Sparse Spectrum Fitting in Array Processing

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jimeng Zheng

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Professor Mostafa Kaveh, Advisor

February, 2013

Acknowledgements

I would like to express my deep gratitude to all the people who have made this thesis possible and my life in Minneapolis full of cheers.

First and foremost, I would like to thank my advisor, Professor Mostafa Kaveh, for giving me the invaluable opportunity to work on interesting projects and acquire knowledge and experience in array signal processing. I am truly indebted to his persistent support, help, patience, encouragement and guidance throughout my Ph.D. studies. His enthusiasm for research also deeply impressed me and directed me in my research work. It has been my honor to work with and learn from such an extraordinary individual.

I would like to thank to Professor Georgios B. Giannakis, Professor Zhi-Quan Luo and Professor Xiaotong Shen for what they taught me not only in the classes but also through the discussions on this research project. I also want thank them for agreeing to serve on my thesis committee and for reviewing the manuscript.

My colleagues at the Electrical and Computer Engineering Department and the Digital Technology Center (DTC) have enriched my graduate life and helped me in many ways, and they deserve a special mention: Chinann Yang, Hossein Zare, Pedram Nezafat, Bryan Perfetti, Jing Wang, Hao Zhu, Mingyi Hong, Ruoyu Sun, Dan Wang, Yu Zhang, Wentao Shi, Vassilis Kekatos, Seung-Jun Kim, Ketan Rajawat, Qiliang Wu, Xiaofei Wang, Xi Chen, Jialie Hu and Feng Cai. My thanks also go to my soccer team members. Playing soccer with them provided much-needed fun during my graduate life.

I also want to thank Dr. Michael Lotti for his efforts in proofreading the draft of this thesis and significantly improving its readability.

I would like to acknowledge financial support from the US Department of Energy for funding parts of my research projects.

It is impossible to remember everyone who has helped me, and I apologize to those

I have inadvertently left out. Thank you all!

Dedication

Dedicate to my parents and all my family members.

Abstract

In this thesis, we focus on the application of sparsity-encouraging regularization techniques to the problem of Direction-Of-Arrival (DOA) estimation using sensor arrays. By developing the sparse representation models for the spatial covariance matrix of correlated or uncorrelated sources respectively, the DOA estimation problem is reformulated under the framework of Sparse Signal Reconstruction (SSR). The L_1-Norm regularization technique is employed to formulate sparsity-exploiting DOA estimation algorithms, which can estimate DOAs and signal powers simultaneously. The algorithm specialized for uncorrelated sources, Sparse Spectrum Fitting (SpSF), is attractive for its computational complexity, resolution capability, low large error threshold and robustness with respect to source correlation (when combined with spatial smoothing and forward-backward smoothing). Despite the similarity between the formulation of SpSF and ordinary SSR algorithms, such as Lasso [1], SpSF can perform much better than predicted by the existing theoretical results in SSR literature, because of the extra positivity constraints in its formulation (which are included since the optimization variables of SpSF represent the signal powers). Thus, we begin this thesis by developing and justifying its formulations. This is followed by a discussion of the existence and uniqueness of the solutions of SpSF, which provides an explicit formula for the maximum number of sources whose DOAs can be reliably estimated by SpSF. Although it is hard to directly relate this maximum number of sources to the number of sensors M, we recognize that this maximum number is actually the degree of freedom of the co-array and, as an exception, we prove that by using Uniform Linear Arrays (ULA), SpSF can work with $M-1$ sources. Further, by combining with the uniqueness result, the estimation consistency of SpSF with respect to infinitely many snapshots and sensors is obtained.

Another focus of this thesis is the selection of the regularization parameter (RP) of SpSF. Inspired by the theoretical analysis in the first part of this thesis, Diagonal Loading technique is used to significantly reduce the sensitivity of SpSF with respect to its RP, which further enables the control of such sensitivity. Based on the Discrepancy Principle, an iterative version of SpSF, Iterative-SpSF (I-SpSF), is formulated as a parameter-free optimization algorithm and shown to achieve similar or even better

performance than SpSF using exhaustively searched but fixed RPs. Following that, by analyzing SpSF's probability of perfect support recovery, an upper bound on such probability is proposed. This upper bound can accurately approximate the probability of perfect DOA estimates of SpSF (perfect DOA estimates is only possible in the context of DOAs belonging to the grid of candidate directions). Based on this upper bound and a Monte Carlo evaluation process, an automatic and data-adaptive RP selector is proposed for the purpose of DOA estimation when the number of snapshots is finite. This technique is robust with respect to its parameters and can help SpSF to achieve almost the same or even better performance than exhaustively searched and fixed RPs for uncorrelated and correlated sources respectively.

The extensions of SpSF to the cases of off-grid DOAs and moving sources are considered and two algorithms SpSF with Modeling Uncertainty (SpSFMU) and Persistently Active Block Sparsity (PABS) are proposed for the two cases, respectively. By using extra variables to parameterize the modeling errors caused by off-grid DOAs, SpSFMU is formulated as a convex optimization problem that can provide continuous DOA estimates. Under the presence of off-grid DOAs, SpSFMU can be used either to improve the estimation performance of SpSF or to reduce its computational complexity. In contrast, PABS is introduced as a general estimator for inconsistent but persistently active sparse models. PABS is formulated by using a novel objective function promoting Block-Level Sparsity and Element-Level Sparsity simultaneously. Then, on the basis of the sparse representation model of array snapshots, PABS is applied to the DOA estimation of moving sources and shown to exhibit significant performance improvement over C-HiLasso [2] and SPICE [3].

Intensive simulation results are included in each chapter to illustrate the effectiveness and performance of the presented algorithms and theoretical analyses.

Contents

List of Figures

Chapter 1

Introduction

In this dissertation, we consider the problem of source localization using sensor arrays. With sensor arrays, source localization, which is a major objective in a number of signal processing applications, can be translated to the Direction-Of-Arrival (DOA) estimation. Among the existing DOA estimation methods, the signal subspace technique has been widely used due to its super-resolution performance balanced with reasonable computational complexity. A fundamental assumption of the methods based on this technique is the existence of a low-dimensional signal subspace, which usually implies a limited number of impinging sources. In combination with the plane wave propagation model, this implication leads to spatial sparsity, which has promoted sparsity-exploiting DOA estimation methods by utilizing recently advanced sparse signal reconstruction algorithms. These methods have been shown to have the potential for performance advantage over signal subspace techniques, including better resolution and robustness to correlation among sources. A new DOA estimation method based on the sparsity of spatial spectrum and its estimation performance and practical applications are the main concerns in this study.

The configuration of array and the basic assumptions on signal and noise are described in Section 1.1. The mathematical models of array observation and spatial covariance are formulated in Section 1.2. Examples of typical non-sparsity-exploiting methods are presented in Section 1.3. In Section 1.4, algorithms of sparse signal reconstruction and related areas are introduced. In Section 1.5, we provide a review of prior sparsity-exploiting DOA estimation methods, some of which are of interest to this dissertation.

Finally, an outline of the thesis with a brief summary of the contributions is presented in Section 1.6.

1.1 Background

Array signal processing is generally concerned with the development of parameter retrieving algorithms from data sequences which are collected by a set of spatially distinct sensors. Figure 1.1 is an example of general configuration of an array which receives signal(s) with inevitable random noise in each sensor. The positions of the sensors are specified by their coordinates (x_m, y_m, z_m), $m = 1, ..., M$, where M is the number of sensors in the array. The direction of any source is specified by its spherical coordinates (ρ, ψ), where, as illustrated in Figure 1.1, $\rho \in (-90°, 90°)$ is the elevation angle and $\psi \in (0°, 360°)$ is the azimuth angle. The signal sources may represent active sources or multipath reflectors. Therefore, the incoming signals can be correlated or even coherent. Further, the systems of array sensors often contain receivers or demodulators, which sample and convert the received continuous signals into digital sequences in base-band and facilitate the implementation of array signal processing algorithms in the succeeding digital processors. Since the focus of this dissertation is the DOA estimation algorithms, we refer the reader to [5] for more details on such conversions. In this dissertation, our formulations and derivations are based on the complex envelop representations (in discrete time) of signals and noise. The objective of DOA estimation methods is to estimate the directions of the sources, with or without knowledge of the number of sources, from the array data sequences, which are the responses to the impinging signals. DOA estimation can be applied to wireless communications, radar, sonar, radio-astronomy and many other domains.

From the theoretical point of view, some reasonable assumptions about a sensor array, signals and noise are needed for developing mathematical models and analyzing DOA estimation methods. Throughout this dissertation, the following assumptions are made:

- The sensors in an array are identical and omnidirectional. The positions of the sensors are known or measured with sufficient accuracy.
- The sources of interest are narrow-band and in the far-field with respect to the

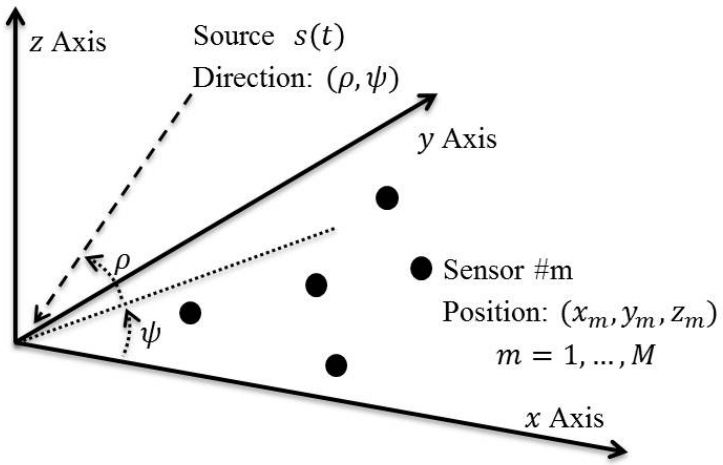


Figure 1.1: Spatial Configuration for General Array Signal Processing

physical size of an array, which means that the plane-wave model for the impinging signals is adopted. DOA estimation methods for narrow-band signals can also be applied to wide-band signals by dividing the wide-band signal into a few narrow-band ones.

- The signals passing through homogeneous wave media result in pure time-delay and attenuation of the original signals.
- The discrete time signals are temporally white and the sources are spatially stable (unless otherwise noted).
- The noise, $n(t) \in \mathbb{C}^{M\times 1}$, is Additive White Gaussian Noise (AWGN) with zero-mean and unknown power level. Further, it is uncorrelated with the sources.

Although these assumptions are necessary to our theoretical analysis, the methods proposed in this thesis can still work if some of these assumptions are invalid.

1.2 Array Observation and Spatial Covariance Models

In this section, we provide the mathematical models of an array snapshot and spatial covariance, based on the definition of the steering vector of sensor arrays. Our interest

is in narrow-band array signal processing. Therefore, all the signals, noise and array outputs are represented by their discrete-time complex envelopes, with the relative propagation delay across the different sensors parameterized by the directional information of the propagating plane waves.

The steering vector of a sensor array is the array response to the unitary signal (unit amplitude and zero phase) from a specified direction and of size $M \times 1$, where M is the number of sensors in the array. We denote the steering vector at direction $\theta = (\rho, \psi)$ as $a(\theta)$. Without loss of generality, one can assume unit gain for every sensor and, thus, the steering vector of an array with general geometry can be modeled as

$$a(\theta) = [e^{j\frac{2\pi}{\lambda_s}u_1^T v_\theta}, e^{j\frac{2\pi}{\lambda_s}u_2^T v_\theta}, ..., e^{j\frac{2\pi}{\lambda_s}u_M^T v_\theta}]^T \tag{1.1}$$

where $j = \sqrt{-1}$, λ_s is the wavelength of signal (at its carrier frequency), $v_\theta \triangleq [cos\rho \cdot cos\psi,\ cos\rho \cdot sin\psi,\ sin\rho]^T$ and, as shown in Figure 1.1, $u_m \triangleq [x_m, y_m, z_m]^T$ is the position vector of the m^{th} sensor. In model (1.1), we set the phase-reference point at the origin, $[0, 0, 0]^T$. However, in practice, array is usually only two-dimensional ($z_m = 0,\ m = 1, ..., M$) or even one-dimensional ($y_m = 0, z_m = 0,\ m = 1, ..., M$). For a two-dimensional array (planar array), one needs to restrict $\rho \geq 0$ to avoid spatial aliasing between DOAs (ρ, ψ) and $(-\rho, \psi)$. Further, one-dimensional arrays cannot distinguish the directions that have the same $cos\rho \cdot cos\psi$. Therefore, for such arrays, we define the DOA $\theta = arcsin(cos\rho \cdot cos\psi) \in (-90°, 90°)$, which is the angle between v_θ and z-y plane. If the sensors of a one-dimensional array are equally separated, the array is called Uniform Linear Array (ULA), which has been widely used in research literature and practical applications. Figure 1.2 shows an example of a 9-sensor ULA, where d_u is the inter-sensor distance. (Note that the y-Axis in Figure 1.2 represents the z-y plane of Figure 1.1.) By defining $d = d_u/\lambda_s$, the steering vector of M-sensor ULA is modeled as

$$a(\theta) = [e^{-j2\pi dsin\theta\cdot(M-1)/2}, e^{-j2\pi dsin\theta\cdot(M-3)/2}, ..., e^{j2\pi dsin\theta\cdot(M-1)/2}]^T. \tag{1.2}$$

Suppose that L signals from the directions $\theta_1, \theta_2, ..., \theta_L$ impinge on an array, and define $\mathbf{A}_x = [a(\theta_1), a(\theta_2), ..., a(\theta_L)]$. The array output at time t can be formulated as

$$y(t) = \mathbf{A}_x x(t) + n(t), \quad t = 1, ..., N, \tag{1.3}$$

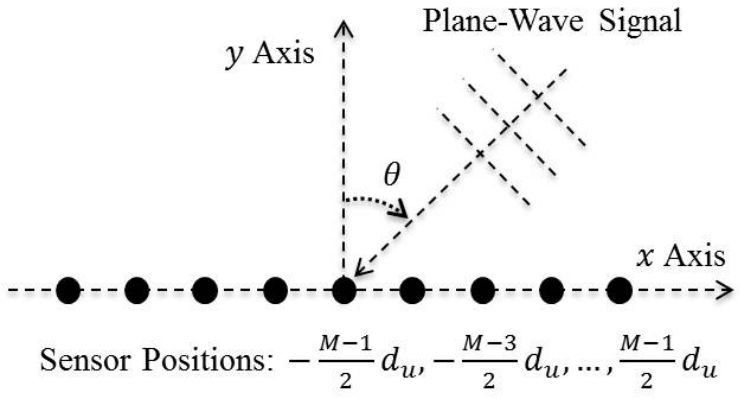


Figure 1.2: Configuration of a Uniform Linear Array

where $y(t) \in \mathbb{C}^{M\times 1}$ is the snapshot of the array output; $x(t) = [x_1(t), x_2(t), ..., x_L(t)]^T$ with $x_i(t) \in \mathbb{C}$, $i = 1, ..., L$, being the t^{th} sample of the signal from θ_i; $n(t) \in \mathbb{C}^{M\times 1}$ is the AWGN with unknown power σ^2; and N is the number of independent snapshots. Based on (1.3), the spatial covariance model can be written as

$$\bar{\mathbf{R}} \triangleq E(yy^H) = \mathbf{A}_x\mathbf{R}_x\mathbf{A}_x^H + \sigma^2\mathbf{I}_M, \tag{1.4}$$

where $E(\cdot)$ denotes the expectation operation; ^{H} is Hermitian Transpose; $\mathbf{R}_x = E(xx^H)$ is the signal covariance; and $\mathbf{I}_M$ is the identity matrix of size $M \times M$. In practice, the true spatial covariance matrix $\bar{\mathbf{R}}$ is unknown and, therefore, replaced or estimated by the sample covariance matrix

$$\mathbf{R} = \frac{1}{N}\sum_{t=1}^{N} y(t)y^H(t), \tag{1.5}$$

where $\mathbf{R}$ is a consistent estimator of $\bar{\mathbf{R}}$ and $\lim_{N\to\infty}\mathbf{R} = \bar{\mathbf{R}}$. Based on (1.3) (1.4) and (1.5), we discuss several typical and popular DOA estimators in the next section.

1.3 Examples of Direction-Of-Arrival Estimation Methods

There are three major categories of non-sparsity-exploiting DOA estimators: Beamforming methods, Signal-Subspace-based methods and Maximum Likelihood methods.

In the following subsections, we take one typical method from each of these three categories as examples of the non-sparsity-exploiting DOA estimators. The performance of these estimators will be compared with those of the proposed methods.

1.3.1 A Beamforming Method - MVDR

Beamforming methods were merely applications of Fourier-based spectral analysis to spatio-temporally sampled data when they were introduced in the second world-war. Later, adaptive beamformers were proposed to enhance the algorithms' capability of resolving closely spaced sources (so-called super-resolution) [6]. Minimum Variance Distortionless Response (MVDR) [7] is a typical adaptive beamformer and widely used in practice because of its high resolution capability, closed-form solution and low computational complexity.

The idea of MVDR is to minimize the output power/variance while keeping signals from the DOA of interest undistorted. Specifically, the weight vector of MVDR is the solution of the following optimization problem:

$$\min_{w \in \mathbb{C}^{M \times 1}} \quad w^H R w \quad s.t.\ w^H a(\phi) = 1, \tag{1.6}$$

where ϕ is the DOA of interest. The solution of (1.6) is $w = (a^H(\phi) R^{-1} a(\phi))^{-1} R^{-1} a(\phi)$, and the corresponding power spectrum is

$$p_{mvdr}(\phi) = \frac{1}{a^H(\phi) R^{-1} a(\phi)}, \tag{1.7}$$

which is evaluated for each of the DOAs of interest (or candidate DOAs). The candidate DOAs corresponding to the L largest peaks in p_{mvdr} are taken as the estimates of the true DOAs $\theta_1, \theta_2, ..., \theta_L$. One of the major concerns about MVDR is its poor DOA estimation performance, in comparison with Signal Subspace-based methods and sparsity-exploiting methods, when the number of snapshots N or Signal-to-Noise-Ratio (SNR) are small or the sources are highly correlated. Further, the resolution performance of MVDR is limited by SNR, even if the sources are uncorrelated and N is infinitely large.

1.3.2 A Signal Subspace Method - MUSIC

The MUSIC [8] method and its Signal Subspace Technique (SST) have been a major focus of DOA estimation research literature. Many methods have been proposed to utilize the SST for DOA estimation under different settings, e.g. wideband signals [9], cyclic signals [10], ULA [11] etc. The underlying idea of SST is to divide the subspace spanned by the eigenvectors of $\bar{\mathbf{R}}$ into signal and noise components based on the special structure of the noise covariance $E(nn^H) = \sigma^2\mathbf{I}_M$. Specifically, if the sources are not coherent, we have the eigenvalue decomposition $\mathbf{A}_x\mathbf{R}_x\mathbf{A}_x^H = \mathbf{V}\mathbf{D}_x\mathbf{V}^H$, where $\mathbf{D}_x = diag(\lambda_1, \lambda_2, ..., \lambda_L, 0, ..., 0)$ with $\lambda_i > 0$ and $\mathbf{V} = [v_1, ..., v_L, v_{L+1}, ..., v_M]$ with v_i being the eigenvectors. Further,

$$\bar{\mathbf{R}} = \mathbf{V} \cdot diag(\lambda_1 + \sigma^2, ..., \lambda_L + \sigma^2, \sigma^2, ..., \sigma^2) \cdot \mathbf{V}^H. \tag{1.8}$$

The subspace spanned by $v_1, ..., v_L$ is identical to that spanned by $a(\theta_1), ..., a(\theta_L)$ and referred to as signal subspace. Therefore, the noise subspace, which is spanned by $v_{L+1}, ..., v_M$, is orthogonal to $a(\theta_i)$, $i = 1, ..., L$. Under the assumption $L < M$ (so that the noise subspace exists), the MUSIC "spectrum" is formulated as a measurement of the orthogonality between $a(\phi)$ and the noise subspace for each candidate direction ϕ. Thus,

$$p_{music}(\phi) = \frac{1}{a^H(\phi)\mathbf{V}_n\mathbf{V}_n^H a(\phi)}, \tag{1.9}$$

where $\mathbf{V}_n = [v_{L+1}, ..., v_M]$. In practice, $\bar{\mathbf{R}}$ is replaced by $\mathbf{R}$ and $v_1, ..., v_L$ are identified as the eigenvectors corresponding to the L largest eigenvalues of $\mathbf{R}$. Similar to MVDR, the true DOAs are estimated by the L largest peaks in p_{music}. As shown in [12], if the sources are uncorrelated, MUSIC provides consistent and efficient (in the sense of estimation theory) estimates of the DOAs as SNR and M go to infinity.

A disadvantage of MUSIC is that it suffers from obvious degradation in DOA estimation performance when the pre-given or estimated L is not accurate. Further, MUSIC is vulnerable to correlations among the impinging sources. When some of the sources are highly correlated or even coherent, MUSIC fails because of the rank deficiency of the signal subspace. In contrast to MVDR, when N is infinitely large and the sources are uncorrelated, the resolution capability of MUSIC is unlimited no matter what SNR is [13].

1.3.3 Maximum Likelihood - Stochastic ML

Maximum Likelihood (ML) DOA estimators have been intensively studied [5][6][12]. In contrast to the ad-hoc manner of MVDR and MUSIC, ML methods directly optimize the likelihood functions by treating DOAs as unknown parameters and more fully exploiting the underlying array-data models. Generally, there are two types of ML DOA estimators: Deterministic ML (DML) and Stochastic ML (SML). In DML and SML, the signals are modeled as deterministic and Gaussian random variables respectively. Due to its better DOA estimation performance than DML, the SML described in the reference [6] is used as an example of ML estimators and as a specific point of comparison for the proposed estimators. The SML takes σ^2, $\mathbf{R}_x$ and $\theta_1, ..., \theta_L$, as parameters for minimizing its negative log-likelihood function

$$L(\theta_1, ..., \theta_L, \mathbf{R}_x, \sigma^2) = \ln |\bar{\mathbf{R}}| + tr[\bar{\mathbf{R}}^{-1}\mathbf{R}], \tag{1.10}$$

where $\bar{\mathbf{R}} = \mathbf{A}_x\mathbf{R}_x\mathbf{A}_x^H + \sigma^2\mathbf{I}_M$, $tr[\cdot]$ is the trace of the matrix and $\mathbf{R}$ is the sample covariance matrix. By minimizing this function with respect to σ^2 and $\mathbf{R}_x$, one can obtain

$$\hat{\sigma}^2 = \frac{1}{M-L} tr[(\mathbf{I}_M - \mathbf{A}_x(\mathbf{A}_x^H\mathbf{A}_x)^{-1}\mathbf{A}_x^H)\mathbf{R}], \tag{1.11}$$

and

$$\hat{\mathbf{R}}_x = (\mathbf{A}_x^H\mathbf{A}_x)^{-1}\mathbf{A}_x^H(\mathbf{R} - \hat{\sigma}^2\mathbf{I}_M)\mathbf{A}_x(\mathbf{A}_x^H\mathbf{A}_x)^{-1}. \tag{1.12}$$

Substituting (1.11) and (1.12) into (1.10), the estimates of the DOAs are given by

$$\{\hat{\theta}_1, ..., \hat{\theta}_L\} = arg\min \quad |\mathbf{A}_x\hat{\mathbf{R}}_x\mathbf{A}_x^H + \hat{\sigma}^2\mathbf{I}_M|. \tag{1.13}$$

Note that $\hat{\sigma}^2$ and $\hat{\mathbf{R}}_x$ are dependent on $\theta_1, ..., \theta_L$. Therefore, when solving the optimization problem (1.13) through exhaustive search, one needs to evaluate (1.11) and (1.12) before the objective function of (1.13) for each point in the search grid. This estimator shows better estimation performance than MVDR and MUSIC, especially for highly correlated or even coherent signals.

Although there are some fast algorithms to solve (1.13) for ULAs [6], solving (1.13) generally involves a L-dimensional search whose computational complexity can be prohibitive when the number of sources is large. Further, as shown in the previous paragraph, SML requires knowledge of L and is sensitive to errors in L.

1.4 Sparse Signal Representation

Since the methods proposed in this dissertation and many existing sparsity-exploiting DOA estimation methods are based on concepts, models and algorithms of sparse representation, we now provide a brief review of the literature of sparsity-encouraging regularization methods. Although there are many works proposing and analyzing such methods, in this section, we adopt the notations, formulations and terms of sparse signal reconstruction, e.g.[14][15], because the settings of the problems considered in this area are more similar to DOA estimation problems than the other related ones, e.g. compressive sensing. We start with the introduction of the sparse representation model and then describe the sparse signal reconstruction methods which will be used in the following chapters.

