\dagger = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H$. The Frobenius norm of $\mathbf{X}$ is $\|\mathbf{X}\|_F$, the ℓ_1-induced norm is $\|\mathbf{X}\|_1 \triangleq \max_j \sum_i |\mathbf{X}(i,j)|$ and the ℓ_∞-induced norm is $\|\mathbf{X}\|_\infty \triangleq \max_i \sum_j |\mathbf{X}(i,j)|$. Given a set S of indices, $|S|$ denotes its cardinality, $\mathbf{A}_S$ is the matrix obtained by considering only the columns with indices in S, $\Pi_{\mathbf{A}_S} \triangleq \mathbf{A}_S\mathbf{A}_S^\dagger$ is the orthogonal projection matrix onto the range of $\mathbf{A}_S^H$, and $\Pi^\perp_{\mathbf{A}_S} \triangleq \mathbf{I} - \Pi_{\mathbf{A}_S}$ is the orthogonal projection matrix onto the null space of $\mathbf{A}_S^H$. Given two set of indices, S and S', $S\setminus S'$ contains the indices of S which are not present in S'. A K-sparse matrix has only K non-zero norm rows, the number of non-zero norm rows of $\mathbf{X}$ is $\|\mathbf{X}\|_0$ and the support of $\mathbf{X}$ collects the indices of such non-zero norm rows. The operator $\mathbb{E}$ denotes expectation and $\psi_x(u) \triangleq \mathbb{E}[\exp(jxu)]$ is the characteristic function of the random variable x. The symbol "$\otimes$" denotes the Kronecker product. The notation $\mathbf{x} \sim \mathcal{CN}(\mu, \mathbf{C})$ means that the vector $\mathbf{x}$ has a circular symmetric complex normal distribution with mean μ and covariance matrix $\mathbf{C}$. $F_{a,b}$ denotes an F distribution with a numerator degrees of freedom and b denominator degrees of freedom, while $F'_{a,b}(\eta)$ denotes a non-central F distribution with a numerator degrees of freedom and

b denominator degrees of freedom, and non-centrality parameter η. For a probability density function X, the right-tail probability at γ is denoted by $P = Q_X(\gamma)$, while $\gamma = Q_X^{-1}(P)$ denotes its inverse function. Finally, $K_\alpha(\cdot)$ denotes the modified Bessel function of the second kind.

The work in the following chapters is mainly based on author's publications [9, 30, 31].

CHAPTER 2

SPARSE RECOVERY PROBLEM AND COMPRESSIVE SENSING

In a noiseless setting, sparse recovery seeks the solution to a linear system

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{2.1}$$

while requiring that the unknown is sparsest [22]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2.2}$$

This setup is known as Single Measurement Vector (SMV), highlighting the fact that a single vector of measurements $\mathbf{y}$ is available. More generally, when multiple measurement vectors have the same sparsity pattern, the setting is known as Multiple Measurement Vectors (MMV) or joint sparse. In this case, the model is

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \tag{2.3}$$

where $\mathbf{Y} \in \mathbb{C}^{m \times l}$ is the observed signal matrix, $\mathbf{A} \in \mathbb{C}^{m \times n}$ is the measurement matrix and the matrix $\mathbf{X} \in \mathbb{C}^{n \times l}$ is the unknown. The linear system in Eq. (2.3) is sparse in the sense that $\mathbf{X}$ has only $K \ll n$ non-zero norm rows. The MMV sparse recovery problem is to estimate the sparse matrix $\mathbf{X}$. It has been shown [1] that, under certain conditions on the matrix $\mathbf{A}$ and the sparsity K, the sparse matrix $\mathbf{X}$ can be recovered from linear measurements $\mathbf{Y}$ by solving the nonconvex l_0-norm problem:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_0 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{A}\mathbf{X}. \tag{2.4}$$

In this dissertation, it is assumed that $\text{spark}\,(\mathbf{A}) > 2K - \text{rank}\,(\mathbf{X}) + 1$, where $\text{spark}\,(\mathbf{A})$ is the smallest number of linearly dependent columns of the matrix $\mathbf{A}$. This is a necessary and sufficient condition for the measurements $\mathbf{Y} = \mathbf{A}\mathbf{X}$ to uniquely determine any K-sparse matrix $\mathbf{X}$ (see [1]).

In the presence of noise, the measurements comply with

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E} \tag{2.5}$$

where $\mathbf{E} \in \mathbb{C}^{m \times l}$ is the noise term. In this scenario, the sparse matrix $\mathbf{X}$ can be recovered by solving a relaxation of the problem in Eq. (2.4), $\min_{\mathbf{X}} \|\mathbf{X}\|_0$ s.t. $\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F \leq \epsilon$, where the Frobenius norm is used if the noise is supposed to be Gaussian distributed. Other formulations can also be used: a Lagrangian formulation, $\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F + \nu \|\mathbf{X}\|_0$, or a cardinality-constrained formulation

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F \quad \text{s.t.} \quad \|\mathbf{X}\|_0 \leq \tau. \tag{2.6}$$

The parameters τ, ν and ϵ depend on prior information, e.g., the signal sparsity K or the noise level $\|\mathbf{E}\|_F$. Unfortunately, the solution to any of these formulations requires an exhaustive search among all combinations of non-zero norm row indices of $\mathbf{X}$, necessitating exponential complexity [1].

The problem (2.6) can be also reformulated in terms of the support S of the solution $\mathbf{X}$. In particular, the problem in Eq. (2.6) is equivalent to

$$\min_{S} \left\| \Pi^{\perp}_{\mathbf{A}_S} \mathbf{Y} \right\|_F \quad \text{s.t.} \quad |S| \leq \tau. \tag{2.7}$$

The reformulation follows by noticing that the minimization with respect to $\mathbf{X}$ in the problem in Eq. (2.6) can be separated into the minimization with respect to the support S and the minimization with respect to the actual non-zero value of $\mathbf{X}$. In particular, assuming that the spark condition is satisfied (i.e., $\text{spark}(\mathbf{A}) > 2K - \text{rank}(\mathbf{X}) + 1$), the optimal non-zero value of $\mathbf{X}$ for a fixed support S is given by the least square solution: $\mathbf{X}^*_S = \mathbf{A}^{\dagger}_S \mathbf{Y}$. This reduces the problem in Eq. (2.6) to the problem in Eq. (2.7).

Finding conditions that guarantee correct recovery (i.e., find the solution to the problem in Eq. (2.6)) with a practical complexity algorithm has been a main topic

of research and one of the underpinnings of compressive sensing theory [1]. A variety of polynomial complexity algorithms have been proposed to obtain an approximate solution to the problem in Eq. (2.6) or one of the other formulations, and the area is still very active. These methods adopt sophisticated convex optimization theory concepts applied to the relaxation of the ℓ_0-norm in Eq. (2.4) with an ℓ_1-norm [32, 33, 34, 35]. This reformulation is known as Basis Pursuit (BP) or LASSO, and it is defined by

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma. \tag{2.8}$$

Unlike the problem in Eq. (2.4), this problem is convex, and thus a global solution can be found in polynomial time. However, since it is a relaxation, the solution obtained could be different from that of the problem in Eq. (2.4). Other algorithms use graphical models to take advantage of additional sparsity priors [36, 37], find local solutions of non-convex relaxations (such as the ℓ_p-norm (with $p < 1$) [38], the reweighting family [39, 40], or the M-FOCUSS algorithm [41]), or leverage simple, but effective, Matching Pursuit (MP) strategies (also known as greedy algorithms from combinatorial optimization) to address the reformulation in (2.7). In particular, the matching pursuit idea is to refine an empty provisional support by adding one element at each iteration. Among the matching pursuit algorithms, the most notable in the SMV setting are Orthogonal Matching Pursuit (OMP) [42], Orthogonal Least Squares (OLS) [43], and CoSaMP [44], where the latter allows to add more than one element at each iteration. For the general MMV setting, examples are the Rank Aware-Orthogonal Recursive Matching Pursuit (RA-ORMP) algorithm [26] and its generalization, Multi-Branch Matching Pursuit (MBMP) [45]. In some extensions of the matching pursuit, at each iteration, more than one index is added to the provisional support at each iteration. Notable examples are CoSaMP [44] and IHT [46].

It has been shown that (see [1] for a review), under suitable conditions on the measurement matrix $\mathbf{A}$, on the sparsity level K, and on the recovery algorithm, the approximate solution coincides with the solution of the problem in Eq. (2.4). The remaining of the dissertation studies these conditions.

CHAPTER 3

SPATIAL COMPRESSIVE SENSING FOR MIMO RADAR

3.1 System Model

3.1.1 MIMO Radar Model

A MIMO radar system (see Figure 3.1)is modeled in which N sensors collect a finite train of P pulses sent by M transmitters and returned from K stationary targets. Transmitters and receivers are assumed to each form a (possibly overlapping) linear array of total aperture Z_{TX} and Z_{RX}, respectively. The quantities Z_{TX} and Z_{RX} are normalized in wavelength units. Defining $Z \triangleq Z_{TX} + Z_{RX}$, the m-th transmitter is at position $Z\xi_m/2$ on the x-axis, while the n-th receiver is at position $Z\zeta_n/2$. Here ξ_m lies in the interval $\left[-\frac{Z_{TX}}{Z}, \frac{Z_{TX}}{Z}\right]$, and ζ_n is in $\left[-\frac{Z_{RX}}{Z}, \frac{Z_{RX}}{Z}\right]$. This definition ensures that when $Z_{TX} = Z_{RX}$, both ξ_m and ζ_n are confined to the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$, simplifying the notation in the sequel.

Let $s_m(t)$ denote the continuous-time baseband signal transmitted by the m-th transmit antenna and let θ denote the location parameter(s) of a generic target, for example, its azimuth angle. Assume that the propagation is nondispersive and that the transmitted probing signals are narrowband (in the sense that the envelope of the signal does not change appreciably across the antenna array). Then the baseband signal at the target location, considering the p-th transmitted pulse, can be described by (see, e.g., [2])

$$\sum_{m=1}^{M} \exp\left(j2\pi f_0 \tau_m(\theta)\right) s_m(t - pT) \triangleq \mathbf{c}^T(\theta)\, \mathbf{s}(t - pT). \tag{3.1}$$

Here f_0 is the carrier frequency of the radar, $\tau_m(\theta)$ is the time needed by the signal emitted by the m-th transmit antenna to arrive at the target, $\mathbf{s}(t) \triangleq [s_1(t), \ldots, s_M(t)]^T$, T denotes the pulse repetition interval, and

$$\mathbf{c}(\theta) = \left[\exp\left(j2\pi f_0 \tau_1(\theta)\right), \ldots, \exp\left(j2\pi f_0 \tau_M(\theta)\right)\right]^T \tag{3.2}$$

is the transmit steering vector. Assuming that the transmit array is calibrated, $\mathbf{c}(\theta)$ is a known function of θ.

To develop an expression for the received signal $r_n(t)$ at the n-th receive antenna, let

$$\mathbf{b}(\theta) = \left[\exp\left(j2\pi f_0 \tilde{\tau}_1(\theta)\right), \ldots, \exp\left(j2\pi f_0 \tilde{\tau}_N(\theta)\right)\right]^T \tag{3.3}$$

denote the receive steering vector. Here $\tilde{\tau}_n(\theta)$ is the time needed for the signal reflected by the target located at θ to arrive at the n-th receive antenna. Define the vector of received signals as $\mathbf{r}(t) \triangleq [r_1(t), \ldots, r_N(t)]^T$. Under the simplifying assumption of point targets, the received data vector is described by [2]

$$\mathbf{r}(t) = \sum_{k=1}^{K} \sum_{p=1}^{P-1} x_{k,p} \mathbf{b}(\theta_k) \mathbf{c}^T(\theta_k) \mathbf{s}(t - pT) + \mathbf{e}(t) \tag{3.4}$$

where K is the number of targets that reflect the signals back to the radar receiver, $x_{k,p}$ is the complex amplitude proportional to the radar cross sections of the k-th target relative to pulse p-th, θ_k are locations, and $\mathbf{e}(t)$ denotes the interference plus-noise term. The targets' positions are assumed constant over the observation interval of P pulses. The target gains $\{x_{k,p}\}$ are assumed to follow a Swerling Case II model, meaning that they are fixed during the pulse repetition interval T, and vary independently from pulse to pulse [47].

Analyzing how to estimate the number of targets K, or the noise level, without prior information is investigated in Chapter 5. In this chapter, the number of targets K and the noise level are assumed to be known.

3.1.2 Problem Formulation

The purpose of the system is to determine the DOA angles to targets of interest. Targets associated with a particular range and Doppler bin are considered, while targets in adjacent range-Doppler bins contribute as interferences to the bin of

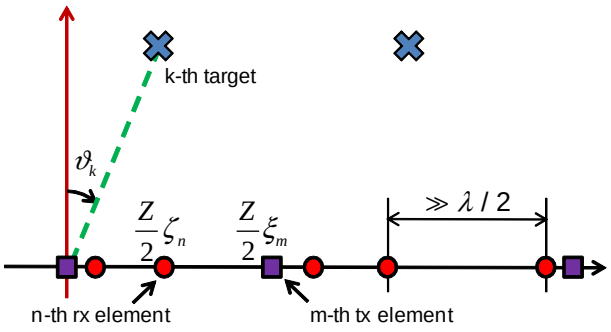

Figure 3.1 MIMO radar system model.

interest. The assumption of a common range bin implies that all waveforms are received with the same time delay after transmission. Since range and Doppler measurements are not of interest, the common time delay and Doppler shift are not explicitly shown in the following model. This approach is justified because angle resolution is essentially independent of range-Doppler resolution in antenna arrays [48]. Being capable to handle targets with non-zero Doppler, the proposed approach is applicable to airborne or ground targets. Targets are assumed in the far-field, meaning that a target's DOA parameter $\theta \triangleq \sin\vartheta$ (where ϑ is the DOA angle) is constant across the array. Under these assumptions, the receiver and transmitter steering vectors, $\mathbf{b}(\theta)$ and $\mathbf{c}(\theta)$, respectively, become

$$\mathbf{b}(\theta) = \left[\exp(j\pi Z\theta\zeta_1), \ldots, \exp(j\pi Z\theta\zeta_N)\right]^T \tag{3.5}$$

and

$$\mathbf{c}(\theta) = \left[\exp(j\pi Z\theta\xi_1), \ldots, \exp(j\pi Z\theta\xi_M)\right]^T. \tag{3.6}$$

By cross-correlating the received signal at each sensor with filters matched to each of the probing waveforms, one obtain

$$\begin{aligned}\mathbf{y}_p &= \mathrm{vec}\left[\int \mathbf{r}\left(t\right)\mathbf{s}^H\left(t-pT\right)dt\right] \\ &= \mathrm{vec}\left[\sum\nolimits_{k=1}^{K}\sum\nolimits_{p=0}^{P-1} x_{k,p}\mathbf{b}\left(\theta_k\right)\mathbf{c}^T\left(\theta_k\right)\mathbf{W} + \right. \\ &\qquad \left. + \int \mathbf{e}\left(t\right)\mathbf{s}^H\left(t-pT\right)dt\right]\end{aligned} \tag{3.7}$$

where the $M \times M$ matrix $\mathbf{W}$ has elements

$$\left[\mathbf{W}\right]_{m,j} = \int s_m\left(t\right)s_j^*\left(t\right)dt. \tag{3.8}$$

The M probing waveforms are assumed to be orthogonal (e.g., pulses modulated by an orthogonal code), therefore $\mathbf{W} = \mathbf{I}$. Defining the $MN \times P$ matrix $\mathbf{Y} \triangleq [\mathbf{y}_1, \ldots, \mathbf{y}_P]$, from Eq. (3.7) it follows that

$$\mathbf{Y} = \tilde{\mathbf{A}}\left(\theta\right)\tilde{\mathbf{X}} + \mathbf{E}. \tag{3.9}$$

Here $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_P]$ is a $K \times P$ matrix with $\tilde{\mathbf{x}}_p = [x_{1,p}, \ldots, x_{K,p}]^T$,

$$\tilde{\mathbf{A}}\left(\theta\right) = [\mathbf{a}\left(\theta_1\right), \ldots, \mathbf{a}\left(\theta_K\right)] \tag{3.10}$$

is a $MN \times K$ matrix with columns

$$\mathbf{a}\left(\theta\right) \triangleq \mathbf{c}\left(\theta\right) \otimes \mathbf{b}\left(\theta\right) \tag{3.11}$$

known as the "virtual array" steering vector, and $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_P]$ is $MN \times P$ with $\mathbf{e}_p = \mathrm{vec}\left[\int \mathbf{e}\left(t\right)\mathbf{s}^H\left(t-pT\right)dt\right]$. The term "virtual array" indicates that $\mathbf{a}\left(\theta\right)$ can be thought of as a steering vector with MN elements.

The aim is to recover θ and $\tilde{\mathbf{X}}$ from $\mathbf{Y}$ using a small number of antenna elements. To do this, a sparse localization framework is employed. Neglecting the discretization error, it is assumed that the target possible locations θ comply with a grid of G

points $\phi_{1:G}$ (with $G \gg K$). Since each element of θ parameterizes one column of $\tilde{\mathbf{A}}(\theta)$, it is possible to define an $MN \times G$ dictionary matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_G]$, where $\mathbf{a}_g = \mathbf{a}(\phi_g)$. From Eq. (3.11), the steering vector $\mathbf{a}_g$ is the Kronecker product of the receive steering vector $\mathbf{b}_g = \mathbf{b}(\phi_g)$ and the transmit steering vector $\mathbf{c}_g = \mathbf{c}(\phi_g)$:

$$\mathbf{a}_g = \mathbf{c}_g \otimes \mathbf{b}_g. \tag{3.12}$$

The received signal is then expressed as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E}, \tag{3.13}$$

where the unknown $G \times P$ matrix $\mathbf{X}$ contains the target locations and gains. Zero rows of $\mathbf{X}$ correspond to grid points without a target. The system model in Eq. (3.13) is sparse in the sense that $\mathbf{X}$ has only $K \ll G$ non-zero rows. Note that in the sparse localization framework, the matrix $\mathbf{A}$ is known, whereas in the array processing model in Eq. (3.9), the matrix $\tilde{\mathbf{A}}(\theta)$ is unknown.

Given the measurements $\mathbf{Y}$ and matrix $\mathbf{A}$, the goal translates into determining the non-zero norm rows' indices of $\mathbf{X}$, i.e., the support of $\mathbf{X}$. The matrix $\mathbf{A}$ is governed by the choice of grid points $\phi_{1:G}$, by the number M of transmitters and their positions, $\xi_{1:M}$, and by the number N of receivers and their positions, $\zeta_{1:N}$. In the following, the transmitter (receiver) elements' positions $\xi_{1:M}$ ($\zeta_{1:N}$) are assumed to be independent and identically distributed (i.i.d.) random variables governed by a probability density function (pdf) $p(\xi)$ ($p(\zeta)$).

3.2 Spatial Compressive Sensing Framework

The aim of spatial compressive sensing is to recover the unknown $\mathbf{X}$ from the measurements $\mathbf{Y}$ (see Eq. (3.13)) using a small number of antenna elements, MN, while fixing the array aperture Z. This section introduces the proposed spatial

compressive sensing framework and overview practical recovery algorithms (the well-known beamforming method as well as compressive sensing based algorithms).

3.2.1 Beamforming

Consider the scenario in which the transmitters and receivers locations support the Nyquist array (virtual ULA) geometry. In this setting, the matrix $\mathbf{A}$ in Eq. (3.13) has a Vandermonde structure, and the aperture scales linearly with the number of antenna elements, $Z = (MN - 1)/2$. If the (uniform) grid of possible target locations $\phi_{1:G}$ is chosen to match the array resolution, that is $G = 2Z + 1$, then the matrix $\mathbf{A}$ becomes a Fourier matrix. In this case, $\mathbf{Q} \triangleq \mathbf{A}^H\mathbf{A} = MN \cdot \mathbf{I}$. It follows that $\mathbf{X}$ can be estimated as $(1/MN) \cdot \mathbf{A}^H\mathbf{Y}$. In array processing, this method is called beamforming. The support of the unknown $\mathbf{X}$ is recovered by looking for peak values of $\left\|\mathbf{a}_g^H\mathbf{Y}\right\|_2$ over the grid points. Beamforming is also applied to estimate the locations of targets not limited to a grid. This is done by finding the peaks of $\left\|\mathbf{a}^H(\theta)\,\mathbf{Y}\right\|_2$, where $\mathbf{a}(\theta)$ is the steering vector in Eq. (3.11), as θ is swept over the angles of interest. The shortcoming of the Nyquist array setup is that the number of elements MN must scale linearly with the array aperture Z and consequently, with the resolution (i.e., such sampling mode requires $MN = G$).

Spatial compressive sensing implies that a sparse $\mathbf{X}$ can be recovered from a number of spatial measurements significantly lower than the Nyquist array, i.e., $MN \ll G$. The idea is to design the sensing procedure so that the matrix $\mathbf{Q}$ is a scalar multiple of the identity matrix *on average*[1], i.e., $\mathbb{E}[\mathbf{Q}] = MN \cdot \mathbf{I}$, and to control the variance of the non-diagonal elements by using a sufficient number of measurements. Intuitively, the more measurements MN are employed, the closer $\mathbf{Q}$ get to a diagonal matrix. Furthermore, because when $MN < G$, each realization of $\mathbf{Q}$ has non-zero off-diagonal terms, the beamforming metric $\left\|\mathbf{a}_g^H\mathbf{Y}\right\|_2$ is affected not only

[1]For instance, this is obtained when using a partial Fourier matrix.

by the g-th row of $\mathbf{X}$ and by the noise, but also by any row of $\mathbf{X}$ that has non-zero norm. This entails that, instead of beamforming, one should resort to compressive sensing recovery algorithms (introduced in Chapter 2), which take advantage of the signal's sparsity to mitigate the mutual interference among non-zero rows of $\mathbf{X}$, and therefore enhance the support recovery capabilities. In the following, an SMV scenario (i.e., $P = 1$, $\mathbf{Y} = \mathbf{y}$, $\mathbf{X} = \mathbf{x}$ and $\mathbf{E} = \mathbf{e}$ in Eq. (3.13)) is considered.