1.4.1 Sparse Representation Model

Consider the following linear equation system

$$y = \sum_{i=1}^{L} s_i b_i + n. \tag{1.14}$$

Here, $y \in \mathbb{C}^{M\times 1}$ is the observation data, $s_i \in \mathbb{C}$ are the input signals, $b_i \in \mathbb{C}^{M\times 1}$ are the so-called "atoms" and $n \in \mathbb{C}^{M\times 1}$ is the additive noise. A common problem setting of signal representation or decomposition is that:

- It is known that the atoms b_i $i = 1, ..., L$ belong to a "dictionary" (set), $\mathcal{A}$, containing K $(> L)$ atoms (vectors of size $M \times 1$).
- All the atoms in $\mathcal{A}$ are known and denoted by $a_k \in \mathbb{C}^{M\times 1}$, $k = 1, ..., K$.
- y is known and anything else in (1.14) is unknown.
- The goal is to find out which atoms of $\mathcal{A}$ are involved in this linear system and estimate the values of s_i $i = 1, ..., L$.

Under this setting, (1.14) can be reformulated as

$$y = \mathbf{A}x + n, \tag{1.15}$$

where $\mathbf{A} = [a_1, a_2, ..., a_K]$ is called observation matrix and $x = [x_1, x_2, ..., x_K]^T$ with $x_k = s_i$ if $\exists b_i = a_k$ and $x_k = 0$ otherwise. Any x satisfying the equation (1.15) is called a *representation* or *decomposition* of y on the dictionary $\mathcal{A}$. Further, the goal is translated to the estimation of x from y, which is a linear inverse problem. In this dissertation, we are only interested in the cases where $K > M$ and $a_1, a_2, ..., a_K$ form an overcomplete basis, $rank(\mathbf{A}) = M$. Such cases have found applications in many different areas, e.g. machine learning and image compression. However, the major difficulty of such a linear inverse problem is the existence of infinitely many solutions (rank deficiency of the representation in 1.15).

The desired solutions can be distinguished by various properties (or "measurements") of x, which are usually used as the objective functions, penalization or regularization terms of some optimization problems. A traditional choice of such measurements is the L_2-Norm of x, $\|x\|_2$, and the corresponding solution can be obtained by using the Moore-Penrose pseudo-inverse of $\mathbf{A}$. Further, L_2-Norm can be viewed as a special case of the Tikhonov regularization term [16], $\|\Gamma x\|_2$, where Γ is the Tikhonov matrix giving priority to solutions of desirable properties. Both $\|x\|_2$ and $\|\Gamma x\|_2$ tend to have a smoothing effect and all the elements in their solutions are generally nonzero. Another relevant measurement of x is total variation [17], $\|Dx\|_1$, where D is a gradient or derivative operator. Total variation has been widely used in image processing since it favors sparsity in gradients and thus encourages sharp edges and piecewise smoothness. However, none of these measurements promote sparse solutions, which are the true or favorable solutions in many applications. (By *sparse*, we mean that most of x_is are zero (or very small) and the number, L, of nonzero (or significant) x_is is much smaller than K. x is said to be *L-sparse*.) In Figure 1.3, we show an illustration of a linear system (1.15) with sparse x. The elements in vectors or matrices are represented by the small squares; the darker the color is, the larger the absolute value is (white=0). Thus, the target of sparsity-encouraging regularization is to find the sparsest x which satisfies the equation (1.15) with given y and $\mathbf{A}$ and "explains" y with fewest atoms.

A natural choice of a measurement of x to encourage sparsity in the solution is the so-called L_0-Norm, which is actually a pseudo-norm and defined as $\|x\|_0 = \sum_{k=1}^{K} |x_k|^0$.

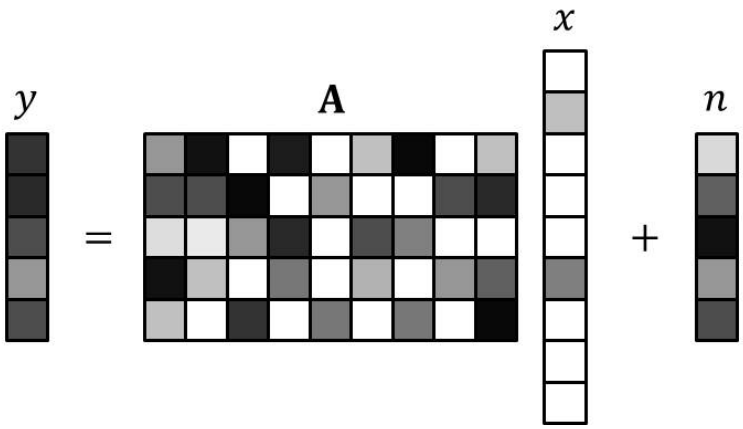


Figure 1.3: Example of Linear System with Sparse x

Obviously, $\|x\|_0$ equals the number of nonzero elements in x; thus, the following optimization problem finds the sparsest solution within its feasible region

$$\min_{x \in \mathbb{C}^{K \times 1}} \|x\|_0 \quad s.t. \ \|y - \mathbf{A}x\|_2 \leqslant \beta, \tag{1.16}$$

where β is a regularization parameter and the constraint is introduced to account for the unknown additive noise n. Although L_0-Norm and (1.16) directly aim at the sparsest solution, it is inherently a combinatorial problem and computationally intractable. Thus, there are many convex or non-convex measurements proposed to approximate the L_0-Norm with much smaller computational complexity. In the next subsection, we provide a review of the most popular approximation, L_1-Norm. Other choices can be found in [18][19][20] and the references therein.

1.4.2 L_1-Norm Regularization

L_1-Norm is defined as: $\|x\|_1 = \sum_{k=1}^{K} |x_k|$, where $|x_k|$ is the absolute value of x_k. Among all the "Norms" defined as $\|x\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + ... + |x_K|^p}$ with $p > 0$, L_1-Norm is the best convex approximation of L_0-Norm and therefore has gained popularity in promoting sparse solutions to the linear inverse problem (1.15), e.g. Compressive Sensing (CS) [21], Sparse Signal Reconstruction (SSR) [14] and LASSO [1]. The algorithms of these approaches are formulated as

$$\min_{x \in \mathbb{C}^{K \times 1}} \|x\|_1 \quad s.t. \ \|y - \mathbf{A}x\|_2 \leqslant \beta, \tag{1.17}$$

or its equivalent optimization problems, which are convex and can be efficiently solved by using either an off-the-shelf software based on the interior point method or the coordinate descent method [22].

Despite the equivalence of these algorithms, the problems faced in CS, SSR and LASSO are not exactly the same. In CS [21][23][24], it is usually assumed that $\mathbf{A}$ is the product of a random projection matrix and a basis matrix. Part of the research foci of CS are the design or the selection of these two matrices and the conditions under which L_1-Norm minimization or regularization can stably recover the sparse signal x, e.g. Restricted Isometry Property (RIP) [21]. In contrast, the literature of SSR [14][25][26] assumes the matrix $\mathbf{A}$ to be arbitrary but fixed, and analyzes the recovery performance based on some properties of $\mathbf{A}$, e.g. upper bounds on the recovery errors [14]. The works about LASSO [1][27][28] have similar problem settings as CS, but are usually applied to machine learning or classification. We are more interested in the analysis and results of SSR than CS and LASSO because the major DOA estimators proposed in this dissertation appear very similar to (1.17), and their observation matrices are determined by the array geometry and thus are non-random. However, as a specific application of SSR, the problem of DOA estimation has some special properties which are not considered in the works of SSR. Therefore, the performance of the proposed DOA estimators can be better than predicted by the theoretical results of SSR, e.g. [14]. More detailed discussions and analysis of these issues can be found in Chapter 3.

1.4.3 Group Sparsity and Group Lasso

When multiple observations (ys) are available and their corresponding representations (xs) share some common properties, one would naturally expect performance improvement by jointly estimating xs (rather than separately). An important member of such properties is the so-called *Group Sparsity*, which was first introduced in reference [29]. Although group sparsity was defined in a very general way in [29], in this subsection, we adapt the settings and formulations for Group Sparsity so that they better fit with the DOA estimation problem.

Assuming a constant observation matrix $\mathbf{A}$, one can have the model for multiple observations

$$y_t = \mathbf{A}x_t + n_t, \quad t = 1, ..., N \tag{1.18}$$

where N is the number of observations and, $y_t \in \mathbb{C}^{M\times 1}$ and $x_t \in \mathbb{C}^{K\times 1}$ denote the t^{th} observation and its representation, respectively. Further, by defining $\mathbf{Y} = [y_1, y_2, ..., y_N]$, $\mathbf{X} = [x_1, x_2, ..., x_N]$ and $\mathbf{N} = [n_1, n_2, ..., n_N]$, (1.18) can be rewritten in a compact form

$$\mathbf{Y} = \mathbf{AX} + \mathbf{N}. \tag{1.19}$$

In the cases where x_t $t = 1, ..., N$ are L-sparse and *have the same support*, matrix $\mathbf{X}$ only contains L nonzero rows. Each row of $\mathbf{X}$ is called a *group*, and thus only L out of K groups are non-trivial. Such sparsity is defined as Group Sparsity; in contrast, the sparsity discussed in Subsections 1.4.1 and 1.4.2 is called Element Sparsity. In the rest of this dissertation, the word "sparsity" is used to represent either Group Sparsity or Element Sparsity and its meaning will be clear from the context. In Figure 1.4, we show an example, with $N = 3$, to illustrate Group Sparsity.

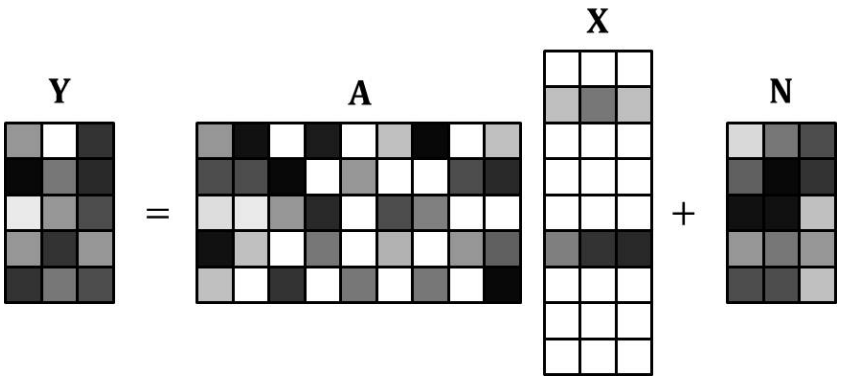


Figure 1.4: Example of Group Sparsity

To exploit the group sparsity of $\mathbf{X}$, Group LASSO [29] is formulated as

$$\min_{\mathbf{X}\in\mathbb{C}^{K\times N}} \|[r_1, r_2, ..., r_K]^T\|_1 \quad s.t.\ \|\mathbf{Y} - \mathbf{AX}\|_f \leqslant \beta,\ and\ r_k = \|\mathbf{X}_{k,:}\|_2 \tag{1.20}$$

where $\mathbf{X}_{k,:}$ is the k^{th} row of $\mathbf{X}$ and $\|\cdot\|_f$ is the Frobenius-Norm. The objective function of Group LASSO is a combination of L_1-Norm and L_2-Norm; therefore, it has two-fold effects: by minimizing the L_1-Norm of vector $[r_1, r_2, ..., r_K]^T$, Group LASSO encourages solutions with group sparsity, and by using the L_2-Norm of the groups (rows) of $\mathbf{X}$, Group LASSO tends to smooth the elements inside each active (nonzero) group. The latter effect is desirable, especially in cases where the elements inside the active

groups are of similar strengths or distributions. To efficiently solve Group LASSO, one can either transform (1.20) to a Second-Order-Cone (SOC) programming or use the specialized algorithm proposed in the reference [30]. For the performance analysis and extensions of Group LASSO, we refer the reader to the works [30][31][32][33][34] and the references therein.

1.5 Review of Sparsity-Exploiting DOA Estimators

The concept of sparsity has motivated many algorithms in spectrum and DOA estimation that explicitly emphasize or exploit either spectral or spatial sparsity. These algorithms promise high resolution capabilities in some cases with few or even a single snapshot. Generally, they can be categorized according to their (equivalent) approximation of L_0-Norm, either convex or non-convex.

Before the wide application of L_1-Norm minimization or regularization in array signal processing, there were a few sparsity-exploiting algorithms related to DOA estimation that were based on non-convex approximations of L_0-Norm. FOCUSS [35], which is an iterative weighted minimum L_2-Norm algorithm, was proposed for a general linear system inverse problem and, as an example, applied to the problem of estimating DOAs by using only one snapshot. Later in [36], it was shown that FOCUSS is an equivalent algorithm achieving local minima of L_p-Norm ($0 < p < 1$), which is a non-convex approximation of L_0-Norm. Therefore, depending on the initial "guess", FOCUSS may converge to non-sparse or non-desirable solutions. After FOCUSS was proposed, an algorithm formulated by explicitly minimizing L_p-Norm ($0 < p < 1$) was proposed in [37], along with a simplex search algorithm to find a global optimum. Although this algorithm can be straightforwardly applied to a DOA estimation problem for the one-snapshot case, the author of [37] used it to design sparse beamforming arrays. Another related algorithm (referred to as SPCG), which approximates L_0-Norm by a ln function, was proposed for estimating line-spectrum in [38]. SPCG assumes Cauchy distribution for Fourier coefficients and was solved by an iterative method [38]. Since the objective function of SPCG is non-convex, this iterative method can only achieve local minima, and thus the solution of SPCG is also dependent on the initial "guess". Similar to the above two algorithms, SPCG can only work with one-snapshot case when applied to

DOA estimation problem

The algorithms utilizing convex approximation of L_0-Norm (L_1-Norm) are of more interest to us than the non-convex ones because such algorithms are more computationally efficient, require no initial "guess" and are based on LASSO or sparse signal reconstruction algorithms whose properties have been well studied. Further, these algorithms jointly use multiple snapshots, which is very desirable under the settings of this dissertation because the signal strength is a random variable and thus can be very weak in some of the snapshots. Therefore, such algorithms can achieve better DOA estimation performance than the ones described in the last paragraph. One of the early efforts of applying L_1-Norm minimization to DOA estimation is the Global Matched Filter (GMF) [39]. GMF exploits spatial sparsity by constructing a sparse representation model in beam-space, which relies on a special property of the Uniform Circular Arrays (UCA). GMF can only work with a UCA. Since the sparse model of GMF is constructed on the basis of a sample covariance matrix, its computational complexity keeps constant with respect to the number of snapshots N (and also the number of sources L). The sparsity-exploiting DOA estimator of most interest to us is L_1-SVD [40], which establishes a sparse representation model of snapshots in element space and for arbitrary array geometry. It efficiently exploits spatial sparsity in multiple snapshots by using Singular-Value-Decomposition (SVD) to "concentrate" the data sets and Group LASSO to jointly estimate the DOAs. Although L_1-SVD can provide lower resolution threshold than MUSIC and is relatively robust with regard to highly correlated sources, it requires knowledge of the number of sources L and its computational complexity grows proportionally with L. An example of recent sparsity-exploiting DOA estimators is SPICE [3], which is based on a sparse model of sample covariance matrix and weighted L_1-Norm of spatial spectrum. SPICE can be efficiently solved by an iterative method with simple updating formulas and, more importantly, is free of regularization parameter and is robust with regard to coherent signals. However, as a price of such advantages, the spatial spectrum estimated by SPICE tends to be non-sparse. Additional examples of L_1-Norm-based DOA estimators can be [41][42][43][44].

Note that, for the underlying sparse representation models of these algorithms to be valid for the DOA estimation problem, one needs to make a few assumptions, including

- The true DOAs belong to a grid of candidate directions.

- There is no calibration error.
- The sources are not moving if the algorithm utilizes multiple snapshots.

The above assumptions will apply to the derivations and analyses in Chapter 2 - 4. Extensions of the proposed DOA estimators will be presented in Chapter 5 to incorporate off-grid DOAs and moving sources into our sparse representation models. Some other methods dealing with these issues and the related analyses can be found in [45][46][47] and the references therein.

1.6 Dissertation Outline

In this dissertation, we develop a sparse representation model for the spatial covariance in element space. This model makes possible the application of sparse signal reconstruction algorithms to DOA estimation problem. Based on the L_1-Norm regularization, we formulate two different optimization problems: one for correlated signals and the other, referred to as Sparse Spectrum Fitting (SpSF), for uncorrelated signals. SpSF is of more interest to us than the other one, since it can achieve similar estimation performance even for correlated sources (in such cases, a decorrelation method is necessary to be combined with SpSF) with much less computational complexity. Because of some special properties of the sparse model, the general theoretical analysis in sparse signal reconstruction and its related literatures is not meaningfully applicable to the proposed algorithms; therefore, some specialized performance analysis is provided in this dissertation. Further, as a regularized estimator, SpSF has a regularization parameter (RP), the value of which is critical to the estimation performance of SpSF. Although there are a few methods for selecting RPs, e.g. cross validation and discrepancy principle, they are either not computationally efficient or heavily dependent on the exact knowledge of the distributions of signals or noise. Thus, following the theoretical analysis, we propose several methods to help select the RP of SpSF that do not increase computational complexity significantly and require very limited prior information. Along with the concept of persistent block sparsity, extensions of SpSF are presented to incorporate various modeling errors into the estimators.

In Chapter 2, we develop the sparse representation model for the spatial covariance. Based on this model, a sparsity-exploiting DOA estimator is proposed for potentially

correlated or even coherent sources. Its specialized version for uncorrelated sources, which is named SpSF and is much more computationally efficient than the former one, is also presented. The application of SpSF to the cases of correlated sources is discussed, and simulation examples are provided to compare the DOA estimation performance of the proposed estimators with MUSIC, MVDR, L_1-SVD, SPICE and SML.

In Chapter 3, we analyze the conditions for the existence and uniqueness of the solution to the optimization problem of SpSF. Then, the maximum number of sources, whose spatial spectrum can be stably recovered by SpSF, is discussed and related to the degree of freedom of the co-array. Further, an explicit formula relating this maximum number to the number of sensors is developed for a uniform linear array, and SpSF's capability of resolving more than M sources by using a nonuniform linear array is discussed. On the basis of the uniqueness result, the estimation consistency of SpSF with respect to infinitely many snapshots or sensors is proven, and simulation examples are used to examine these theoretical results.

In Chapter 4, the problem of selecting a "good" RP for SpSF is considered. First, diagonal loading is utilized in an alternative formulation of SpSF to alleviate its sensitivity to the choice of RP. Then, based on the discrepancy principle, an iterative version of SpSF is proposed that automatically selects the RP for SpSF only requiring the number of sources L but increases the computational complexity of SpSF by a few times. Further, by relaxing the optimality conditions of SpSF, an automatic and adaptive RP selector is developed. This method adds minimal computational burden and only assumes very limited prior information. The effectiveness of these methods is illustrated through the comparison between the DOA estimation performance of SpSF using these methods and that of using exhaustively searched optimal RPs.

In Chapter 5, the modeling errors introduced by off-grid DOAs and moving sources are taken into consideration. We first develop a convex DOA estimator SpSF with Modeling Uncertainty (SpSFMU) for off-grid DOA problem by relying on the group sparsity between the spatial spectrum and the so-called correction parameters. Then, after introducing the concepts of Block Sparsity and Persistent Activity in a sparse representation framework, a general-purpose sparse signal reconstruction algorithm, which is based on a novel objective function, is proposed to exploit such concepts and properties. By recognizing block-sparsity and persistent activity in the sparse model of array snapshots,

this algorithm is used to estimate the DOAs of moving sources. Performances of these estimators are evaluated through simulations.

In Chapter 6, conclusions derived from the previous chapters are summarized and the discussion on future work is provided.

Chapter 2

Sparse Spectrum Fitting

The goal of this chapter is to propose and develop sparsity-exploiting DOA estimators Sparse Spectrum Fitting (SpSF) and its more general version SpSF-C (C stands for "correlated"), which are formulated for uncorrelated and correlated sources, respectively. In Section 2.1, we first develop the sparse representation model in element space for array snapshots and, then, extend it to sparse models for spatial covariance matrix. These models will serve as the basis for all the algorithms and analyses presented in this dissertation. Based on these models and L_1-Norm regularization or penalization, SpSF and SpSF-C are proposed and their advantages and disadvantages are briefly discussed in Section 2.2. Here, for simplicity of presentation, SpSF is formulated as a specialized version of SpSF-C. Although SpSF-C is more generally applicable than SpSF, SpSF is of more interest to us because its computational complexity is much smaller than that of SpSF-C. Thus, in Section 2.3 we discuss the combination of SpSF and decorrelation methods so that it can achieve similar DOA estimation performance as SpSF-C while retaining its computational advantage for highly correlated sources. In Section 2.4, simulation results are presented to evaluate the estimation performance of SpSF and SpSF-C.

2.1 Sparse Models in Signal and Spatial Covariance Domains

In this section, we construct the sparse representation models for array snapshots and the spatial covariance matrix on the basis of standard models (1.3) and (1.4). These models facilitate the application of Sparse Signal Reconstruction (SSR) algorithms to the DOA estimation problem.

For completeness, we recite the array snapshot model (1.3)

$$y(t) = \mathbf{A}_x x(t) + n(t), \quad t = 1, ..., N, \tag{2.1}$$

where $\mathbf{A}_x = [a(\theta_1), a(\theta_2), ..., a(\theta_L)]$ and $x(t) = [x_1(t), x_2(t), ..., x_L(t)]^T$. In order to take advantage of the steering vector models, most of the DOA estimation algorithms employ a grid of candidate directions, and their DOA estimates are selected from such grid. To achieve acceptable estimation accuracy, the grid has to be sufficiently dense, and hence the number of candidate directions, K, is usually much larger than the number of sources L and the number of sensors M. We denote these candidate directions as $\{\phi_1, \phi_2, ..., \phi_K\}$ and their corresponding steering vectors as $\{a(\phi_1), a(\phi_2), ..., a(\phi_K)\}$. If the true DOA θ_i belongs to this grid, $\theta_i \in \{\phi_1, \phi_2, ..., \phi_K\}$, it is called an *on-grid DOA*; otherwise it is an *off-grid DOA*. Although it is almost impossible to have on-grid DOAs since the true DOAs are continuous variables but the candidate directions are fixed and discrete, we can still make an assumption that $\{\theta_1, \theta_2, ..., \theta_L\} \subset \{\phi_1, \phi_2, ..., \phi_K\}$. This is because if the grid of candidate directions is sufficiently dense, the error terms caused by the mismatch between the true DOAs and the candidate directions are dominated by the noise contributions. Under this assumption, model (2.1) can be rewritten as

$$y(t) = \mathbf{A}_s s(t) + n(t), \quad t = 1, ..., N, \tag{2.2}$$

where $\mathbf{A}_s = [a(\phi_1), a(\phi_2), ..., a(\phi_K)]$ and $s(t) = [s_1(t), s_2(t), ..., s_K(t)]^T$ with $s_k(t) = x_i(t)$ if $\exists \theta_i = \phi_k$ and $s_k(t) = 0$ otherwise. Because the vector $s(t) \in \mathbb{C}^{K\times 1}$ only contains L non-zero elements, it can be viewed as a sparse signal and (2.2) is a sparse representation model. Further, the target of estimating the true DOAs can now be translated to finding the indices/positions of the non-zero elements in $s(t)$. In Figure 2.1, we use ULA as an example to illustrate the spatial sparse model (2.2). Please

note that if it is a planar array, the grid should be two dimensional and this model can be obtained by constructing (arbitrary) one-to-one mapping between the candidate directions and the indices.

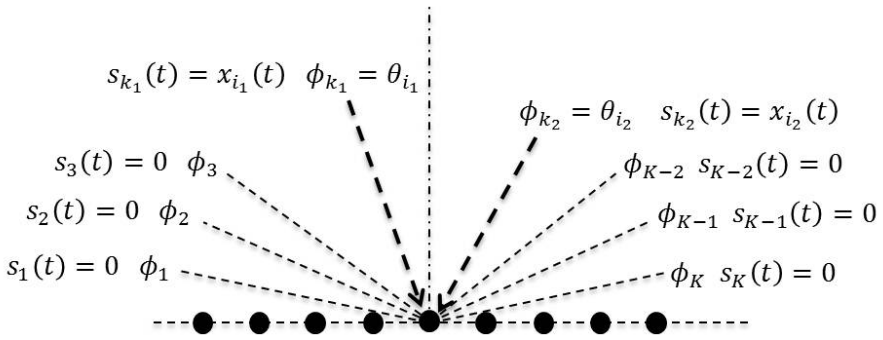


Figure 2.1: Spatial Sparse Model

Although one can apply the SSR algorithms to each snapshot separately, we prefer to jointly utilize the snapshots for better estimation performance (since the absolute value of $x_i(t)$ is a random variable, in some snapshots the signal strength is very weak) and lower computational complexity. If the observation time, N, is short enough, one can assume spatially stationary sources; therefore, one way to "simultaneously process" multiple snapshots is to estimate the DOAs from the spatial (sample) covariance matrix. Following the model (2.2), the spatial covariance matrix is defined as

$$\bar{\mathbf{R}} \triangleq E(yy^H) = \mathbf{A}_s \mathbf{R}_s \mathbf{A}_s^H + \sigma^2 \mathbf{I}_M, \tag{2.3}$$

where $\mathbf{R}_s = E(ss^H)$. According to this definition, the k^{th} diagonal element of $\mathbf{R}_s$, $\mathbf{R}_s(k,k)$, is just the power of the signal from ϕ_k, and the $(i,k)^{th}$ element, $\mathbf{R}_s(i,k)$, is the correlation between the signals from ϕ_i and ϕ_k. Therefore, by finding the positions and the values of the non-zero diagonal elements of $\mathbf{R}_s$, we can estimate the DOAs and the powers of the incoming signals, and their corresponding subdiagonal elements provide the estimates of the correlation terms. Since $\mathbf{R}_s$ is a Hermitian matrix of size $K \times K$ and contains at most L^2 non-zero elements, (2.3) is a sparse representation model of the spatial covariance matrix $\bar{\mathbf{R}}$ for (possibly) correlated sources. This model can be further simplified if the sources are known to be uncorrelated, which makes $\mathbf{R}_s$ to be

diagonal and (2.3) can be rewritten as

$$\bar{\mathbf{R}} = \sum_{k=1}^{K} p_k a(\phi_k) a^H(\phi_k) + \sigma^2 \mathbf{I}_M, \tag{2.4}$$

where $p_k = \mathbf{R}_s(k,k)$ is the power of the signal from ϕ_k. Because $p = [p_1, p_2, ..., p_K]^T$ contains only L non-zero (positive) elements, (2.4) is a sparse representation model of $\bar{\mathbf{R}}$ especially for *uncorrelated sources*.

However, the true spatial covariance $\bar{\mathbf{R}}$ is unknown and, in practice, the sample covariance matrix, $\mathbf{R}$, is used to replace it:

$$\mathbf{R} \triangleq \frac{1}{N} \sum_{t=1}^{N} y(t) y^H(t). \tag{2.5}$$

Substituting (2.1) into (2.5) gives

$$\mathbf{R} = \frac{1}{N} \mathbf{A}_x [\sum_{t=1}^{N} x(t) x^H(t)] \mathbf{A}_x^H + \frac{1}{N} \sum_{t=1}^{N} \{n(t) n^H(t) + \mathbf{A}_x x(t) n^H(t) + n(t) x^H(t) \mathbf{A}_x^H\}. \tag{2.6}$$

Since signals and noise are uncorrelated, the second term on the right-hand-side (RHS) of (2.6) converges to $\sigma^2 \mathbf{I}_M$ when $N \to \infty$, and $\lim_{N\to\infty} \mathbf{R} = \bar{\mathbf{R}}$. The signal subspace technique of the MUSIC algorithm can be further applied to remove some of the noise contributions in $\mathbf{R}$. Specifically, by denoting the eigenvalue decomposition of $\mathbf{R}$ as $\mathbf{R} = [v_1, v_2, ..., v_M] \mathbf{\Lambda} [v_1, v_2, ..., v_M]^H$, where $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_M)$ with $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_M$, the signal subspace reconstruction of $\mathbf{R}$ (assuming the knowledge of L) can be defined as:

$$\tilde{\mathbf{R}} = \sum_{i=1}^{L} \lambda_i v_i v_i^H. \tag{2.7}$$

Although both $\mathbf{R}$ and $\tilde{\mathbf{R}}$ can be used for the proposed algorithms of this dissertation and, in some cases, $\tilde{\mathbf{R}}$ can provide better estimation performance than $\mathbf{R}$, for simplicity of developments and analysis, we focus on cases where $\mathbf{R}$ is used.

2.2 SpSF Formulation

In this section, based on the models (2.3), (2.4) and (2.6), we derive and propose two sparsity-exploiting DOA estimation algorithms, SpSF-C and SpSF. These estimators are

formulated as convex optimization problems and utilize the L_1-Norm-based SSR algorithms. However, because of the underlying models (2.3) and (2.4), extra constraints are added to SpSF-C and SpSF, which make them different from ordinary SSR algorithms. Despite the better generality of SpSF-C, we will mainly focus on the formulation, analysis and extensions of SpSF since its computational complexity is much smaller than that of SpSF-C. In the following two subsections, we derive the formulations of SpSF-C and SpSF and compare their computational complexities.

2.2.1 SpSF-C for Correlated Sources

In general, the impinging signals can be arbitrarily correlated and, even if the sources are uncorrelated, very limited amount of snapshots will make them "seem to be" correlated. Therefore, in order to obtain an algorithm which can be robust with respect to source correlation and lack of snapshots, we start with the general model (2.3). Because of the unavailability of the spatial covariance matrix $\bar{\mathbf{R}}$, we need to work on the following model for the sample covariance matrix:

$$\mathbf{R} = \mathbf{A}_s \mathbf{R}_s \mathbf{A}_s^H + \mathbf{E}, \tag{2.8}$$

where the error term $\mathbf{E}$ summarizes the second term of the RHS of (2.6) and $\mathbf{R}_s = \sum_{t=1}^{N} [s(t)s^H(t)]/N$. As observed in the last section, (2.8) is a sparse representation model of $\mathbf{R}$, and the SSR algorithms can be applied to estimate $\mathbf{R}_s$ from $\mathbf{R}$. Thus, we propose to formulate the Sparse Spectrum Fitting - Correlated (SpSF-C) as

$$\begin{aligned} \min_{\mathbf{R}_s \in \mathbb{C}^{K\times K}} \quad & \|\mathbf{R} - \mathbf{A}_s \mathbf{R}_s \mathbf{A}_s^H\|_f^2 + \lambda \|\mathbf{R}_{s,v}\|_1 \\ & s.t.\ \mathbf{R}_s = \mathbf{R}_s^H \ and\ \mathbf{R}_s(k,k) \geqslant 0,\ k = 1, ..., K, \end{aligned} \tag{2.9}$$

where $vec(\cdot)$ denotes vectorization operation, $\mathbf{R}_{s,v} = vec(\mathbf{R}_s)$, $\|\cdot\|_f$ is the Frobenius norm and $\lambda > 0$ is a regularization parameter to encourage sparsity in $\mathbf{R}_s$. The larger λ is, the sparser the solution $\mathbf{R}_s$ will be. Note that the constraints in (2.9) are not included in ordinary SSR algorithms, but they are added in SpSF-C because $\mathbf{R}_s$ is a signal covariance matrix.

To efficiently solve SpSF-C, one can transform it to a Second-Order Cone Programming (SOCP). Specifically, by defining $\mathbf{R}_v = vec(\mathbf{R})$, $a_v(\phi_i, \phi_k) = vec(a(\phi_i)a^H(\phi_k))$

and

$$\mathbf{A}_{s,v} = [a_v(\phi_1,\phi_1), a_v(\phi_2,\phi_1), ..., a_v(\phi_K,\phi_1), a_v(\phi_1,\phi_2), ..., a_v(\phi_K,\phi_K)], \tag{2.10}$$

the objective function of (2.9) can be rewritten as $\|\mathbf{R}_v - \mathbf{A}_{s,v}\mathbf{R}_{s,v}\|_2^2 + \lambda\|\mathbf{R}_{s,v}\|_1$. Then, the equivalent SOCP formulation of (2.9) can be

$$\begin{aligned}
\min_{\mathbf{R}_{s,v},\tau_1,...,\tau_{K^2}} \quad & \sum_{k=1}^{K^2} \tau_k \\
s.t. \quad & \|\Re[\mathbf{R}_{s,v}(k)], \Im[\mathbf{R}_{s,v}(k)]\|_2 \leqslant \tau_k, \ k = 1, ..., K^2, \\
& \|\mathbf{R}_v - \mathbf{A}_{s,v}\mathbf{R}_{s,v}\|_2 \leqslant \beta, \\
& \Re[\mathbf{R}_{s,v}(i+(m-1)K)] = \Re[\mathbf{R}_{s,v}(m+(i-1)K)], \\
& \Im[\mathbf{R}_{s,v}(i+(m-1)K)] = -\Im[\mathbf{R}_{s,v}(m+(i-1)K)], \\
& and \ \ \Re[\mathbf{R}_{s,v}(i+(i-1)K)] \geqslant 0, \ i, m = 1, ..., K,
\end{aligned} \tag{2.11}$$

where $\mathbf{R}_{s,v}(k)$ is the k^{th} element of $\mathbf{R}_{s,v}$, $\Re[\cdot]$ and $\Im[\cdot]$ are the real and imaginary parts respectively, and $\mathbf{R}_{s,v}(i+(m-1)K) = \mathbf{R}_s(i,m)$. Here, $\beta > 0$ is a regularization parameter which has a one-to-one mapping relationship to λ of (2.9). Now, the formulation (2.11) is ready to be efficiently solved by many off-the-shelf optimization softwares. Therefore, in this dissertation we always use this formulation for SpSF-C in the simulations.

As shown by the derivation of SpSF-C, its "dictionary" contains atoms for any possible correlation terms and, by fitting with the signal covariance matrix $\mathbf{R}_s$, the solution of SpSF-C can provide estimates of DOAs, signal powers and correlation factors. Further, by looking at SpSF-C from a SSR point of view, the first term of the RHS of (2.6) is fully described by the fitting model of SpSF-C and thus viewed as "signal" in SSR, while the second term is viewed as "noise". Therefore, SpSF-C fully utilizes the power of the signals and can be expected to be robust with respect to source correlation. However, if we solve SpSF-C in a SOCP framework using an interior point method/software, its computational complexity will be $O(K^6)$, which is huge (compared to L_1-SVD $O(L^3K^3)$) since K has to be large so that the assumption $\{\theta_1, ..., \theta_L\} \subset \{\phi_1, \phi_2, ..., \phi_K\}$ can be (at least roughly) satisfied. A simple way to circumvent this disadvantage is to first use some simple estimator (e.g. MVDR) to find the areas where the sources are from and then construct $\mathbf{A}_{s,v}$ only using the candidate directions in those areas. In this way, the

number of candidate directions K and the computational complexity of SpSF-C can be significantly decreased. In Section 2.4, the DOA estimation performance of SpSF-C is compared with several other methods, and its robustness with respect to source correlations is illustrated.

2.2.2 SpSF for Uncorrelated Sources

Despite the generality of SpSF-C, its computational complexity (even if combined with the method suggested in the last paragraph) always makes it impractical for (near) real time applications. Since this complexity heavily depends on the number of optimization variables, we seek to reduce its order by considering uncorrelated sources, which eliminates the non-diagonal optimization variables in $\mathbf{R}_s$.

When the incoming signals are uncorrelated, by using (2.4), (2.8) can be simplified to

$$\mathbf{R} = \sum_{k=1}^{K} p_k a(\phi_k) a^H(\phi_k) + \mathbf{E}, \tag{2.12}$$

where $\mathbf{E}$ contains not only the second term of the RHS of (2.6) but also the correlation terms if the number of snapshots is very small. Similar to correlated sources, $\lim_{N\to\infty} \mathbf{E} = \sigma^2 \mathbf{I}_M$. Following the definitions of Subsection 2.2.1, we define $\mathbf{E}_v = vec(\mathbf{E})$, $p \triangleq diag(\mathbf{R}_s) = [p_1, p_2, ..., p_K]^T$, $a_v(\phi_k) \triangleq a_v(\phi_k, \phi_k) = vec(a(\phi_k)a^H(\phi_k))$ and

$$\mathbf{A}_K \triangleq [a_v(\phi_1), a_v(\phi_2), ..., a_v(\phi_K)]. \tag{2.13}$$

Thus (2.12) can be rewritten as

$$\mathbf{R}_v = \mathbf{A}_K p + \mathbf{E}_v. \tag{2.14}$$

Obviously, (2.14) is a sparse representation of sample covariance matrix $\mathbf{R}$ for uncorrelated sources. Therefore, we propose the Sparse Spectrum Fitting (SpSF) method as

$$\min_{p\in\mathbb{R}^{K\times 1}} \ \|\mathbf{R}_v - \mathbf{A}_K p\|_2^2 + \lambda\|p\|_1, \quad s.t. \ \ p_k \geqslant 0, \ k = 1, ..., K, \tag{2.15}$$

which is adapted from the formulation (2.9) of SpSF-C by restricting the non-diagonal elements of $\mathbf{R}_s$ to be zero. As a counterpart of (2.11), an equivalent formulation of SpSF can be

$$\min_{p\in\mathbb{R}^{K\times 1}} \ \|p\|_1, \quad s.t. \ \ \|\mathbf{R}_v - \mathbf{A}_K p\|_2 \leqslant \beta \ and \ \ p_k \geqslant 0, \ k = 1, ..., K. \tag{2.16}$$

By solving SpSF, one can obtain an estimate of the spatial spectrum, whose non-zero (or large) entries provide estimates of DOAs and signal powers. Note that, because of the constraints in (2.15) and (2.16), $\|p\|_1$ in their objective functions is equivalent to $1_K^T p$, where $1_K = [1, 1, ..., 1]^T \in \mathbb{R}^{K\times 1}$. Thus, SpSF is actually a Quadratic Programming (QP), whose computational complexity, if solved by interior point method, is $O(K^3)$ and much smaller than that of SpSF-C and L_1-SVD (for $L > 1$). Further, in contrast to L_1-SVD, this complexity does not increase with L and remains constant once K is fixed.

Although SpSF looks almost the same (except for the constraints) as the ordinary SSR algorithm, the existing SSR theories fail in predicting its estimation performance. Here, by ignoring the constraints, we apply the stability result [14] proposed for ordinary SSR algorithms to SpSF as an example to show such failure. In the reference [14], the *mutual coherence* of $\mathbf{A}_K$ is defined as

$$\mathcal{M}(\mathbf{A}_K) \triangleq \max_{1\leqslant i,k\leqslant K, i\neq k} |\mathbf{C}(i,k)|, \tag{2.17}$$

where $\mathbf{C} = \mathbf{A}_K^H \mathbf{A}_K$. Under the assumption that $L < (1+\mathcal{M}(\mathbf{A}_K)^{-1})/4$, the theoretical results of [14] provide an upper bound

$$\|p^* - p^\circ\|_2^2 \leqslant \frac{(\varepsilon+\beta)^2}{1-(4L-1)\mathcal{M}(\mathbf{A}_K)}, \tag{2.18}$$

where p^* and p° are the solution of SpSF and the true sparse spatial spectrum respectively, and $\|\mathbf{E}_v\|_2 \leqslant \varepsilon$. If we consider an 8-element ULA with $d = 0.5$ and set $\{\phi_1, \phi_2, ..., \phi_K\} = \{-80°, -79.5°, ..., 80°\}$ ($K = 321$), it can be calculated that $\mathcal{M}(\mathbf{A}_K) \approx 0.9999$. Then, the assumption becomes $L < (1+\mathcal{M}(\mathbf{A}_K)^{-1})/4 \approx 0.5$ and the bound (2.18) becomes negative for any integer $L > 0$, which indicates that SpSF cannot stably recover the spatial spectrum for even only one source. However, according to the simulation results and theoretical analysis presented in this chapter and the next chapter, SpSF can perform well with ULA for up to $M-1$ sources. Thus, the upper bound, as well as many other analyses for an ordinary SSR algorithm, fails to be meaningfully applicable to SpSF (because the constraints of SpSF are not considered), and some specialized theoretical analysis for SpSF is lacking. In Chapter 3, we present specialized theoretical analysis on the estimation performance of SpSF, specifically discussing the limit on the number of sources and its estimation consistency.

As a regularized estimator, the selection of β is critical to the performance of SpSF. The most famous and systematic methods of choosing β are Cross Validation (CV) and Discrepancy Principle (DP). With CV, a set of candidate βs is available, and one has to solve SpSF a large number of times to evaluate the "generality" for each of the βs. This process can be extremely computationally intensive and is not realistic for practical applications because β has to be updated every few seconds as the sources are moving and their powers are changing. According to DP, which has been widely used by statisticians for a long time, β should be selected such that

$$\beta \geqslant \|\mathbf{E}_v\|_2. \tag{2.19}$$

Unfortunately, in practice, $\|\mathbf{E}_v\|_2$ is unknown and its variance can be large when the number of snapshots is small. Therefore, an efficient and automatic β-selector, which relies on no or very limited prior information of the sources and noise, is highly desirable. Further, even in the simulations of the same scenario, different realizations of $\mathbf{E}$ would favor different values of β. Thus, if the selector can adaptively select β with respect to each realizations, one would expect a performance improvement over the fixed βs. More detailed discussion on the selection of β, along with several β-selectors, is provided in Chapter 4.

2.3 Application of SpSF to Correlated Sources

Although SpSF is proposed for cases of uncorrelated signals, one would still be interested in applying it to correlated cases because of its advantage in computational complexity among the sparsity-exploiting DOA estimators. In this section, we discuss the effect of source correlation on DOA estimation performance of SpSF and propose the combination of decorrelation methods and SpSF.

When the sources are correlated (or N is very small), the error term $\mathbf{E}$ of (2.12) will be

$$\mathbf{E} = \sum_{i,l=1,i\neq l}^{L} \rho_{il} a(\theta_i) a^H(\theta_l) + \frac{1}{N} \sum_{t=1}^{N} \{n(t)n^H(t) + \mathbf{A}_x x(t) n^H(t) + n(t)x^H(t)\mathbf{A}_x^H\}, \tag{2.20}$$

where $\rho_{il} = \sum_{t=1}^{N} x_i(t)x_l^*(t)/N$ is the correlation factor between signals from θ_i and θ_l, and $\|\mathbf{E}_v\|_2$ is dependent on the power of the signals. If $\mathbf{E}$ is dominated by the second

term of the RHS of (2.20), e.g. in the low SNR cases, the correlation terms can be ignored, and we can expect limited degradation of the estimation performance of SpSF. However, the problem becomes serious when $\mathbf{E}$ is dominated by the correlation terms, e.g. in high SNR or high correlation cases. In such cases, $\mathbf{E}$ is structured and cannot be sparsely represented by the dictionary $\mathbf{A}_K$. Through simulations, we have observed that, when the sources are closely spaced, their correlations will introduce significant bias to the DOA estimates of SpSF, even if the regularization parameter is chosen to give minimum estimation errors. Further, such bias does not decrease as SNR increases, so it becomes the major source of estimation errors when SNR is large. As an effect, the DOA estimation performance of SpSF reaches a "floor" as SNR becomes large enough, which is highly undesirable in most applications.

A natural way to mitigate such degradation is the preprocessing of some decorrelators on either the snapshots or spatial covariance. Here, we use ULA and Spatial Smoothing (SS) as an example to discuss the combination of SpSF and a decorrelator. SS (along with Forward-Backward Smoothing (FBS) [48]) has been studied intensively in many works, e.g. [49][50][51][52], and widely used to improve the performance of signal-subspace-based algorithms especially when the sources are highly correlated or even coherent. The basic idea of SS algorithms is to trade spatial degree-of-freedoms for source decorrelation by exploiting the special structure of the ULA steering vector. Specifically, SS algorithms divide ULA into many successive, overlapped and equal-size subarrays and then linearly combine the corresponding covariance matrices to form a "smoothed" spatial covariance matrix whose size is smaller than the original one. Further, by adaptively choosing the weights used in the combination, Weighted Spatial Smoothing (WSS) can achieve even better decorrelation performance. Some WSS algorithms, e.g. [49], optimize the weights by exploiting only the Toeplitz structure of the covariance matrix of ULA. Assuming the availability of rough estimates of the DOAs, optimization of the weights can be cast as a (robust) beamforming problem and therefore many WSS algorithms, e.g. [50][51][52], have been proposed to utilize such rough estimates and the existing beamforming algorithms. However, to achieve satisfactory suppression of the correlation terms, each WSS algorithm requires a minimum number of subarrays (which are dependent on L) and thus puts an upper bound on the size of subarrays. On the other hand, smaller subarrays decrease the resolution capability of

the DOA estimation algorithms (even if the WSS algorithm can completely decorrelate the sources) and their limits on the number of sources. Therefore, one needs to carefully select the subarray size to balance between the gain of decorrelation and the loss of reduced aperture size.

In the next section, a simulation example is used to illustrate the DOA estimation performance of SpSF combined with WSS and FBS, when the sources are closely spaced and highly correlated. As a benchmark, the performance of SpSF-C is also shown for comparison.

2.4 Simulation Results

In this section, we present simulation and experimental examples illustrating the performance of SpSF-C and SpSF for both correlated and uncorrelated sources. Through all these examples, we consider a ULA with $M = 8$, $d = 0.5$ and two equal-power sources with $\theta_1 = -5^\circ$ and $\theta_2 = 5^\circ$. SpSF and L_1-SVD always use a set of candidate directions $\{-90^\circ, -89.9^\circ, ..., 90^\circ\}$. However, in order to reduce computational complexity, the candidate directions of SpSF-C are set to $\{-90^\circ, -89^\circ, ..., 90^\circ\}$ and $\{-10^\circ, -9.9^\circ, ..., 10^\circ\}$ for SNR$\leqslant$ 0dB and SNR$>$ 0dB, respectively. The regularization parameters of SpSF, SpSF-C and L_1-SVD are *manually* selected, which means that, for each of these methods, we evaluate its Root-Mean-Squared-Error (RMSE) on many candidate regularization parameters and select the best one. (We use "(Manual)" in the figures to indicate that the corresponding methods choose their regularization parameters in such way.) Here, RMSE is defined as $\sqrt{E\{(\theta_1 - \hat{\theta}_1)^2 + (\theta_2 - \hat{\theta}_2)^2\}/2}$, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the estimates of θ_1 and θ_2 respectively. Further, $L = 2$ is assumed to be available to MUSIC and L_1-SVD.

2.4.1 Example of Spectra Estimated by SpSF-C and SpSF

In this subsection, we show examples of the spectra estimated by SpSF and SpSF-C and compare them to MUSIC, MVDR and L_1-SVD. For better illustration, the spectra are normalized such that their largest value is equal to 1.

First, we consider uncorrelated sources with $N = 300$ and SNR$= -5dB$; and the corresponding spectra are plotted in Figure 2.2. Since the sources are uncorrelated,

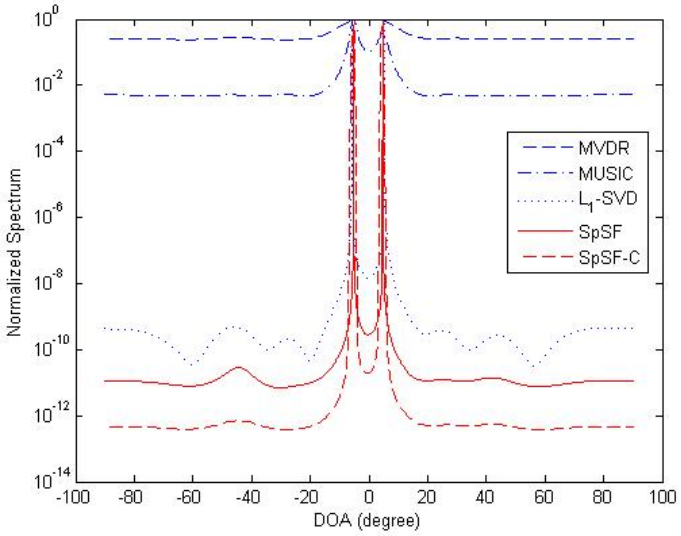


Figure 2.2: Normalized Spectra for Uncorrelated Sources

the spectrum ($\mathbf{R}_s$) estimated by SpSF-C is not shown but represented by its diagonal. As shown by Figure 2.2, MVDR cannot resolve the two sources in such cases while the other methods can clearly provide two peaks at the correct DOAs. However, the spectra estimated by SpSF, SpSF-C and L_1-SVD have sharper and narrower peaks and lower noise floors, which indicate better resolution capabilities of these sparsity-exploiting algorithms than those of MUSIC.

In the following, highly correlated sources (the correlation factor being 0.99) are considered, and N and SNR are increased to 3000 and $0dB$, respectively. In Figure 2.3, we compare the spectra estimated by the five algorithms ("SpSF-C" represents the diagonal of $\mathbf{R}_s$ estimated by SpSF-C). As shown in this case, MUSIC and MVDR fail to resolve the two sources, and the DOA estimates provided by L_1-SVD are obviously biased, which is also reported in [40]. In Figure 2.4, the two-dimensional spectra $\mathbf{R}_s$, estimated by SpSF-C, is illustrated as an image with each of its points representing the absolute value of the correlation between the sources from the corresponding DOAs. Combining Figure 2.3 and 2.4, we can see that, when the sources are highly correlated, SpSF-C can achieve better DOA estimation performance than L_1-SVD, and it can provide estimates for the correlations among the sources.