3.3 Recovery Guarantees

In this section, recovery guarantees for sparse localization with MIMO random arrays are derived. In detail, it is shown how to choose the grid-points $\phi_{1:G}$, the number of elements MN and the distributions governing the element positions $p(\xi)$ and $p(\zeta)$, in order to guarantee target localization by spatial compressive sensing using the solution of the LASSO problem in Eq. (2.8).

Two kinds of recovery guarantees are defined in compressive sensing: uniform and non-uniform. A uniform recovery guarantee (addressed below by Theorem 3.3.5) means that for a fixed instantiation of the random measurement matrix $\mathbf{A}$, all possible K-sparse signals are recovered with high probability. In contrast, a non-uniform recovery result (addressed by Theorem 3.3.7) captures the typical recovery behavior for a random measurement matrix $\mathbf{A}$. Specifically, suppose an arbitrary K-sparse vector $\mathbf{x}$ is given, and then $\mathbf{A}$ is drown at random (independent of $\mathbf{x}$). Non-uniform recovery details under what conditions an algorithm will recover $\mathbf{x}$ with high probability. Note that, for a non-uniform guarantee, $\mathbf{A}$ is being asked to recover only a specific $\mathbf{x}$, not *any* K-sparse vectors. Therefore, uniform recovery implies non-uniform recovery, but the converse is not true.

Loosely speaking, a uniform recovery guarantee can be obtained if, with high probability, the matrix $\mathbf{A}$ has small coherence [1]. The coherence is defined as the

maximum inner product between the normalized columns of $\mathbf{A}$,

$$\mu \triangleq \max_{i \neq l} \frac{\left|\mathbf{a}_i^H \mathbf{a}_l\right|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_l\|_2}. \tag{3.14}$$

Alternatively, uniform recovery is guaranteed if $\mathbf{A}$ satisfies the Restricted Isometry Property (RIP) [1] with high probability. Non-uniform recovery follows if a specific property of the random measurement matrix $\mathbf{A}$, called isotropy, holds [20]. The isotropy property states that the components of each row of $\mathbf{A}$ have unit variance and are uncorrelated, i.e.,

$$\mathbb{E}\left[\mathbf{A}^H(t,:)\,\mathbf{A}(t,:)\right] = \mathbf{I} \tag{3.15}$$

for every t.

Both Eq. (3.14) and Eq. (3.15) suggest that the matrix $\mathbf{Q} \triangleq \mathbf{A}^H\mathbf{A}$ plays a key role in establishing recovery guarantees. Indeed, because in the considered setting the rows of $\mathbf{A}$ are identically distributed, a simple calculation shows that $\mathbb{E}\left[\mathbf{A}^H(t,:)\,\mathbf{A}(t,:)\right] = \frac{1}{MN}\mathbb{E}[\mathbf{Q}]$, thus the isotropy property requires $\mathbb{E}[\mathbf{Q}] = MN \cdot \mathbf{I}$. Furthermore, as evident by the definition of coherence in Eq. (3.14), μ is the maximum absolute value among normalized off-diagonal elements of $\mathbf{Q}$.

Due to the role of the matrix $\mathbf{Q}$ to obtain (uniform and non-uniform) recovery guarantees, the statistics of the matrix $\mathbf{Q}$ are first studied.

3.3.1 Statistics of $\mathbf{Q} \triangleq \mathbf{A}^H\mathbf{A}$

To study the statistics of $\mathbf{Q}$, first it is analyzed its relationship to the *array pattern* [49], a quantity well known to radar practitioners. In array processing, the array pattern $\beta(u_{i,l})$ is the system response of an array beamformed in direction ϕ_l to a unit amplitude target located in direction ϕ_i. In other words, $\beta(u_{i,l})$ is the inner

product between two normalized columns of the measurement matrix:

$$\begin{aligned}\beta\left(u_{i,l}\right) &\triangleq \frac{\mathbf{a}_i^H \mathbf{a}_l}{\left\|\mathbf{a}_i\right\|_2 \left\|\mathbf{a}_l\right\|_2} \\ &= \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \exp\left[j u_{i,l} \left(\zeta_n + \xi_m\right)\right],\end{aligned} \tag{3.16}$$

where

$$u_{i,l} \triangleq \pi Z \left(\phi_l - \phi_i\right). \tag{3.17}$$

The peak of the absolute value of the array pattern for a target colinear with the beamforming direction, $|\beta(0)|$, is called the *mainlobe*. Peaks of $|\beta(u)|$ for $u \neq 0$, are known as *sidelobes,* and the highest among all the sidelobes is called the *peak sidelobe*. Thus the terms $\mathbf{a}_i^H \mathbf{a}_l$ in (3.16) play the role of sidelobes.

The relation between coherence, isotropy and array pattern is apparent. Indeed, from Eq. (3.14), Eq. (3.16), and the definition of sidelobes, the coherence, in array processing parlance, is the peak sidelobe associated with the matrix $\mathbf{A}$. Similarly, from Eq. (3.15) and Eq. (3.16), the isotropy can be related to the mean array pattern

$$\eta\left(u_{i,l}\right) \triangleq \mathbb{E}\left[\beta\left(u_{i,l}\right)\right], \tag{3.18}$$

where the expectation $\mathbb{E}\left[\beta\left(u_{i,l}\right)\right]$ is taken with respect to the ensemble of element locations. In particular, isotropy requires that $\eta\left(u_{i,l}\right) = 0$ for any $i \neq l$.

For a system with randomly placed sensors, the array pattern $\beta\left(u_{i,l}\right)$ is a stochastic process. Naturally, statistics of the array pattern of a random array depend on the pdf of the sensor locations. In [7], the authors derive the means and the variances of the real and imaginary parts of $\beta\left(u_{i,l}\right)$. The following proposition formalizes pertinent results from [7]. For the sake of brevity, the dependency on i and l is dropped, and denote the array pattern as $\beta(u)$. Define $z \triangleq \zeta + \xi$, and assume that the pdf of z, $p(z)$, is an even function (so that $\operatorname{Im} \eta(u) = 0$). Further, define

the variances of the array pattern $\sigma_1^2(u) \triangleq \mathrm{var}[\mathrm{Re}\,\beta(u)]$, $\sigma_2^2(u) \triangleq \mathrm{var}[\mathrm{Im}\,\beta(u)]$ and $\sigma_{12}(u) \triangleq \mathbb{E}[(\mathrm{Re}\,\beta(u) - \eta(u))\,\mathrm{Im}\,\beta(u)]$.

Proposition 3.3.1. *Let the locations ξ of the transmit elements be i.i.d., drawn from a distribution $p(\xi)$, and the locations ζ of the receive elements be i.i.d., drawn from a distribution $p(\zeta)$. Then, for a given u, the following holds:*

1. *The mean array pattern is the characteristic function of z, i.e.,*

$$\eta(u) = \psi_z(u). \tag{3.19}$$

2. *If ξ and ζ are identically distributed, then Eq. (3.20), Eq. (3.21) (at the bottom of this page), and $\sigma_{12}(u) = 0$ hold.*

Proof. See Appendix A.0.1. □

Proposition 3.3.1 links the probability distributions $p(\xi)$ and $p(\zeta)$ (via $\psi_z(u)$ and $\psi(u)$) to the mean and variances of each element of the matrix $\mathbf{Q}$, i.e., $\beta(u_{i,l}) = \frac{1}{MN}\mathbf{a}_i^H\mathbf{a}_l$. As shown below, this result is used to obtain non-uniform recovery guarantees.

To characterize the statistics of the coherence μ (defined in Eq. (3.14)), the distribution of the maximum absolute value among normalized off-diagonal elements of $\mathbf{Q}$ is needed. It is now shown that, by imposing specific constraints on the grid points $\phi_{1:G}$ and on the probability distributions $p(\xi)$ and $p(\zeta)$, the distributions of

---

$$\sigma_1^2(u) = \frac{1}{2MN}(1 + \psi_z(2u)) + \psi_z(u)\left[\frac{N+M-2}{2MN}(1 + \psi_\xi(2u)) - \psi_z(u)\frac{N+M-1}{MN}\right] \tag{3.20}$$

$$\sigma_2^2(u) = \frac{1}{2MN}(1 - \psi_z(2u)) + \psi_z(u)\frac{N+M-2}{2MN}(1 + \psi_\xi(2u)) \tag{3.21}$$

the elements of $\mathbf{Q}$ can be characterized. To do this, it is required an intermediate result about the structure of the matrix $\mathbf{Q}$ when $\phi_{1:G}$ is a uniform grid:

Lemma 3.3.2. *If $\phi_{1:G}$ is a uniform grid, $\mathbf{Q}$ is a Toeplitz matrix.*

Proof. See Appendix A.0.2. □

Thanks to Lemma 3.3.2, whenever $\phi_{1:G}$ is a uniform grid, $\mathbf{Q}$ is described completely by the elements of the first row of $\mathbf{A}$, $\mathbf{a}_1^H\mathbf{a}_i$ for $i = 1, \ldots, G$. From the definition of $\mathbf{A}$, its columns all have squared-norm equal to MN. Therefore the elements on the main diagonal of $\mathbf{Q}$ are equal to MN. Thus, it is needed to investigate the remaining random elements, $\mathbf{a}_1^H\mathbf{a}_i$ for $i = 2, \ldots, G$. By exploiting the Kronecker structure of the columns of $\mathbf{A}$ in Eq. (3.12), the elements of $\mathbf{Q}$ can be expressed as:

$$\begin{aligned}\mathbf{a}_i^H\mathbf{a}_j &\triangleq \left(\mathbf{c}_i^H \otimes \mathbf{b}_i^H\right)\left(\mathbf{c}_j \otimes \mathbf{b}_j\right)\\ &= \mathbf{c}_i^H\mathbf{c}_j\mathbf{b}_i^H\mathbf{b}_j,\end{aligned} \tag{3.22}$$

where $\mathbf{b}$ and $\mathbf{c}$ are the steering vectors of the receiver and transmitter arrays, respectively.

From Eq. (3.22), the random variable $\beta\left(u_{1,i}\right) \triangleq \frac{1}{MN}\mathbf{a}_1^H\mathbf{a}_i$ is the product between the random variables $\beta_\zeta\left(u_{1,i}\right) \triangleq \frac{1}{N}\mathbf{b}_1^H\mathbf{b}_i$ and $\beta_\xi\left(u_{1,i}\right) \triangleq \frac{1}{M}\mathbf{c}_1^H\mathbf{c}_i$. As such, the distribution of $\beta\left(u_{1,i}\right)$ (or equivalently, $\mathbf{a}_1^H\mathbf{a}_i$) can be characterized from the distributions of $\beta_\zeta\left(u_{1,i}\right)$ and $\beta_\xi\left(u_{1,i}\right)$. Following the approach in [6], in Appendix A.0.3 is detailed that the real and imaginary parts of $\beta_\zeta\left(u_{1,i}\right)$ (or $\beta_\xi\left(u_{1,i}\right)$) have an asymptotic joint Gaussian distribution, but, in general, the variances of real and imaginary parts of such variables are not equal. Interestingly, a closed form expression for the cumulative density function (cdf) of the product of $\beta_\zeta\left(u_{1,i}\right)$ and $\beta_\xi\left(u_{1,i}\right)$ (i.e., the cdf of $\frac{1}{MN}\mathbf{a}_1^H\mathbf{a}_i$) exists in the special case when $\operatorname{var}\left[\operatorname{Re}\beta_\zeta\left(u_{1,i}\right)\right] = \operatorname{var}\left[\operatorname{Im}\beta_\zeta\left(u_{1,i}\right)\right]$ and $\operatorname{var}\left[\operatorname{Re}\beta_\xi\left(u_{1,i}\right)\right] = \operatorname{var}\left[\operatorname{Im}\beta_\xi\left(u_{1,i}\right)\right]$. By meeting these conditions, the following

theorem derives an upper bound on the sidelobes' complementary cdf (ccdf), i.e., $\Pr\left(\frac{1}{MN}\left|\mathbf{a}_1^H\mathbf{a}_i\right| > q\right)$, and show that sidelobes have uniformly distributed phases.

Two MIMO radar setups are addressed: (1) M transmitters and N receivers, where ξ and ζ are independent, and (2) N transceivers, where $\xi_n = \zeta_n$, for all n and $M = N$.

Theorem 3.3.3. *Let the locations ξ of the transmit elements be drawn i.i.d. from a distribution $p(\xi)$, and the locations ζ of the receive elements be drawn i.i.d. from a distribution $p(\zeta)$. Assume that $p(\xi)$, $p(\zeta)$ and the uniform grid $\phi_{1:G}$ are such that the transmitter and receiver characteristic functions satisfy*

$$\psi_\xi(u_{1,i}) = \psi_\xi(2u_{1,i}) = \psi_\zeta(u_{1,i}) = \psi_\zeta(2u_{1,i}) = 0 \tag{3.23}$$

for $i = 2, \ldots, G$, where $u_{1,i} = \pi Z(\phi_i - \phi_1)$. Then for $i = 2, \ldots, G$:
1) If ξ and ζ are independent:

$$\Pr\left(\frac{1}{MN}\left|\mathbf{a}_1^H\mathbf{a}_i\right| > q\right) < x \cdot K_1(x), \tag{3.24}$$

where $x \triangleq 2\sqrt{MN}q$.
2) If $\xi_n = \zeta_n$ for all n:

$$\Pr\left(\frac{1}{N^2}\left|\mathbf{a}_1^H\mathbf{a}_i\right| > q\right) < \exp(-Nq). \tag{3.25}$$

3) In both scenarios, the phase of $\mathbf{a}_1^H\mathbf{a}_i$ is uniformly distributed on the unit circle, i.e.,

$$\measuredangle\mathbf{a}_1^H\mathbf{a}_i \sim \mathcal{U}[0, 2\pi]. \tag{3.26}$$

Proof. See Appendix A.0.3. □

This theorem characterizes the distribution of $\frac{1}{MN}\mathbf{a}_1^H\mathbf{a}_i$ for the M transmitters N receivers setup, and for the N transceivers setup. Subsection 3.3.4 provides a

practical setup that satisfies Eq. (3.23). As shown below, this allows to obtain a uniform recovery guarantee for spatial compressive sensing.

3.3.2 Uniform Recovery

The following corollary of Theorem 3.3.3 bounds the probability that the matrix $\mathbf{A}$ has high coherence, or equivalently, the probability of a peak sidelobe:

Corollary 3.3.4. *Let the locations of the transmit elements ξ be drawn i.i.d. from a distribution $p(\xi)$, and the locations of the receivers ζ be drawn i.i.d. from a distribution $p(\zeta)$. Assume that the distributions $p(\xi)$ and $p(\zeta)$ and the uniformly spaced grid-points $\phi_{1:G}$ are such that Eq. (3.23) holds for $i = 2, \ldots, G$. Then:*
1) If ξ and ζ are independent:

$$\Pr(\mu > q) < 1 - [1 - x \cdot K_1(x)]^{G-1}, \tag{3.27}$$

where $x \triangleq 2\sqrt{MN}q$.
2) If $\xi_n = \zeta_n$ for all n:

$$\Pr(\mu > q) < 1 - [1 - \exp(-Nq)]^{G-1}. \tag{3.28}$$

Proof. See Appendix A.0.4. □

Since μ can be interpreted as the peak sidelobe of the array pattern, Eq. (3.27) (Eq. (3.28)) characterizes the probability of having a peak sidelobe higher than q in a system with M transmitters and N receivers (N transceivers). These results are not asymptotic (i.e., they do not need the number of measurements M and N to tend to infinity). To further explore this point, in numerical results these bounds are compared against empirical simulations.

The coherence μ plays a major role in obtaining uniform recovery guarantees for compressive sensing algorithms, as well as guaranteeing the uniqueness of the

sparsest solution to Eq. (2.4). For instance, using the coherence μ, it is possible to obtain a bound on the RIP constant, $\delta_K \leq (K-1)\,\mu$ [50]. This ensures stable and robust recovery by l_1-minimization (i.e., using LASSO in Eq. (2.8)) from noisy measurements. By building on Corollary 3.3.4, the following theorem establishes the number of elements MN needed to obtain uniform recovery with high probability using (2.8):

Theorem 3.3.5 (Uniform recovery guarantee)**.** *Let the locations ξ of the transmit elements be drawn i.i.d. from a distribution $p(\xi)$, and the locations ζ of the receivers be drawn i.i.d. from a distribution $p(\zeta)$. Let the distributions $p(\xi)$ and $p(\zeta)$, and the uniform grid $\phi_{1:G}$ be such that relations (3.23) hold for $i = 2, \ldots, G$. Further, let*

$$MN \geq C\left(K - \frac{1}{2}\right)^2 \left[\log\gamma + \frac{1}{2}\log\left(2\log\gamma\right)\right]^2 \tag{3.29}$$

where $\gamma \triangleq \sqrt{\pi}G/(2\epsilon)$, and the constant $C = \left(43 + 12\sqrt{7}\right)/16 \approx 4.6718$. Then, with probability at least $1-\epsilon$, for any K-sparse signal $\mathbf{x} \in \mathbb{C}^G$ measured from MN MIMO radar measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, with $\|\mathbf{e}\|_2 \leq \sigma$, the solution $\hat{\mathbf{x}}$ of (2.8) satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq c\sigma, \tag{3.30}$$

where c is a constant that depends only on ϵ.

Proof. See Appendix A.0.5. □

The significance of (3.29) is to indicate the number of elements necessary to control the peak sidelobe. This is used to obtain a uniform recovery guarantee for spatial compressive sensing. In addition, the previous theorem ensures exact recovery of any K-sparse signal using (2.8) in the noise-free case $\sigma = 0$.

It is important to point out that the number of grid points G is not a free variable since $\phi_{1:G}$ must satisfy (3.23). This point will be explored in subsection

3.3.4, where it is shown that the resolution G must be linearly proportional to the "virtual" array aperture Z.

Uniform recovery guarantees capture a worst case recovery scenario. Indeed, the average performance is usually much better than that predicted by uniform recovery guarantees. In the following section, it is shown that if a non-uniform recovery guarantee is considered, then the zero mean conditions (3.23) can be relaxed, and recovery guarantees that scale linearly with K are obtained.

3.3.3 Non-uniform Recovery

In this subsection non-uniform recovery guarantees are investigated. In recent work [20], it has been shown that for a sufficient number of i.i.d. compressive sensing measurements, non-uniform recovery is guaranteed if isotropy holds. However, the result in [20] cannot be directly used in the proposed framework since the MN rows of the matrix $\mathbf{A}$, following (3.11), are not i.i.d. This scenario is addressed in [21] in which non-uniform recovery is guaranteed for a MIMO radar system with N transceivers if the isotropy property (under the name *aperture condition*) holds. The following theorem derives conditions on grid points $\phi_{1:G}$ and probability distributions $p(\xi)$ and $p(\zeta)$, in order for the random matrix $\mathbf{A}$ to satisfy the isotropy property:

Theorem 3.3.6. *Let the locations ξ of the transmit elements be drawn i.i.d. from a distribution $p(\xi)$, and the locations ζ of the receivers be drawn i.i.d. from a distribution $p(\zeta)$. For every t, the t-th row of $\mathbf{A}$ in (3.13) satisfies the isotropy property [20], i.e.,*

$$\mathbb{E}\left[\mathbf{A}^H(t,:)\,\mathbf{A}(t,:)\right] = \mathbf{I}, \tag{3.31}$$

iff $p(\xi)$, $p(\zeta)$ and $\phi_{1:G}$ are chosen such that, for $i = 2, \ldots, G$,

$$\psi_z(u_{1,i}) = 0, \tag{3.32}$$

where $z \triangleq \zeta + \xi$ *and* $u_{1,i} \triangleq \pi Z\left(\phi_i - \phi_1\right)$.

Proof. See Appendix A.0.6. □

Theorem 3.3.6 links grid points $\phi_{1:G}$ and probability distributions $p\left(\xi\right)$ and $p\left(\zeta\right)$ (through the characteristic function of z) with the isotropy property of $\mathbf{A}$. When (3.32) holds, it can be shown that the aperture condition used in [21] holds too. Therefore, using the same approach as in [21], non-uniform recovery of K targets via (2.8) is guaranteed in the proposed spatial compressive sensing framework. The following Theorem customizes Theorem 2.1 in [21] to the proposed framework:

Theorem 3.3.7 (Non-uniform recovery guarantee)**.** *Consider a* K*-sparse* $\mathbf{x} \in \mathbb{C}^G$ *measured from* MN *MIMO radar measurements* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$**,** *where* $\left\|\mathbf{e}\right\|_2 \leq \sigma$*. Let* $\varepsilon > 0$ *be an arbitrary scalar, and suppose that the random matrix* $\mathbf{A}$ *satisfies the isotropy property,* $\mathbb{E}\left[\mathbf{A}^H\left(t,:\right)\mathbf{A}\left(t,:\right)\right] = \mathbf{I}\ \forall t$*. Then with probability at least* $1-\varepsilon$*, the solution* $\hat{\mathbf{x}}$ *to (2.8) obeys*

$$\left\|\hat{\mathbf{x}} - \mathbf{x}\right\|_2 \leq C_1 \sigma \sqrt{\frac{K}{MN}}, \tag{3.33}$$

provided that the number of rows of $\mathbf{A}$ *meets*

$$MN \geq CK \log^2\left(\frac{cG}{\varepsilon}\right), \tag{3.34}$$

where C_1*,* C *and* c *are constants.*

Proof. The theorem is obtained from Theorem 2.1 in [21] by performing the following substitutions: K for s (sparsity), MN for n^2 (number of rows of $\mathbf{A}$), and G for N (number of columns of $\mathbf{A}$). Since only K-sparse signals are considered, in (3.33) the term that accounts for nearly-sparse signals present in [21] is discarded. □

Theorem 3.3.7 shows that, when the isotropy property is satisfied, the proposed framework is capable to localize K targets using about $MN = K\left(\log G\right)^2$ MIMO radar measurements.