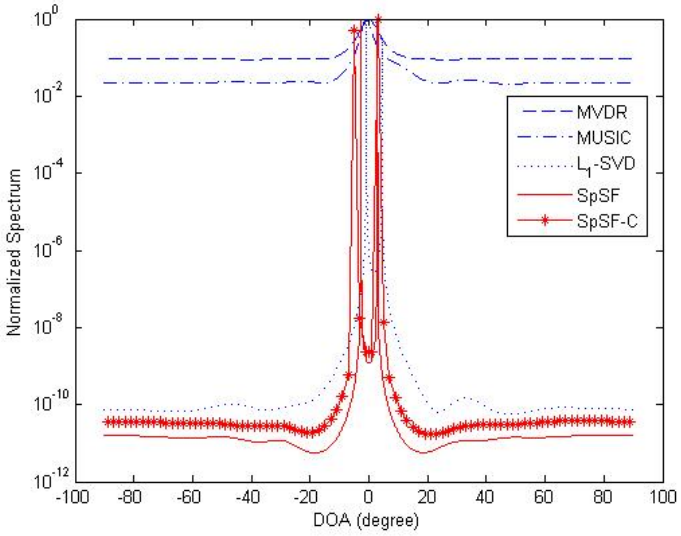


Figure 2.3: Normalized Spectra for Correlated Sources

2.4.2 Estimation Performance for Uncorrelated Sources

In this subsection, we compare the RMSE of SpSF with that of MUSIC, MVDR, L_1-SVD, stochastic Cramer-Rao Lower Bound (CRB) [12][53] and stochastic Maximum Likelihood estimator (MLE) for uncorrelated sources (since SpSF-C is especially for correlated sources and is computationally very expensive to evaluate its performance, we do not include it in this subsection but will compare it in the next subsection considering correlated sources). For each SNR between $-15dB$ and $5dB$ (with $5dB$ stepsize), 1000 independent trials are utilized to evaluate the RMSEs of these methods and, during each trial, $N = 300$ snapshots are assumed to be available. As shown in Figure 2.5, the large error thresholds of SpSF and L_1-SVD are $5dB$ lower than that of MUSIC. When SNR is large, SpSF, MUSIC and MLE become close to CRB and perform better than L_1-SVD. Further, when SNR is low, SpSF can perform similar to or even better than MLE.

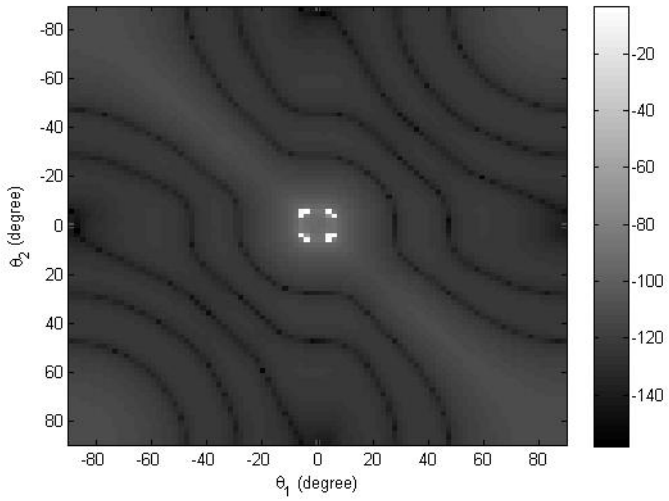


Figure 2.4: $\mathbf{R}_s$ Estimated by SpSF-C

2.4.3 Estimation Performance for Correlated Sources

In this subsection, we consider correlated sources and compare the RMSE of SpSF and SpSF-C with that of MUSIC, MVDR and L_1-SVD. We will especially show the effectiveness of WSS and FBS on mitigating the performance degradation of SpSF for source correlation. Again, the correlation coefficient is assumed to be 0.99. Five hundred independent trials are used to evaluate the RMSEs, and $N = 3000$ snapshots are assumed for each trial (we use such large number of snapshots so that the selected WSS method can well suppress the correlation terms).

We choose the WSS method proposed in [49] not only because of its simplicity but also because it does not require any prior information about the sources (e.g. L or rough estimates of the DOAs). Note that when the scenario is not so difficult, other WSS methods, such as the method of [50], may provide better suppression of the correlation terms if rough estimates of the DOAs are available. The sub-array size is set to be 5 since this WSS method requires the number of sub-arrays to be larger than $L^2 - L + 1$. The WSS method and FBS are applied to the sample covariance matrix of the full array $\mathbf{R}$, and the result is denoted by $\mathbf{R}_{SS} \in \mathbb{C}^{5\times 5}$. Since the effective correlation of the two sources represented in $\mathbf{R}_{SS}$ is reduced significantly, we apply SpSF to $\mathbf{R}_{SS}$ just as was done in Subsection 2.4.2.

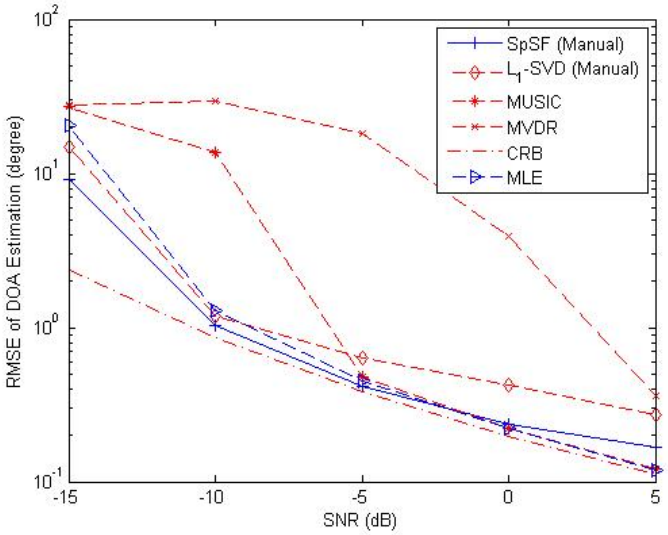


Figure 2.5: RMSE for Uncorrelated Sources

In Figure 2.6, we show the estimation performance of the aforementioned DOA estimation methods. "SpSF (Full-Array)" and "SpSF (Sub-Array)" are obtained by applying SpSF to $\mathbf{R}$ and $\mathbf{R}_{SS}$, respectively. "SpSF-C (Full-Array)" is the RMSE of SpSF-C using $\mathbf{R}$. "MUSIC (Full-Array)", "MUSIC (Sub-Array)", "MVDR (Sub-Array)" and "L_1-SVD (Full-Array)" respectively represent the RMSEs of DOA estimation of applying MUSIC to $\mathbf{R}$, MUSIC to $\mathbf{R}_{SS}$, MVDR to $\mathbf{R}_{SS}$ and L_1-SVD to the data matrix of the full array. According to Figure 2.6, the performance of MUSIC can be significantly improved by WSS and FBS, a fact that has been reported in many related works. As described in last section, signal correlation introduces bias to the DOA estimates of SpSF, which in turn makes the performance of "SpSF (Full-Array)" better than the others when SNR is low but worse when SNR is high. In addition, "SpSF (Full-Array)" reaches a floor at high SNRs (this floor is even increasing with SNR). However, if WSS and FBS are combined with SpSF ("SpSF (Sub-Array)"), its performance consistently improves along with SNR. Further, SpSF (Sub-Array) can perform much better than MUSIC and provides about $5dB$ gain over L_1-SVD because of the bias of L_1-SVD due to source correlation [40]. Not surprisingly, SpSF-C performs best among all these methods and shows about $3dB$ gain over "SpSF (Sub-Array)" (at a price of much higher

computational complexity, however).

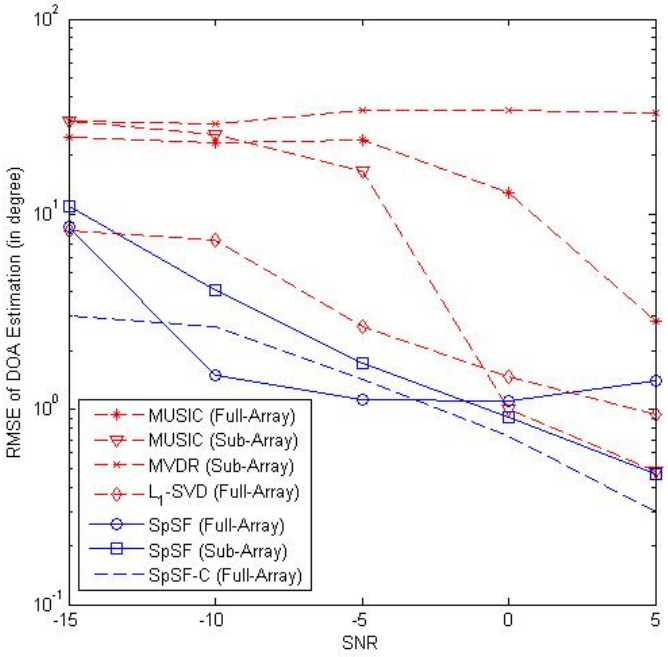


Figure 2.6: RMSE for Correlated Sources

2.4.4 Experimental Result Using Equilateral Triangular Array

In this subsection, we apply SpSF to the data set from the experiment described in [4] and present the spectra estimated by SpSF to illustrate its effectiveness when used with two-dimensional arrays and real-life data.

In this experiment, an Equilateral Triangular Array (ETA), which is depicted in Figure 2.7 (in this figure, λ is the signal wavelength), is carried by an airship and three narrow-band signals impinge on the array. More detailed description of the experiment setting can be found in [4]. Since ETA is a planar array, the DOAs are constituted of elevation and azimuth angles. The grid of candidate directions is also two-dimensional and thus K is very large. Further, because of its prohibitive computational complexity L_1-SVD is not considered in this example, and we use MUSIC for a comparison with SpSF (the grids of its candidate elevation and azimuth angles are both set as

$\{-90°, -89°, ..., 90°\}$). In Figure 2.8 and 2.9 are the spectra estimated by MUSIC and SpSF, respectively. As shown by these figures, SpSF can be effectively applied to two-dimensional arrays and retains sharp peaks, low noise floor and spatial sparsity in its spectrum.

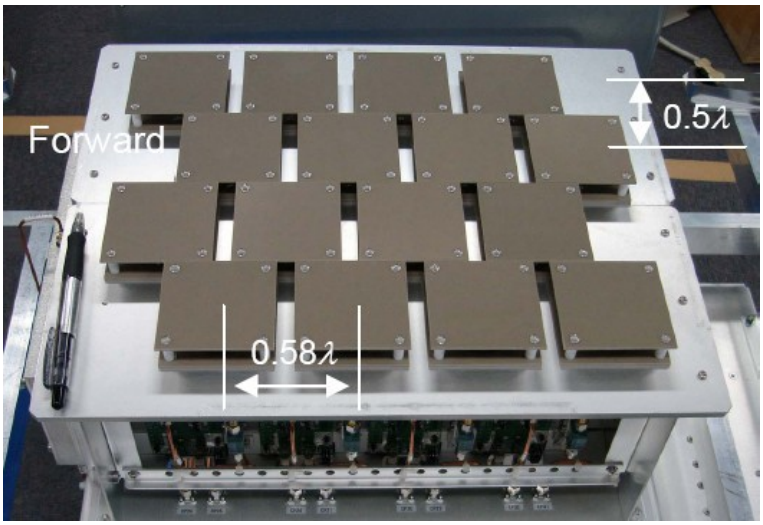


Figure 2.7: Equilateral Triangular Array Used in the Experiment

In Figure 2.10, we compare the DOA estimates of SpSF (2.16) using different values of β to that of MUSIC. As shown in this figure, when $\beta \in [0.7, 2.6]$, the DOA estimates provided by SpSF are stable and very similar to those of MUSIC. This result shows the sensitivity of SpSF with respect to its regularization parameter.

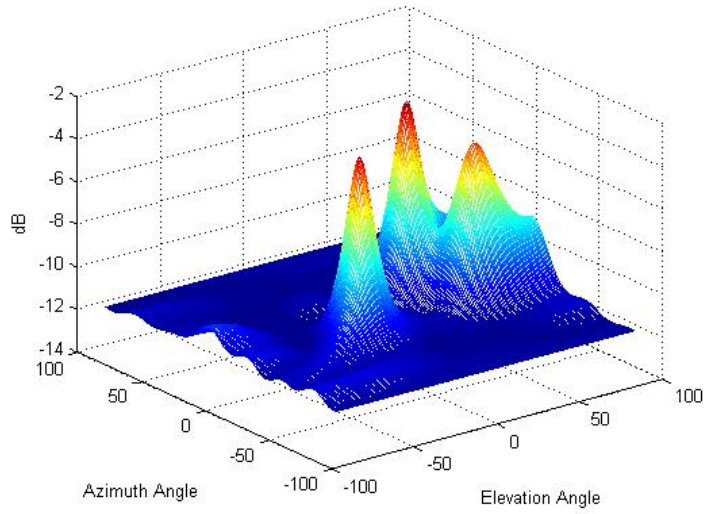


Figure 2.8: Spectrum Estimated by MUSIC

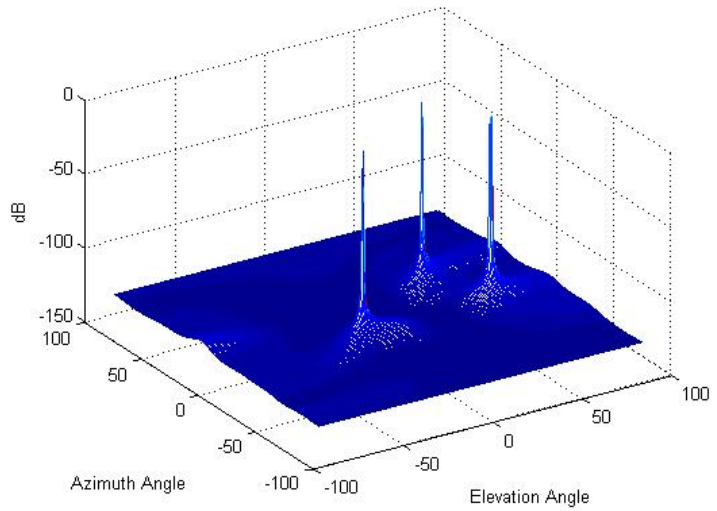


Figure 2.9: Spectrum Estimated by SpSF

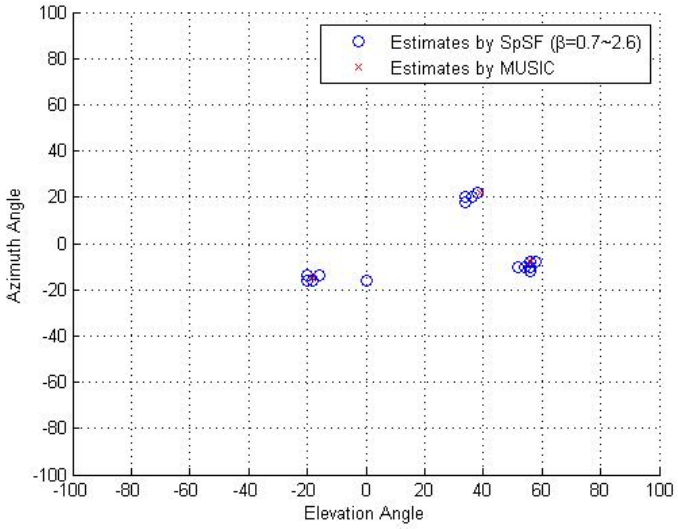


Figure 2.10: DOA Estimates of SpSF Versus MUSIC

Chapter 3

Theoretical Analysis of SpSF

In this chapter, we provide some theoretical results based on the optimality conditions of SpSF which, as described in the last chapter, are different from those of LASSO and ordinary sparse signal reconstruction algorithms. These analyses focus on the DOA estimation performance of SpSF, and the results on its power estimation performance are provided as by-products.

In Section 3.1, we begin the analyses by considering the conditions under which SpSF can achieve non-trivial solution(s). Under such conditions, SpSF generally has multiple solutions and, based on the properties of the corresponding solution set, the uniqueness of "Model Selection Consistent" (which will be defined later) solution is discussed, which is critical to the success of SpSF. Further, since the maximum number of sources whose DOAs can be reliably estimated is usually very important to the selection of DOA estimation algorithms, we provide an explicit formula for such a maximum number for SpSF based on the uniqueness result. Although this formula is dependent on the observation matrix $\mathbf{A}_K$ and is very difficult to evaluate for fixed but arbitrary array geometry (or its associated $\mathbf{A}_K$), linear arrays are exceptions. We prove that: 1) by using ULAs such maximum number of SpSF is $M-1$, which is the same maximum as MUSIC and L_1-SVD, and 2) by using non-uniform linear arrays, SpSF can successfully work with more than $M-1$ sources. The results presented in this section form the basis of the analyses and discussions in the rest of this chapter.

In Section 3.2, we present our main theoretical results: the asymptotic DOA and signal power estimation consistencies of SpSF for both infinitely many snapshots and

sensors. In the derivations, we focus on SpSF's DOA estimation performance, and the consistencies of the signal power estimates are provided as a by-product. Specifically, by considering a stronger condition of Model Selection Consistency, we first prove that, for infinitely many snapshots or sensors, there exists a Model Selection Consistent solution in the solution set of SpSF. Then, by combing such existence result with the uniqueness property of Model Selection Consistent solutions (Section 3.1), the estimation consistencies of SpSF are obtained. The analyses and derivations provided in this section are based on the optimality conditions of SpSF, which further inspire an automatic regularization parameter selector proposed in Chapter 4.

Simulation examples verifying these theoretical results are presented in Section 3.3. In Subsection 3.3.1, we show examples of SpSF successfully estimating the spatial spectrum of $M-1$ sources while failing to resolve M well-separated sources. Following that, the statistical DOA estimation performance (RMSE) of SpSF is evaluated with respect to increasing number of snapshots or sensors. These simulation results validate our theoretical results for the asymptotic consistency of SpSF.

We consider two equivalent formulations of SpSF, (2.15) and (2.16), which are represented below:

$$\min_{p\in\mathbb{R}^{K\times 1}} \quad \|\mathbf{R}_v - \mathbf{A}_K p\|_2^2 + \lambda 1_K^T p, \quad s.t. \;\; p_k \geqslant 0, \;\; k = 1, ..., K, \tag{3.1}$$

$$\min_{p\in\mathbb{R}^{K\times 1}} \quad 1_K^T p, \quad s.t. \;\; \|\mathbf{R}_v - \mathbf{A}_K p\|_2 \leqslant \beta \; and \;\; p_k \geqslant 0, \;\; k = 1, ..., K. \tag{3.2}$$

Further, we denote S_λ and S_β as the sets of the solutions of (3.1) and (3.2) respectively, where the subscripts represent the different formulations and the dependency of the solution sets on the values of the corresponding regularization parameters. As mentioned above, the analysis presented in this chapter is based on the concept of Model Selection Consistent [28], which is a term borrowed from machine learning literature and defined as:

Definition 1. [28] *A solution p^* of SpSF is* Model Selection Consistent *(MSC) if* $sign(p_i^*) = sign(p_i^\circ), \;\; i = 1, 2, ..., K,$

where $p^* = [p_1^*, p_2^*, ..., p_K^*]^T$ and $sign(x) = 1$, if $x > 0$; $sign(x) = -1$, if $x < 0$; and $sign(x) = 0$, if $x = 0$. The existence of a MSC solution of SpSF is dependent on

the selection of the regularization parameter. Specifically, for some realizations, SpSF cannot find any MSC solution no matter what value it uses for the regularization parameter. However, for the other realizations, SpSF is able to find MSC solution(s) if the regularization parameter is carefully selected. Therefore, we define:

Definition 2. *A regularization parameter β^* or λ^* of SpSF is also called MSC if and only if $\exists p^* \in S_{\beta^*}$ or S_{λ^*} such that p^* is Model Selection Consistent.*

Now, based on these definitions, our analyses and discussions can be presented.

3.1 Uniqueness of MSC Solution

In this section, the uniqueness of SpSF's MSC solutions is discussed. Through the analysis of the properties of the solution set of SpSF, the conditions guaranteeing such uniqueness is analyzed and, since SpSF generally has multiple solutions, such uniqueness is very important for the success of SpSF. In the following discussion, for simplicity of derivations, we may prove some properties for only one of (3.1) and (3.2) and use the same properties for the other formulation in some following derivations since the two formulations are equivalent.

First, we consider the existence of any optimizers (MSC or not) of SpSF. With the formulation (3.2), if $\beta \geqslant \|\mathbf{R}_v\|_2$, $p^* = 0$ is in the feasible region and thus is the only solution of SpSF, $S_\beta = \{0\}$. Further, if we define

$$\beta_{min} \triangleq \min_p \|\mathbf{R}_v - \mathbf{A}_K p\|_2, \quad s.t.\ p_i \geqslant 0, \quad i = 1, ..., K,$$

and select $\beta < \beta_{min}$, then SpSF has no solution since the feasible region is empty, $S_\beta = \varnothing$. However, if $\beta_{min} < \beta < \|\mathbf{R}_v\|_2$, SpSF has at least one solution because its objective function is lower-bounded and its feasible region is compact. In contrast, S_λ is always non-empty and equal to $\{0\}$ only when $\lambda = \infty$. Therefore, unless otherwise stated, we always assume that β (or λ) is in its interesting range where SpSF has non-zero solution(s), $\beta \in (\beta_{min}, \|\mathbf{R}_v\|_2)$ (or $\lambda < \infty$).

With the solution set being non-empty, we have:

Theorem 1. *S_β (S_λ) is a convex set.*

Proof. Suppose two solutions $p^*,\ q^* \in S_\beta$, then $\|p^*\|_1 = \|q^*\|_1 = \|\rho p^* + (1-\rho)q^*\|_1$ for $\forall \rho \in [0,1]$. At the same time,

$$\begin{aligned}\|\mathbf{R}_v - \mathbf{A}_K(\rho p^* + (1-\rho)q^*)\|_2 =& \|\rho(\mathbf{R}_v - \mathbf{A}_K p^*) + (1-\rho)(\mathbf{R}_v - \mathbf{A}_K q^*)\|_2 \\ \leqslant& \rho\|\mathbf{R}_v - \mathbf{A}_K p^*\|_2 + (1-\rho)\|\mathbf{R}_v - \mathbf{A}_K q^*\|_2 \\ \leqslant& \beta.\end{aligned}$$

Thus, we get $\rho p^* + (1-\rho)q^* \in S_\beta$. S_β is a convex set. □

Therefore, if there is any MSC solution in the solution set, the elements in (at least part of) its neighborhood also belong to the solution set. Then, one would naturally ask how the MSC solutions could be distinguished from the others in the solution set, at least from those in its neighborhood? Based on the convexity of the solution set, we can get the following theorem, which is the basis of the uniqueness of MSC solutions and makes the optimization algorithms capable of locating them.

Theorem 2. *If any N_A columns of $\mathbf{A}_K$ are linearly independent and there are two solutions $p^*,\ q^* \in S_\beta$ satisfying $p^* \neq q^*$, then $\|p^*\|_0 + \|q^*\|_0 > N_A$.*

Proof. Since the solution set of SpSF is a convex set, $p^*_\rho = \rho p^* + (1-\rho)q^*,\ \forall \rho \in [0,1]$ is also a solution of SpSF. Thus:

$$\|\mathbf{R}_v - \mathbf{A}_K p^*_\rho\|_2 = \|\mathbf{R}_v - \mathbf{A}_K p^*\|_2 = \|\mathbf{R}_v - \mathbf{A}_K q^*\|_2. \tag{3.3}$$

By (3.3):

$$\begin{aligned}\|\mathbf{R}_v - \mathbf{A}_K p^*_\rho\|_2 &= \|\rho(\mathbf{R}_v - \mathbf{A}_K p^*) + (1-\rho)(\mathbf{R}_v - \mathbf{A}_K q^*)\|_2 \\ &= \rho\|\mathbf{R}_v - \mathbf{A}_K p^*\|_2 + (1-\rho)\|\mathbf{R}_v - \mathbf{A}_K q^*\|_2.\end{aligned}$$

Then, by triangular inequality, we have

$$\rho(\mathbf{R}_v - \mathbf{A}_K p^*) = \tau(1-\rho)(\mathbf{R}_v - \mathbf{A}_K q^*)$$

with some $\tau > 0$. Combined with (3.3), $\tau = \rho/(1-\rho)$ and

$$\mathbf{R}_v - \mathbf{A}_K p^* = \mathbf{R}_v - \mathbf{A}_K q^*$$

Thus, $\mathbf{A}_K p^* = \mathbf{A}_K q^*$ which leads to

$$\mathbf{A}_K(p^* - q^*) = \mathbf{0}.$$

Since any N_A columns of $\mathbf{A}_K$ are linearly independent and $p^* - q^* \neq 0$, we have $\|p^* - q^*\|_0 > N_A$. At the same time, $\|p^* - q^*\|_0 \leqslant \|p^*\|_0 + \|q^*\|_0$, thus, $\|p^*\|_0 + \|q^*\|_0 > N_A$. □

Theorem 2 says that there is a lower limit on the sparsity (the number of non-zero elements) of any two solutions. Therefore, if there exists one solution in S_λ which is highly sparse, the other solutions will be relatively non-sparse. This gives the following corollary.