Some comments are in order. First, it is important to stress that in (3.34), the number of elements scales linearly with the sparsity K. This is in contrast with uniform recovery bounds based on coherence (e.g., (3.29)), which scale quadratically with K. Moreover, the significance of the logarithmic dependence on G is that the proposed framework enables high resolution with a small number of MIMO radar elements. This is in contrast with a filled virtual MIMO array where the product MN scales linearly with G. Again, it is crucial to point out that the number of grid points G is not a free variable, because the grid points $\phi_{1:G}$ must satisfy (3.32). Second, differently from (3.33), in (3.30) the error did not depend on K, M and N. Third, (3.33) shows that reconstruction is stable even when the measurements are noisy. Additionally, it can be seen from (3.33) that when $\sigma = 0$, Theorem 3.3.7 guarantees exact reconstruction with high probability, when (3.34) holds. Both results above can be extended to approximately sparse vectors, in which case an extra term appears in the right hand-side of (3.30) and (3.33). This situation may emerge when targets are not exactly on a grid, however, the analysis of such scenario is outside the scope of this paper. Finally, to suggest some intuition into the above conditions, notice that recovery can be guaranteed by requiring the matrix $\mathbf{A}$ to satisfy the isotropy property, $\mathbb{E}[\mathbf{Q}] = MN \cdot \mathbf{I}$, and by controlling the variances of the non-diagonal elements of $\mathbf{Q}$(which, according to (3.20) and (3.21), scale with $1/MN$) through the use of a sufficient number of measurements MN.

3.3.4 Element Locations and Grid-points

Here an example of $p(\xi)$, $p(\zeta)$ and $\phi_{1:G}$ that meet the requirements of Theorem 3.3.3 and Theorem 3.3.6 is provided.

The conditions needed by each theorem constraint the characteristic function of the random variables ξ, ζ. Let, $Z_{TX} = Z_{RX} = Z/2$, such that the random variables ξ and ζ are both confined to the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$. The characteristic function of a

uniform random variable $\zeta \sim \mathcal{U}\left[-\frac{1}{2}, \frac{1}{2}\right]$ is the sinc function, i.e.,

$$\psi_\zeta(u) = \frac{\sin(u/2)}{u/2}. \tag{3.35}$$

Therefore, when ζ is uniformly distributed, by choosing $\phi_{1:G}$ as a uniform grid of $2/Z$-spaced points in the range $[-1, 1]$, it results $\psi_\zeta(u_{i,l}) = \psi_\zeta(2u_{i,l}) = 0$ for any $i \neq l$ (since $u_{i,l} \triangleq \pi Z(\phi_l - \phi_i) = 2\pi|i-l|$). The number of grid points G is not a free variable, because the grid points $\phi_{1:G}$ must satisfy (3.23) or (3.32). For instance, in the example above, $\phi_{1:G}$ must be a uniform grid of $2/Z$-spaced points between $[-1, 1]$, and, assuming that Z is an integer, the number of grid points is $G = Z + 1$.

The dependence between the number of grid points G and the virtual array aperture Z can be understood by noticing that both (3.23) and (3.32) impose that grid points are placed at the zeros of the characteristic function of the relative random variable (i.e. the sinc function). The spacing of the zeros is dictated by the virtual array aperture Z. The bigger the aperture the more grid points fit in the range $[-1, 1]$.

Summarizing, choose $\phi_{1:G}$ as a uniform grid of $2/Z$-spaced points in the range $[-1, 1]$. Then:

1. If *both* ζ and ξ are uniformly distributed, relations (3.23) hold, and Theorem 3.3.3 (for uniform recovery) can be invoked;

2. If *either* ζ or ξ are uniformly distributed, relation (3.32) holds, and Theorem 3.3.6 (for non-uniform recovery) can be invoked.

Note that non-uniform recovery, i.e., (3.32), requires only one density function, say $p(\zeta)$, to be uniform, while the other distribution, $p(\xi)$, can be arbitrarily chosen, e.g., it can be even deterministically dependent on ζ. For instance, (3.32) is satisfied in a MIMO radar system with N transceivers, i.e., when $\zeta_{1:N}$ are i.i.d. uniform distributed and it is set deterministically $\xi_n = \zeta_n$.

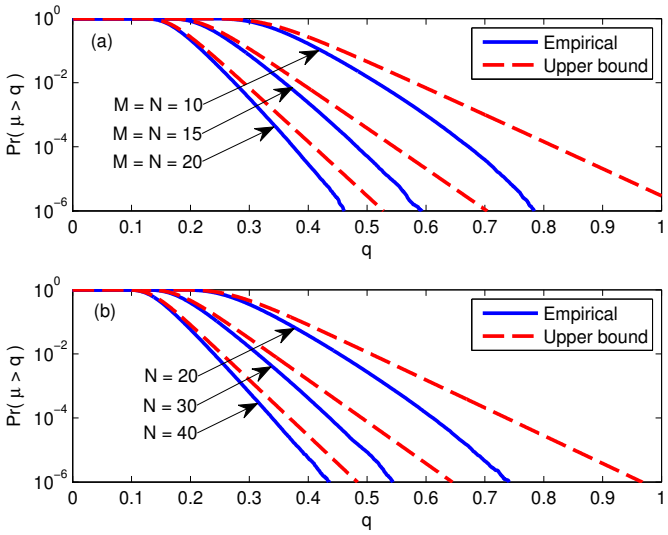

Figure 3.2 Empirical ccdf of the coherence of the measurement matrix **A** and its upper bound as a function of the number of elements. (a) considers the M transmitters and N receivers setup and the upper bound is given in (3.27); (b) considers the N transceivers setup and the upper bound is given in (3.28).

As a final remark, the analysis provided in this section regarding the statistics of the matrix **A** may be used with block-sparsity results in the compressive sensing literature [1] to obtain guarantees for the general MMV scenario.

3.4 Numerical Results

In this section, numerical results illustrating the proposed spatial compressive sensing framework are presented.

It is designed an example to follow Theorem 3.3.3, in which $p(\xi)$ and $p(\zeta)$ are both uniform distributions, and $\phi_{1:G}$ represents a uniform grid of $2/Z$-spaced points in the interval $[-1, 1]$, which implies that the number of grid points is $G = Z+1$. The system transmits a total of P pulses. When expressed in discrete form, each pulse consists of M orthogonal codes composed by M symbols. In particular, the codes are selected to be the rows of the $M \times M$ Fourier matrix. Equal length apertures were assumed for the transmit and receive arrays, i.e., $Z_{TX} = Z_{RX} = Z/2$. The target gains were given by $x_k = \exp(-j\varphi_k)$, with φ_k drawn i.i.d., uniform over $[0, 2\pi)$, for all

$k = 1, ..., K$ (where K is the number of targets). The noise (see (3.13)) was assumed to be distributed as $\text{vec}(\mathbf{E}) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ and the SNR is defined as $-10\log_{10}\sigma^2$. From the definition of the measurement matrix $\mathbf{A}$, its columns all have squared-norm equal to MN. Throughout the numerical results, the columns of $\mathbf{A}$ are normalized in order to have unit norm.

First, the statistics of the matrix $\mathbf{Q}$ discussed in Section 3.2 are investigated. In particular, the coherence μ of the measurement matrix $\mathbf{A}$ is compared to the result given in Theorem 3.3.3. The virtual aperture was $Z = 250$ (thus $G = 251$). In Figure 3.2, the ccdf of the coherence μ, i.e. $\Pr(\mu > q)$, is plotted as a function of the number of elements for (a) the M transmitters and N receivers setup and (b) the N transceivers setup. As a reference, it is also plotted the upper bound given in (3.27) and (3.28), respectively. It can be seen how the upper bound becomes tighter and tighter as the number of elements increases. In addition, it is interesting to notice that the coherence of the matrix $\mathbf{A}$ for the N' transceivers setup is very close to the coherence of the matrix $\mathbf{A}$ for the the M transmitters and N receivers setup when $M = N = N'/2$.

Next, localization performance using practical algorithms are presented. Target localization is implemented using LASSO (following the algorithm proposed in [21]) to solve problem (2.8). In addition, Beamforming, OLS, OMP, CoSaMP, FOCUSS [41] and MBMP are implemented. In the MMV setup, MBMP, RA-ORMP [26], M-FOCUSS [41], and MUSIC [28] are compared. Concerning MBMP, it requires as input a K length branch vector $\mathbf{d}$, which define the algorithm's complexity (see [45] for details on setting parameters for MBMP). The output of MBMP is the estimated support. Notice that, when $\mathbf{d} = [1, \ldots, 1]$, MBMP reduces to OLS in the SMV scenario, and to RA-ORMP in the MMV scenario. A support recovery error is defined as the event when the estimated support does not coincide with the true one. For algorithms that return an estimate $\hat{\mathbf{x}}$ of the sparse vector $\mathbf{x}$ (e.g., LASSO, FOCUSS

and MUSIC), the support was then identified as the K largest modulo entries of the signal $\hat{\mathbf{x}}$.

Analyzing how to set the noise parameter σ in (2.4), or the sparsity K, without prior information is the topic of current work [31], but outside the scope of this paper. Therefore, it is assumed that the noise level is available and that the number of targets K is known (notice that this information is needed by all the algorithms including MBMP). The virtual aperture was $Z = 250$ (thus $G = 251$), and tests were carried out for $K = 5$ targets. The SNR was 20 dB throughout.

The main focus of the paper is to reduce the number of antenna elements while avoiding sidelobes errors and while preserving the high-resolution provided by the virtual array aperture Z (i.e., recovering $2/Z$-spaced targets). Therefore, to account for errors due to sidelobes (an erroneous target is estimated at a sidelobe location) and unresolved targets (the responses of two targets in consecutive grid-points is merged in only one grid-point), it is considered as performance metric the support recovery error probability, defined as the error event when at least one target is estimated erroneously.

First, consider the non-uniform guarantee setting. Monte Carlo simulations were carried out using independent realizations of target gains, targets locations, noise and element positions. Figure 3.3 illustrates the probability of support recovery error as a function of the number of measurements MN. From the figure, it can be seen that compressive sensing algorithms enable better performance (smaller probability of sidelobe error and better resolution) than beamforming, which is not well-suited for the sparse recovery framework. Among compressive sensing algorithms, two main groups appear: on one side, OLS and OMP, which both have practically the same performance; on the other side, L1-norm opt., CoSaMP, FOCUSS and MBMP. Among the latter group, it is important to point out that, although the recovery guarantee

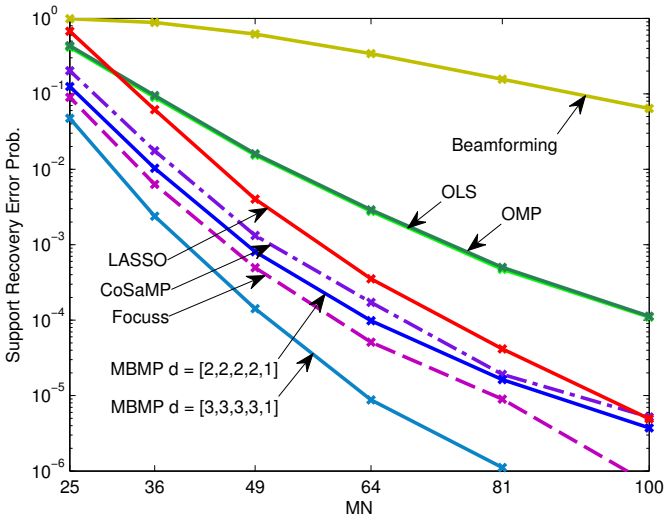

Figure 3.3 Probability of support recovery error as a function of the number of rows MN of $\mathbf{A}$. Non-uniform SMV setup. The system settings are $Z = 250$, $G = 251$, $P = 1$ and $K = 5$ targets with $|x_k| = 1$ for all k. The SNR is 20db.

established in Theorem 3.3.7 requires the solution of (2.8), and thus, using LASSO, MBMP provides a viable and competitive way to perform target localization.

Next, consider uniform guarantees. In this setup, first a realization of the matrix $\mathbf{A}$ is generated by drawing at random the element positions. Maintaining the matrix $\mathbf{A}$ fixed, 500 Monte Carlo simulations are performed using independent realizations of target gains, targets locations and noise. For each recovery method, a support recovery error is defined if an error occur in any of the 500 simulations. The results are then averaged throughout element positions realizations. Figure 3.4 illustrates the probability of support recovery error as a function of the number of measurements MN. The difference among OLS/OMP and the more sophisticated methods (i.e., LASSO, CoSaMP, FOCUSS and MBMP) is even more evident in this setup (e.g., at $MN = 81$, the probability of OLS/OMP error is greater than 0.1), confirming the theoretical finding [51] of OLS/OMP unfitness to deliver uniform recovery. On the other hand, MBMP, an extension of OLS, still provides competitive performance. In particular, MBMP with $\mathbf{d} = [3, 3, 3, 3, 1]$ outperform the other methods.

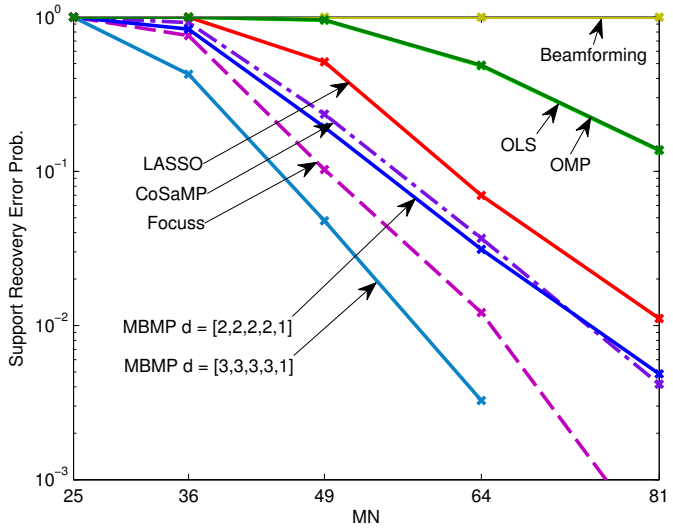

Figure 3.4 Probability of support recovery error as a function of the number of rows MN of $\mathbf{A}$. Uniform SMV setup. The system settings are $Z = 250$, $G = 251$, $P = 1$ and $K = 5$ targets with $|x_k| = 1$ for all k. The SNR is 20db.

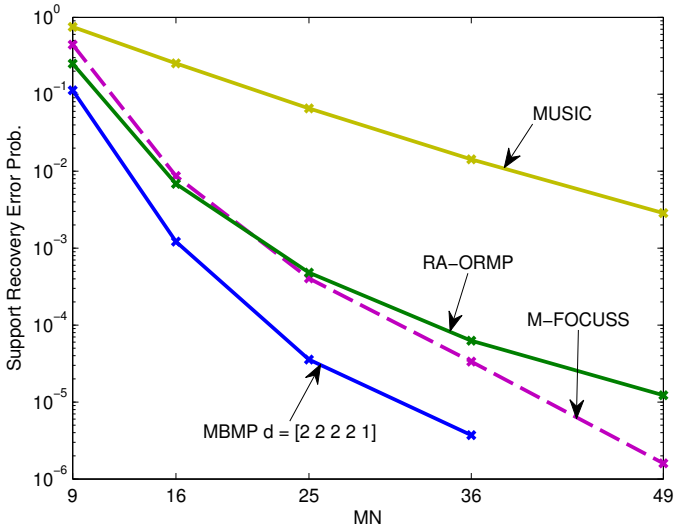

Figure 3.5 Probability of support recovery error as a function of the number of rows MN of $\mathbf{A}$. Non-uniform MMV setup. The system settings are $Z = 250$, $G = 251$, $P = 5$ and $K = 5$ targets with $|x_{k,p}| = 1$ for all k, p. The SNR is 20db.

The theoretical results presented in this work focus on the SMV setting. However, in practice several snapshots can be available. To explore the benefits of the proposed MIMO random array framework in such case, in Figure 3.5 an MMV setting ($P = 5$) is considered and sparse recovery methods are compared against the well-known MUSIC algorithm. Five different elements configurations are considered: $[M, N] = [3, 3]$, $[4, 4]$, $[5, 5]$, $[6, 6]$ and $[7, 7]$. The figure illustrates the probability of support recovery error as a function of the number of measurements MN (nonuniform setup). Sparse recovery algorithms have better performances than MUSIC, and the availability of multiple snapshots allows to considerably reduce the number of antenna elements. Moreover, in the MMV setting, algorithms which are able to exploit the signal subspace information (e.g., MBMP and RA-ORMP) posses a clear advantage over those algorithms that are unable (e.g., M-FOCUSS). For instance, this can be appreciated by the difference in performance of FOCUSS and MBMP with $\mathbf{d} = [2, 2, 2, 2, 1]$ when comparing the SMV (Figure 3.3) and MMV (Figure 3.5) settings. The numerical simulations presented in this paper considered a medium SNR level and show a superior performance of sparse recovery methods over classical methods (e.g., beamforming or MUSIC) in the proposed framework. Since the sparsity property, upon which sparse recovery methods rely, is independent from the SNR, a similar behavior is also expected at low SNR (e.g., SNR = 0 dB or lower).

3.5 Concluding Remarks

The proposed sparse framework for source localization using random array MIMO radar links system design quantities, i.e., the probability distributions $p(\xi)$ and $p(\zeta)$ of the tx/rx sensors location and the sparse localization grid points $\phi_{1:G}$, with the statistics of the Gram matrix $\mathbf{Q}$ and the related coherence of the matrix $\mathbf{A}$. Based on this result, uniform and non-uniform recovery guarantees for spatial compressive sensing are developed. Within the proposed framework, it is possible to localize K

targets using about $MN = K(\log G)^2$ MIMO radar noisy measurements, where G is proportional to the array aperture and determines the angle resolution. In other words, the proposed framework supports the high-resolution provided by the virtual array aperture while using a reduced number of MIMO radar elements. This is in contrast with a filled virtual MIMO array for which the product MN scales linearly with G. Moreover, since the results characterize the product of the number of transmit and receive elements, MIMO random array implementation further reduces the total number of antenna elements needed. From numerical simulations it emerges that, in the proposed framework, compressive sensing recovery algorithms (e.g., MBMP) are capable of better performance (i.e., smaller probability of sidelobe errors and better resolution) than classical methods, such as beamforming and MUSIC.

CHAPTER 4

MULTI-BRANCH MATCHING PURSUIT

This chapter introduces the Multi-Branch Matching Pursuit (MBMP) algorithm, which belongs to the matching pursuit family and aims to solve problem in Eq. (2.7), which is restated here:

$$\min_{C} \left\| \Pi^{\perp}_{\mathbf{A}_C} \mathbf{Y} \right\|_F \quad \text{s.t.} \quad |C| \leq \tau \tag{4.1}$$

where τ is a given parameter. In the following, state-of-the-art algorithms are overviewed and then MBMP is detailed.

4.0.1 Matching Pursuit

Matching pursuit [1] is a strategy that starts with an empty provisional support $C = \emptyset$, and then adds a new index to C at each iteration based on a selection strategy. For example, in OMP, the index g that maximizes $\left\| \mathbf{a}_g^H \Pi^{\perp}_{\mathbf{A}_C} \mathbf{Y} \right\|_2$ is selected. This selection strategy can be refined in two ways: a dictionary refinement and a subspace refinement.

The *dictionary refinement* applies when a non-empty provisional support C is already available. In this case, instead of using the original dictionary's atoms, the current dictionary is projected on the orthogonal subspace of $\mathbf{A}_C$, i.e., $\breve{\mathbf{a}}_g^C \triangleq \Pi^{\perp}_{\mathbf{A}_C} \mathbf{a}_g$, and each atom is renormalized according to,

$$\bar{\mathbf{a}}_g^C \triangleq \begin{cases} \breve{\mathbf{a}}_g^C / \left\| \breve{\mathbf{a}}_g^C \right\|_2 & \text{if } \left\| \breve{\mathbf{a}}_g^C \right\|_2 > 0 \\ \mathbf{0} & \text{otherwise} \end{cases} . \tag{4.2}$$

The dictionary refinement procedure (4.2) distinguishes OLS from OMP [52]: OLS evaluates the inner product between the residual and the modified atoms $\bar{\mathbf{a}}_g^C$, while OMP computes the inner product using $\breve{\mathbf{a}}_g^C$.

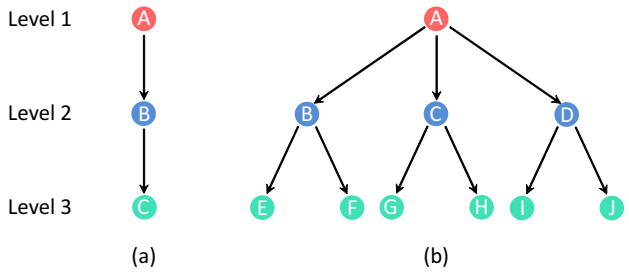

Figure 4.1 Graph of MBMP algorithm: (a) for a branch vector $\mathbf{d} = [1, 1]$ - MBMP reduces to RA-ORMP; (b) for a branch vector $\mathbf{d} = [3, 2]$.

The *subspace refinement* is possible in a MMV scenario (when $\text{rank}\,(\mathbf{X}) > 1$). In such case, rather than evaluating the norm of the inner product $\left\|\mathbf{a}_g^H \Pi_{\mathbf{A}_C}^{\perp} \mathbf{Y}\right\|_2$ using the residual $\Pi_{\mathbf{A}_C}^{\perp} \mathbf{Y}$, one can use an orthonormal basis $\mathbf{U}$ of $\Pi_{\mathbf{A}_C}^{\perp} \mathbf{Y}$, and compute $\left\|\mathbf{a}_g^H \mathbf{U}\right\|_2$. The matrix $\mathbf{U}$ is also known as the signal subspace. When the signal subspace can be accurately estimated, this procedure eliminates the mutual-interference between signals.

Depending on how refinement strategies are combined (dictionary refinement and/or subspace refinement), four different algorithms can be obtained. Three of them have been already introduced in the literature [26]: if dictionary and subspace refinements are not used, Simultaneous Orthogonal Matching Pursuit (SOMP) is obtained, which extends OMP to the general MMV scenario; if dictionary refinement is not used, but subspace refinement is used, the RA-OMP (which is not fully rank-aware); finally, if both dictionary and subspace refinements are used, the best algorithm is obtained, namely RA-ORMP, which is fully rank-aware. In particular, rank awareness means that, assuming $\text{spark}\,(\mathbf{A}) > 2K - \text{rank}\,(\mathbf{X}) + 1$ and considering a noiseless scenario, whenever the received signal $\mathbf{Y}$ is full rank the correct support is recovered by RA-ORMP with probability one. As a final remark, RA-ORMP can be contrasted with the MUSIC algorithm. The difference between them is the selection strategy: while RA-ORMP performs an iterative selection of nonzero elements, MUSIC identifies nonzero elements as the K largest values of $\left\|\mathbf{a}_g^H \mathbf{U}\right\|_2$.

4.0.2 Multi-Branch Matching Pursuit

The proposed MBMP algorithm, which generalizes RA-ORMP by including a multi-branch structure, is now detailed. It is possible to visualize RA-ORMP as a linear network of nodes, depicted in Figure 4.1-(a). Node A is tagged with an empty support. A new index is selected following a chosen selection strategy, and it becomes the provisional support of node B. To solve problem in Eq. (4.1), this procedure is repeated until level $\tau + 1$ is reached.

Instead of a linear network of nodes, the MBMP algorithm may be visualized as a tree of nodes as shown in Figure 4.1-(b), where each node is allowed to have multiple children (node A is the parent of nodes B, C and D; B is the parent of E and F). For instance, in Figure 4.1-(b), node A has three branches, resulting in three nodes at level 2. Node A is tagged with an empty support. Then, the index with the largest value according to the chosen selection strategy becomes the provisional support of node B. While RA-ORMP stops here, with MBMP, the index with the *second* largest value is assigned to the provisional support of node C. Similarly, the index with the *third* largest value is assigned to the provisional support of node D. One of these atom indices will necessarily be part of the solution reached by the algorithm. Then, MBMP removes from the dictionary the three atom indices selected at level 2, and continues to populate nodes at level 3. For example, consider node B. Since node B has two branches, it has *two* children. Following the selection strategy, *two* new indices are selected (the indices with the largest and the second largest values in the selection strategy metric). Each of these is added to the provisional support of node B and used to tag node E and F, respectively. This procedure is performed for all nodes at level 2 (i.e., nodes C and D), thus populating nodes G, H, I and J. The process stops when all nodes at level $\tau + 1$ have been populated. The provisional support S achieving the minimum data-fit $\left\|\Pi^{\perp}_{\mathbf{A}_S}\mathbf{Y}\right\|_F$ is elected as the solution to problem in Eq. (4.1).

The MBMP tree depends on the number of levels and on the number of branches at each level (assumed constant for nodes within the same level of the tree). The MBMP structure can be specified using a vector $\mathbf{d} \triangleq [d_1, \ldots, d_\tau]$ referred to as *branch vector*: d_i represents the number of branches of each node at level i. For instance, the tree in Figure 4.1-(a) has $\mathbf{d} = [1, 1]$ while the tree in Figure 4.1-(b) has $\mathbf{d} = [3, 2]$ (node A at level 1 has $d_1 = 3$ branches, and each node at level 2 (i.e., B, C, and D) possesses $d_2 = 2$ branches).

The pseudo-code of the MBMP algorithm is detailed in the following table. A *leaf* node is defined as a node without children, and $\text{orth}\,(\mathbf{Y})$ indicates an estimate of the signal subspace of the matrix $\mathbf{Y}$.

Algorithm 1 Multi-branch matching pursuit algorithm

Input: $\mathbf{Y} \in \mathbb{C}^{m \times P}$, $\mathbf{A} \in \mathbb{C}^{m \times G}$, and $\mathbf{d} \in \mathbb{N}^\tau$

Output: Support of approximate solution to problem in Eq. (4.1)

1: Set $\mathbf{U} = \mathbf{Y}$ and tag root node with $S = \emptyset$ and $\bar{S} = \emptyset$

2: **while** $\exists$ leaf node with $|S| < \tau$

3: Select a leaf node with $|S| < \tau$

4: Set $\Gamma = S$, $\bar{\Gamma} = \bar{S}$ and $i = |S| + 1$

5: if $P > 1$ set $\mathbf{U} = \text{orth}\left(\Pi^{\perp}_{\mathbf{A}_\Gamma} \mathbf{Y}\right)$

5: $\{\hat{g}_{1:d_i}\}$ is the set of d_i indices $g \in \bar{\Gamma}$ that maximize $\left\|\mathbf{U}^H \bar{\mathbf{a}}_g^\Gamma\right\|_2$

6: **for** $d = 1, \ldots, d_i$

8: Tag a new child node with:

$$S = \Gamma \cup \hat{g}_d,\ \bar{S} = \bar{\Gamma} \cup \{\hat{g}_{1:d}\} \text{ and } f = \left\|\Pi^{\perp}_{\mathbf{A}_S} \mathbf{Y}\right\|_F$$

9: **end**

10: **end**

11: Return a support S that minimizes f

Finally, it is noted that nodes at level τ need only $d_\tau = 1$ branch. This is because any additional branch would result in a provisional support S with higher cost function and thus would not be selected at step 11 of the MBMP algorithm.

4.0.3 Computational Complexity

Given an $m \times G$ matrix $\mathbf{A}$ and an $m \times P$ matrix[1] $\mathbf{Y}$, the computational requirements of MBMP depend on the specific implementation details, the structure of the measurement matrix $\mathbf{A}$ and the branch vector $\mathbf{d} = [d_1, \ldots, d_\tau]$ (MBMP has 1 node at level 1 and $\prod_{j<i} d_j$ nodes at level i). Due to the variability of the computation costs of applying the transform $\mathbf{A}^H$ (ranging from $\mathcal{O}(G \log(m) P)$ for an FFT-type operations to $\mathcal{O}(mGl)$ for unstructured matrices), F denotes the computational cost associated with performing $\mathbf{A}^H\mathbf{U}$ without specifying an associated number of flops. Furthermore, to perform subspace refinement, a practical implementation of MBMP would also need to incorporate an estimate of the signal subspace, and R denotes the relative cost. For a node at level i, other operations performed by MBMP are: selecting the d_i largest inner products, which is known as the "selection problem" [53] and can be solved using $\mathcal{O}(G)$ flops; the dictionary refinement, which costs $2m(G - i + 1)$ flops; the update of the projection matrix $\Pi^{\perp}_{\mathbf{A}_S}$, which requires $2m$ flops, and the computation of the residual, that needs mP flops. An efficient implementation of both the dictionary refinement and the projection matrix update is obtained by using a QR factorization [1].

Summarizing, the first node requires $F + R + \mathcal{O}(G)$ flops, since the dictionary refinement and the projection matrix update are not performed. Any node at level i (with $2 \leq i \leq \tau$) requires $F + R + \mathcal{O}(G) + 2m(G - i + 2) + mP$ flops. Finally, a node at level $\tau + 1$ requires $m(P + 2)$ flops to update the projection matrix and to compute the residual norm.

[1]It is assumed that $P \leq m$. When $P > m$, $\mathbf{Y}$ can be substituted with any square root of $\mathbf{Y}\mathbf{Y}^H$ (an $m \times m$ matrix) without affecting the solution to problem in Eq. (4.1).

This result can be contrasted with the complexity of the beamforming estimator $F + \mathcal{O}(G)$ (which selects the τ indices with the largest metric $\left\|\mathbf{a}_g^H \mathbf{Y}\right\|_2$ among $g = 1, \ldots, G$), or with the MUSIC estimator $F+R+\mathcal{O}(G)$ (which selects the τ indices with the largest metric $\left\|\mathbf{a}_g^H \mathbf{U}\right\|_2$ among $g = 1, \ldots, G$, where $\mathbf{U}$ is an estimate of the signal subspace of $\mathbf{Y}$). As discussed previously, when $d_i = 1$ for all i, the MBMP algorithm reduces to the RA-ORMP. It is worth mentioning that, due to the tree-structure, the proposed MBMP algorithm lends itself to a parallel implementation. Indeed, although the total number of MBMP operations remains the same, they can be performed in parallel, reducing the algorithm's execution time.

4.1 Recovery Guarantees for MBMP

This section provides recovery guarantees for MBMP in a noiseless scenario $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Throughout this section, the measurement matrix $\mathbf{A}$ is a given deterministic matrix, with no assumptions imposed on its structure. The sparsity level K is assumed to be known and, thus $\tau = K$ for the problem in Eq. (4.1), and MBMP is executed with a branch vector $\mathbf{d} \triangleq [d_1, \ldots, d_K]$ of length K. The information available to the recovery algorithm includes $\mathbf{Y}$, $\mathbf{A}$, and K. Moreover, as MBMP is executed, provisional supports, denoted C_i, are available at all nodes of level i. By convention, $C_1 = \emptyset$, since at level 1, no provisional support is available. MBMP succeeds in recovering a K-sparse $\mathbf{X}$ if one MBMP node at level $K+1$ is tagged with the correct support of $\mathbf{X}$, denoted S^*.

The road map of this section is as follows: First, Tropp's ERC [25] is overviewed. This condition considers signals with a *specific* support S^*. This restriction enables to obtain recovery guarantees for pursuit algorithms (e.g., OMP, OLS, BP, and RA-ORMP). By generalizing ERC to a multi-branch algorithm, in Definition 4.1.1 the MB-ERC is formulated. Theorem 4.1.2 relies on the MB-ERC to provide a sufficient condition to guarantee successful recovery with MBMP. Similar to ERC, MB-ERC

is non-constructive, since it focuses only on signals with a specific support S^*. To overcome this limitation, in Definition 4.1.3 the MB-coherence condition for multi-branch algorithms is introduced. With the help of the MB-coherence, Theorem 4.1.4 specifies a sufficient condition that guarantees the recovery of any K-sparse signal $\mathbf{X}$ using MBMP. Interestingly, the MB-coherence condition is the generalization to a multi-branch algorithm of the Neuman ERC (or weak ERC) [54] and it improves on the cumulative coherence condition proposed in [25].

4.1.1 MB-ERC

The ERC characterizes the ability of practical algorithms to recover sparse signals supported on a *specific* support S^*. For a given support S^* and for a matrix $\mathbf{A}$, the ERC is defined as [25]

$$\max_{g \notin S^*} \left\| \mathbf{A}_{S^*}^{\dagger} \mathbf{a}_g \right\|_1 < 1. \tag{4.3}$$

This condition addresses linear systems of equations of the form $\mathbf{A}_{S^*}\mathbf{x} = \mathbf{a}_g$, where $\mathbf{a}_g$ is a column from $\mathbf{A}$ that is outside the support S^*. The ERC states that the minimum (ℓ_2-)energy solution to all these systems should have an ℓ_1-length smaller than 1. The importance of the ERC comes from its strong connection to the success of pursuit techniques. In particular, ERC is a sufficient condition for successful recovery via RA-ORMP (as shown in [26]) and thus for MBMP as well as. ERC is also sufficient for correct recovery via OMP, OLS and BP in the SMV setup (see [25] and [55]).

Next MB-ERC is introduced, which generalizes ERC to a multi-branch algorithm and leads to a stronger sufficient condition to guarantee the success of MBMP. In contrast to RA-ORMP, in which each node has only one child, the number of children of each node of MBMP is specified by the branch vector $\mathbf{d} = [d_1, \ldots, d_K]$, where d_i is the *number of branches* of each node at level i. As a result, MB-ERC is a function of d_i. To proceed, it is convenient to define the d_i_max operator. Explicitly, given a

real vector $\mathbf{z}$ and a positive integer d_i, $d_i\_\max_{g\notin S^*}(\mathbf{z})$ selects the d_i-ranked maximum among the indices of $\mathbf{z}$ outside the support S^*. For instance, if $\mathbf{z} = [.7, 1.4, 1.1, .8, .9]^T$ and $S^* = \{2\}$, then index 2 should be excluded, and $1\_\max_{g\notin S^*}(\mathbf{z}) = 1.1$ (the largest entry outside S^*), while $2\_\max_{g\notin S^*}(\mathbf{z}) = .9$ (the second largest entry outside S^*), and so on.

Now MB-ERC can be defined. Consider level i of MBMP. Given a correct provisional support C_i (with $C_i \subset S^*$ and $|C_i| = i-1$), the dictionary refinement modification can be implemented, and a refined dictionary $\bar{\mathbf{A}} \triangleq \left\{\bar{\mathbf{a}}_g^{C_i}, g \notin C_i\right\}$ is obtained (see also Eq. (4.2)). Denoting $S \triangleq S^* \setminus C_i$ the support's indices yet to be identified, consider a sub-matrix of $\bar{\mathbf{A}}$ obtained by collecting only atoms $\bar{\mathbf{a}}_g^{C_i}$ such that $g \in S$, i.e., $\bar{\mathbf{A}}_S \triangleq \left\{\bar{\mathbf{a}}_g^{C_i}, g \in S\right\}$. The MB-ERC is defined as follows:

Definition 4.1.1. *[MB-ERC]Consider a support S^*, a matrix $\mathbf{A}$, a positive integer d_i, and a correct provisional support C_i. Let $S \triangleq S^* \setminus C_i$ be the set of indices yet to be identified. The MB-ERC is defined as*

$$d_i\_\max_{g\notin S^*}\left(\left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_g^{C_i}\right\|_1\right) < 1. \tag{4.4}$$

MB-ERC generalizes ERC to a multi-branch algorithm. By using MB-ERC, success of MBMP can guaranteed for any signal $\mathbf{X}$ supported on S^*:

Theorem 4.1.2 (Recovery of any signal supported on S^*)**.** *Given $\mathbf{Y} = \mathbf{A}\mathbf{X}$ where $\mathbf{X}$ is an unknown, K-sparse matrix of rank r and known to be supported on S^*. If the MB-ERC in Eq. (4.4) is met for all nodes at levels $i = 1, \ldots, K-r$, then MBMP with branch vector $\mathbf{d} = [d_1, \ldots, d_{K-r}, 1, \ldots, 1]$ is guaranteed to recover $\mathbf{X}$ successfully.*

Proof. See Appendix B.0.7. □

Theorem 4.1.2 formulates a sufficient condition for MBMP successful recovery of sparse signals supported on a *specific* support S^*. In the next subsection, another condition that guarantees MBMP successful recovery for *any* K-sparse signal is

obtained. Before proceeding, it is worth contrasting MB-ERC and ERC. Consider level 1 of MBMP. Since the provisional support $C_1 = \emptyset$, MB-ERC reduces to

$$d_1\_\max_{g \notin S^*} \left(\left\| \mathbf{A}_{S^*}^{\dagger} \mathbf{a}_g \right\|_1 \right) < 1, \tag{4.5}$$

where d_1 is the number of branches at level 1. The difference between Eq. (4.5) and Eq. (4.3) is the d_1_max operator in Eq. (4.5), which replaces the max operator in Eq. (4.3). This embodies the effect of seeking the d_1-largest (instead of the maximum) value of $\left\| \mathbf{A}_{S^*}^{\dagger} \mathbf{a}_g \right\|_1$ among indices g outside the support S^*. In other words, up to $d_1 - 1$ indices $\{g\}$ outside the support S^* with ℓ_1-length of the minimum (ℓ_2-)energy solution of $\mathbf{A}_{S^*}\mathbf{x} = \mathbf{a}_g$ *greater* than 1 can be tolerated. Furthermore, since ERC is obtained as a special case of Eq. (4.5) when $d_1 = 1$, for $d_1 > 1$ Eq. (4.5) is easier to meet than ERC (i.e., higher tolerance). As a concrete example, assume $S^* = \{2\}$ and $\left\| \mathbf{A}_{S^*}^{\dagger} \mathbf{a}_g \right\|_1 = z_g$ for $g = 1, \ldots, 5$, with $\mathbf{z} = [.7, 1.4, 1.1, .8, .9]^T$. The values of $\left\| \mathbf{A}_{S^*}^{\dagger} \mathbf{a}_g \right\|_1$ outside S^* are $\{.7, 1.1, .8, .9\}$. Therefore, while ERC is not satisfied (since $\max_{g \notin S^*}(\mathbf{z}) = 1.1 > 1$), MB-ERC is satisfied, if MBMP is designed with $d_1 \geq 2$ branches at level 1 (since $2\_\max_{g \notin S^*}(\mathbf{z}) = .9 < 1$).

4.1.2 MB-coherence Condition

A disadvantage of MB-ERC and ERC is that both require the knowledge of the true support S^*, hardly available in practice. This implies that to check if a measurement matrix $\mathbf{A}$ satisfies MB-ERC (or ERC), one has to compute the conditions for all $\binom{n}{K}$ possible supports S^* of cardinality K, which is usually prohibitive even for small values of K. To overcome this limitation, a practical condition that guarantees recovery via MBMP for *any* K-sparse signal $\mathbf{X}$ is developed. The main problem with MB-ERC and ERC is the presence of the pseudo-inverse. As shown in [25], by using standard norm inequalities to upper bound ERC, it is possible to obtain practical conditions that include only inner products rather than the pseudo-inverse

operator. These conditions rely on the notion of *coherence* of a measurement matrix $\mathbf{A}$, defined in [25] as

$$\mu(\mathbf{A}) \triangleq \max_{i \neq j} \left| \mathbf{a}_i^H \mathbf{a}_j \right|, \tag{4.6}$$

and on the notion of cumulative coherence (also known as *Babel's function* [56]), defined in [25] as

$$\bar{\mu}(k, \mathbf{A}) \triangleq \max_{S, |S|=k} \max_{g \notin S} \left\| \mathbf{A}_S^H \mathbf{a}_g \right\|_1, \tag{4.7}$$

for a positive integer k. Using these definitions, it was shown in [25] that the ERC holds for any K-sparse signal $\mathbf{X}$, if either the coherence satisfies

$$\mu(\mathbf{A}) < \frac{1}{2K-1} \tag{4.8}$$

or the cumulative coherence satisfies

$$\bar{\mu}(K-1, \mathbf{A}) + \bar{\mu}(K, \mathbf{A}) < 1. \tag{4.9}$$

Inspection of the definition of the coherence and cumulative coherence shows that $\bar{\mu}(k, \mathbf{A}) \leq k\mu(\mathbf{A})$ for any k. This is because the term $\left\| \mathbf{A}_{S^*}^H \mathbf{a}_g \right\|_1$ in the definition of $\bar{\mu}(k, \mathbf{A})$ is the sum of the absolute value of k elements from $\mathbf{A}^H\mathbf{A} - \mathbf{I}$, whereas $\mu(\mathbf{A})$ is the maximum among them. Since the cumulative coherence condition in Eq. (4.9) sums $2K-1$ such terms, it follows that the coherence condition in Eq. (4.8) implies the cumulative coherence condition in Eq. (4.9), i.e., Eq. (4.8) $\Rightarrow$ Eq. (4.9).

Condition in Eq. (4.9) (or Eq. (4.8)) implies ERC, and it is thus sufficient to guarantee recovery of any K-sparse $\mathbf{X}$ using MBMP or RA-ORMP. However, condition in Eq. (4.9) does not capture the multi-branch structure of MBMP. Thus a stronger condition for multi-branch algorithms, dubbed MB-coherence condition, is proposed. This condition is sufficient to guarantee recovery of any K-sparse $\mathbf{X}$

using MBMP. Considering a provisional support C_i, let $\bar{\mathbf{A}}^{C_i} \triangleq \left\{\bar{\mathbf{a}}_g^{C_i}, g \notin C_i\right\}$ be the associated refined measurement matrix.

Definition 4.1.3. *[MB-coherence]Consider a matrix* $\mathbf{A}$*, positive integers* K *and* d_i*, and a correct provisional support* C_i *(with* $|C_i| = i-1$*). The MB-coherence condition is defined as*

$$\max_{S,|S|=K-i+1}\left(\max_{g\in S}\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1+ d_i \max_{g\notin S\cup C_i}\left(\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1\right)\right) < 2. \tag{4.10}$$

A key aspect of the MB-coherence condition is that it includes only inner products among columns of the matrix $\bar{\mathbf{A}}^{C_i}$ (as opposed to MB-ERC in Eq. (4.4) which includes the pseudo-inverse operator). As discussed in Appendix B.0.9, this enables to practically compute the smallest integer d_i such that the MB-coherence condition Eq. (4.10) is met.