Corollary 1. *If any N_A columns of $\mathbf{A}_K$ are linearly independent and there are two solutions $p^*,\ q^* \in S_\beta$ with $\|p^*\|_0,\ \|q^*\|_0 \in (0, N_A/2]$, then $p^* = q^*$.*

Proof. Suppose $p^* \neq q^*$, then, by Theorem 2, $\|p^*\|_0 + \|q^*\|_0 > N_A$, which contradicts the assumption that $\|p^*\|_0,\ \|q^*\|_0 \in (0, N_A/2]$. Thus, $p^* = q^*$. □

By the above theoretical analysis, we get the uniqueness of the optimizers with fewer than $N_A/2$ non-zero elements. Thus, the condition for the uniqueness of MSC solution is that the true spatial spectrum p° has less than $N_A/2$ non-zero elements (or equivalently $0 < L \leqslant N_A/2$), assuming that β or λ is *MSC*. Because the optimization algorithms always reach the sparsest solution of the solution set, the MSC solution can be distinguished from the others and obtained by solving SpSF. According to Corollary 1, $N_A/2$ can be viewed as the maximum number of sources whose spatial spectrum can be stably recovered by SpSF. Note that this limit, $N_A/2$, is determined by the observation matrix $\mathbf{A}_K$, which is only dependent on the array geometry and the grid of candidate directions. Further, as shown later in this section, under very weak and common assumptions, this limit is even independent of the grid of candidate directions. Therefore, it represents the intrinsic estimation capability of SpSF.

Although it is desirable to relate such a limit to the number of sensors M, there is, unfortunately, no simple method to determine N_A for an arbitrary array geometry. Finding N_A from an arbitrary matrix $\mathbf{A}_K$ is actually a combinatorial problem and computationally very expensive, which may be worth solving through an exhaustive search if the same $\mathbf{A}_K$ is going to be used for a large number of applications. Further, if the array has regular geometry, e.g. a linear array, explicit formulas for N_A can be obtained which relate N_A to M. In the following, we first consider Uniform Linear Array (ULA) with $d \leqslant 0.5$ and then extend the result to ULA with $d > 0.5$ and Nonuniform Linear Array.

Assume a ULA with $d \leqslant 0.5$ and the candidate directions arbitrarily selected from the range $(-90°, 90°)$. The i^{th} element of its steering vector $a(\theta)$ is $e^{-j2\pi dsin\theta\cdot(M-2i+1)/2}$ for $i = 1, ..., M$. The $(i,k)^{th}$ element of $a(\theta)a^H(\theta)$ is $e^{-j2\pi dsin\theta\cdot(k-i)}$. Thus, in $a(\theta)a^H(\theta)$, the total number of distinct elements, $e^{-j2\pi dsin\theta\cdot i}$ $i = -M+1, ..., M-1$, is $2M-1$. By removing the repeated elements in $a_v(\theta)$ and re-arranging the surviving ones, $\mathbf{A}_K$ can be rewritten as

$$\mathbf{A}_K = \begin{bmatrix} e^{-j2\pi dsin\phi_1\cdot(1-M)} & e^{-j2\pi dsin\phi_2\cdot(1-M)} & \dots & e^{-j2\pi dsin\phi_K\cdot(1-M)} \\ e^{-j2\pi dsin\phi_1\cdot(2-M)} & e^{-j2\pi dsin\phi_2\cdot(2-M)} & \dots & e^{-j2\pi dsin\phi_K\cdot(2-M)} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-j2\pi dsin\phi_1\cdot(M-1)} & e^{-j2\pi dsin\phi_2\cdot(M-1)} & \dots & e^{-j2\pi dsin\phi_K\cdot(M-1)} \end{bmatrix}.$$

Observe that each column of $\mathbf{A}_K$ is exactly a discrete Fourier series of length $2M-1$ at distinct (since $d \leqslant 0.5$) frequencies $w_i = 2\pi dsin(\phi_i)$, and $\mathbf{A}_K$ itself is a discrete Fourier Transformation matrix of size $2M-1 \times K$. Therefore, any $2M-1$ columns of $\mathbf{A}_K$ are linearly independent and any $2M$ columns are linearly dependent. Thus, for ULA with $d \leqslant 0.5$, $N_A = 2M-1$ and the limit on the number of sources is $N_A/2 = M - 1/2$, which is the same limit for MUSIC and L_1-SVD. When $d > 0.5$, it is generally required to assume absence of spatial ambiguity for any DOA estimation algorithm to work correctly. This can be achieved by either restricting the range of candidate directions $(sin^{-1}(-1/2d) < \phi_1 < \phi_2 < \cdots < \phi_K < sin^{-1}(1/2d))$ or carefully selecting them so that there is no repeated spatial frequencies among w_i $i = 1, ..., K$. Based on this weak and common assumption, we can still have $N_A = 2M-1$ for ULA with $d > 0.5$, because the columns of $\mathbf{A}_K$ are still Fourier series at different frequencies. From the above derivations, it can be seen that the columns of $\mathbf{A}_K$ are actually the steering vectors of the co-array. Furthermore, for linear arrays, N_A is equivalent to the degree of freedom of the co-array. The design of Nonuniform Linear Arrays have been discussed in many works, e.g. [54][55][56][57], and increasing the degree of freedom of its co-array is one of the major research directions. Two special designs of nonuniform linear arrays, the minimum redundancy array and the optimal nested array, are discussed or proposed in [55][57] and [54] respectively. These designs can achieve $O(M^2)$ degree of freedom in their co-arrays by using M sensors. For example, if the nested array proposed in [54] is used, the co-array would be a larger Uniform Linear Array and, even with only two levels of nesting, we can achieve $N_A = M + (M^2-1)/2$ (if M is odd) or

$N_A = M + (M^2 - 2)/2$ (if M is even). Accordingly, the limit on the number of sources becomes $M^2/4 + M/2 - 1/4$ and $M^2/4 + M/2 - 1/2$ respectively. Thus, by using such nonuniform linea arrays, SpSF can stably work with many more than M sources.

3.2 Asymptotic Estimation Consistency

In the last section, we analyzed the conditions for the uniqueness of MSC solutions which enable the optimization softwares to locate the MSC solutions if there exists such solution(s) in the solution set. In this section, we address a more important question: does a MSC solution exist in the solution set? By analyzing the optimality conditions of SpSF, it is proved that, when either the number of sources N or the number of sensors M becomes infinitely large, an MSC solution always exists in the solution set and, further, such solutions can provide perfect recovery of not only the DOAs but also the signal powers. Combined with the uniqueness result of the previous section, the theoretical results of this section prove the asymptotic estimation consistency of SpSF in the sense of infinitely large N or M.

3.2.1 Asymptotic Consistency for Infinitely Large N

In this subsection, we consider the case of $N \to \infty$. First, we state our main result and then provide discussion of it.

Theorem 3. *Under the assumptions:*

1. *any N_A columns of $\mathbf{A}_K$ are linearly independent,*
2. *$\{\theta_1, \theta_2, ..., \theta_L\} \subset \{\phi_1, \phi_2, ..., \phi_K\}$, $L \leqslant N_A$, $K \gg L$ and $K > M^2$,*
3. *the sources are zero-mean, white and mutually uncorrelated,*
4. *$n(t)$ is AWGN with power σ^2,*

when $N \to \infty$ and $\lambda = 2\sigma^2 M$, the solution set of SpSF contains the true spatial spectrum: $p^\circ \in S_\lambda$.

Proof. In the following proof, both v_i or $[v]_i$ are used to denote the i^{th} entry of vector v. Further, without loss of generality, we assume $p_i^\circ > 0, 1 \leqslant i \leqslant L$ and $p_i^\circ = 0, L+1 \leqslant$

$i \leqslant K$ (since we can re-arrange the columns of $\mathbf{A}_K$). Following the techniques used in [28], we define

$$\begin{aligned} u^* = p^* - p^\circ, \;\; u = p - p^\circ \\ \mathbf{R}_\Re = \Re[\mathbf{R}_v], \;\; \mathbf{R}_\Im = \Im[\mathbf{R}_v] \\ \mathbf{A}_\Re = \Re[\mathbf{A}_K], \;\; \mathbf{A}_\Im = \Im[\mathbf{A}_K] \\ \mathbf{E}_\Re = \Re[\mathbf{E}_v], \;\; \mathbf{E}_\Im = \Im[\mathbf{E}_v]. \end{aligned}$$

where $\Re[\cdot]$ and $\Im[\cdot]$ represent the real and the imaginary part respectively. Then, SpSF (3.1) can be rewritten as:

$$p^* = arg\min_u \;\; \lambda 1_K^T(u + p^\circ) + \|\mathbf{E}_\Re - \mathbf{A}_\Re u\|_2^2 + \|\mathbf{E}_\Im - \mathbf{A}_\Im u\|_2^2 \;\; s.t. \;\; u \geqslant -p^\circ,$$

where $1_K \in \mathbb{R}^{K\times 1}$ is the vector of all 1s and the inequality between the vectors is element-wise. Before proceeding, we need to make some additional definitions:

$$p^\circ = \begin{bmatrix} p^\circ_{(1)} \\ p^\circ_{(2)} \end{bmatrix}, \;\; p^* = \begin{bmatrix} p^*_{(1)} \\ p^*_{(2)} \end{bmatrix}, \;\; u^* = \begin{bmatrix} u^*_{(1)} \\ u^*_{(2)} \end{bmatrix},$$

where $p^\circ_{(1)}$, $p^*_{(1)}$ and $u^*_{(1)}$ are the first L entries of p°, p^* and u^*, respectively, and $p^\circ_{(2)}$, $p^*_{(2)}$ and $u^*_{(2)}$ are self-explanatory. Similarly, by dividing the matrices into two parts containing the first L columns and the others respectively, we can define:

$$\begin{aligned} &\mathbf{A}_\Re = [\mathbf{A}_{\Re,(1)}, \mathbf{A}_{\Re,(2)}], \;\; \mathbf{A}_\Im = [\mathbf{A}_{\Im,(1)}, \mathbf{A}_{\Im,(2)}], \\ &\mathbf{C}_{ik} = \mathbf{A}^T_{\Re,(i)}\mathbf{A}_{\Re,(k)} + \mathbf{A}^T_{\Im,(i)}\mathbf{\Phi}_{\Im,(k)}, \;\; i,k \in \{1,2\} \\ &b_i = \mathbf{A}^T_{\Re,(i)}\mathbf{E}_\Re + \mathbf{A}^T_{\Im,(i)}\mathbf{E}_\Im, \;\; i \in \{1,2\} \\ &b = [b_1^T, b_2^T]^T \in \mathbb{R}^{K\times 1}. \end{aligned}$$

Denoting γ as the vector of Lagrangian Multipliers, by the KKT conditions, λ is *MSC* if and only if $\exists u^*$ which satisfies:

$$u^*_{(1)} > -p^\circ_{(1)} \tag{3.4a}$$

$$u^*_{(2)} = 0_{K-L} \tag{3.4b}$$

$$\mathbf{C}_{11}u^*_{(1)} - b_1 = -0.5\lambda 1_L \tag{3.4c}$$

$$\mathbf{C}_{21}u^*_{(1)} - b_2 \geqslant -0.5\lambda 1_{K-L}, \tag{3.4d}$$

where $0_{K-L}, 1_{K-L} \in \mathbb{R}^{K-L\times 1}$ are the vectors of all 0s and 1s respectively, and these equations and inequalities are defined in element-wise. Note: the second condition is trivial because there is no other restriction on $u^*_{(2)}$ and thus we can set $u^*_{(2)} = 0_{K-L}$ and drop the second condition.

Since, by assumption, any N_A columns of $\mathbf{A}_K$ are linearly independent and $L \leqslant N_A$, $\mathbf{A}_{(1)} = \mathbf{A}_{\Re,(1)} + j\mathbf{A}_{\Im,(1)}$ is of full-column-rank. Thus, $\mathbf{A}^H_{(1)}\mathbf{A}_{(1)}$ is a positive-definite Hermitian matrix and, since $\mathbf{C}_{11} = \Re[\mathbf{A}^H_{(1)}\mathbf{A}_{(1)}]$, its inverse $\mathbf{C}^{-1}_{11}$ exists. By solving (3.4c) we can get:

$$u^*_{(1)} = \mathbf{C}^{-1}_{11}b_1 - 0.5\lambda\mathbf{C}^{-1}_{11}1_L. \tag{3.5}$$

Then after substituting this solution into (3.4a) and (3.4d), the conditions in (3.4) can be further simplified to:

$$\begin{aligned} &\mathbf{C}^{-1}_{11}b_1 - 0.5\lambda\mathbf{C}^{-1}_{11}1_L > -p^\circ_{(1)}, \\ &\mathbf{C}_{21}\mathbf{C}^{-1}_{11}b_1 - b_2 \geqslant 0.5\lambda\mathbf{C}_{21}\mathbf{C}^{-1}_{11}1_L - 0.5\lambda 1_{K-L}. \end{aligned} \tag{3.6}$$

Now, we define three events:

Δ_1: λ is *MSC*,

Δ_2: $\mathbf{C}^{-1}_{11}b_1 - 0.5\lambda\mathbf{C}^{-1}_{11}1_L > -p^\circ_{(1)}$,

Δ_3: $[\mathbf{C}_{21}\mathbf{C}^{-1}_{11}, -\mathbf{I}_{K-L}]b \geqslant 0.5\lambda[\mathbf{C}_{21}\mathbf{C}^{-1}_{11}, -\mathbf{I}_{K-L}]1_K$,

where $\mathbf{I}_{K-L}$ is the identity matrix of size $K-L \times K-L$. By the conditions in (3.6), we have:

Lemma 1. *$P(\Delta_1)$=$P(\Delta_2\cap\Delta_3)$=$P(\Delta_2)$+$P(\Delta_3)$-$P(\Delta_2\cup\Delta_3)$,*

where $P(\cdot)$ stands for the possibility of the event. According to the definition of b_1 and b_2, we have:

$$E(b) = E\{\Re[\mathbf{A}^H_K\mathbf{E}_v]\} = \Re[\mathbf{A}^H_K E(\mathbf{E}_v)].$$

Under our assumptions that sources are uncorrelated and zero-mean, it is easy to check that $E(\mathbf{E}) = E(n(t)n^H(t)) = \sigma^2\mathbf{I}_M$. Thus, $E(b) = \Re[\sigma^2\mathbf{A}^H_K\mathbf{I}_{v,M}]$. Since the k^{th} column of $\mathbf{A}_K$ is $a_v(\phi_k)$, and the sensors are assumed to be unit-gain, $a^H_v(\phi_k)\mathbf{I}_{v,M} = M$. Therefore,

$$E(b) = \sigma^2 M 1_K. \tag{3.7}$$

After some straightforward but non-trivial manipulations, we can obtain

$$\lim_{N\to\infty} E(\mathbf{E}_v\mathbf{E}_v^T) = \lim_{N\to\infty} E(\mathbf{E}_v\mathbf{E}_v^H) = \sigma^4\mathbf{\Pi},$$

where

$$\mathbf{\Pi} = \Big[\mathbf{I}_{v,M}, \mathbf{0}_{M^2\times M}, \mathbf{I}_{v,M}, \mathbf{0}_{M^2\times M}, ..., \mathbf{I}_{v,M}\Big]$$

with $\mathbf{\Pi} \in \mathbb{R}^{M^2\times M^2}$ and $\mathbf{0}_{M^2\times M} \in \mathbb{R}^{M^2\times M}$ is the matrix of all 0s. With these two results in hand, we can now proceed

$$\begin{aligned} E(bb^T) &= E\Big(\Re[\mathbf{A}_K^H\mathbf{E}_v] \,.\, \Re[\mathbf{A}_K^H\mathbf{E}_v]^T\Big) \\ &= \frac{1}{2}\Re\Big[\mathbf{A}_K^H E(\mathbf{E}_v\mathbf{E}_v^H)\mathbf{A}_K + \mathbf{A}_K^H E(\mathbf{E}_v\mathbf{E}_v^T)\mathbf{A}_K^*\Big], \end{aligned}$$

where * denotes the conjugate operation. Therefore,

$$\begin{aligned} \lim_{N\to\infty} E(bb^T) &= \frac{\sigma^4}{2}\Re\Big[\mathbf{A}_K^H\mathbf{\Pi}\mathbf{A}_K + \mathbf{A}_K^H\mathbf{\Pi}\mathbf{A}_K^*\Big] \\ &= \sigma^4\Re\Big(\mathbf{A}_K^H\mathbf{\Pi}\mathbf{A}_\Re\Big) \\ &= \sigma^4\mathbf{A}_\Re^T\mathbf{\Pi}\mathbf{A}_\Re. \end{aligned} \tag{3.8}$$

As mentioned in (3.7), $\mathbf{A}_\Re^T\mathbf{I}_{v,M} = M1_K$. Thus,

$$\mathbf{A}_\Re^T\mathbf{\Pi}\mathbf{A}_\Re = M\Big[1_K, \mathbf{0}_{K\times M}, 1_K, \mathbf{0}_{K\times M}, ..., 1_K\Big]\mathbf{A}_\Re.$$

where $\mathbf{0}_{K\times M}$ is the all-zero matrix of size $K \times M$, and

$$\Big[1_K, \mathbf{0}_{K\times M}, 1_K, \mathbf{0}_{K\times M}, ..., 1_K\Big] = \Big[\mathbf{I}_{v,M}, \mathbf{I}_{v,M}, ..., \mathbf{I}_{v,M}\Big]^T.$$

Therefore, we have:

$$\mathbf{A}_\Re^T\mathbf{\Pi}\mathbf{A}_\Re = M^2 1_K 1_K^T \tag{3.9}$$

Substituting (3.9) into (3.8) gives

$$\lim_{N\to\infty} E(bb^T) = \sigma^4 M^2 1_K 1_K^T. \tag{3.10}$$

Combining with (3.7), we get:

$$\begin{aligned}\lim_{N\to\infty} Var(b) &= \lim_{N\to\infty} E\left[\left(b - E(b)\right)\left(b - E(b)\right)^T\right] \\ &= \lim_{N\to\infty} E(bb^T) - E(b)E(b^T) \\ &= \sigma^4 M^2 1_K 1_K^T - \sigma^4 M^2 1_K 1_K^T \\ &= \mathbf{0}_{K\times K}\end{aligned}$$

where $\mathbf{0}_{K\times K}$ is the all-zero matrix of size $K \times K$. Thus, when N goes to infinity, b becomes a constant vector which is $\sigma^2 M 1_K$.

So, if we set $\lambda = 2\sigma^2 M$, we can have:

$$\begin{aligned}\lim_{N\to\infty} P(\Delta_2) &= \lim_{N\to\infty} P\left(\mathbf{C}_{11}^{-1} b_1 - \frac{\lambda}{2}\mathbf{C}_{11}^{-1} 1_L > -p^\circ_{(1)}\right) \\ &= P\left(\sigma^2 M \mathbf{C}_{11}^{-1} 1_L - \sigma^2 M \mathbf{C}_{11}^{-1} 1_L > -p^\circ_{(1)}\right) \\ &= 1.\end{aligned}$$

Similarly we can show that $\lim_{N\to\infty} P(\Delta_3) = 1$. Then by Lemma 1, we prove Theorem 3: for $\lambda = 2\sigma^2 M$, $\lim_{N\to\infty} P(\Delta_1) = 1$. Further, according to (3.5), if we set $\lambda = 2\sigma^2 M$, then:

$$\begin{aligned}\lim_{N\to\infty} u^*_{(1)} &= \lim_{N\to\infty} \mathbf{C}_{11}^{-1} b_1 - \sigma^2 M \mathbf{C}_{11}^{-1} 1_L \\ &= \sigma^2 M \mathbf{C}_{11}^{-1} 1_L - \sigma^2 M \mathbf{C}_{11}^{-1} 1_L \\ &= 0_L.\end{aligned}$$

Therefore, $\lim_{N\to\infty} p^* = p^\circ$, which means that when $N \to \infty$ and $\lambda = 2\sigma^2 M$, the true spatial spectrum is in the solution set. □

Note that the *MSC* value of λ (when $N \to \infty$) only depends on the noise power σ^2 and the number of sensors M. This result agrees with intuition, for $N \to \infty$, the error term $\mathbf{E}$ converges to the covariance matrix of the noise $\sigma^2\mathbf{I}_M$. Despite the unknown mapping between λ and β, our simulations show that the corresponding *MSC* value of β (when $N \to \infty$) is $\sigma^2\sqrt{M}$. More importantly, when $N \to \infty$ and $L \leqslant N_A/2$, by Theorem 3, there is an MSC regularization parameter λ (or β) such that the true spatial spectrum p° is inside the corresponding solution set of SpSF. Further, by Corollary 1,

p° is the only MSC solution and sparsest in the solution set and thus can be obtained by solving SpSF. Now, we have proved the consistency of SpSF in both DOA and signal power estimation for an infinitely large number of snapshots.

Although the results of Theorem 3 still hold when $N_A/2 \leqslant L \leqslant N_A$, the MSC solution may no longer be the sparsest one in S_λ or S_β, and the problem of how to identify it from the solution set arises. Here, we provide brief discussion about a potential DOA estimator for the cases where $N_A/2 \leqslant L \leqslant N_A$. Defining $supp(p)$ as the support of p, we obtain a corollary from Theorem 2:

Corollary 2. *For any two solutions $p^*,\ q^* \in S_\beta$ and $p^* \neq q^*$, if $\|p^*\|_0 \leqslant N_A$ then $supp(q^*)$ is not a subset of $supp(p^*)$: $supp(q^*) \nsubseteq supp(p^*)$.*

Proof. Following the proof of Theorem 2, we have that $p^* - q^*$ has more than N_A non-zero elements. At the same time, if $supp(q^*) \subseteq supp(p^*)$, then $\|p^* - q^*\|_0 \leqslant \|p^*\|_0 \leqslant N_A$. The two statements contradict each other. □

Under the assumptions that $L \leqslant N_A$ and a MSC solution exists in S_β, Corollary 2 has the following important implication. There exists only one MSC solution in S_β. Thus, the MSC solution is a local minimizer of the following optimization problem

$$\min_p \|p\|_0 \quad s.t.\ p \in S_\beta,$$

which may be achieved by some non-convex approximations of the L_0-norm, e.g. L_g-norm with $0 < g < 1$ [18] and Capped-L_1 [19].

Another key assumption of Theorem 3 is that $\{\theta_1, ..., \theta_L\} \subset \{\phi_1, ..., \phi_K\}$, which can hardly be valid in practice because the candidate directions are discrete, but the true DOAs vary continuously. The modeling error caused by such mismatch can dramatically degrade the estimation performance of many sparsity-exploiting DOA estimation algorithms, including L_1-SVD and SpSF. We will address this problem and propose more complicated DOA estimation algorithms based on SpSF in Chapter 5.

3.2.2 Asymptotic Consistency for Infinitely Large M

In this subsection, we prove the asymptotic estimation consistency of SpSF in the sense of infinitely large number of sensors M. In contrast to Theorem 3, if the candidate

directions are kept fixed and K is finite as $M \to \infty$, the observation matrix $\mathbf{A}_K$ becomes full-column-rank and the inverse problem becomes over-determined rather than under-determined. Although when $M \to \infty$ p° is still sparse (assuming $K \gg L$), such sparsity is not necessary for the recovery of p° and, even without any regularization ($\lambda = 0$), one would still expect SpSF to perfectly recover the spatial spectrum and hence DOAs and signal powers. By the following Theorem 4, we prove such expected consistency of SpSF.

The derivations and discussions of this subsection basically follow the same guideline of Subsection 3.2.1, the analyses and notations of which are also intensively used in this subsection. For simplicity, we do not repeat the definitions but refer the readers to the last subsection, especially the proof of Theorem 3.

Theorem 4. *Based on the assumptions of Theorem 3, if the number of sources L is finite and the grid of candidate directions keeps constant with finite K, then when $M \to \infty$, any λ which is nonnegative and grows slower than M^2 ($\lambda < O(M^2)$) will have the true spatial spectrum inside its solution set, $p^\circ \in S_\lambda$.*

Proof. Again, we start with the optimality conditions of MSC solutions of SpSF,

$$u^*_{(1)} > -p^\circ_{(1)} \tag{3.11a}$$

$$\mathbf{C}_{11} u^*_{(1)} - b_1 = -0.5\lambda 1_L \tag{3.11b}$$

$$\mathbf{C}_{21} u^*_{(1)} - b_2 \geqslant -0.5\lambda 1_{K-L}. \tag{3.11c}$$

Since the $(i,j)^{th}$ element of

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

is $[\mathbf{C}]_{i,j} = \|a^H(\theta_i)a(\theta_j)\|_2^2$, where $a_v(\theta_i) = vec(a(\theta_i)a^H(\theta_i))$ is the i^{th} column of $\mathbf{A}_K$, $\mathbf{C}_{11}$ converges to $M^2\mathbf{I}_L$ ($\lim_{M\to\infty} \mathbf{C}_{11} = M^2\mathbf{I}_L$) and $\mathbf{C}_{21}$ converges to a null matrix ($\lim_{M\to\infty} \mathbf{C}_{21} = \mathbf{0}$) when M goes to infinity. Thus, as M becomes infinitely large, the optimality conditions reduce to

$$0.5\lambda 1_L < b_1 + M^2 p^\circ_{(1)} \tag{3.12a}$$

$$b_2 \leqslant 0.5\lambda 1_{K-L}, \tag{3.12b}$$

where $u^*_{(1)} = \frac{1}{M^2}(b_1 - 0.5\lambda 1_L)$. Apparently, if

$$max(b_2/M^2) < min(p^\circ_{(1)} + b_1/M^2),$$

then any λ (actually $0.5\lambda/M^2$) falling in between are MSC.

According to the definitions of b_1 and b_2, as $M \to \infty$, the k^{th} element of b_1/M^2 and b_2/M^2 become

$$[b_1/M^2]_k = \frac{1}{NM^2}\sum_{t=1}^{N} \|a^H(\theta_k)n(t)\|_2^2 + 2M\Re[x_k^*(t)a^H(\theta_k)n(t)],\ k = 1, ..., L$$

$$[b_2/M^2]_k = \frac{1}{NM^2}\sum_{t=1}^{N} \|a^H(\theta_{L+k})n(t)\|_2^2,\ k = 1, ..., K-L.$$

As $M \to \infty$, $a(\theta_i)$ and $a(\theta_j)$ with $i \neq j$ become orthogonal to each other, $a^H(\theta_i)n(t)$ and $a^H(\theta_j)n(t)$ become independent and identically distributed. Therefore, by the *law of large numbers*, for $\forall k = 1, ..., K$

$$\begin{aligned}\lim_{M\to\infty} P\left\{|\frac{1}{M}\|a^H(\theta_k)n(t)\|_2^2 - \sigma^2| < \epsilon\right\} \\ &= \lim_{M\to\infty} P\left\{\left|\frac{1}{M}\sum_{i=1}^{M}|[a^*(\theta_k)]_i[n(t)]_i|^2 - \sigma^2\right| < \epsilon\right\} \\ &= 1,\end{aligned}$$

where $\sigma^2 = E(|[n(t)]_i|^2) = E(|[a^*(\theta_k)]_i[n(t)]_i|^2)$ and ϵ is a positive but arbitrarily small number. Thus,

$$\lim_{M\to\infty}\frac{1}{NM^2}\sum_{t=1}^{N}\|a^H(\theta_k)n(t)\|_2^2 = \lim_{M\to\infty}\frac{\sigma^2}{M} = 0, \quad k = 1, ..., K,$$

$\lim_{M\to\infty} b_2/M^2 = 0$ and

$$\lim_{M\to\infty}\left[p^\circ_{(1)} + b_1/M^2\right]_k = [p^\circ_{(1)}]_k + \lim_{M\to\infty}\frac{2}{NM}\Re[x_k^*(t)a^H(\theta_k)n(t)].$$

Note that $a^H(\theta_k)n(t) = \sum_{i=1}^{M}[a^*(\theta_k)]_i[n(t)]_i$ and

$$\lim_{M\to\infty} P\left\{\left|\frac{1}{M}a^H(\theta_k)n(t)\right| < \epsilon\right\} = 1.$$

Therefore, $\lim_{M\to\infty}\left[p^{\circ}_{(1)}+b_1/M^2\right]_k=[p^{\circ}_{(1)}]_k$ and

$$0=\lim_{M\to\infty} max(b_2/M^2)<\lim_{M\to\infty} min(p^{\circ}_{(1)}+b_1/M^2)=min(p^{\circ}_{(1)}).$$

Further, as $max(2b_2)\leqslant\lambda<min(2b_1+2M^2p^{\circ}_{(1)})$, when $M\to\infty$, any λ between 0 and $2M^2min(p^{\circ}_{(1)})$ are MSC. More importantly, since $\lim_{M\to\infty} b_1/M^2=0$ and $u^*_{(1)}=\frac{1}{M^2}(b_1-0.5\lambda 1_L)$, *for any MSC* λ *which grows slower than* M^2, $\lim_{M\to\infty} p^*=p^{\circ}$. This means that, when $M\to\infty$ and $0\leqslant\lambda<O(M^2)$, the true spatial spectrum is in the solution set, $p^{\circ}\in S_{\lambda}$. □

Similar to the arguments presented after Theorem 3, we can obtain the asymptotic estimation consistency of SpSF in the sense of an infinitely large number of sensors.

As M increases, the array becomes larger, and therefore one would expect its resolution performance becomes better. Further, one would also expect larger arrays to be able to resolve larger number of sources. Thus, despite the consistency proved in Theorem 4, a more interesting case is that, while preserving the sparsity and the underdetermined assumptions, K and L also grow to infinity along with M. Such cases are studied in [28] for Lasso. In that work, it is stated that, under the Strong Irrepresentable Condition (SIC) and the conditions on the decaying rate (vs M) of the lower limit of the signal powers and the increasing rate of the upper limit of the number of sources L, Lasso is asymptotically consistent in model selection for an infinitely large number of sensors (after translating the notations of [28] to the settings of SpSF). Although SpSF has an additional constraint of positivity over Lasso, the consistency result of [28] is still applicable to SpSF since Lasso minimizes the same objective function with a larger feasible region than SpSF and any solution, which is MSC for $p^{\circ}\geqslant 0$, falls into the feasible region of SpSF. However, the SIC unfortunately cannot hold even for Uniform Linear Array, not to mention an arbitrary scenario of array geometry, candidate directions and true DOAs. Thus, to obtain similar asymptotic consistency for SpSF (if it exists), one needs to propose some other conditions. In the following, as an example, we show the failure of SIC for Uniform Linear Arrays with $L=2$ and leave more general results for future work.

First, we recite the definition of SIC in terms that are specific to the DOA estimation problem and the settings of SpSF:

Definition 3. *Strong Irrepresentable Condition* [28]*: There exists a positive constant vector* $\eta \in \mathbb{R}_{+}^{K-L\times 1}$*, s.t.*

$$\left|\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L\right| \leqslant 1_{K-L} - \eta.$$

As stated in the proof of Theorem 4, $[\mathbf{C}]_{i,j} = \|a^H(\theta_i)a(\theta_j)\|_2^2$ and, for simplicity, we denote $\delta_{ij} = \|a^H(\theta_i)a(\theta_j)\|_2^2$ with $\delta_{ii} = M^2$ and $\delta_{ij} = \delta_{ji}$. Then,

$$\mathbf{C}_{11} = \begin{bmatrix} M^2 & \delta_{12} \\ \delta_{12} & M^2 \end{bmatrix},$$

and

$$\mathbf{C}_{11}^{-1} = \begin{bmatrix} \frac{M^2}{M^4-\delta_{12}^2} & \frac{-\delta_{12}}{M^4-\delta_{12}^2} \\ \frac{-\delta_{12}}{M^4-\delta_{12}^2} & \frac{M^2}{M^4-\delta_{12}^2} \end{bmatrix}.$$

Thus, the $(k-L)^{th}$ element of $\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L$ would be

$$[\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L]_{k-L} = \frac{\delta_{1k}+\delta_{2k}}{M^2+\delta_{12}}.$$

Let's assume an 8-element ULA with $d = 0.5$ and two sources from directions -5° and 5°, and take $\theta_k = 0^\circ$. Then $M^2 = 64$, $\delta_{12} = 9.0639$, $\delta_{1k} = \delta_{2k} = 42.4346$, and $[\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L]_{k-L} = 1.162$. Thus, SIC does not hold for SpSF and its observation matrix $\mathbf{A}_K$.

3.3 Simulation Results

In this section, we present simulation examples to validate the theoretical results of this chapter. Again, we consider a ULA with $d = 0.5$. The sources are assumed to be uncorrelated and the candidate directions are set as $\{-90^\circ, -89.9^\circ, ..., 90^\circ\}$. In Subsection 3.3.1, the discussion and results on the maximum number of sources are examined and examples of spatial spectra estimated by SpSF are provided to show the success of SpSF in resolving $M-1$ sources (for ULA) and its failure with even M sources. In Subsections 3.3.2 and 3.3.3, the asymptotic consistency of SpSF is checked and simulation results of Root-Mean-Squared-Error (RMSE) varying with N and M, respectively, are presented.

3.3.1 Example of the Maximum Number of Sources

In this subsection, we choose $M = 8$ and use SpSF to estimate the spatial spectra of $L = 7$ and $L = 8$ uncorrelated sources, with SNR=$0dB$. In the $L = 7$ case, the sources are placed at $-60°$, $-40°$, $-20°$, $0°$, $20°$, $40°$ and $60°$. To show the effectiveness of the upper limit on the number of sources $L \leqslant N_A/2 = M - 1/2$, in the $L = 8$ case, we place one more source, which is still uncorrelated with the others, at $-80°$. To avoid the impact of finite sample estimates, in this subsection, we assume availability of infinitely many snapshots and therefore supply the true spatial covariance to SpSF rather than the sample covariance matrix. Further, according to Theorem 3, the regularization parameter λ of SpSF, formulation (3.1), is fixed at $2\sigma^2 M$. In Figure 3.1 and 3.2, the spatial spectra estimated by SpSF for the two cases are presented. As shown by these two figures, SpSF can very accurately estimate the spatial spectrum of $M - 1$ ($N_A/2$) uncorrelated sources. However, even with only one more source which is still uncorrelated and well-separated from the other sources, SpSF cannot correctly resolve the M sources (in Figure 3.2, there are less than M peaks).

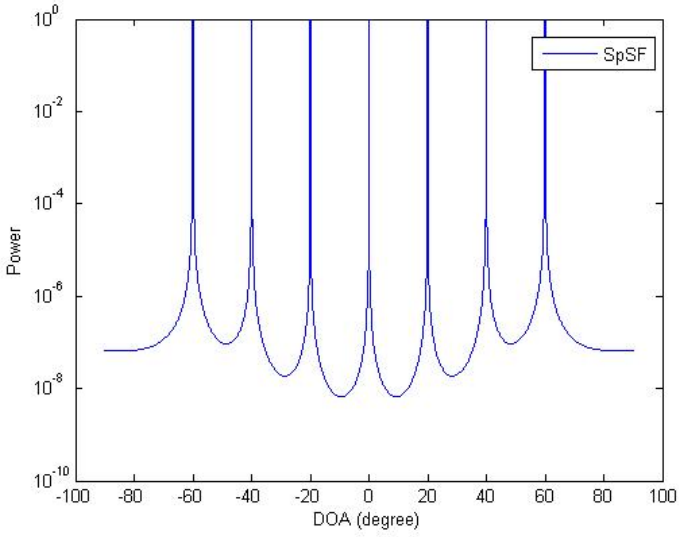


Figure 3.1: Example of an Estimated Spatial Spectrum of $M - 1$ Sources

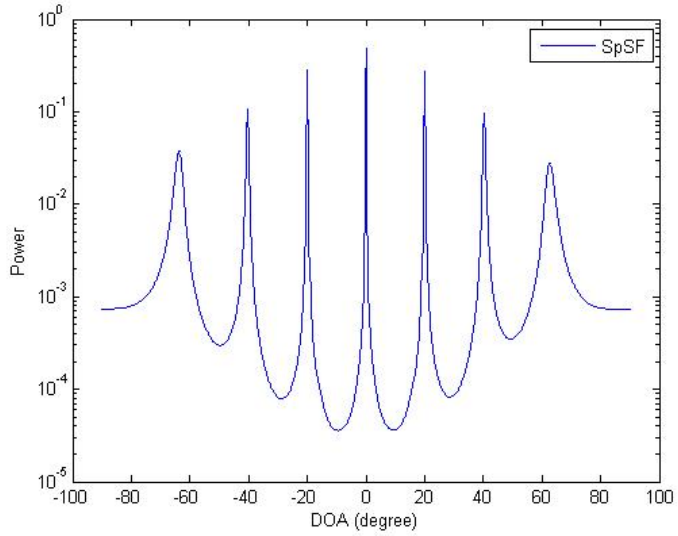


Figure 3.2: Example of an Estimated Spatial Spectrum of M Sources

3.3.2 Consistency of SpSF with Respect to N

In this subsection, we show the behavior of SpSF's RMSE when $N \to \infty$ in Figure 3.3. As in Chapter 2, we consider $M = 8$ ULA and $L = 2$ uncorrelated sources from $\theta_1 = -5^\circ$ and $\theta_2 = 5^\circ$. For each combination of SNR and N, 500 independent trials are used and the regularization parameter λ is fixed at $2\sigma^2 M$ through all these trials. In Figure 3.3, we plot the RMSE vs SNR for $N = 10, 10^2, 10^3, 10^4$ and 10^5. As shown by the figure, the performance of SpSF is monotonically increasing with N and such improvement is very consistent, especially when N is large enough. Further, when N is relatively small, the performance of SpSF at high SNRs is worse than that of SpSF at low SNRs. This is because when N is small but SNR is large, the contributions of signal correlation (due to limited number of snapshots N) can be much larger than that of noise and, thus, significantly degrade the DOA estimation performance of SpSF. In these cases, the spatial smoothing technique, discussed in Section 2.3, can be utilized to compensate for such performance loss. Due to computational complexity of calculating the sample covariance matrix, we can only show the results of finite N. The improving trend shown in Figure 3.3 implies the asymptotic DOA estimation consistency of SpSF in the sense of infinitely large number of snapshots.

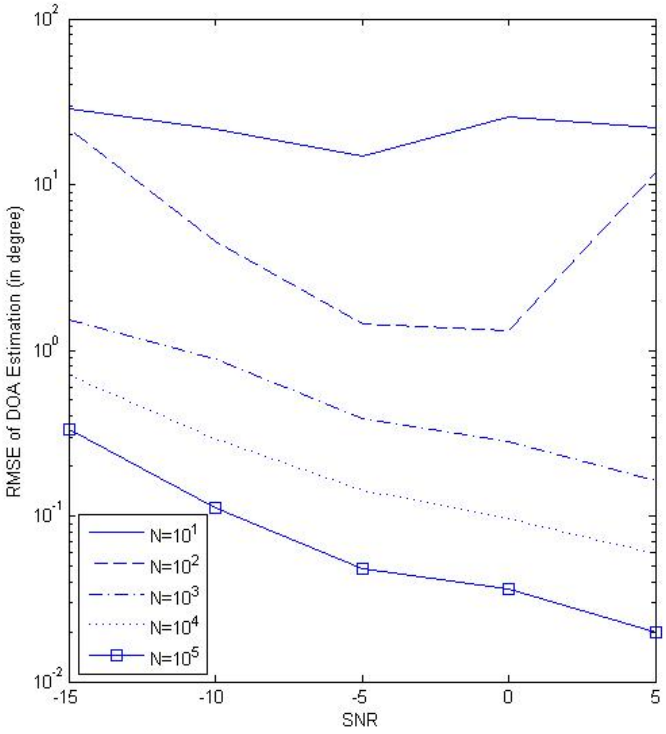


Figure 3.3: Consistency with Respect to N

3.3.3 Consistency of SpSF with Respect to M

Similar to Subsection 3.3.2, the asymptotic consistency of SpSF with respect to M is illustrated through simulation examples of SpSF's RMSE versus M, in which λ is fixed at $2\sigma^2 M$ for each M. In Figure 3.4, the RMSE performance is plotted for $M = 4, 8, 16$ and 32, where the performance improvement along with increasing M implies the asymptotic consistency. As shown by this figure, as M becomes larger, the same ratio of increase in M will lead to less improvement in estimation performance. Further, comparing Figure 3.4 with Figure 2.5, one would notice the performance loss in the case of $M = 8$. Such difference is because, in Figure 2.5, the regularization parameter of SpSF is exhaustively searched for each SNR but, in Figure 3.4, the regularization parameter is fixed at $2\sigma^2 M$. This fact suggests the need for adaptive parameter selection methods, which will be proposed in next chapter.

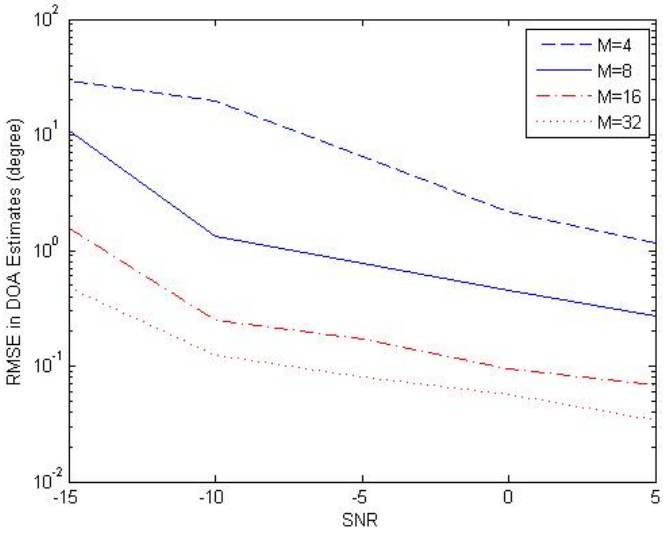


Figure 3.4: Consistency with Respect to M

Chapter 4

Regularization Parameter Selection

In the last two chapters, it has been shown that sparsity-exploiting DOA estimation algorithms, SpSF, SpSF-C and L_1-SVD, can provide significant estimation performance improvements over well-established algorithms. One of their main drawbacks, however, is the difficulty in finding the regularization parameters which can lead to good DOA estimation performances. This drawback is inherited by these algorithms from their basis, i.e. the sparse signal reconstruction or L_1-Norm regularization methods. These algorithms and methods use their regularization parameters to control the balance between the observation fitting residual and the sparsity of the estimated signals or models. Thus, their estimation performance generally depends on such balance and, hence, the selection of the regularization parameters. However, since the true signals and their sparsity measurements (L_1-Norms) are unknown, it is not clear whether the result provided by any given regularization parameter is good or not. Therefore, there is no obvious or direct way to distinguish a good regularization parameter from a bad one.

In order to circumvent such difficulties, many different principles or criteria have been proposed for selecting the parameters. A simple but very widely used one is the Discrepancy Principle, which prefers the solutions (and the corresponding parameters) with the norm of their fitting residual equal to the norm of the error terms. This principle

can be helpful only in simulations since the error terms are generally unknown in practical applications. Another widely-used regularization parameter selection method is the Cross Validation (CV), which distinguishes the good estimates (and their corresponding parameters) by evaluating their generality within the set of available observation data. Such generality is a measurement of the capability of the current estimates in predicting future observations. Since CV relies on no prior information and is generally applicable, it has been widely used in regularization-based algorithms (e.g. machine learning and sparse signal reconstruction) and shown to provide good estimation performance. Further, CV is data-adaptive, which means that it can adjust its estimates of the "best" regularization parameters according to the observation data (or ideally the realizations). However, CV requires the estimation algorithm to be solved many times for each of the candidate regularization parameters, a process which can be computationally very expensive, especially when the algorithm itself is not computationally efficient. This disadvantage of CV is not critical for off-line processing and potential performance evaluation, but it prohibits the usage of CV in many real-time applications. Thus, in this chapter, we address the problem of regularization parameter selection for SpSF and propose several methods to either alleviate the sensitivity of SpSF to the parameter or automatically and adaptively select the parameter for the purpose of DOA estimation.

In Section 4.1, using Diagonal Loading for SpSF is proposed to reduce its sensitivity to the selection of the regularization parameter for a specific formulation of SpSF. Such method is inspired by the asymptotic consistency of SpSF for an infinitely large number of snapshots, which indicates that the identity matrix, no matter how large it is scaled, in the covariance matrix does not introduce any estimation error to SpSF if its regularization parameter is set at a specific value. Thus, by adding an extra identity matrix into the sample covariance and relying on the approximately linear behavior of SpSF around the loaded covariance, this diagonal loading method can not only reduce but also enable the control of the sensitivity of SpSF to its regularization parameter.

In Section 4.2, we propose an iterative version of SpSF, Iterative SpSF (I-SpSF). This method starts with the minimum L_2-Norm fitting result and iteratively estimates the size of the error term $\mathbf{E}_v$ based on the sparsity assumption and the linear decomposition model (2.14). Then, according to the Discrepancy Principle, the value of the

estimated $\|\mathbf{E}_v\|_2$ is used as the regularization parameter for the next iteration. This method only requires the knowledge of the number of sources L and is completely free of a regularization parameter (in contrast to the Diagonal Loading method). Although this iteration process needs to solve SpSF repeatedly, it can provide satisfactory DOA estimation performance in just a few iterations.

In Section 4.3, an automatic and data-adaptive regularization parameter selector for SpSF is developed on the basis of the optimality conditions of Model Selection Consistent (MSC) and DOA Estimation Correct (DOAEC) solutions. Since any "good" regularization parameter should be data dependent and its value (or even existence) is random, we start this section by analyzing the probability of the existence of MSC regularization parameters based on the optimality conditions of corresponding solutions. This analysis leads to several inequalities describing such a probability, which, however, are very hard to analytically evaluate. We propose to approximate this probability by Monte Carlo Evaluation process. Following that, by comparing the optimality conditions of DOAEC solutions with those of MSC ones, we present an upper bound on the probability of the existence of MSC solutions. Such an upper bound, through simulations, is shown to not only approximate the probability of the existence of DOAEC solutions but also to provide an estimate of the distribution of the regularization parameters giving correct DOA estimates for any given scenarios. Based on this upper bound, we propose an automatic regularization parameter selector which, although it requires a limited amount of prior information, is data adaptive and very computationally efficient. Through simulations, this selector is shown to be robust with respect to the errors in the prior information and able to provide almost the same DOA estimation performance as the exhaustively searched and fixed regularization parameters. Further, after applying this selector (with SpSF) to the cases of highly correlated sources, its advantage of data adaptation is highlighted by its DOA estimation performance, which is not only better than any fixed regularization parameter but also the same as SpSF-C using an exhaustively search parameter.

In Section 4.4, simulation results are presented to illustrate the effectiveness, estimation performance and robustness of the methods proposed in this chapter.

4.1 Diagonal Loading

In this section, we propose to use Diagonal Loading (DL) for alleviating the sensitivity of the estimation performance of SpSF to its regularization parameter. Although this method cannot estimate the "good" values of the regularization parameter, it can extend the range of such values, inside which SpSF achieves its best estimation performance among all the possible values of the regularization parameter. Such an advantage can make any fixed value to be "good" for more scenarios and thus reduce the difficulty in parameter selection.

In order to utilize DL, the following equivalent formulation of SpSF is now considered:

$$p^* = arg\min_{p} \ \|\mathbf{R}_v - \mathbf{A}_K p\|_2 + \delta\|p\|_1, \quad s.t. \ \ p_i \geqslant 0, \quad i = 1, 2, ...K, \tag{4.1}$$

where δ is the regularization parameter to be tuned. When the number of snapshots, N, is finite, we can partition the error term $\mathbf{E} = \tilde{\mathbf{E}} + \alpha\mathbf{I}_v$ and thus

$$\mathbf{R}_v = \mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v + \alpha\mathbf{I}_v, \tag{4.2}$$

where α is positive and represents the scale of the identity matrix in $\mathbf{E}$ and $\tilde{\mathbf{E}}$ is the deviation of the estimated covariance matrix from the ideal model. If $\mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v$ can be decomposed into fewer than M columns of $\mathbf{A}_K$ (thus their summation can be viewed as $\mathbf{A}_K\tilde{p}$ where the number of non-zero terms in $\tilde{p}$ is less than M), then according to Theorem 3 presented in the previous chapter, there always exists a special regularization parameter (for the equivalent formulation (4.1)) which can perfectly recover $\tilde{p}$. Although it is almost impossible for this sparsity assumption of $\tilde{p}$ to be valid and such a special regularization parameter is a function of both σ^2 and M, we have observed through simulations that, for the formulation (4.1), if $\tilde{p}$ is truly sparse, $\delta = \sqrt{M}$ is the special value and its corresponding fitting residual is identical to $\alpha\mathbf{I}_v$. Further, if $\mathbf{A}_K p^\circ$ and $\tilde{\mathbf{E}}_v$ are fixed and δ is set at $\sqrt{M} - \epsilon$ with ϵ being positive and sufficiently small, then one can observe the following relationships between the fitting residual and varying but sufficiently large α:

$$\|\mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v - \mathbf{A}_K p^*\|_2 \cong \alpha\sqrt{M} + c_1 \tag{4.3a}$$

$$\mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v + \alpha\mathbf{I}_v - \mathbf{A}_K p^* \cong c_2\mathbf{I}_v, \tag{4.3b}$$

where c_1 and c_2 are independent from α, $c_2 < \alpha$ and p^* is the solution of

$$p^* = arg\min_p \ \|\mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v + \alpha \mathbf{I}_v - \mathbf{A}_K p\|_2 + (\sqrt{M} - \epsilon)\|p\|_1, \quad s.t. \ \ p_i \geqslant 0.$$

Such relationship should not be surprising, for when ϵ is very small, one would expect approximately linear behavior of SpSF and the equation (4.3a). When α is sufficiently large, the scaled identity matrix component of $\mathbf{E}$ becomes dominating and thus the algorithm SpSF cannot "see" any other components in $\mathbf{E}$. Therefore, one can approximately have the relationship (4.3b). Similar to the equations of (4.3), for $\delta = \sqrt{M}+\epsilon$, we have:

$$\|\mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v - \mathbf{A}_K p^*\|_2 \approxeq f_1\alpha^2 + f_2\alpha + f_3 \tag{4.4a}$$

$$\|\mathbf{A}_K p^\circ + \tilde{\mathbf{E}}_v + \alpha\mathbf{I}_v - \mathbf{A}_K p^*\|_2 \approxeq (\sqrt{M} + f_4)\alpha + f_5, \tag{4.4b}$$

where f_1, f_2, and f_4 are very small (e.g. $10^{-3} \backsim 10^{-6}$). Interestingly, if we have α in a proper range and denote $\delta_1 = \sqrt{M} + \epsilon$ and $\delta_2 = \sqrt{M} + \gamma\epsilon$ with ϵ sufficiently small and $\gamma > 0$, then

$$f_4^{(2)} \approx \gamma f_4^{(1)},$$

where $f_4^{(i)}$ is the coefficient f_4 corresponding to δ_i $i = 1, 2$. Thus, as shown by these observations, if γ is very small, then the size of the fitting residual grows very slowly with α. Therefore, by controlling γ, we can control the length of the range of "good" α and hence the sensitivity of SpSF to its parameter.