By using the MB-coherence condition, it is possible to obtain a sufficient condition to guarantee that MBMP recovers any K-sparse signal $\mathbf{X}$:

Theorem 4.1.4 (Recovery of *any* K-sparse signal)**.** *Given* $\mathbf{Y} = \mathbf{AX}$ *where* $\mathbf{X}$ *is an unknown,* K*-sparse matrix of rank* r*. If the MB-coherence condition in Eq. (4.10) is met for all nodes at levels* $i = 1, \ldots, K-r$*, then MBMP with branch vector* $\mathbf{d} = [d_1, \ldots, d_{K-r}, 1, \ldots, 1]$ *is guaranteed to recover* $\mathbf{X}$ *successfully.*

Proof. See Appendix B.0.8. □

Theorem 4.1.4 guarantees correct recovery of any K-sparse signal using MBMP (without requiring the knowledge of its support S^*). The special case of Theorem 4.1.4 for a single branch algorithm (when $\mathbf{d} = [1, \ldots, 1]$), i.e.,

$$\max_{S,|S|=K}\left(\max_{g\in S}\left\|\mathbf{A}_S^H\mathbf{a}_g\right\|_1 + \max_{g\notin S}\left\|\mathbf{A}_S^H\mathbf{a}_g\right\|_1\right) < 2 \tag{4.11}$$

is addressed as Neuman ERC or weak ERC and it was proposed in [54]. This condition guarantees correct recovery using RA-ORMP (the special case of MBMP when $\mathbf{d} = [1, \ldots, 1]$), BP, OMP or OLS.

Given a matrix $\mathbf{A}$, the number of branches of MBMP may be designed to guarantee recovery of any K-sparse signals for some targeted sparsity level K. An application of Theorem 4.1.4 is to provide an upper bound on the number of branches needed by each node of MBMP. Consider level 1 of MBMP. By choosing d_1 as the smallest integer such that Eq. (4.10) holds at level 1, it is guaranteed that at least one node at level 2 has a support C_2 such that $C_2 \subset S^*$. Therefore, there is no need to increase d_1 further. Similarly, for each node at level 2, the refined measurement matrix $\bar{\mathbf{A}}^{C_2} \triangleq \left\{\bar{\mathbf{a}}_g^{C_2}, g \notin C_2\right\}$ is computed, and the number of branches d_2 is chosen as the smallest integer that satisfies condition in Eq. (4.10) at level 2. By doing so, it is guaranteed that at least one node at level 3 has support C_3 such that $C_3 \subset S^*$. In general, for each node at level i, the refined measurement matrix $\bar{\mathbf{A}}^{C_i} \triangleq \left\{\bar{\mathbf{a}}_g^{C_i}, g \notin C_i\right\}$ is computed, and d_i is selected to satisfy Eq. (4.10) at level i. The process continues until d_{K-r} is set at level $K-r$, since, as shown in Lemma B.0.2 in Appendix B, nodes at level $i > K - r$ need only $d_i = 1$ branch. Furthermore, it can be shown that the number of branches d_i needed to satisfy Eq. (4.10) is non-increasing with respect to the level i. Therefore, if at some node, Eq. (4.10) holds with a given d_i, then, at any children of such node, the number of branches d_j needed to met Eq. (4.10) obeys $d_j \leq d_i$. This implies that if at some node, Eq. (4.10) holds with $d_i = 1$, then it can be set $d_j = 1$ branch for all children of such node.

4.2 Numerical Results

This section presents numerical results to illustrate the guarantees obtained in Section 4.1 and to investigate the performance of the proposed MBMP algorithm for MIMO radar direction-of-arrival (DOA) localization, utilizing the MIMO radar

spatial compressive sensing setup, introduced in Chapter 3. Below, the main settings of the Spatial Compressive Sensing setup are restated.

4.2.1 MIMO Radar Setup

A MIMO radar system where N sensors collect a finite train of P pulses is considered. Each pulse consists of M orthogonal spread spectrum waveforms of length M chips. Each one of the waveforms is sent by one of the M transmitters and returned from K stationary targets. Transmitters and receivers are assumed to form (possibly overlapping) linear arrays of equal aperture $Z/2$, respectively (Z is normalized in wavelength units): the i-th transmitter is at position ξ_i, where $\xi_i \in [0, Z/2]$ for $i = 1, \ldots, M$ on the x-axis; the i-th receiver is at position ζ_i, where $\zeta_i \in [0, Z/2]$ for $i = 1, \ldots, N$. The targets' positions are assumed constant over the observation interval of P pulses.

The purpose of the system is to determine the DOA angles to targets of interest. Targets associated with a particular range and Doppler bin are considered. Targets in adjacent range-Doppler bins contribute as interference to the bin of interest. The assumption of a common range bin implies that all waveforms are received with the same time delay after transmission. Targets are assumed in the far-field, meaning that a target's DOA parameter $\theta \triangleq \sin\vartheta$ (where ϑ is the DOA angle) is constant across the array. Following [9], the DOA estimation problem can be cast within a sparse localization framework. Neglecting the discretization error, it is assumed that the target possible locations comply with a grid of G points $\phi_{1:n}$ (with $G \gg K$). By defining the $MN \times G$ matrix

$$\mathbf{A} = [\mathbf{a}(\phi_1), \ldots, \mathbf{a}(\phi_n)] \tag{4.12}$$

where $\mathbf{a}(\theta) \triangleq \mathbf{c}(\theta) \otimes \mathbf{b}(\theta)$ with $\mathbf{b}(\theta) = [\exp(j2\pi Z\theta\zeta_1), \ldots, \exp(j2\pi Z\theta\zeta_N)]^T$ the receiver steering vector and $\mathbf{c}(\theta) = [\exp(j2\pi Z\theta\xi_1), \ldots, \exp(j2\pi Z\theta\xi_M)]^T$ the

transmitter steering vector (see [9] for further details), the signal model is expressed as Eq. (2.5). In particular, the unknown matrix $\mathbf{X} \in \mathbb{C}^{G\times P}$ contains the targets locations and gains. Non-zero rows of $\mathbf{X}$ correspond to grid points with a target.

Spatial compressive sensing assumes that the elements' positions are random variables (described by the probability density functions (pdf) $p(\xi)$ and $p(\zeta)$). Following the setup discussed in [9], $p(\xi)$ and $p(\zeta)$ are chosen as uniform distributions, and $\phi_{1:G}$ as a uniform grid of $2/Z$-spaced points in the range $[-1, 1]$. This implies that the number of grid points is $G = Z + 1$ (columns of the measurement matrix $\mathbf{A}$).

The target gains are given by $x_{k,p} = \exp(-j\varphi_{k,p})$, with $\varphi_{k,p}$ drawn i.i.d., uniform over $[0, 2\pi)$, for all $k = 1, ..., K$ (where K is the number of targets) and $p = 1, \ldots, P$ (where P is the number of snapshots). In other words, the target gains follow a Swerling Case II model, meaning that they are fixed during the pulse repetition interval, and vary independently from pulse to pulse [47]. The noise (see Eq. (2.5)) is assumed to be distributed as $\mathrm{vec}(\mathbf{E}) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ (where $\mathrm{vec}(\cdot)$ is the vectorization operator) and the SNR is defined $-10\log_{10}\sigma^2$. From the definition of the measurement matrix $\mathbf{A}$, its columns all have norms equal to $\sqrt{MN}$. Throughout the numerical results, the columns of $\mathbf{A}$ are normalized to unit norm.

4.2.2 Numerical Experiments

First, the guarantees obtained in Section 4.1 using the MB-coherence are explored. The trade-off between the number of measurements and number of branches d_1 at level 1 of MBMP (which relates to the algorithm's complexity) in order to meet the MB-coherence condition (4.10) at level 1 is investigated, i.e.,

$$\max_{S,|S|=K} \left(\max_{g\in S} \left\| \mathbf{A}_S^H \mathbf{a}_g \right\|_1 + d_1 \max_{g\notin S} \left(\left\| \mathbf{A}_S^H \mathbf{a}_g \right\|_1 \right) \right) < 2, \tag{4.13}$$

where $\bar{\mathbf{A}}^{C_1} = \mathbf{A}$ (since $C_1 = \emptyset$). The probabilities of meeting the coherence condition in Eq. (4.8) and the cumulative coherence condition in Eq. (4.9) are also plot as

references. Several realizations of the MIMO radar measurement matrix $\mathbf{A} \in \mathbb{C}^{MN \times n}$ (as defined in Eq. (4.12)) are generated, and for each realization it is tested whether Eq. (4.13) holds. Figure 4.2 plots the probability of meeting condition in Eq. (4.13) as a function of the number of measurements MN and parametrized by the number of branches d_1. Let $K = 3$, $Z = 500$, $p(\xi)$ and $p(\zeta)$ be uniform distributions, and $\phi_{1:n}$ be a uniform grid of $2/Z$-spaced points in the range $[-1, 1]$. This implies that the number of grid points is $n = Z + 1 = 501$. The main insight of the figure is the fewer measurements needed by the proposed MB-coherence condition with $d_1 = 1$ compared to the previous conditions (coherence and cumulative coherence). For instance, while the cumulative coherence condition needs almost $MN = n = 501$ to guarantee recovery with high probability, the proposed MB-coherence is met with high probability for $MN = 400$ (i.e., $M = N = 20$ elements). Furthermore, increasing the number of branches d_1 of MBMP enables to further reduce the number of measurements needed to guarantee recovery. For example, if $d_1 = 3$ branches are used, only $MN = 289$ measurements (i.e., $M = N = 17$ elements) are needed to obtain the guarantee with high probability, saving 6 antenna elements with respect to the $d_1 = 1$ setup.

In addition to the MIMO radar measurement matrix, a Gaussian measurement matrix, which has been widely studied in compressive sensing [1], is also investigated. The matrix $\mathbf{A} \in \mathbb{C}^{m \times G}$ is formed by generating mG i.i.d. random samples from the complex Gaussian distribution (arranged in matrix form), and subsequently normalizing each column of $\mathbf{A}$. In Figure 4.3, the probability of meeting condition in Eq. (4.13) is plotted as a function of the number of measurements m for different value of d_1. As before, $K = 3$ and $G = 501$. The advantage of the proposed MB-coherence condition, with $d_1 = 1$, over previous conditions is even more marked than in the MIMO radar setting. For instance, whereas the proposed MB-coherence is met with high probability for $m = 180$, the cumulative coherence condition is never

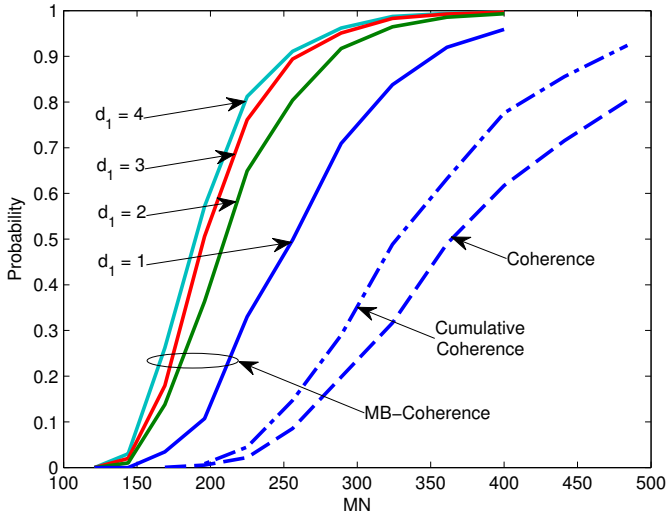

Figure 4.2 Probability of meeting condition in Eq. (4.13) as a function of the number of measurements MN for different value of d_1. The MIMO radar measurement matrix $\mathbf{A} \in \mathbb{C}^{MN \times G}$ defined in Eq. (4.12) is employed. Signal sparsity is $K = 3$ and $G = 501$.

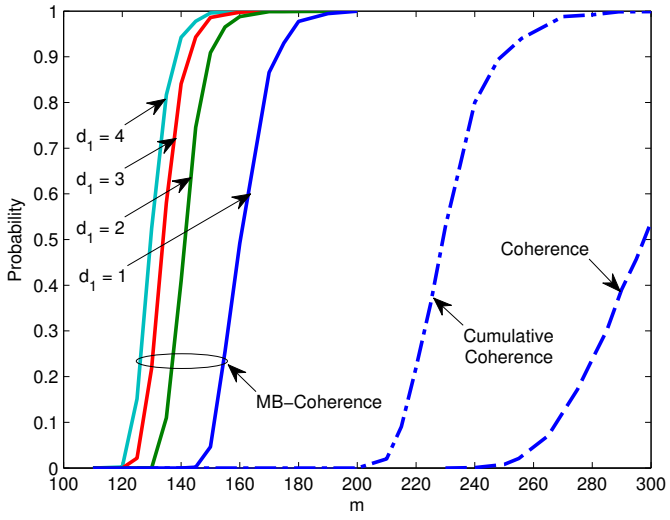

Figure 4.3 Probability of meeting condition in Eq. (4.13) as a function of the number of measurements m for different value of d_1. The complex Gaussian measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times G}$ is employed. Signal sparsity is $K = 3$ and $G = 501$.

met at $m = 180$ requiring around $m = 270$ measurements to guarantee recovery with high probability. The reduction in the number of measurements when the number of branches d_1 of MBMP is increased can be seen from the figure. For instance, while, using $d_1 = 1$ branch, about $m = 200$ measurements are needed to guarantee with high probability that MBMP correctly identify an atom in the support at level 1, by using $d_1 = 2$ branches the same guarantee is obtained with only $m = 160$ measurements, and the measurements can be further reduced to $m = 140$ if $d_1 = 4$ branches are employed.

The MB-coherence condition, investigated in Figures 4.2 and 4.3, relates to the correct recovery of any K-sparse signal using MBMP. A uniform recovery guarantee certifies that, given a fixed instantiation of the random measurement matrix $\mathbf{A}$, all possible K-sparse signals are recovered correctly [1]. Uniform recovery conditions capture the worst-case behavior of a measurement matrix $\mathbf{A}$. However, if one focuses on typical recovery, the conditions to obtain successful (non-uniform) recovery with high probability can be relaxed significantly, as shown in the numerical examples below.

To investigate the typical recovery behavior of MBMP, numerical results for non-uniform recovery (i.e., at each realization, the matrix $\mathbf{A}$ and the signal $\mathbf{X}$ are draw independently at random) are presented and the localization performance in the presence of noise is explored by comparing MBMP with other SMV and MMV algorithms. For the SMV setting, target localization is implemented using LASSO (applying the algorithm proposed in [21]). In addition, beamforming, OLS and FOCUSS [41] are implemented. For the MMV scenario, MBMP is compared against RA-ORMP, M-FOCUSS, and MUSIC. Notice that when $\mathbf{d} = [1, \ldots, 1]$, MBMP reduces to RA-ORMP (which equals OLS in the SMV scenario).

A support recovery error is defined as the event when the estimated support does not coincide with the true one. For algorithms that return an estimate $\hat{\mathbf{X}}$ of

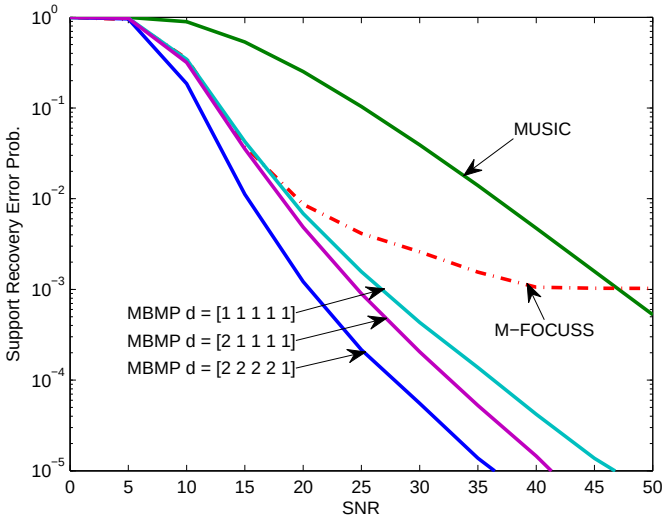

Figure 4.4 Probability of support recovery error as a function of the SNR. The system settings are $Z = 250$, $G = 251$, $M = N = 4$, $P = 5$ and $K = 5$ targets with $|x_{k,p}| = 1$ for all k and p.

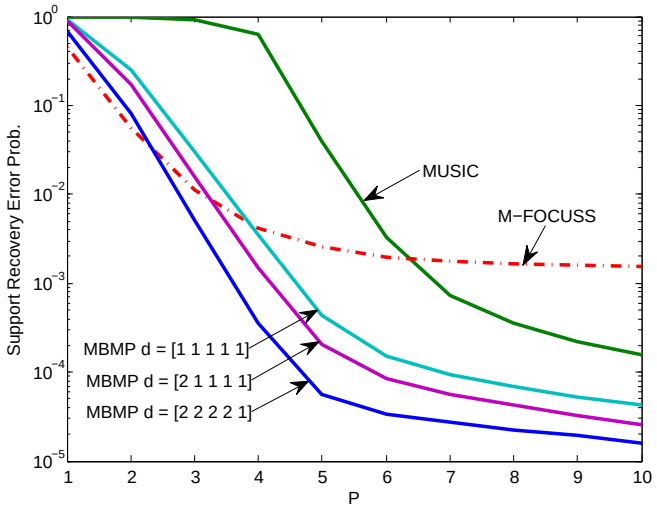

Figure 4.5 Probability of support recovery error as a function of the number of snapshots P. The system settings are $Z = 250$, $G = 251$, $M = N = 4$ and $K = 5$ targets with $|x_{k,p}| = 1$ for all k and p. SNR is 30 dB.

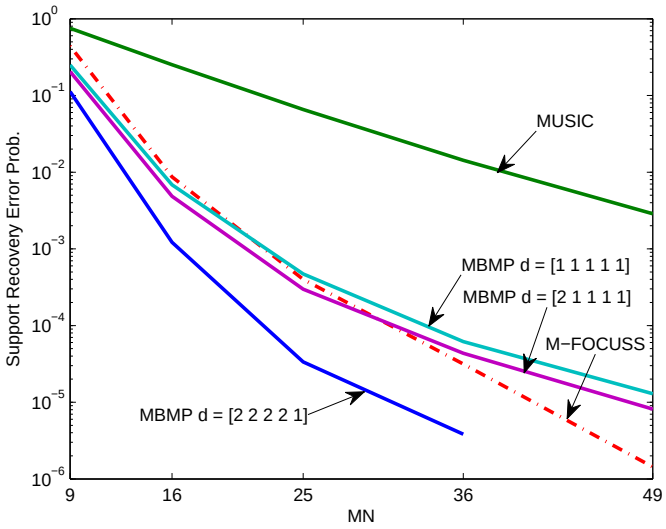

Figure 4.6 Probability of support recovery error as a function of the number of rows MN of $\mathbf{A}$. MMV setup ($P = 5$). The system settings are $Z = 250$, $G = 251$ and $K = 5$ targets with $|x_{k,p}| = 1$ for all k and p. SNR is 20 dB.

the sparse signal $\mathbf{X}$ (e.g., LASSO and M-FOCUSS), the support is identified as the K largest norm rows of the signal $\hat{\mathbf{X}}$. Concerning MBMP, it requires as input a K length branch vector $\mathbf{d}$. The output of MBMP is the estimated support. Therefore in this chapter, K is assumed to be known (notice that this information is needed by all the algorithms including MBMP). It is further assumed that the noise parameter σ in Eq. (2.4) is known, since this information is needed by LASSO and M-FOCUSS. The virtual aperture was $Z = 250$ (thus $G = 251$), and numerical results were obtained for $K = 5$ targets.

In Figure 4.4, it is addressed an MMV setting ($P = 5$) and it is investigated the probability of support recovery error as a function of the SNR. The number of antenna elements are $M = N = 4$. From the figure, it can be seen that MBMP has performance superior to both M-FOCUSS and MUSIC. The floor incurred by M-FOCUSS is due to the inability of this method to exploit the signal subspace information (since M-FOCUSS utilizes $\mathbf{Y}$ instead of $\mathbf{U}$). In addition, MBMP requires a much smaller SNR than MUSIC: for instance, to achieve a probability of error of 10^{-3}, MUSIC requires SNR = 47 dB, while MBMP with $\mathbf{d} = [2, 2, 2, 2, 1]$ achieves the

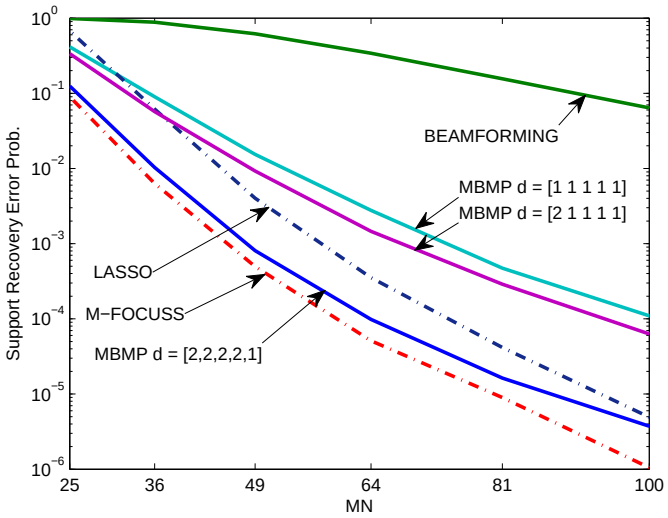

Figure 4.7 Probability of support recovery error as a function of the number of rows MN of $\mathbf{A}$. SMV setup ($P = 1$). The system settings are $Z = 250$, $G = 251$ and $K = 5$ targets with $|x_k| = 1$ for all k. SNR is 20 dB.

same probability of error with just 20 dB. This gain is ascribed to the iterative signal support estimation performed by MBMP, which differs from the one-shot support estimation performed by MUSIC (target locations correspond to the K highest peaks of $\left\|\mathbf{a}_g^H \mathbf{U}\right\|_F^2$). In addition it can be seen that increasing the number of branches for MBMP translates into further SNR gain.