Now, we propose the following transformed formulation of SpSF which uses DL to alleviate the difficulty of finding "good" regularization parameters.

Diagonal Loading for SpSF:

$$\begin{aligned} p^* = arg\min_p \ & \|\mathbf{R}_v + \alpha\mathbf{I}_v - \mathbf{A}_K p\|_2 + (\sqrt{M} + \epsilon)\|p\|_1 \\ & s.t. \ \ p_i \geqslant 0, \quad i = 1, 2, ...K, \end{aligned} \tag{4.5}$$

where ϵ is sufficiently small and α is the parameter to adjust. Generally, ϵ could be as small as possible, and the only limit is the numerical accuracy of the computing system. Based on the Discrepancy Principle, if the norm of the residual $\|\mathbf{R}_v+\alpha\mathbf{I}_v-\mathbf{A}_K p^*\|_2$ is on the order of but larger than $\|\tilde{\mathbf{E}}+\alpha\mathbf{I}_v\|_2$, then we can expect SpSF to successfully estimate the DOAs of the sources. If we set $\delta = \sqrt{M}+\epsilon$ with ϵ sufficiently small and denote α_{min}

(α_{max}) as the smallest (largest) α that can give correct DOA estimates, then according to (4.4b), $r_\alpha = \alpha_{max} - \alpha_{min}$ can be very large because f_4 is very small. For different scenarios, such extended ranges of "good" parameters are more likely to have overlaps than that of δ of SpSF (4.1); therefore, for any given parameter value α, it is more generally "good". By using the Diagonal Loading version of SpSF, (4.5), the robustness of SpSF with respect to parameter selection is improved and, more importantly, such robustness can be enhanced by decreasing ϵ. Another advantage of DL is that there is no increase in computational complexity than the original SpSF. Simulation results illustrating the effectiveness of this method are presented in Section 4.4.

4.2 Iterative SpSF

In practice, any algorithm that is free of a regularization parameter can be favorable because it saves significant effort for tuning the parameters and avoids performance uncertainty caused by different parameter values. In this section, we propose a DOA estimator, Iterative SpSF (I-SpSF), which only requires the knowledge of the number of sources L and has no parameter to be tuned. I-SpSF repeatedly solves SpSF and updates its regularization parameter with a fixed initialization. Specifically, for each iteration, I-SpSF solves the formulation (2.16) of SpSF, which is re-presented here:

$$\min_{p\in\mathbb{R}^{K\times 1}} \ \|p\|_1, \quad s.t. \ \ \|\mathbf{R}_v - \mathbf{A}_K p\|_2 \leqslant \beta \ and \ \ p_k \geqslant 0, \ k = 1, ..., K, \tag{4.6}$$

and then updates its estimate of "good" β according to the fitting residual of the current estimate of the spatial spectrum. In what follows, we present the development of I-SpSF by starting with an approximated lower-bound for Model Selection Consistent (MSC) βs.

Suppose that we have the sample covariance matrix $\mathbf{R}$ and we know that the L uncorrelated sources are from $\theta_1, \theta_2, ..., \theta_L$. Then, if a regularization parameter β_0 is MSC, the corresponding feasible region defined by $\|\mathbf{R}_v - \mathbf{A}_K p\|_2^2 \leqslant \beta_0^2$ should have intersection with the subspace spanned by $\{a_v(\theta_1), a_v(\theta_2), ..., a_v(\theta_L)\}$. Denoting the distance between $\mathbf{R}_v$ and this subspace by d_0, then any MSC β_0 should satisfy

$$\beta_0 \geqslant d_0. \tag{4.7}$$

However, $\{\theta_i,\ i = 1, 2, ..., L\}$ are unknown and d_0 cannot be calculated. But if we have estimates $\{\phi_{t_j},\ j = 1, 2, ..., L\}$ of the directions $\{\theta_i,\ i = 1, 2, ..., L\}$, then we can approximate d_0 by the distance between $\mathbf{R}_v$ and $Span\{a_v(\phi_{t_1}), a_v(\phi_{t_2}), ..., a_v(\phi_{t_L})\}$. By denoting this approximate distance by d_T and the set $T = \{t_1, t_2, ..., t_L\}$, d_T is the minimum value of d such that the following system has solution:

$$\|\mathbf{R}_v - \mathbf{A}_K p\|_2^2 \leqslant d^2, \tag{4.8}$$

where $p_i = 0$ if $i \notin T$. If the rank of $\mathbf{A}_K$ is k, then the Singular Value Decomposition of $\mathbf{A}_K$ can be written as:

$$\mathbf{A}_K = \mathbf{U} \begin{bmatrix} \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^H, \tag{4.9}$$

where $\mathbf{\Omega} = diag(\omega_1, \omega_2, ..., \omega_k)$, with $|\omega_1| \geqslant |\omega_2| \geqslant ... \geqslant |\omega_k| > 0$. Based on (4.9), (4.8) can be transformed to:

$$\left\| \begin{bmatrix} \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^H p - q \right\|_2^2 \leqslant d^2 \tag{4.10}$$

where $q = \mathbf{U}^H \mathbf{R}_v$. By partitioning $\mathbf{V} = [\mathbf{V}_s^H, \mathbf{V}_n^H]$ and $q^H = [q_s^H, q_n^H]$, where $\mathbf{V}_s$: $k \times K$, $\mathbf{V}_n$: $(K - k) \times K$, q_s: $k \times 1$ and q_n: $(M^2 - k) \times 1$, we can further simplify (4.8) to:

$$\left\| \mathbf{\Omega}(\mathbf{V}_s p - \mathbf{\Omega}^{-1} q_s) \right\|_2^2 \leqslant d^2 - \|q_n\|_2^2, \tag{4.11}$$

where

$$\left\| \mathbf{\Omega}(\mathbf{V}_s p - \mathbf{\Omega}^{-1} q_s) \right\|_2^2 \geqslant |\omega_k|^2 \left\| \mathbf{V}_s p - \mathbf{\Omega}^{-1} q_s \right\|_2^2.$$

Thus, the d defined by the following equation will be smaller than d_T and hence a lower-bound of β_0:

$$d^2 = \|q_n\|_2^2 + \min_p |\omega_k|^2 \left\| \mathbf{V}_s p - \mathbf{\Omega}^{-1} q_s \right\|_2^2, \tag{4.12}$$

where the support of p is restricted to be a subset of T. It is easy to show that the second term of the right-hand-side of (4.12) equals:

$$\min_p |\omega_k|^2 \left\| \mathbf{V}_s p - \mathbf{\Omega}^{-1} q_s \right\|_2^2 = |\omega_k|^2 \left\| \left[\mathbf{V}_{s,T} (\mathbf{V}_{s,T}^H \mathbf{V}_{s,T})^{-1} \mathbf{V}_{s,T}^H - \mathbf{I} \right] \mathbf{\Omega}^{-1} q_s \right\|_2^2,$$

where $\mathbf{V}_{s,T}$ constitutes of the columns of $\mathbf{V}_s$ which are indexed by the elements of T. Now, we obtain the lower-bound for the MSC regularization parameter β_0:

$$\beta_0^2 \geqslant |\omega_k|^2 \cdot \left\| \left[\mathbf{V}_{s,T}(\mathbf{V}_{s,T}^H \mathbf{V}_{s,T})^{-1} \mathbf{V}_{s,T}^H - \mathbf{I} \right] \mathbf{\Omega}^{-1} q_s \right\|_2^2 + \|q_n\|_2^2. \tag{4.13}$$

Now that SpSF is used to find the estimates $\{\phi_{t_j},\ j = 1, 2, ..., L\}$ (or equivalently the set T), *Iterative SpSF* (I-SpSF) is formulated as alternating between estimating the DOAs and updating its regularization parameter β. We choose to initialize SpSF by providing rough estimates of the L DOAs, which can be generated either by some conventional algorithms, e.g. MUSIC, or by solving SpSF with $\beta = 0$. The second choice sounds more interesting to us because 1) it reduces the dependency of SpSF on any other algorithms; 2) it does not introduce uncertainty of different initialization algorithms; and 3) when SNR is not very low, SpSF with $\beta = 0$ still has competitive resolution capability and its DOA estimates will not be far away from the true DOAs (though the corresponding spatial spectrum is generally not sparse). In the following, we summarize the steps of I-SpSF

Step 1: initiate at $l = 0$ and $\beta^{(0)} = 0$

Step 2: solve SpSF with $\beta^{(l)}$ and find the L largest peaks, which give the DOA estimates $\{\phi_{t_j}^{(l)},\ j = 1, 2, ..., L\}$

Step 3: use these estimates and calculate $\beta^{(l+1)}$ by (4.13)

Step 4: if I-SpSF converges, i.e. $\beta^{(l+1)} = \beta^{(l)}$, oscillates or the maximum number of iteration is achieved, go to step 6; otherwise go to step 5

Step 5: $l = l + 1$, go back to Step 2

Step 6: if it oscillates, output the spectrum by β_{max} which is the largest β in the oscillation; otherwise, output the spectrum by $\beta^{(l)}$

Although in most instances I-SpSF can converge in just a few iterations; in some other cases, its $\beta^{(l)}$ as a function of l may show periodic behavior (oscillate) after the first 3 to 5 iterations. In such cases, $\beta^{(l)}$ oscillates between two extreme values which are denoted by β_{max} and β_{min}, and the period of such oscillation is usually

within 10 iterations. This periodic behavior of I-SpSF can be distinguished from its true convergence by keeping track of $\beta^{(l)}$ and detecting $\beta^{(l)} = \beta^{(l+\tau)}$, $l, \tau > 0$. Such oscillation is because I-SpSF updates $\beta^{(l)}$ according to the lower bound (4.13), which is conservative (meaning that, even when the DOA estimates provided in l^{th} iteration are all correct, the estimated $\beta^{(l+1)}$ is just the distance between $\mathbf{R}_v$ and the corresponding subspace and hence too small to provide truly sparse solution). Therefore, in Step 6, we choose the final β as β_{max}, which is more likely to provide sparse spatial spectrum than the other βs in the period. Further, we have observed in our simulations that I-SpSF usually converges very fast when SNR is above the resolution threshold of SpSF (using exhaustively searched but fixed regularization parameter), and the oscillation behavior only happens when SNR is around the threshold. Compared to the Diagonal Loading method of Section 4.1 and the automatic regularization parameter selector to be proposed in Section 4.3, the advantages of I-SpSF are that it is free of a regularization parameter and requires only the knowledge of the number of sources. Its disadvantage, however, is that it needs to solve SpSF for many times (though still much less than Cross Validation) and can be less favorable for some computational-complexity-sensitive applications.

4.3 Automatic Regularization Parameter Selector

In this section, we present an automatic and data-adaptive Regularization Parameter (RP) selector for SpSF (3.2) which is developed on the basis of its optimality conditions. Even if the scenario is fixed, the range of "good" RPs of SpSF is generally different for each realization, the purpose of this section is to find (possibly different) RPs, which can provide good DOA estimation performance, of SpSF with respect to each realization. Although the optimality conditions of the DOAEC solutions can describe the functional relationship between such RPs, the signal and noise samples, DOAs and array geometries, they cannot be directly used because the DOAs and signal and noise samples are unknown and such conditions cannot be evaluated if the DOAEC solution itself is unknown. Therefore, in order to obtain such a selector, in Subsection 4.3.1, we first analyze the optimality conditions of MSC solutions of SpSF which are further translated to several inequalities defining the range of "good" RPs for any given

realization. Following that, a Monte Carlo evaluation process is proposed to efficiently evaluate the probability of the existence of MSC solutions, which will also be utilized for evaluation of the upper bound and the automatic selector. Then, in Subsection 4.3.2, by comparing the optimality conditions of MSC solutions with those of DOAEC ones, these inequalities are relaxed and an upper bound on the probabilities of the existence of MSC RPs is obtained. On the basis of this upper bound and the Monte Carlo evaluation process, our automatic selector is presented in Subsection 4.3.3. This automatic selector is computationally efficient and requires limited amount of information.

In this section, we consider SpSF in the formulation (3.2), which is re-presented here for reader's reference:

$$\min_{p\in\mathbb{R}^{K\times 1}} \quad \|\mathbf{R}_v - \mathbf{A}_K p\|_2^2 + \lambda\|p\|_1, \quad s.t. \;\; p_k \geqslant 0, \;\; k = 1, ..., K. \tag{4.14}$$

Further, for simplicity of description, we define that

Definition 4. P_{MSC}: *the probability that, no matter what regularization parameter it uses, SpSF can find a Model Selection Consistent solution for any given but fixed scenario (including DOAs, array geometries, signal and noise distributions and etc.),*

and

Definition 5. P_θ: *the probability that, no matter what regularization parameter it uses, SpSF can perfectly estimate the DOAs for any given but fixed scenario.*

Here, we note that perfect estimation of the DOAs is only possible in the context of DOAs belonging to the grid of candidate directions. Further, for the given and fixed scenario, since the signal and noise samples are random variables, these two probabilities are dependent on the distributions of these random variables and thus are desired to be mathematically formulated through the conditions or inequalities on these random variables.

4.3.1 Optimality Conditions and Monte Carlo Evaluation Process

Model Selection Consistency, as defined in Chapter 3, is a strictly stronger requirement than correctly estimating the DOAs since there may be some solutions (estimated spatial spectrum) of SpSF which are non-sparse but do have their peaks located at the correct

positions. (Actually, the success of I-SpSF relies on the existence of such solutions.) Thus, as will be shown in Section 4.4, the probability of the existence of MSC λ is usually much smaller than that of DOAEC ones, which may cause the effort of analyzing the optimality conditions of MSC solutions and finding the MSC λs seem meaningless for practical applications. However, because the support of the MSC solutions are exactly known (supposing that the true DOAs are available for analysis), such solutions can be explicitly formulated and their optimality conditions can be considerably simplified. In contrast, there is no explicit formula for DOAEC solutions even if the realizations of the signal and noise samples are known. Further, the corresponding optimality conditions of these solutions cannot be simplified because their supports are generally unknown and varying with the realizations. Therefore, in this subsection, we present and analyze the optimality conditions of MSC solutions and show their relationship to the conditions of the existence of MSC λs. In the following, we first recite some definitions from Chapter 3, which are the basis of the derivations presented in the rest of this section.

The true spatial spectrum and any solution of SpSF are defined as p° and p^*, respectively. Without loss of generality, the first L elements of p° are assumed to be non-zero: $p_i^\circ > 0, 1 \leqslant i \leqslant L$ and $p_i^\circ = 0, L < i \leqslant K$ (by re-arranging the columns of $\mathbf{A}_K$). p° and p^* are partitioned into subvectors:

$$p^\circ = \begin{bmatrix} p^\circ_{(1)} \\ p^\circ_{(2)} \end{bmatrix}, \quad p^* = \begin{bmatrix} p^*_{(1)} \\ p^*_{(2)} \end{bmatrix}, \quad u^* = \begin{bmatrix} u^*_{(1)} \\ u^*_{(2)} \end{bmatrix},$$

where $u^* = p^* - p^\circ$, $p^\circ_{(1)}$, $p^*_{(1)}$ and $u^*_{(1)}$ are the first L entries of p°, p^* and u^*, respectively, and $p^\circ_{(2)}$, $p^*_{(2)}$ and $u^*_{(2)}$ are self-explanatory. The observation matrix $\mathbf{A}_K$ is partitioned into the sub-matrices

$$\mathbf{A}_K = [\mathbf{A}_{K,(1)}, \mathbf{A}_{K,(2)}],$$

where $\mathbf{A}_{K,(1)}$ and $\mathbf{A}_{K,(2)}$ consist of the first L and the last $K - L$ columns, respectively. Based on these sub-matrices, for $i, j = 1, 2$, one can define:

$$\mathbf{C}_{ij} = \Re[\mathbf{A}^H_{K,(i)} \mathbf{A}_{K,(j)}]$$
$$b_i = \Re[\mathbf{A}^H_{K,(i)} \mathbf{E}_v],$$

where $\Re[\cdot]$ denotes the real part and $b = [b_1^T, b_2^T]^T$. 1_L and 1_{K-L} are the vectors of all 1s and of size $L \times 1$ and $K - L \times 1$ respectively. According to these definitions, any

MSC solution has to have its $p^*_{(2)} = 0$, $u^*_{(2)} = 0$, $p^*_{(1)} > 0$ and $u^*_{(1)} > -p^*_{(1)}$, where the inequalities are element-wise.

By the derivations in Chapter 3, the necessary and sufficient conditions of an MSC solution are that there exists $u^*_{(1)}$ satisfying the following element-wise inequalities

$$u^*_{(1)} > -p^\circ_{(1)} \tag{4.15a}$$

$$u^*_{(2)} = 0 \tag{4.15b}$$

$$\mathbf{C}_{11}u^*_{(1)} - b_1 = -0.5\lambda 1_L \tag{4.15c}$$

$$\mathbf{C}_{21}u^*_{(1)} - b_2 \geqslant -0.5\lambda 1_{K-L}. \tag{4.15d}$$

Since there is no additional restriction on $u^*_{(2)}$, we can just let $u^*_{(2)} = 0$ and drop the second condition. Then, the inequalities in (4.15) can be further simplified to

$$\lambda \mathbf{C}_{11}^{-1}1_L \leqslant 2\mathbf{C}_{11}^{-1}b_1 + 2p^\circ_{(1)} \tag{4.16a}$$

$$\lambda(\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L - 1_{K-L}) \leqslant 2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}b_1 - 2b_2, \tag{4.16b}$$

where, by (4.15c), $u^*_{(1)} = \mathbf{C}_{11}^{-1}(b_1 - 0.5\lambda 1_L)$. Now that there is explicit formula for $u^*_{(1)}$ (and hence $p^*_{(1)}$) and it is not involved in the inequalities of (4.16), for any given $\mathbf{C}$ and b, whether the conditions (4.16) can be satisfied or not is now independent of the optimization process of SpSF but only relies on the value of λ. More accurately speaking, if we can find any λ satisfying the inequalities of (4.16), the corresponding solution, given by $u^*_{(1)} = \mathbf{C}_{11}^{-1}(b_1 - 0.5\lambda 1_L)$, will be the solution of SpSF and Model Selection Consistent. Otherwise, no matter what regularization parameter it uses, SpSF cannot find MSC solution. Therefore, the probability P_{MSC} is equivalent to the probability of the existence of such λ, which is obviously a function of the distributions of the involved random variables b and $p^\circ_{(1)}$ (if the DOAs and $\mathbf{A}_K$ are fixed).

Note that the inequalities of (4.16) are between vectors and element-wise. Thus, each of these inequalities represents a series of constraints on λ, where only at most four of them are effective. Specifically, let's partition the elements of $\mathbf{C}_{11}^{-1}1_L$ into positive ones

and negative ones and denote the indices of the i^{th} positive and the j^{th} negative elements by $+_i$ and $-_j$ respectively. Then, (4.16a) is equivalent to the following inequalities:

$$\lambda \leqslant [2\mathbf{C}_{11}^{-1}b_1 + 2p_{(1)}^{\circ}]_{+_i}/[\mathbf{C}_{11}^{-1}1_L]_{+_i}, \quad i = 1, ..., K_+ \tag{4.17}$$

and

$$\lambda \geqslant [2\mathbf{C}_{11}^{-1}b_1 + 2p_{(1)}^{\circ}]_{-_j}/[\mathbf{C}_{11}^{-1}1_L]_{-_j}, \quad j = 1, ..., K_-, \tag{4.18}$$

where K_+ and K_- are the number of positive and negative elements in $\mathbf{C}_{11}^{-1}1_L$, respectively. Similarly, (4.16b) can be rewritten as

$$\lambda \leqslant [2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}b_1 - 2b_2]_{+_i}/[\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L - 1_{K-L}]_{+_i}, \quad i = 1, ..., Z_+, \tag{4.19}$$

and

$$\lambda \geqslant [2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}b_1 - 2b_2]_{-_j}/[\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L - 1_{K-L}]_{-_j}, \quad j = 1, ..., Z_-. \tag{4.20}$$

The above four inequalities fully describe the functional relationship between MSC λs and the DOAs, signal and noise samples, array geometry and candidate directions. Given any specific realization of $\{b_1, b_2, p_{(1)}^{\circ}, \theta_1, ..., \theta_L\}$, the boundaries on MSC λs can be calculated by (4.17), (4.18), (4.19) and (4.20). These boundaries along with $\lambda \geqslant 0$, define a range, S_{MSC}. When S_{MSC} is non-empty, any λ inside it is MSC; if it is empty, SpSF cannot find MSC regularization parameter and solutions. Therefore, P_{MSC} can be further translated to the probability that S_{MSC} is non-empty.

Although an explicit formula of P_{MSC} seems to be within the reach by these inequalities, it is generally very hard to obtain since the distributions of the involved random variables, especially b_1 and b_2, are very difficult (if not impossible) to analyze (for special cases, e.g. $L = 1$, theoretical results can be found in [58]). Thus, instead of an explicit formula, we propose to estimate P_{MSC} through Monte Carlo evaluation for any given scenario, a proposal which is based on the idea of Monte Carlo simulation to evaluate the probability by a large number of independent trials. Specifically, once the DOAs, the distributions and $\mathbf{A}_K$ are fixed and given, N_T independent trials are employed, during which the signal and the noise samples are randomly generated according to their distribution functions. Then, using the DOAs and $\mathbf{A}_K$, b_1, b_2 and $p_{(1)}^{\circ}$ are computed, which, after substituting into the inequalities (4.17), (4.18), (4.19) and (4.20), can provide S_{MSC} for this realization. The number of trials in which S_{MSC} is non-empty is

divided by N_T, providing an estimate of P_{MSC}. N_T is chosen to be sufficiently large so that this estimate becomes accurate enough. Note that this evaluation process is proposed to replace the explicit formula for P_{MSC}, and thus it requires exact knowledge of the distributions and DOAs, which makes it not usable for practical applications. However, the optimality conditions and the idea of this Monte Carlo evaluation process presented in this subsection serve as the basis of the automatic λ-selector proposed in Subsection 4.3.3.

4.3.2 Upper Bound on P_{MSC}

Although through the analysis presented in the last subsection P_{MSC} and the range of MSC λs can be evaluated, the actual aim of SpSF is to correctly recover the DOAs (usually along with non-sparse or less sparse spatial spectra), which is much more possible than perfect support recovery, $P_{MSC} \ll P_\theta$ (especially when N is finite). Therefore, in this subsection, we propose an upper bound on P_{MSC} which will be based on the optimality conditions of DOAEC solutions and later shown to be able to approximate P_θ well. This upper bound, along with the Monte Carlo evaluation process, forms the basis of the automatic λ-selector proposed in Subsection 4.3.3.