Maintaining the antenna configuration of Figure 4.4, i.e., $M = N = 4$, the SNR is fixed to 30 dB, and Figure 4.5 plots the probability of support recovery error as a function of the number of snapshots P. Although with this figure, all the methods hit a floor as P increase, M-FOCUSS has clearly the worst performance. Comparing MBMP and MUSIC, two main regions emerge: when the number of snapshots P is small, the signal subspace estimation is not sufficiently accurate, and any additional snapshot considerably reduces the error probability. Once the signal subspace estimation is sufficiently accurate (P number of snapshots $\geq K$ number of targets), the value of additional snapshots diminishes. Nevertheless, MBMP requires a smaller number of snapshots P than MUSIC to deliver the same performance.

Moreover, increasing the MBMP number of branches (i.e., algorithm complexity) enables to further decrease the error probability.

In Figure 4.6, the number of snapshots is $P = 5$, the SNR is 20 dB, and the probability of support recovery error is plotted as a function of the number of measurements MN (number of rows of the matrix $\mathbf{A}$). Five different elements configurations are evaluated: $(M, N) = (3, 3)$, $(4, 4)$, $(5, 5)$, $(6, 6)$ and $(7, 7)$. It can be seen that, by increasing the complexity of MBMP, the probability of error can be decreased even when a limited number of antenna elements is used (e.g., MBMP with $\mathbf{d} = [2, 2, 2, 2, 1]$ achieves a probability of error close to 10^{-5} with $MN = 25$). Moreover, in all cases, MBMP performs much better than MUSIC.

In Figure 4.7, the probability of support recovery error is analyzed as a function of the number of measurements MN in an SMV setting ($P = 1$). Six different configurations are considered: $(M, N) = (5, 5)$, $(6, 6)$, $(7, 7)$, $(8, 8)$, $(9, 9)$ and $(10, 10)$, and keep the SNR = 20 dB. In an SMV setting, MUSIC cannot be applied since the signal is not full-rank ($\text{rank}\,(\mathbf{X}) = 1 < K$). In addition to MBMP and M-FOCUSS, target DOA recovery using beamforming and LASSO is performed. From Figure 4.7 it can be seen that beamforming is not well suited to the sparse recovery framework, incurring a very high probability of error compared to sparse recovery methods. Moreover, although in a SMV scenario the signal subspace is not available, MBMP still provides competitive performance. Comparing Figure 4.6 and Figure 4.7, it can be appreciated that by having multiple snapshots ($P > 1$), the number of antenna elements can be dramatically reduced.

4.3 Concluding Remarks

The MBMP algorithm for sparse recovery is develop, and a sufficient condition under which MBMP can recover any sparse signal belonging to a given support is derived. The MB-coherence is then introduced, and it is applied to derive a second

sufficient condition (easier to check) under which MBMP can recover *any* K-sparse signal. This condition enables to guarantee the success of the proposed MBMP for dictionaries that do not satisfy previously known conditions based on coherence or on cumulative coherence. Furthermore, it is demonstrated by numerical examples that MBMP supports trading off measurements (e.g., antenna elements) for computational complexity. Both theoretical guarantees and numerical results illustrate that MBMP enables recovery with fewer measurements than other practical algorithms such as RA-ORMP or OMP.

CHAPTER 5

DETECTION VIA MULTI-BRANCH MATCHING PURSUIT

This chapter addresses target detection from a set of compressive sensing radar measurements corrupted by additive white Gaussian noise. The previous chapters focused on target localization using compressive sensing in the spatial domain, i.e., the use of an undersampled MIMO radar array, and discussed the Multi-Branch Matching Pursuit (MBMP) algorithm, which requires knowledge of the number of targets. Generalizing the MBMP algorithm, a framework for target detection is proposed, which has several important advantages over previous methods: (*i*) it is fully adaptive; (*ii*) it addresses the general multiple measurement vector (MMV) setting; (*iii*) it provides a finite data records analysis of false alarm and detection probabilities, which holds for any measurement matrix. Using numerical simulations, it is shown that the proposed algorithm is competitive with respect to state-of-the-art compressive sensing algorithms for target detection.

5.1 System Model

In the proposed spatial compressive sensing framework for MIMO radar (see Chapter 3), N sensors collect a finite train of P pulses sent by M transmitters and returned from K stationary targets. In this chapter, it is assumed that transmitters and receivers each form a linear arrays of aperture $Z/2$: the m-th transmitter is at position ξ_m, $\xi_m \in [0, Z/2]$ $\forall m$, on the x-axis; the n-th receiver is at position ζ_n, $\zeta_n \in [0, Z/2]$ $\forall n$. Targets are assumed in the far-field, meaning that a target's aspect angle[1] θ_k is constant across the array. The purpose of the system is to detect the presence of targets and determine their DOA angles. Neglecting the discretization error, it is assumed that the target possible locations comply with a grid of G points

[1] θ_k is defined as the sine of the k-th target's DOA angle.

$\phi_{1:G}$ (with $G \gg K$). By defining the $MN \times G$ matrix $\mathbf{A} = [\mathbf{a}(\phi_1), \ldots, \mathbf{a}(\phi_n)]$ where $\mathbf{a}(\theta) \triangleq \mathbf{c}(\theta) \otimes \mathbf{b}(\theta)$ with $\mathbf{b}(\theta) = \left[\exp\left(-j2\pi \frac{Z\theta}{\lambda}\zeta_1\right), \ldots, \exp\left(-j2\pi \frac{Z\theta}{\lambda}\zeta_N\right)\right]^T$ the transmitter steering vector and $\mathbf{c}(\theta) = \left[\exp\left(-j2\pi \frac{Z\theta}{\lambda}\xi_1\right), \ldots, \exp\left(-j2\pi \frac{Z\theta}{\lambda}\xi_M\right)\right]^T$ the receiver steering vector (see [45] for further details), the signal model is expressed:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E} \tag{5.1}$$

where $\mathbf{E} \in \mathbb{C}^{MN \times P}$ represents the noise, which is assumed to be independent and identically distributed (i.i.d.) complex Gaussian, i.e., $\mathrm{vec}(\mathbf{E}) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, with unknown σ^2. The unknown matrix $\mathbf{X} \in \mathbb{C}^{G \times P}$ contains the targets locations and gains. Zero rows of $\mathbf{X}$ correspond to grid points without a target. The problem (5.1) is sparse in the sense that $\mathbf{X}$ has only $K \ll G$ non-zero rows.

The properties of the measurement matrix $\mathbf{A}$ are governed by the grid-points $\phi_{1:G}$ and by the sensors' number and positions $\xi_{1:M}$ and $\zeta_{1:N}$. Since the sensors' positions are assumed random (described by the probability density functions (pdf) $p(\xi)$ and $p(\zeta)$), the elements of the measurement matrix $\mathbf{A}$ are also random. In the following discussion, $p(\xi)$ and $p(\zeta)$ are chosen to be uniform distributions, and $\phi_{1:G}$ is a uniform grid of $2\lambda/Z$-spaced points in the range $[-1, 1]$.

5.2 Detection Using MBMP

The goal of the detection problem is to identify the non-zero norm rows of $\mathbf{X}$ (i.e., its support) given the measurements $\mathbf{Y}$ in Eq. (5.1). It has been shown [26] that, under certain conditions on the matrix $\mathbf{A}$ and the sparsity K, the matrix $\mathbf{X}$ in Eq. (5.1) can be recovered by solving the nonconvex noisy l_0-norm problem (see also Chapter 2):

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \nu \|\mathbf{X}\|_0 \tag{5.2}$$

where ν is a *regularization parameter* which depends on prior information, e.g., the number of targets K or the noise level σ^2. In the following, the proposed framework for target detection is detailed. The MBMP algorithm (introduced in Chapter 4) is extended to handle an unknown number of targets K.

To present the proposed framework, it is instructive to focus on the equivalent reformulation of problem in Eq. (5.2) in terms of the support S of the solution $\mathbf{X}$ using similar steps as in Eq. (2.7), i.e.,

$$\min_{S} \left\| \Pi^{\perp}_{\mathbf{A}_S} \mathbf{Y} \right\|_F^2 + \nu \left| S \right| . \tag{5.3}$$

The regularization parameter ν in Eq. (5.3) affects only the cardinality of the solution's support S. Therefore, for a fixed cardinality, the support can be found from

$$S_j = \arg\min_{S} \left\| \Pi^{\perp}_{\mathbf{A}_S} \mathbf{Y} \right\|_F^2 \text{ s.t. } |S| = j \tag{5.4}$$

which does not require the knowledge of ν. In general, finding the global optimal solution to this problem requires combinatorial complexity, however the solution can be efficiently approximated by using the MBMP algorithm presented in Chapter 4.

5.2.1 MBMP with Unknown Sparsity K

In this section, the MBMP is extended to handle an unknown number of targets K. While K is unknown, it is assumed that an upper limit $\bar{K}$ is available. Note that K has to be lower than the number of rows of $\mathbf{A}$, i.e., $K < MN$, since otherwise the uniqueness of the solution is not guaranteed, even in a noiseless scenario [26].

The MBMP algorithm is extended as follows: (1) MBMP is applied using the upper limit $\bar{K}$, to obtain $\bar{K}$ supports, $S_1, \ldots, S_{\bar{K}}$; (2) by relying on results from detection theory, one support among the $\bar{K}$ supports is chosen. The idea is to approximate the solution to (5.4) for $j = 1$ to $\bar{K}$, to obtain $\bar{K}$ supports,

$S_1, \ldots, S_{\bar{K}}$, with cardinality ranging from 1 to $\bar{K}$. This entire process can be efficiently approximated by using the MBMP algorithm to solve problem in Eq. (5.4) for $j = \bar{K}$. The provisional support achieving the minimum data-fit ($\left\|\Pi^{\perp}_{\mathbf{A}_S}\mathbf{Y}\right\|_F^2$), among nodes at level j, can be used to approximate S_j. In the following, the detection process is analyzed, meaning determining which of the supports $S_1, \ldots, S_{\bar{K}}$ is the true one.

The idea of the detection process is to check whether a test statistic is higher than a threshold. For a given support S, consider the data model $\mathbf{Y} = \mathbf{A}_S\tilde{\mathbf{X}} + \mathbf{E}$, where $\mathbf{A}_S \in \mathbb{C}^{MN\times|S|}$ $(MN > |S|)$ is a known measurement matrix of rank $|S|$, $\tilde{\mathbf{X}} \in \mathbb{C}^{|S|\times P}$ is a matrix of unknown parameters, and the noise term $\mathbf{E} \in \mathbb{C}^{MN\times P}$ satisfies $\mathrm{vec}\,(\mathbf{E}) \sim \mathcal{CN}\,(\mathbf{0}, \sigma^2\mathbf{I})$, where σ^2 is unknown. The goal of the detection process is to decide which of the rows of $\tilde{\mathbf{X}}$ are non-zero. The SMV setting ($P = 1$) for a real measurement matrix was addressed in [5, p. 345]. In the following theorem the general MMV ($P \geq 1$) in complex case is addressed:

Theorem 5.2.1. ***(GLRT for MMV Model - σ^2 Unknown)*** *The hypothesis testing problem of whether a specific row i of $\tilde{\mathbf{X}}$ is non-zero given that the other rows $l \neq i$ are known to be non-zero, is formulated*

$$\begin{aligned} \mathcal{H}_{i,0} &: \left\|\tilde{\mathbf{X}}\,(i,:)\right\| = 0, \sigma^2 > 0 \\ \mathcal{H}_{i,1} &: \left\|\tilde{\mathbf{X}}\,(i,:)\right\| \neq 0, \sigma^2 > 0 \end{aligned} \tag{5.5}$$

Then:

1) The Generalized Log-likelihood Ratio Test (GLRT) for deciding $\mathcal{H}_{i,1}$ is

$$T_i\,(\mathbf{Y}, S) = \frac{\left\|\hat{\mathbf{X}}\,(i,:)\right\|_2^2}{\frac{\left\|\Pi^{\perp}_{\mathbf{A}_S}\mathbf{Y}\right\|_F^2}{MN-|S|}\left[\left(\mathbf{A}_S^H\mathbf{A}_S\right)^{-1}\right]_{i,i}} > \gamma \tag{5.6}$$

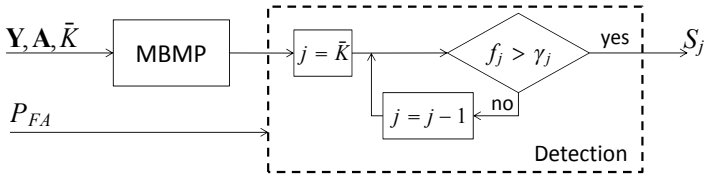

Figure 5.1 Block diagram of the proposed architecture for CS-radar detection.

where $\hat{\mathbf{X}} \triangleq \mathbf{A}_S^{\dagger}\mathbf{Y}$ *is the MLE of* $\tilde{\mathbf{X}}$ *under* $\mathcal{H}_{i,1}$.

2) The exact probability of false alarm for finite data records is given by

$$P_{FA} = Q_{F_{2P,2P(MN-|S|)}}(\gamma). \tag{5.7}$$

3) If the elements of the i-th row of $\tilde{\mathbf{X}}$ *have constant modulo, i.e.,* $\left|\left[\tilde{\mathbf{X}}\right]_{i,t}\right| = \beta$ *for every t, the exact probability of detection for finite data records is given by*

$$P_D = Q_{F'_{2P,2P(MN-|S|)}(\eta_i)}(\gamma) \tag{5.8}$$

where the non-centrality parameter is given by $\eta_i = P\beta^2 / \left(\frac{\sigma^2}{2}\left[\left(\mathbf{A}_S^H\mathbf{A}_S\right)^{-1}\right]_{i,i}\right)$.

Proof. See Appendix C. □

The theorem is applicable when one wants to test the support $S \triangleq S_{K-1} \cup i$ for $i = 1, \ldots, G$ and $i \notin S_{K-1}$, where S_{K-1} is a subset of the true support with cardinality $K-1$. The case when i matches the remaining non-zero row index (i.e., the true support index not in S_{K-1}) is equivalent to the $\mathcal{H}_{i,1}$ hypothesis. Furthermore, the theorem also applies when one wants to test for $S \triangleq S^{true} \cup i$ for $i = 1, \ldots, G$ and $i \notin S^{true}$, where S^{true} is the true support: this case matches the $\mathcal{H}_{i,0}$ hypothesis. The importance of this theorem is that it characterizes the test statistic $T_i(\mathbf{Y}, S)$ and its distribution under $\mathcal{H}_{i,0}$. In other words, the theorem implies that, assuming S contains the true support S^{true}, the test statistics for a row's index i with only noise (i.e., $i \in S \backslash S^{true}$) follows the $F_{2P,2P(MN-|S|)}$ distribution.

In MBMP, when the correct support is estimated and a new index is added to the support, the new index is the one that correlates the most with the noise realization $\mathbf{E}$ among the $G-K$ columns with indices outside the support. Therefore the test statistic, at cardinality $K+1$, is $\max_{i \notin S^{true}} T_1\left(\mathbf{Y}, i \cup S^{true}\right)$, and its distributions is the maximum among $G-K$ random variables, each having a $F_{2P,2P(MN-K-1)}$ distribution. The dependency among these random variables is hard to analyze, hence a closed form seams difficult to obtain. In the numerical results, it is shown that a reasonable assumption is to approximate them as independent. Using this assumption, the test statistic $\max_{i \notin S^{true}} T_1\left(\mathbf{Y}, i \cup S^{true}\right)$ is distributed as the maximum of $G-K$ i.i.d. $F_{2P,2P(MN-K-1)}$ random variables (which for ease of reference is denoted as $g\left(K+1\right)$). Therefore, the threshold for the test statistic at cardinality j can be set as

$$\gamma_j \triangleq Q_{g(j)}^{-1}\left(P_{FA}\right) \tag{5.9}$$

for a given probability of false alarm P_{FA}. Notice that the threshold γ_j depends on the cardinality of the support under test. Focusing at cardinality j, the detection process aims to detect whether all the indices in S_j are non-zero, given that the indices in S_{j-1} are non-zero (i.e., contain targets plus noise). This translates to check whether the minimum test statistic among the indices of S_j not contained in S_{j-1}, i.e.,

$$f_j \triangleq \min_{i \in S_j \setminus S_{j-1}} T_i\left(\mathbf{Y}, S_j\right) \tag{5.10}$$

is higher then the given threshold γ_j. A high value of the metric f_j indicates that S_j is likely to be the true support, whereas a small value of this metric indicates that S_j is likely to contain at least one index that contains only noise.

Summarizing, in order to detect the support among $S_1, \ldots, S_{\bar{K}}$, obtained from MBMP, let $j = \bar{K}$, compute γ_j using Eq. (5.9) and f_j using Eq. (5.10), and check whether $f_j > \gamma_j$. In this case the process stops and $S_{\bar{K}}$ is picked as the support of the

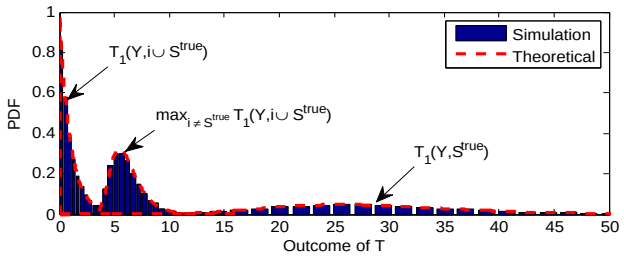

Figure 5.2 Simulated and theoretical pdf of the test statistic $T_i(\mathbf{Y}, S_j)$. System settings: $G = 181$, $K = 8$, $P = 1$, $M = N = 7$, SNR $= 15$db.

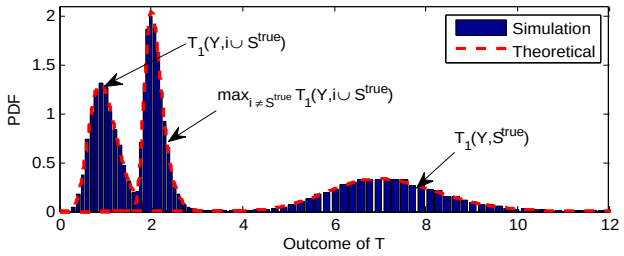

Figure 5.3 Simulated and theoretical pdf of the test statistic $T_i(\mathbf{Y}, S_j)$. System settings: $G = 181$, $K = 8$, $P = 10$, $M = N = 7$, SNR $= 8$db.

detected targets; otherwise set $j = j - 1$ until $f_j > \gamma_j$ or $j = 0$ (which corresponds to the case of detecting no targets). Figure 5.1 depicts a block diagram of the process.

5.3 Numerical Results

In this section, numerical results are presented to demonstrate the potential of the MBMP algorithm for detection using the spatial compressive sensing signal model in Eq. (5.1). To produce each figure, first a random realization of the array sensors' positions is drawn, which is maintained fixed throughout independent Monte-Carlo realizations of the noise ($\text{vec}(\mathbf{E}_p) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})\ \forall p$) and of the targets' positions and responses ($x_{k,p} = \exp(-j\varphi_k)$, $\forall p$, $p = 1, ..., P$, and $\varphi_k \sim \mathcal{U}(0, 2\pi)\ \forall k$, $k = 1, ..., K$). The signal-to-noise ratio (SNR) is defined as $10\log_{10}\sigma^2$.

To start, the simulated and theoretical probability density functions of the test statistic in Eq. (5.6), over the ensemble of random targets and noise, are analyzed in three different cases: (i) the test statistic $T_1(\mathbf{Y}, i \cup S^{true})$, where i is a *fixed* index

outside the support; (ii) the maximum test statistic over indices i outside the support, i.e., $\max_{i \notin S^{true}} T_1\left(\mathbf{Y}, i \cup S^{true}\right)$; (iii) the test statistic $T_1\left(\mathbf{Y}, S^{true}\right)$ for the first index in the true support. Consider $K = 8$ targets and an array aperture of $Z = 180\lambda$ (where λ is the transmitted signal wavelength). The grid size is $G = 181$ grid-points. Figure 5.2 plots the results for an SMV scenario ($P = 1$), while Figure 5.3 shows an MMV scenario ($P = 10$). The theoretical distributions for cases (i) and (iii) (used in Eq. (5.7) and Eq. (5.8), respectively) are exact. It can be seen that the theoretical distributions in case (ii), obtained by assuming independent random variables, closely matches the Monte-Carlo simulations results.

Next, the performance of the proposed detection scheme based on MBMP (i.e., Figure 5.1) are investigated. MBMP is compared with the detection scheme proposed in [29] based on the CAMP algorithm. The MBMP algorithm's complexity can be adjusted by varying the branch vector $\mathbf{d}$. To give an idea of how the complexity of the algorithm affects the performance, the MBMP is run with two different choices of the branch vector $\mathbf{d}$, both of length $\bar{K} = 20$: the first uses $\mathbf{d} = [5, 4, 3, 2, 1, 1, \ldots, 1]$, while the second uses $\mathbf{d} = [1, 1, \ldots, 1]$. Consider $K = 9$ targets and an array aperture of $Z = 250\lambda$. The grid size is $G = 181$ grid-points. In order to compare with the architecture proposed in [29], an SMV scenario is simulated. Denoting with S^{true} the true support, and with $\widehat{S}$ the support determined by a detection algorithm, let $\left(\widehat{S} \backslash S^{true}\right) / (G - K)$ be the empirical probability of false alarm and $\left(\widehat{S} \cap S^{true}\right) / K$ be the empirical probability of detection.

Figure 5.4 plots the ROC curve for the different algorithms. It can be seen how the proposed architecture is able to achieve higher probability of detection for the same false alarm probability as compared to the CAMP algorithm. Further tests would be required to establish which algorithm performs better and under which conditions.

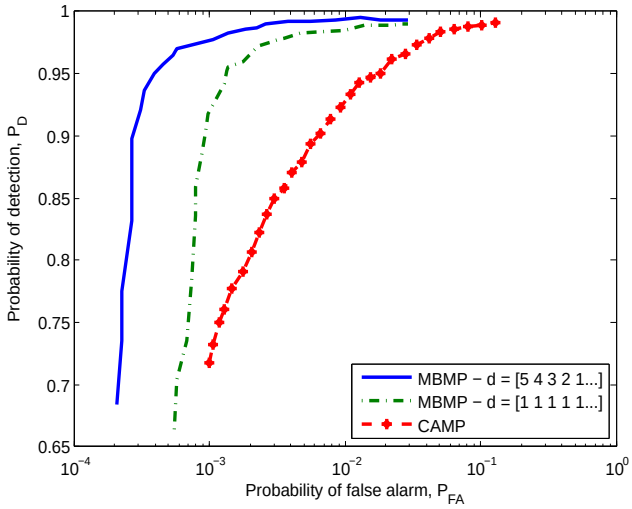

Figure 5.4 ROC curves for the two MBMP architectures and for CAMP. System settings: $G = 251$, $K = 9$, $P = 1$, $M = N = 7$, SNR = 15db.

5.4 Concluding Remarks

This chapter addresses target detection from compressive sensing radar measurements corrupted by additive white Gaussian noise. By taking a detection point-of-view, the MBMP algorithm proposed in [45] is generalized to allow for an unknown number of targets. The resulting architecture for the sparse recovery problem is fully adaptive, i.e., it does not require knowledge of the number of targets or the noise variance. In addition, the false alarm and detection probabilities for the proposed architecture are analyzed. Using numerical simulations, the proposed algorithm is compared against a state-of-the-art compressive sensing algorithm for target detection.

APPENDIX A

PROOFS OF CHAPTER 3

A.0.1 Proof of Proposition 3.3.1

Mean The mean $\eta(u)$ is by definition the expectation of the random array pattern, i.e., $\mathbb{E}[\beta(u)]$, over $z_{mn} = \xi_m + \zeta_n$. The expectation and the summations can be interchanged obtaining

$$\eta(u) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \mathbb{E}\left[\exp\left(juz_{mn}\right)\right]. \tag{A.1}$$

Moreover, the average of $\exp\left[ju_{i,l}\left(\zeta_n + \xi_m\right)\right]$ does not depend on the index n and m, since $\zeta_{1:N}$ are identically distributed, and so are $\xi_{1:M}$. By dropping the indexes of ζ_n and ξ_m and using $z = \xi + \zeta$, it is obtained the sum of MN identical terms, divided by MN. Thus $\eta(u)$ equals $\mathbb{E}[\exp(juz)]$, the characteristic function of the random variable z.

Variance Let $\sigma_1^2(u) = \operatorname{Re}\beta(u)$ and $\sigma_2^2(u) = \operatorname{Im}\beta(u)$. For brevity of notation, the dependency on u is dropped. First notice that since $p(z)$ is even, its characteristic function is real, thus so is the mean value of the array pattern η. Moreover, $\mathbb{E}\left[(\operatorname{Re}\beta - \eta)\operatorname{Im}\beta\right] = \mathbb{E}\left[\operatorname{Re}\beta \operatorname{Im}\beta\right] - \eta\mathbb{E}\left[\operatorname{Im}\beta\right] = 0$, since the real and imaginary parts are uncorrelated and because $\mathbb{E}[\operatorname{Im}\beta] = 0$. Next, focus on $\sigma_1^2 \triangleq \mathbb{E}\left[\left(\operatorname{Re}\beta - \eta\right)^2\right]$ and $\sigma_2^2 \triangleq \mathbb{E}\left[\left(\operatorname{Im}\beta\right)^2\right]$. In order to derive these quantities, consider the expectations given by $\mathbb{E}\left[\left(\beta - \eta\right)^2\right]$ and $\mathbb{E}\left[|\beta - \eta|^2\right]$. It can be shown that,

$$\mathbb{E}\left[\left(\beta - \eta\right)^2\right] = \sigma_1^2 - \sigma_2^2 + j2\sigma_{12} \tag{A.2}$$

and

$$\mathbb{E}\left[|\beta - \eta|^2\right] = \sigma_1^2 + \sigma_2^2. \tag{A.3}$$

Substituting the definition of the random array pattern $\beta(u)$ in Eq. (3.16) and Eq. (3.19) inside Eq. (A.2) and Eq. (A.3), Eq. (3.20) and Eq. (3.21) are obtained.

A.0.2 Proof of Lemma 3.3.2

From Eq. (3.16), it follows that $\mathbf{a}_i^H \mathbf{a}_l = MN \cdot \beta(u_{i,l})$, where $u_{i,l} \triangleq \pi Z(\phi_i - \phi_l)$. When $\phi_{1:G}$ is a uniform grid, $\phi_i - \phi_l$ is constant whenever $i - l$ is constant, i.e., along every diagonal of the matrix $\mathbf{Q}$. Since $\beta(u_{i,l})$ depends only on the term $\phi_i - \phi_l$ (not on the actual ϕ_i and ϕ_j), $\mathbf{Q}$ is a Toeplitz matrix.

A.0.3 Proof of Theorem 3.3.3

The array pattern associated with the transmitter is defined as

$$\beta_\zeta(u_{i,l}) \triangleq \frac{1}{N}\sum_{n=1}^{N} \exp[j u_{i,l}\zeta_n] = \frac{1}{N}\mathbf{b}_i^H \mathbf{b}_l \tag{A.4}$$

and with the receiver arrays as:

$$\beta_\xi(u_{i,l}) \triangleq \frac{1}{M}\sum_{m=1}^{M} \exp[j u_{i,l}\xi_m] = \frac{1}{M}\mathbf{c}_i^H \mathbf{c}_l. \tag{A.5}$$

Statistical properties of random arrays were analyzed in [6] in the case of passive localization (i.e., an array with only receiving elements). The following lemma customizes useful results from [6]:

Lemma A.0.1. *Let the locations $\zeta_{1:N}$ of the receiving array be i.i.d., drawn from an even distribution $p(\zeta)$ and consider a given u. Then $\beta_\zeta(u)$ is asymptotically jointly Gaussian distributed (neglecting the dependency on u):*

$$\begin{bmatrix} \operatorname{Re}\beta_\zeta \\ \operatorname{Im}\beta_\zeta \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \operatorname{Re}\psi_\zeta \\ \operatorname{Im}\psi_\zeta \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right) \tag{A.6}$$

where $\sigma_1^2(u) = \frac{1}{2N}[1 + \psi_\zeta(2u)] - \frac{1}{N}\psi_\zeta^2(u)$ and $\sigma_2^2(u) = \frac{1}{2N}[1 - \psi_\zeta(2u)]$.

Proof. See [6]. □

The joint distribution of $\mathrm{Re}\,\beta_\xi(u)$ and $\mathrm{Im}\,\beta_\xi(u)$ can be obtained similarly.

For a given $i \in \{2,\ldots,G\}$, using Lemma 3.3.1 and the assumption that the mean patterns of both the transmitter and receiver arrays satisfy Eq. (3.23), i.e., $\psi_\xi(u_{1,i}) = \psi_\xi(2u_{1,i}) = \psi_\zeta(u_{1,i}) = \psi_\zeta(2u_{1,i}) = 0$, thus, for both transmitter and receiver arrays, the array pattern evaluated at any grid point is being drawn from an asymptotically complex normal distribution with variance defined by the number of transmit and receive elements, i.e., $\beta_\xi(u_{1,i}) \sim \mathcal{CN}\left(0, \frac{1}{M}\right)$ and $\beta_\zeta(u_{1,i}) \sim \mathcal{CN}\left(0, \frac{1}{N}\right)$. It follows that the random variable $q = \frac{1}{N}\left|\mathbf{b}_1^H \mathbf{b}_i\right|$ can be approximated as belonging to Rayleigh distribution, i.e., $p(q) = (q/\sigma^2)\exp\left(-q^2/2\sigma^2\right)$, where $\sigma^2 = 1/(2N)$, and similarly the random variable $\frac{1}{N}\left|\mathbf{c}_1^H \mathbf{c}_i\right|$ is governed by a Rayleigh distribution with variance $\sigma^2 = 1/(2M)$.

If ξ and ζ are independent (part 1), the two random variables $\frac{1}{N}\left|\mathbf{b}_1^H \mathbf{b}_i\right|$ and $\frac{1}{N}\left|\mathbf{c}_1^H \mathbf{c}_i\right|$ are independent. Using Eq. (3.22), the distribution of $\frac{1}{MN}\left|\mathbf{a}_1^H \mathbf{a}_i\right|$ is the product of two independent Rayleigh distributed variables. The cumulative density function of such a variable is given in [57]. It follows that the ccdf of $\frac{1}{MN}\left|\mathbf{a}_1^H \mathbf{a}_i\right|$ satisfies

$$\Pr\left(\frac{1}{MN}\left|\mathbf{a}_1^H \mathbf{a}_i\right| > q\right) < x \cdot K_1(x), \tag{A.7}$$

where $x \triangleq 2\sqrt{MN}q$.

If $\xi_n = \zeta_n$ for all n (part 2), by using Eq. (3.22), it results $\frac{1}{N^2}\left|\mathbf{a}_1^H \mathbf{a}_i\right| = \left(\frac{1}{N}\left|\mathbf{b}_1^H \mathbf{b}_i\right|\right)^2$. Since the random variable $\frac{1}{N}\left|\mathbf{b}_1^H \mathbf{b}_i\right|$ has a Rayleigh distribution, $\frac{1}{N^2}\left|\mathbf{a}_1^H \mathbf{a}_i\right|$ is distributed as the square of a Rayleigh distribution, which has cdf $1 - \exp(-Nq)$. As such, its ccdf satisfies

$$\Pr\left(\frac{1}{N^2}\left|\mathbf{a}_1^H \mathbf{a}_i\right| > q\right) < \exp(-Nq). \tag{A.8}$$

Part 3 follows because, from Eq. (3.22), the phase of $\mathbf{a}_1^H\mathbf{a}_i$ is the sum of the phases of $\mathbf{b}_1^H\mathbf{b}_i$ and $\mathbf{c}_1^H\mathbf{c}_i$. In the case of transceivers the phase of $\mathbf{a}_1^H\mathbf{a}_i$ is evidently uniform since it is the same phase of $\mathbf{b}_1^H\mathbf{b}_i$. In the case of M transmitter and N receivers, since both $\mathbf{b}_1^H\mathbf{b}_i$ and $\mathbf{c}_1^H\mathbf{c}_i$ are two independent circular symmetric complex normal variables, the sum of the phases is itself uniformly distributed over $[0, 2\pi)$.

A.0.4 Proof of Corollary 3.3.4

Assume that the $G-1$ random variables $\left|\frac{1}{MN}\mathbf{a}_1^H\mathbf{a}_i\right|$, for $i = 2, \ldots, G$, are independent. If ξ and ζ are independent (part 1), from Eq. (3.24), the ccdf of the maximum among $G-1$ such variables (which gives the coherence), is upper bounded by

$$\Pr\left(\max_{i>1}\left|\frac{1}{MN}\mathbf{a}_1^H\mathbf{a}_i\right| > q\right) < 1 - \left[1 - x \cdot K_1(x)\right]^{G-1}, \tag{A.9}$$

where $x \triangleq 2\sqrt{MN}q$.

If $\xi_n = \zeta_n$ for all n (part 2), by using Eq. (3.25), the ccdf of the maximum among $G-1$ such variables, is upper bounded by

$$\Pr\left(\max_{i>1}\left|\frac{1}{N^2}\mathbf{a}_1^H\mathbf{a}_i\right| > q\right) < 1 - \left[1 - \exp(Nq)\right]^{G-1}. \tag{A.10}$$

This concludes the proof.

A.0.5 Proof of Theorem 3.3.5

The theorem follows by combining the claims of Theorem 2.7 in [50] and Corollary 3.3.4. Theorem 2.7 in [50] provides stable recovery guarantees for any K-sparse signal if the measurement matrix $\mathbf{A}$ has RIP $\delta_{2K} < 2/\left(3 + \sqrt{7/4}\right) \triangleq \alpha$. The goal is therefore to bound the RIP δ_{2K} of the spatial compressive sensing measurement matrix $\mathbf{A}$ with probability higher than $1 - \epsilon$. In other words, the goal is to find how many measurements MN are needed to satisfy $\Pr(\delta_{2K} \leq \alpha) \geq 1 - \epsilon$. By using

$\delta_{2K} \leq (2K-1)\,\mu$ [50], it results $\delta_{2K} \leq \alpha$ if $(2K-1)\,\mu \leq \alpha$. Moreover, the condition $\Pr\left((2K-1)\,\mu \leq \alpha\right) \geq 1-\epsilon$ is equivalent to $\Pr\left(\mu > \alpha/\left(2K-1\right)\right) < \epsilon$. Therefore by invoking Eq. (3.27) in Corollary 3.3.4, it follows

$$\Pr\left(\delta_{2K} > \alpha\right) < 1 - \left[1 - x \cdot K_1\left(x\right)\right]^{G-1} \tag{A.11}$$

where, by combining $x \triangleq 2\sqrt{MN}q$ and $q = \frac{\alpha}{2K-1}$, $x = \frac{2\sqrt{MN}\alpha}{2K-1}$. The aim is to find value MN that makes the right hand-side equal to ϵ.

First approximate the modified Bessel function of the second kind $K_1\left(x\right)$ for a large absolute value and small phase of the argument (in the condidered setting, the argument x is real) [58]: $K_1\left(x\right) \approx \sqrt{\frac{\pi}{2x}}\exp\left(-x\right)$. Thus, it should be enforced that $1 - \left[1 - \sqrt{\frac{\pi x}{2}}\exp\left(-x\right)\right]^{G-1} = \epsilon$. Defining $t \triangleq \sqrt{\frac{\pi x}{2}}\exp\left(-x\right)$ and linearizing the function $\left(1-t\right)^{G-1}$ around $t=0$, it is obtained $G\sqrt{\frac{\pi x}{2}}\exp\left(-x\right) = \epsilon$, where, for simplicity, G has been used in place of $G-1$. This equation can be rewritten in the form $-2x\exp\left(-2x\right) = -\gamma^{-2}$, where $\gamma \triangleq \frac{\sqrt{\pi}G}{2\epsilon}$. The inverse function of such equation is called the Lambert W function [59]. For real arguments, it is not injective, therefore it is divided in two branches: $x > 1/2$ or $x \leq 1/2$. Since in the setup $x \leq 1/2$, the lower branch, denoted W_{-1}, is considered, and the solution satisfies $-2x = W_{-1}\left(-\gamma^{-2}\right)$. By using the asymptotic expansion $W_{-1}\left(-\gamma^{-2}\right) \approx -2\ln\gamma - \ln\left(2\ln\gamma\right)$ and solving for MN, Eq. (3.29) is obtained. The claim of the theorem follows from Theorem 2.7 in [50]. Finally, since in this work K-sparse signals are considered, in the error term (3.30), the term for nearly-sparse signals present in [50] is discarded.

A.0.6 Proof of Theorem 3.3.6

Because the variables $\zeta_{1:N}$ are identically distributed, and so are $\xi_{1:M}$, the average $\mathbb{E}\left[\mathbf{A}^H\left(t,:\right)\mathbf{A}\left(t,:\right)\right]$ does not depend on the index $t = N\left(m-1\right)+n$, where the last relation follows from the definition of $\mathbf{a}_g$. Therefore, $\mathbb{E}\left[\mathbf{A}^H\left(t,:\right)\mathbf{A}\left(t,:\right)\right] = \frac{1}{MN}\sum_{t=1}^{MN}\mathbb{E}\left[\mathbf{A}^H\left(t,:\right)\mathbf{A}\left(t,:\right)\right] = \frac{1}{MN}\mathbb{E}\left[\mathbf{Q}\right]$. Thanks to Lemma 3.3.2, only the first

row of $\frac{1}{MN}\mathbb{E}[\mathbf{Q}]$ must be analyzed. Using Eq. (3.16), the elements of the first row of such matrix are $\eta(u_{1,i})$ for $i = 1, \ldots, G$. From Eq. (3.19) in Proposition 3.3.1, $\eta(u_{1,i}) = \psi_z(u_{1,i})$. Thus, requiring Eq. (3.32), i.e., $\psi_z(u_{1,i}) = 0$ for $i = 2, \ldots, G$, together with the fact that $\exp(jzu_{1,1}) = 1$ (because $u_{1,1} \triangleq \pi Z(\phi_1 - \phi_1) = 0$), gives the "if" direction of the claim.

The "only if" direction follows by noticing that when Eq. (3.32) is not satisfied there will be at least one i such that $\eta(u_{1,i}) \neq 0$. Therefore, the matrix $\mathbf{A}$ does not satisfy the isotropy property, showing that Eq. (3.32) is also a necessary condition.

APPENDIX B

PROOFS OF CHAPTER 4

In order to prove Theorem 4.1.2 and Theorem 4.1.4, the following Lemma is needed:

Lemma B.0.2. *Given a K-sparse matrix $\mathbf{X}$ with support S^*. Consider a MBMP node at level $i > K - r$ tagged with a correct provisional support C_i. If $\text{rank}(\mathbf{X}) = r$ then, by setting $d_i = 1$, the one branch of the considered node successfully selects an atom $\bar{\mathbf{a}}_g^{C_i}$ from the set of indices yet to be identified $g \in S^* \setminus C_i$.*

Proof. The proof follows from the assumption made in Chapter 2 that $\text{spark}(\mathbf{A}) > 2K - \text{rank}(\mathbf{X}) + 1$ and by invoking Theorem 3 in [26]. □

B.0.7 Proof of Theorem 4.1.2

First, it is proved that, given a node at level i tagged with a correct provisional support $C_i \subset S^*$ and $S \triangleq S^* \setminus C_i$, if MB-ERC in Eq. (4.4) holds, i.e.,

$$d_i\_\max_{g \notin S^*} \left(\left\| \bar{\mathbf{A}}_S^{\dagger} \bar{\mathbf{a}}_g^{C_i} \right\|_1 \right) < 1,$$

then at least one of the d_i branches of the considered node successfully selects an atom $\bar{\mathbf{a}}_g^{C_i}$ from the correct support set $g \in S$. The proof follows closely the proof that ERC is sufficient for RA-ORMP, given in [26]. The only differences are the use the $d\_\max$ operator and, since a provisional support C_i is known, a refined dictionary $\bar{\mathbf{A}} \triangleq \left\{ \bar{\mathbf{a}}_g^{C_i}, g \notin C_i \right\}$ is considered.

Similar to other MP techniques, but with the key difference of the $d\_\max$ operator, in order to guarantee that at least one of the d_i branches of the considered node successfully selects an atom $\bar{\mathbf{a}}_g^{C_i}$ from the remaining correct indices $g \in S$, it is required that

$$\frac{d_i\_\max_{g \notin S^*} \left(\left\| \mathbf{U}^H \bar{\mathbf{a}}_g^{C_i} \right\|_2 \right)}{\max_{g \in S} \left\| \mathbf{U}^H \bar{\mathbf{a}}_g^{C_i} \right\|_2} < 1, \tag{B.1}$$

where $\mathbf{U} = \text{orth}\left(\Pi^{\perp}_{\mathbf{A}_{C_i}}\mathbf{Y}\right)$. The norm $d_{i\_}\max_{g\notin S^*}\left(\left\|\mathbf{U}^H\bar{\mathbf{a}}_g^{C_i}\right\|_2\right)$ can be bounded for any $\mathbf{U}$, by using standard norm inequalities. Let $g^* \triangleq \arg d_{i\_}\max_{g\notin S^*}\left(\left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_g^{C_i}\right\|_1\right)$, it can be wrote:

$$\begin{aligned} d_{i\_}\max_{g\notin S^*}\left(\left\|\mathbf{U}^H\bar{\mathbf{a}}_g^{C_i}\right\|_2\right) &= \max_{\mathbf{x}\neq\mathbf{0}}\frac{\left|\left(\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_{g^*}^{C_i}\right)^H\bar{\mathbf{A}}_S^H\mathbf{U}\mathbf{x}\right|}{\|\mathbf{x}\|_2} \\ &\leq \left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_{g^*}^{C_i}\right\|_1\max_{\mathbf{x}\neq\mathbf{0}}\frac{\left\|\bar{\mathbf{A}}_S^H\mathbf{U}\mathbf{x}\right\|_\infty}{\|\mathbf{x}\|_2} \\ &= \left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_{g^*}^{C_i}\right\|_1\max_{g\in S}\left\|\mathbf{U}^H\bar{\mathbf{a}}_g^{C_i}\right\|_2. \end{aligned} \tag{B.2}$$

where the inequality follows since

$$\frac{\left|\left(\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_{g^*}^{C_i}\right)^H\bar{\mathbf{A}}_S^H\mathbf{U}\mathbf{x}\right|}{\|\mathbf{x}\|_2} = \frac{\left|\left(\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_{g^*}^{C_i}\right)^H\bar{\mathbf{A}}_S^H\mathbf{U}\mathbf{x}\right|}{\left\|\bar{\mathbf{A}}_S^H\mathbf{U}\mathbf{x}\right\|_\infty}\frac{\left\|\bar{\mathbf{A}}_S^H\mathbf{U}\mathbf{x}\right\|_\infty}{\|\mathbf{x}\|_2}. \tag{B.3}$$

This shows that

$$\frac{d_{i\_}\max_{g\notin S^*}\left(\left\|\mathbf{U}^H\bar{\mathbf{a}}_g^{C_i}\right\|_2\right)}{\max_{g\in S}\left\|\mathbf{U}^H\bar{\mathbf{a}}_g^{C_i}\right\|_2} \leq d_{i\_}\max_{g\notin S^*}\left(\left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_g^{C_i}\right\|_1\right). \tag{B.4}$$

Thus it is concluded that, if Eq. (4.4), i.e.,

$$d_{\_}\max_{g\notin S^*}\left(\left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_g^{C_i}\right\|_1\right) < 1,$$

holds for any C_i, Eq. (B.1) is guaranteed to hold too and, thus, at least one of the d_i branches of the considered node successfully selects an atom $\bar{\mathbf{a}}_g^{C_i}$ from the correct support set $g \in S$.

It remain to prove that if, for any node at level $i = 1, \ldots, K-1$, MB-ERC$(\mathbf{A}, S^*, d_i, C_i)$ holds for every possible $C_i \subset S^*$ (with $|C_i| = i-1$), then MBMP with branch vector $\mathbf{d} = [d_1, \ldots, d_K]$ is guaranteed to recover $\mathbf{X}$ from measurements $\mathbf{Y} = \mathbf{A}\mathbf{X}$ (i.e., one node at level $K+1$ contains the correct support S^*). To prove this, note that if MB-ERC$(\mathbf{A}, S^*, d_i, C_i)$ holds for any node at level $i = 1, \ldots, K-r$,

it follows that a chain of correct decisions is made along the tree: MB-ERC holds for the first node, thus one node at level 2 has a correct provisional support. Considering such node, since MB-ERC holds, it will select a correct index, and there exists a node at level 3 with correct provisional support, and so on up to level $K-r$. Since $\text{rank}\,(\mathbf{X}) = r$, by invoking Lemma B.0.2, the chain of correct decisions is guaranteed also for successive level $i > K-r$. This concludes the proof.

B.0.8 Proof of Theorem 4.1.4

First, it is shown that, given a node at level i tagged with a correct provisional support C_i, the MB-coherence condition in Eq. (4.10) implies MB-ERC($\mathbf{A}, S^*, d_i, C_i$) in Eq. (4.4), i.e.,

$$d_i\_\max_{g \notin S^*} \left(\left\| \bar{\mathbf{A}}_S^{\dagger} \bar{\mathbf{a}}_g^{C_i} \right\|_1 \right) < 1, \tag{B.5}$$

for any support $S^* = S \cup C_i$ of cardinality K (since $|C_i| = i-1$ and $S \cap C_i = \emptyset$). Consider the left hand-side of Eq. (B.5). By using the definition of pseudo-inverse $\bar{\mathbf{A}}_S^{\dagger} \triangleq \left(\bar{\mathbf{A}}_S^H \bar{\mathbf{A}}_S\right)^{-1} \bar{\mathbf{A}}_S^H$, for any g, it can be wrote

$$\left\| \bar{\mathbf{A}}_S^{\dagger} \bar{\mathbf{a}}_g^{C_i} \right\|_1 \leq \left\| \left(\bar{\mathbf{A}}_S^H \bar{\mathbf{A}}_S\right)^{-1} \right\|_1 \left\| \bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_g^{C_i} \right\|_1. \tag{B.6}$$

By introducing the $d\_\max$ operator and by noticing that $\left\| \left(\bar{\mathbf{A}}_S^H \bar{\mathbf{A}}_S\right)^{-1} \right\|_1$ does not depend on g, it follows that

$$d_i\_\max_{g \notin S \cup C_i} \left(\left\| \bar{\mathbf{A}}_S^{\dagger} \bar{\mathbf{a}}_g^{C_i} \right\|_1 \right) \leq \left\| \left(\bar{\mathbf{A}}_S^H \bar{\mathbf{A}}_S\right)^{-1} \right\|_1 d_i\_\max_{g \notin S \cup C_i} \left(\left\| \bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_g^{C_i} \right\|_1 \right). \tag{B.7}$$

Using standard arguments as in [25], the first term on the right-hand side can be upper bounded as

$$\left\| \left(\bar{\mathbf{A}}_S^H \bar{\mathbf{A}}_S\right)^{-1} \right\|_1 \leq \frac{1}{2 - \max_{g \in S} \left\| \bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_g^{C_i} \right\|_1}. \tag{B.8}$$

Therefore, for any support S, the left hand-side of Eq. (B.5) can be upper bounded as

$$d_{i\_}\max_{g\notin S^*}\left(\left\|\bar{\mathbf{A}}_S^{\dagger}\bar{\mathbf{a}}_g^{C_i}\right\|_1\right)\leq\frac{d_{i\_}\max_{g\notin S\cup C_i}\left(\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1\right)}{2-\max_{g\in S}\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1}. \tag{B.9}$$

It follows that MB-ERC holds (for *any* support $S^*=S\cup C_i$ of cardinality K) if

$$\max_{S,|S|=K-i+1}\frac{d_{i\_}\max_{g\notin S\cup C_i}\left(\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1\right)}{2-\max_{g\in S}\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1}<1. \tag{B.10}$$

This can be manipulated to obtain the MB-coherence condition in Eq. (4.10), i.e.,

$$\max_{S,|S|=K-i+1}\left(\max_{g\in S}\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1+d_{i\_}\max_{g\notin S\cup C_i}\left(\left\|\bar{\mathbf{A}}_S^H\bar{\mathbf{a}}_g^{C_i}\right\|_1\right)\right)<2 \tag{B.11}$$

The claim of the theorem follows by invoking Theorem 4.1.2, completing the proof.

B.0.9 Testing for MB-coherence

Here, a practical way to find the smallest integer d_i such that the MB-coherence condition in Eq. (4.10) is met is outlined. The proposed strategy does not require an exhaustive search over all $\binom{n}{k}$ possible supports of cardinality $k\triangleq K-i+1$. In detail, by exploiting the one-to-one correspondence between a support S of cardinality k (out of n elements) and a binary vector with k ones and $n-k$ zeros, the following proposition relates condition in Eq. (4.10) to an integer program [53]:

Proposition B.0.3. *The smallest integer d_i such that the MB-coherence condition in Eq. (4.10) holds is given by the optimal objective value of the integer program*

$$\max_{\mathbf{s},\mathbf{y},\mathbf{z}} \quad 1+\sum_{l=1}^{n} z_l \tag{B.12}$$

$$s.t. \begin{cases} (\mathbf{q}_j+\mathbf{q}_g)^T(\mathbf{s}+\mathbf{y}) \geq y_j+z_g & \forall g \neq j \\ \sum_{l=1}^{n} s_l = K-i & \\ \sum_{l=1}^{n} y_l = 1 & \\ y_l+s_l+z_l \leq 1 & \forall l \\ s_l, y_l, z_l \in \{0,1\} & \forall l \end{cases},$$

where $\mathbf{q}_g$ is the g-th column of $\mathbf{Q} \triangleq \left|\left(\bar{\mathbf{A}}^{C_i}\right)^H \bar{\mathbf{A}}^{C_i}\right|$ ($|\cdot|$ denotes the element-wise absolute value).

Proof. Let d_i be the smallest integer such that MB-coherence condition in Eq. (4.10) holds. Since Eq. (4.10) does not hold for d_i-1, there exist a support S of cardinality k, an index $j \in S$, and a set of indices $\mathcal{G}$ with $|\mathcal{G}| = d_i - 1$, such that $\left\|\bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_j^{C_i}\right\|_1 + \left\|\bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_g^{C_i}\right\|_1 \geq 2\ \forall g \in \mathcal{G}$. Define the binary vectors $\mathbf{s}$, $\mathbf{y}$, and $\mathbf{z}$, as $s_l = 1$ for $l \in S$, $y_j = 1$, and $z_l = 1$ for $l \in \mathcal{G}$). It can be shown that $\mathbf{s}$, $\mathbf{y}$, and $\mathbf{z}$ maximize problem in Eq. (B.12). Indeed, these vectors are feasible ($\left\|\bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_j^{C_i}\right\|_1 + \left\|\bar{\mathbf{A}}_S^H \bar{\mathbf{a}}_g^{C_i}\right\|_1 = 1+(\mathbf{q}_j+\mathbf{q}_g)^T\mathbf{s} \geq 2 = y_j+z_g\ \forall g \in \mathcal{G}$) and there cannot be feasible $\mathbf{s}$, $\mathbf{y}$, and $\mathbf{z}$ with $1+\sum_{l=1}^{n} z_l = d_i$, since Eq. (4.10), by assumption, does not hold for d_i-1. Conversely, the optimal solution of the problem in Eq. (B.12) can be translated into a support S of cardinality k, an index $j \in S$, and a set of indices $\mathcal{G}$, such that Eq. (4.10) holds for $d_i = |\mathcal{G}|+1$, but not for $d_i = |\mathcal{G}|$. As a result, finding the smallest integer d_i such that the MB-coherence condition in Eq. (4.10) is met tantamounts to find feasible vectors $\mathbf{s}$, $\mathbf{y}$, and $\mathbf{z}$ to maximize the problem in Eq. (B.12) and $d_i = 1+\sum_{l=1}^{n} z_l$, as claimed. □

Problem (B.12) can be solved using techniques for integer programming [53]. Since $d_i = 1+\sum_{l=1}^{n} z_l$ governs the MBMP complexity, only instances with a small

objective value are of interest. As a result, it is usually relatively fast to check whether Eq. (4.10) is met for practical values of the objective function $d_i = 1 + \sum_{l=1}^{n} z_l$.

APPENDIX C

PROOF OF CHAPTER 5

C.0.10 Proof of Theorem 5.2.1

Following the same steps as [5, pp. 371-372], it can be shown that the GLRT can be written as $\left(\hat{\sigma}_0^2 - \hat{\sigma}_1^2\right)/\hat{\sigma}_1^2$ where $\hat{\sigma}_l^2$ is the MLE of the noise level σ^2 under $\mathcal{H}_{i,l}$ for $l = 1, 2$. Extending [60, pp. 176-177], it can be shown that $\hat{\sigma}_1^2 = \left\|\Pi_{\mathbf{A}_S}^{\perp}\mathbf{Y}\right\|_F^2/MNP$, while, following similar steps as [5, Appendix 7B], $\hat{\sigma}_0^2 - \hat{\sigma}_1^2$ reduces to $\left\|\hat{\mathbf{X}}\left(i,:\right)\right\|_2^2/MNP\left[\left(\mathbf{A}_S^H\mathbf{A}_S\right)^{-1}\right]_{i,i}$ where $\hat{\mathbf{X}} \triangleq \mathbf{A}_S^{\dagger}\mathbf{Y}$ is the MLE of $\tilde{\mathbf{X}}$ under $\mathcal{H}_{i,1}$. Therefore, Eq. (5.6) can be written as $\left(MN - |S|\right)\left(\hat{\sigma}_0^2 - \hat{\sigma}_1^2\right)/\hat{\sigma}_1^2$, which, except from a scaling factor, is the GLRT. Now, consider the random variable $\left\|\hat{\mathbf{X}}\left(i,:\right)\right\|_2^2 / \left(\left[\left(\mathbf{A}_S^H\mathbf{A}_S\right)^{-1}\right]_{i,i}\frac{\sigma^2}{2}\right)$. Under $\mathcal{H}_{i,1}$ and assuming $\beta = \left|\left[\tilde{\mathbf{X}}\right]_{i,t}\right|$, for every t, it has a non-central Chi-Squared distribution with $2P$ degrees of freedom and non-centrality parameter given by η_i, and, under $\mathcal{H}_{i,0}$, it has a central Chi-Squared distribution with $2P$ degrees of freedom. Equivalently, it can be shown that the random variable $\left\|\Pi_{\mathbf{A}_S}^{\perp}\mathbf{Y}\right\|_F^2/\frac{\sigma^2}{2}$ has a Chi-Squared distribution with $2P\left(MN - |S|\right)$ degrees of freedom under either $\mathcal{H}_{i,0}$ and $\mathcal{H}_{i,1}$. These two random variables are independent. Notice that Eq. (5.6) can be obtained by normalizing each of the random variables by the number of degrees of freedom, and by taking their ratio. By definition, Eq. (5.6) follows a $F'_{2P,2P(MN-|S|)}\left(\eta_i\right)$ distribution, under $\mathcal{H}_{i,1}$, and a $F_{2P,2P(MN-|S|)}$ distribution, under $\mathcal{H}_{i,0}$.

BIBLIOGRAPHY

[1] Y. C. Eldar and G. Kutyniok, *Compressed sensing: Theory and applications*. Cambridge, UK: Cambridge University Press, 2012.

[2] A. M. Haimovich, R. Blum, and L. Cimini, "MIMO radar with widely separated antennas," IEEE Signal Processing Mag., vol.25, no.1, pp. 116-129, 2008.

[3] P. Stoica and J. Li, "MIMO radar with colocated antennas," IEEE Signal Processing Mag., vol. 24, no. 5, pp. 106-114, Sep. 2007.

[4] H. L. VanTrees, *Detection, estimation and modulation theory: Optimum array processing Vol. 4*. New York: Wiley, 2002.

[5] S. M. Kay, *Fundamentals of statistical signal processing vol. II: Detection theory*. Englewood Cliffs, NJ: Prentice Hall, 1998.

[6] Y. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," IEEE Trans. on Antennas and Propagation, vol. 12, no. 3, pp. 257-268, May 1964.

[7] M. A. Haleem and A. M. Haimovich, "On the distribution of ambiguity levels in MIMO radar," in Proc. 42nd Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA, Oct. 2008.

[8] Y. Yu, A. P. Petropulu, and H. V. Poor, "MIMO radar using compressive sampling ," IEEE J. Sel. Topics in Sig. Proc., vol.4, no.1, pp. 146-163, Feb. 2010.

[9] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for MIMO radar," IEEE Trans. Signal Process., vol. 62, no. 2, pp. 419-430, Jan. 2014.

[10] T. Strohmer, and B. Friedlander, "Analysis of sparse MIMO radar," Available online: http://arxiv.org/pdf/1203.2690, Last Access: Apr. 2014.

[11] J. H. G. Ender, "On compressive sensing applied to radar," Elsevier J. Signal Process., vol. 90, pp. 1402-1414, Nov 2009.

[12] K. Gedalyahu and Y. C. Eldar, "Time-delay estimation from low-rate samples: A union of subspaces approach," IEEE Trans. Signal Process., vol. 58, no. 6, pp. 3017-3031, Jun. 2010.

[13] W. U. Bajwa, K. Gedalyahu, and Y. C. Eldar, "Identification of parametric underspread linear systems with application to super-resolution radar," IEEE Trans. Signal Process., vol. 59, no. 6, pp. 2548-2561, Aug. 2011.

[14] O. Bar-Ilan and Y. C. Eldar, "Sub-nyquist radar via doppler focusing," submitted to IEEE Trans. Signal Process., Nov. 2012.

[15] E. Baransky, G. Itzhak, I. Shmuel, N. Wagner, E. Shoshan and Y. C. Eldar, "A sub-nyquist radar prototype: Hardware and algorithms," submitted to IEEE Trans. on Aerospace and Electronic Systems, special issue on compressed sensing for radar, Aug. 2012.

[16] S. Gogineni and A. Nehorai, "Target estimation using sparse modeling for distributed MIMO radar," IEEE Trans. Signal Process., vol. 59, pp. 5315-5325, Nov. 2011.

[17] T. Strohmer and B. Friedlander, "Compressed sensing for MIMO radar - algorithms and performance," in Proc. 43rd Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2009.

[18] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Trans. Signal Process., vol. 53, no. 8, pp. 3010-3022, 2005.

[19] L. Carin, "On the relationship between compressive sensing and random sensor arrays," IEEE Antennas Propagat. Mag., vol. 5, no. 5, pp. 72-81, Oct. 2009.

[20] E. J. Candes and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," IEEE Trans. Inf. Theory, vol. 57, no. 11, pp. 7235-7254, 2011.

[21] M. Hugel, H. Rauhut, and T. Strohmer, "Remote sensing via ℓ_1 minimization," Available online: http://arxiv.org/pdf/1205.1366, Last Access: Apr. 2014.

[22] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," Proc. IEEE, vol. 98, no. 6, pp. 948-958, June 2010.

[23] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," IEEE Trans. Acoust., Speech, Signal Process., vol. 37, pp. 720-741, May 1989.

[24] D. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp. 1289-1306, 2006.

[25] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. Inf. Theory, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.

[26] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1135-1146, Jun. 2012.

[27] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays ," IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 7, pp. 996-1000, Jul 1989.

[28] R. Schmidt, "Multiple emitter location and signal parameter estimation," Antennas and Propagation, IEEE Transactions on, vol. 34, no. 3, pp. 276-280, 1986.

[29] L. Anitori, A. Maleki, M. Otten, R. G. Baraniuk, and P. Hoogeboom, "Design and analysis of compressive sensing radar detectors," Available online: *https://dsp.rice.edu/sites/dsp.rice.edu/files/publications/journal-article/9998/cfarcs.pdf*, Mar. 2012.

[30] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Multi-Branch Matching Pursuit with applications to MIMO radar," submitted to IEEE Transactions on Signal Processing. Available online: http://arxiv.org/pdf/1312.5765, Last Access: Apr. 2014.

[31] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Compressive sensing with unknown parameters," in Proc. 46th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 4-7, 2012.

[32] E. J. Candes and M.B. Wakin, "An introduction to compressive sampling," IEEE Signal Processing Mag., vol. 25, no. 2, pp. 21-30, Mar. 2008.

[33] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," IEEE J. Sel. Topics Signal Process., vol. 1, no. 4, pp. 586-597, Dec. 2007.

[34] S. Becker, J. Bobin, and E. J. Candes, "NESTA: A fast and accurate first-order method for sparse recovery," SIAM J. Imaging Sci, vol. 4, no. 1, pp. 1-39, 2011.

[35] S. Becker, E. J. Candes, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," in Mathematical Programming Computation, W. J. Cook, Ed., 2012, vol. 3, no. 3, pp. 165-218.

[36] D. L. Donoho, A. Maliki, and A. Montanari, "Message-passing algorithms for compressed sensing," Proc. Nat. Acad. Sci., vol. 106, no. 45, pp. 18 914-18 919, 2009.

[37] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian matching pursuit: Model uncertainty and parameter estimation for sparse linear models," IEEE Trans. Signal Process., Mar. 2009.

[38] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," IEEE Signal Process. Lett., vol. 14, pp. 707-710, Oct. 2007.

[39] E. J. Candes, M. B. Wakin and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," Journal of Fourier Analysis and Applications, vol. 14, no. 5, pp. 877-905, 2008.

[40] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," IEEE J. Select. Topics Signal Process., vol. 4, no. 2, pp. 317-329, 2010.

[41] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," IEEE Trans. Signal Process., vol. 53, no. 7, pp. 2477-2488, July 2005.

[42] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in Proc. 27th Asilomar Conf. on Signals, Systems and Computes, Pacific Grove, CA, Nov. 1993.

[43] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," Int. Journal of Control, vol. 50, no. 5, pp. 1873-1896, 1989.

[44] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.," Applied Computational Harmonic Analysis, vol. 26, pp. 301-321, 2008.

[45] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial Compressive Sensing in MIMO Radar with Random Arrays," in Proc. CISS 2012, Princeton, NJ, Mar. 21-23, 2012.

[46] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," Applied and Computational Harmonic Analysis, vol. 27, no. 3, pp. 265-274, 2009.

[47] M. Skolnik, *Introduction to radar systems*. 3rd ed. New York: McGraw-Hill, 2002.

[48] J. V. DiFranco and W. L. Rubin, "Spatial ambiguity and resolution for array antenna systems," IEEE Trans. on Military Electronics, July-Oct. 1965.

[49] D. H. Johnson and D. E. Dudgeon, *Array signal processing concepts and techniques*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[50] H. Rauhut, "Compressive sensing and structured random matrices," Theoretical Foundations and Numerical Methods for Sparse Recovery, vol. 9, pp. 1-92, 2010.

[51] H. Rauhut, "On the impossibility of uniform sparse reconstruction using greedy methods," Sampl. Theory in Signal and Image Process., vol. 7, no. 2, pp. 197-215, 2008.

[52] T. Blumensath and M. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," [online] unpublished manuscript, available at: *http://eprints.soton.ac.uk/142469/1/BDOMPvsOLS07.pdf*, 2007.

[53] B. Korte, and J. Vygen, *Combinatorial optimization: theory and algorithms*. Vol. 21. New York: Springer, 2008.

[54] D. A. Lorenz, S. Schiffler, and D. Trede, "Beyond convergence rates: Exact recovery with the Tikhonov regularization with sparsity constraints," Inverse Problems, vol. 27, no. 8, pp. 085009, Nov. 2011.

[55] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," IEEE Trans. Inf. Theory, vol. 59, no. 5, pp. 3158-3174, May 2013.

[56] M. Elad, *Sparse and redundant representations – from theory to applications in signal and image processing*. New York: Springer, 2010.

[57] M. K. Simon, *Probability distributions involving Gaussian random variables: A handbook for engineers and scientists*. New York: Springer, 2002.

[58] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Washington, DC: National Bureau of Standards, U.S. Dept. of Commerce, 1972.

[59] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," Adv. Comput. Math., vol. 5, pp. 329-359, 1996.

[60] S. M. Kay, *Fundamentals of statistical signal processing vol. I: estimation theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.