Before the analysis can be presented, we have to first mathematically define the DOAEC solutions. Most of the DOA estimation algorithms, e.g. MUSIC and MVDR, do not directly provide DOA estimates, but generate a "measurement value" for each of the candidate directions. These measurements can be power(MVDR), orthogonality(MUSIC) etc. Then, the estimates of the DOAs are identified by finding the peaks in these measurement values and, if these peaks are associated with the correct candidate directions, the corresponding solution is considered to be DOAEC. In this subsection, for simplicity of analysis, we use a slightly different definition for DOAEC solutions:

Definition 6. DOAEC Solution: *any solution p^* satisfying $min\{p^*_{(1)}\} > max\{p^*_{(2)}\}$ is called DOAEC, and the probability of the existence of such a solution is P_θ,*

where $min\{p^*_{(1)}\}$ and $max\{p^*_{(2)}\}$ are the minimum and the maximum element of $p^*_{(1)}$ and $p^*_{(2)}$, respectively.

As we have mentioned, in most of times, P_{MSC} is much smaller than P_θ, which is caused by (4.15d) (or equivalently (4.16b)). This is because, as shown in the proof of

Theorem 3, $E(b) = \sigma^2 M 1_K$; hence

$$E[2\mathbf{C}_{11}^{-1}b_1 + 2p_{(1)}^{\circ}] = 2\sigma^2 M\mathbf{C}_{11}^{-1}1_L + 2E[p_{(1)}^{\circ}] \tag{4.21a}$$

$$E[2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}b_1 - 2b_2] = 2\sigma^2 M(\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L - 1_{K-L}). \tag{4.21b}$$

Comparing (4.21) and (4.16), we can see that, if λ is set to be $2\sigma^2 M$, the left-hand-side of (4.16a) is always smaller than the mean of its right-hand-side (since $E[p_{(1)}^{\circ}] > 0$). In contrast, the left-hand-side of (4.16b) is equal to the mean of its right-hand-side. By viewing the right-hand-side of these two inequalities as random vectors and assuming their distributions are symmetric around their means, the probability that (4.16a) can be satisfied is larger than that of (4.16b). Therefore, in order to obtain an upper bound, we try to relax the constraint (4.16b) by looking at the optimality conditions of DOAEC solutions.

The general optimality conditions of the solutions of SpSF are the following

$$u_{(1)}^* \geqslant -p_{(1)}^{\circ} \tag{4.22a}$$

$$u_{(2)}^* \geqslant 0 \tag{4.22b}$$

$$[\mathbf{C}u^*]_{i_k} - [b]_{i_k} = -0.5\lambda, \quad k = 1, ..., U_+ \tag{4.22c}$$

$$[\mathbf{C}u^*]_{j_k} - [b]_{j_k} \geqslant -0.5\lambda, \quad k = 1, ..., U_0, \tag{4.22d}$$

where U_+ and U_0 are the number of non-zero and zero entries in p^*, respectively, and i_k and j_k are the corresponding indices, respectively. For DOAEC solutions, such conditions cannot be significantly simplified since $\{i_k,\ k = 1, ..., U_+\}$ and $\{j_k,\ k = 1, ..., U_0\}$ are unknown, but we can still obtain the result that the DOAEC solutions have to at least satisfy the following conditions:

$$min\{u_{(1)}^* + p_{(1)}^{\circ}\} > max\{u_{(2)}^*\} \geqslant min\{u_{(2)}^*\} \geqslant 0 \tag{4.23a}$$

$$2\mathbf{C}_{21}u_{(1)}^* + 2\mathbf{C}_{22}u_{(2)}^* - 2b_2 \geqslant -\lambda 1_{K-L}. \tag{4.23b}$$

If the solution p^* is DOAEC but not MSC, $u^*_{(2)}$ is non-zero and $\mathbf{C}_{22}u^*_{(2)} \geqslant 0$ since all the elements of $\mathbf{C}_{22}$ are non-negative (proved in Section 3.2). Thus, the condition (4.23b) is a relaxation of (4.15d) and, by replacing (4.15d) with (4.23b), we can obtain an upper bound on P_{MSC}, which, unfortunately, cannot be evaluated since $u^*_{(2)}$ is unknown. *Therefore, the key idea of our upper bound is to use a randomly generated non-negative vector, g, to replace the unknown $u^*_{(2)}$ and remove the randomness of g through the Monte Carlo evaluation process of the upper bound.* Specifically, by leaving the number, G, of non-zero elements in g as a parameter chosen by the user, we generate the random vector $g = [g_1, g_2, ..., g_{K-L}]^T$ in the following way:

Step 1: With equal possibilities, randomly select G candidate directions from those separated from any of $\{\theta_1, ..., \theta_L\}$ by at least half null-to-null beamwidth

Step 2: Denoting the indices of their corresponding elements in $p^*_{(2)}$ by $r_1, ..., r_G \in [1, K-L]$, set $e_G \in \mathbb{R}^{K-L\times 1}$ as a vector of 0s except for its $r_1^{th}, ..., r_G^{th}$ elements being 1

Step 3: Construct $g = (min\{\mathbf{C}_{11}^{-1}b_1 + p^\circ_{(1)}\} - 0.5\lambda min\{\mathbf{C}_{11}^{-1}1_L\})e_G$.

In Step 3, we assign the non-zero elements in g to be the same value, which is inspired by (4.23a) and

$$max\{u^*_{(2)}\} < min\{u^*_{(1)} + p^\circ_{(1)}\} \leqslant min\{\mathbf{C}_{11}^{-1}b_1 + p^\circ_{(1)}\} - 0.5\beta min\{\mathbf{C}_{11}^{-1}1_L\}.$$

To further simplify the formulas, we define

$$m_{bp} = min\{\mathbf{C}_{11}^{-1}b_1 + p^\circ_{(1)}\}$$
$$m_{C1} = min\{\mathbf{C}_{11}^{-1}1_L\}$$

and thus

$$g = (m_{bp} - 0.5\lambda m_{C1})e_G.$$

Substituting g into (4.23b), (4.15d) can be relaxed as

$$2\mathbf{C}_{21}u^*_{(1)} + (2m_{bp} - \lambda m_{C1})\mathbf{C}_{22}e_G - 2b_2 \geqslant -\lambda 1_{K-L}. \tag{4.24}$$

Replacing (4.24) into (4.16) gives the inequalities for the upper bound:

$$\lambda \geqslant 0 \tag{4.25a}$$

$$\lambda \mathbf{C}_{11}^{-1} 1_L \leqslant 2\mathbf{C}_{11}^{-1} b_1 + 2p^{\circ}_{(1)} \tag{4.25b}$$

$$\lambda(\mathbf{C}_{21}\mathbf{C}_{11}^{-1} 1_L + m_{C1}\mathbf{C}_{22} e_G - 1_{K-L}) \leqslant 2\mathbf{C}_{21}\mathbf{C}_{11}^{-1} b_1 + 2m_{bp}\mathbf{C}_{22} e_G - 2b_2. \tag{4.25c}$$

Similar to (4.16), we denote the range of λ defined by (4.25) as S_{UP} and, for any realization, $S_{MSC} \subseteq S_{UP}$. Therefore, when $G > 0$, $P_{UP} \geqslant P_{MSC}$, where P_{UP} is the probability that S_{UP} is non-empty. When $G = 0$, the random vector e_G is always 0 and hence this upper bound reduces back to P_{MSC}, $P_{UP} = P_{MSC}$. Because b, p°, $\{\theta_1, \theta_2 ..., \theta_L\}$ and e_G are random vectors and their distributions are difficult to obtain, this upper bound P_{UP} cannot be explicitly formulated, but needs to be evaluated through the Monte Carlo evaluation process (note that m_{bp} is also dependent on b and p°). Specifically, for any given $G > 0$, $\{\theta_1, \theta_2, ..., \theta_L\}$ and the distributions of signal and noise samples, N_T independent trials are employed, during each of which the signal and the noise samples are randomly generated according to their distribution functions. The indices $\{r_1, r_2, ..., r_G\}$ should also be re-generated to remove the randomness of vector e_G. Then, using the DOAs and $\mathbf{A}_K$, b_1, b_2, $p^{\circ}_{(1)}$ and m_{bp} are computed, which in turn give the value of S_{UP} for this realization. The number of trials in which S_{UP} is non-empty is divided by N_T, providing an estimate of P_{UP}. N_T is chosen to be sufficiently large so that this estimate becomes accurate enough.

Since this upper bound is obtained by relaxing the optimality conditions of MSC solutions towards those of DOAEC solutions, one can expect it to be able to approximate P_θ to some extend. In Section 4.4, simulation examples are presented to show that, if G is properly selected, P_{UP} can provide very accurate approximation to P_θ for a large range of SNR. Further, as discussed in the last paragraph, when evaluating P_{UP} through the Monte Carlo process, one also obtains N_T realizations of S_{UP}. Since in practical applications, for any given observation data, the corresponding realization of the signal and noise samples are unknown, the probability of the distribution of S_{UP} is more meaningful than any of its realization for the purpose of parameter selection. However, since the evaluation of P_{UP} relies on the knowledge of the DOAs and the distributions, which are generally unknown, this upper bound and its S_{UP}s cannot be directly utilized to select the regularization parameter of SpSF, and the idea of using rough estimates to

replace this knowledge forms the automatic and data-adaptive regularization parameter selector proposed in next subsection.

4.3.3 Automatic Regularization Parameter Selector

Based on the upper bound, in this subsection, we present the automatic selector of λ for the purpose of DOA estimation. Once the array geometry and the grid of candidate directions are fixed, then for any given data set, the range of DOAEC λs is a function of the DOAs and signal and noise samples. Therefore, since the DOAs and the realization of the random variables are unknown, it is impossible to exactly locate the DOAEC λs for the given realization. Our way to circumvent to such difficulty is to use rough estimates of the DOAs instead and utilize the Monte Carlo evaluation process to find the candidate regularization parameter λ which most likely falls into the range of DOAEC λs. During the Monte Carlo evaluation process, the distributions (or some of their parameters) are also replaced by rough estimates. Thus, the regularization parameter found by this selector is most likely to be DOAEC than any other choices if the rough estimates are accurate enough and the relaxation of the upper bound, described below, can well approximate the conditions of the DOAEC solutions. Another difficulty in developing such a selector is that, because the support of the DOAEC solutions is unknown, there is additional uncertainty in the conditions of DOAEC λs beside the DOAs and random variables. As described in the last subsection, this difficulty is overcome by utilizing the upper bound, which uses a random vector g to replace $u^*_{(2)}$. This vector g is generated according to fixed rules and its randomness is further removed by Monte Carlo process, and this helps the λ-selector to get rid of the aforementioned uncertainty.

For simplicity of the description of the selector, a few definitions need to be made. $f_i(\zeta_i)\ i = 1, ..., L$ and $f_n(\zeta_n)$ are the distribution functions of sources and noise, respectively, and ζs are the parameters of these distributions (e.g. mean and variance for Gaussian distributions). We assume the availability of the number of sources L, G, $f_i, i = 1, ..., L$ and f_n. Any other information, including θ_i, $\zeta_i, i = 1, ..., L$ and ζ_n, is replaced by rough estimates obtained by a lower resolution spectrum estimator such as MVDR [7]. Using $\lambda_c,\ c = 1, ..., K_\lambda$ to denote all the candidate choices of λ (where the candidate λs are usually uniformly distributed in a range of interest and K_λ is the number of such candidate choices and very large), the procedure for the selector is:

Step 1. Use a spectrum estimator, such as MVDR, to estimate θ_i, ζ_i and ζ_n, $i = 1, ..., L$, and denote these estimates by $\tilde{\theta}_i$, $\tilde{\zeta}_i$ and $\tilde{\zeta}_n$ respectively.

Step 2. Set $Z_c = 0,\ c = 1, ..., K_\lambda$

Step 3. for $t = 1, ..., N_T$

a. Randomly generate vector e_G (use $\tilde{\theta}$s instead of θs)

b. Generate signal and noise samples by distributions $f_i(\tilde{\zeta}_i)$ and $f_n(\tilde{\zeta}_n)$

c. Calculate S_{UP}

d. If $\lambda_c \in S_{UP}$, $Z_c = Z_c + 1$, $c = 1, ..., K_\lambda$

end for

Step 4. Find the maximum of $Z_c,\ c = 1, ..., K_\lambda$ and denote its index by c^*.

Step 5. Output λ_{c^*} for SpSF.

Note that this automatic selector, rather than being fixed, adaptively chooses the regularization parameter λ based on actual data-generated covariance. Further, as shown in the next section, this method is not sensitive to the values of L and G, and very rough estimates of θs and ζs are generally effective (combined with SpSF) to achieve satisfactory DOA estimation performance. More importantly, the computational cost of this selector is negligible compared with that of Cross Validation and even the cost of the optimization of SpSF itself.

4.4 Simulation Results

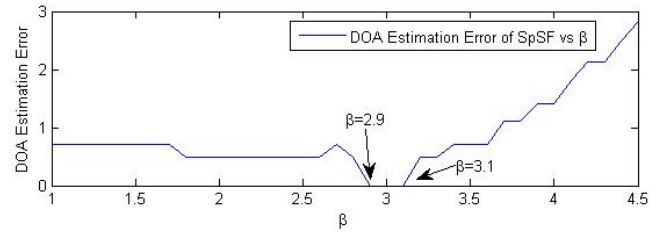
In this section, we present simulation results illustrating the effectiveness and performance of the parameter selection methods proposed in this chapter. Again, a ULA with $M = 8$ sensors and $d = 0.5$ normalized inter-element distance is considered and $L = 2$ sources from -5° and 5° are assumed. Further, $L = 2$ is assumed to be known to MUSIC and L_1-SVD. The sources and the noise are both zero-mean Gaussian with the noise power being $\sigma^2 = 1$, and the sources are assumed to be mutually uncorrelated (except in Subsection 4.4.5).

In Subsection 4.4.1, a simulation example is presented to illustrate the effectiveness of the Diagonal Loading method in reducing the sensitivity of SpSF to its regularization parameter. The DOA estimate RMSE of I-SpSF is compared with that of SpSF, MUSIC and L_1-SVD in Subsection 4.4.2. This shows the success of I-SpSF in retaining the DOA estimation performance of SpSF while removing its dependency on the selection of regularization parameter. Further, I-SpSF is applied to the experimental data [4], which is also used in Subsection 2.4.4. In Subsection 4.4.3, the upper bound proposed in Subsection 4.3.2 is evaluated and shown to be able to well approximate the probability of the existence of DOAEC solutions P_θ. Further, through the same example, the effectiveness of S_{UP} in approximating the range of DOAEC λs is illustrated, which justifies the idea of the automatic λ-selector. The DOA estimation performance of SpSF, using a fixed λ and the automatic λ selector, is compared with those of MUSIC, MVDR, L_1-SVD and Cramer-Rao lower bound in Subsection 4.4.4, and the robustness (in the sense of the DOA estimation performance of SpSF) of the λ-selector with respect to the errors in L and G is also shown. In Subsection 4.4.5, we follow the setting of correlated sources considered in Subsection 2.4.3 and show that by combining the λ-selector with SpSF, Spatial Smoothing and Forward-Backward Smoothing, it can achieve a better estimation performance than an exhaustively searched but fixed regularization parameter and, in some range of SNR, almost the same performance as SpSF-C.

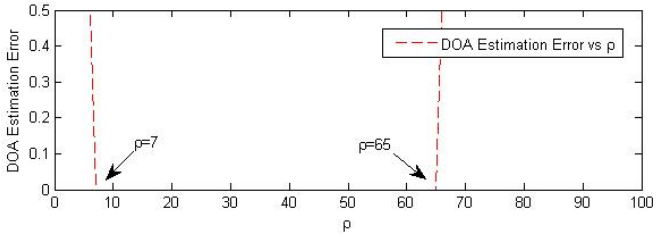
4.4.1 Effectiveness of Diagonal Loading

In this subsection, two examples are given to illustrate the effectiveness of Diagonal Loading method proposed in Section 4.1. We consider $N = 100$, SNR$= -5dB$ and the grid of candidate directions to be $\{-90°, -89.5°, ..., 90°\}$. ϵ is set at 0.01 for the Diagonal Loading formulation of SpSF (4.5). Figure 4.1(a) and 4.1(b) are the DOA estimation error of SpSF versus β (2.16) and ρ (4.5), respectively. As shown in these figures, the range of proper β is from 2.9 to 3.1. In contrast, the range of proper ρ is from 7 to 65, which is much larger than that of β. Thus it will be much easier to find a proper ρ than β. (The DOA estimation error is defined as $\sqrt{\frac{1}{L}\sum_{i=1} L(\theta_i - \hat{\theta}_i)^2}$.)

With the same array setting, consider the case where $L = 4$ uncorrelated sources from $-23°$, $-7°$, $5°$ and $20°$ with SNR$=5dB$, $10dB$, $5dB$ and $0dB$, respectively. In this case, $N = 1000$ and $\epsilon = 0.01$. As shown in Figure 4.2(a) and 4.2(b), the range of proper



(a) DOA Estimation Error vs β (2.16)



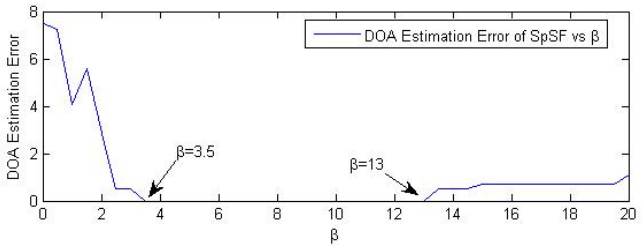
(b) DOA Estimation Error vs ρ (4.5)

Figure 4.1: First Example of the Effectiveness of Diagonal Loading for SpSF

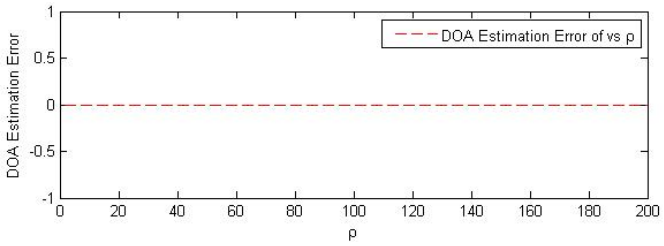
ρ is from 0 to more than 200. But the corresponding range for β is from about 3 to 13.

4.4.2 DOA Estimation Performance of I-SpSF

In this subsection, we still assume two uncorrelated sources with equal power and from $-5°$ and $5°$. The grid of candidate directions is $\{-90°, -89.5°, ..., 90°\}$. 500 independent trials are utilized to evaluate the RMSEs of SpSF, I-SpSF, L_1-SVD and MUSIC, during which $L = 2$ is assumed to be known to all the methods and $N = 3000$ snapshots are available. The regularization parameters of SpSF and L_1-SVD are empirically and independently chosen, and I-SpSF uses the signal subspace reconstruction of $\mathbf{R}$ (which is discussed in Section 2.1) and is initialized with 0 regularization parameter. Figure 4.3 compares the RMSEs of DOA estimation of I-SpSF with SpSF, MUSIC, L_1-SVD and stochastic Maximum Likelihood estimator (MLE) (note that MUSIC and I-SpSF generate no error in the DOA estimates during the 500 trials for SNR= $5dB$, and MLE also has perfect DOA estimates for SNR=$0dB$ and $5dB$). According to Figure 4.3, SpSF, I-SpSF and L_1-SVD have similar large error thresholds which are 5dB better than that of MUSIC. Especially, without any additional information on signal and noise, I-SpSF



(a) DOA Estimation Error vs β (2.16)



(b) DOA Estimation Error vs ρ (4.5)

Figure 4.2: Second Example of the Effectiveness of Diagonal Loading for SpSF

can even achieve better performance than SpSF when SNR is relatively high.

Figure 4.4 is the spectrum estimated by I-SpSF using the experimental data of [4]. More detailed description of the experiment setup can be found in [4]. SpSF is also applied to this data set in Subsection 2.4.4 of this dissertation, where the grids of candidate directions for both elevation and azimuth angles are set to be $\{-90°, -89°, ..., 90°\}$. In contrast, we use grids of $\{-90°, -88°, ..., 90°\}$ for I-SpSF in this subsection (because the SVD process required in (4.9) is not feasible in MATLAB if the denser grid is used). In this example, I-SpSF is initialized with 0 regularization parameter and converges in 3 iterations. Comparing Figure 4.4 to Figure 2.9, one can see that: 1) I-SpSF provides almost the same DOA estimates as SpSF using exhaustively searched regularization parameter; and 2) the peaks of SpSF is sharper than those of I-SpSF.

4.4.3 Effectiveness of the Upper Bound P_{UP}

In this subsection, we evaluate the upper bound P_{UP} for different values of G and show the capability of P_{UP} in approximating P_θ when G is properly selected. To evaluate the upper bound, the availability of the exact DOAs and the distributions is assumed and the

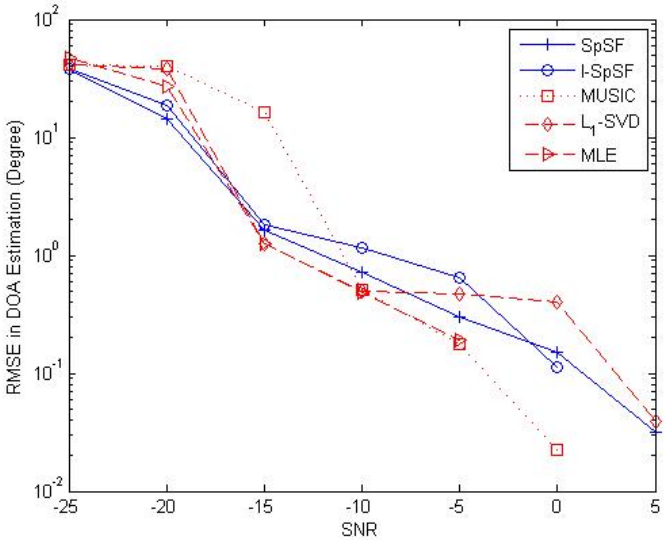


Figure 4.3: Performance Comparison for I-SpSF

number of snapshots is set as $N = 300$. The candidate directions are $\{\phi_1, \phi_2, ..., \phi_K\} = \{-80°, -79°, -78°, ..., 80°\}$. For each SNR between $-20dB$ and $5dB$ (with a $5dB$ step-size), SpSF is solved over a dense grid of candidate λs (more than 1000 grid points) in 1000 independent trials, and the corresponding P_θ is obtained by counting the number of trials in which (at least) one candidate λ provides correct DOA estimates. Note that the time cost for this simulation process is many orders greater than the time for evaluating P_{UP}. When evaluating the upper bound, P_{UP}, we set $N_T = 1000$. In Figure 4.5, P_θ (red dashed line) is shown and P_{UP} is plotted for $G = 0, 2, 3, 4, 5$. According to this figure, P_{UP} can approximate P_θ very well for a large range of SNR when G is properly chosen ($G = 4$ in this case). Further, P_{UP} can very accurately predict the threshold behavior of P_θ, and such prediction is important for the performance evaluation purpose. Note that when $G = 0$, e_G is always a zero vector, and thus the upper bound is identical to P_{MSC} ($P_{UP} = P_{MSC}$ with $G = 0$). Therefore, according to Figure 4.5, P_{MSC} is much smaller than P_θ, a fact which presents the need of the upper bound and supports the use of the automatic λ-selector.

When evaluating P_θ for Figure 4.5, for each candidate regularization parameter, we count the number of trials in which it provides correct DOA estimates. This number,

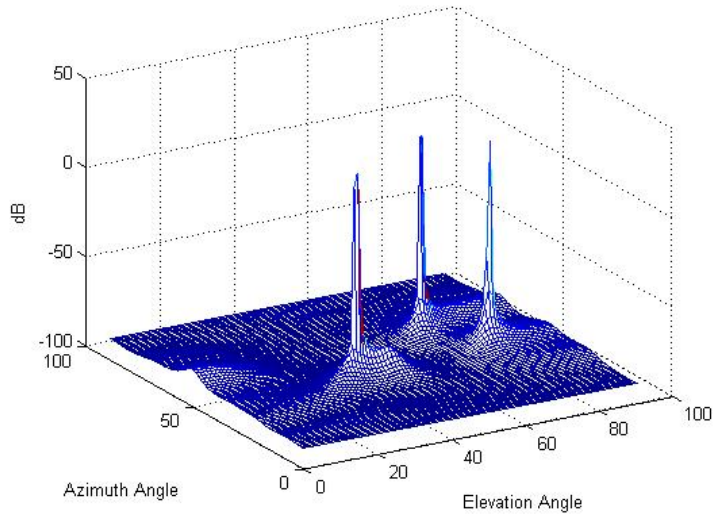


Figure 4.4: Spectrum Estimated by I-SpSF Using the Experimental Data [4]

divided by the total number of trials, becomes an estimate of the probability that this specific regularization parameter is DOAEC, and the regularization parameters with the largest such probabilities are generally better than the others. In Figure 4.6, we plot such probabilities in red dashed lines for $SNR = -5dB$ and $0dB$. The blue solid lines are Z_c/N_T versus λ_c, obtained as described above. As shown by these figures, the range of values for λ where the blue plots achieve their maxima coincide with those of the red lines except for $G = 0$ ($P_{UP} = P_{MSC}$). Therefore, the relaxation of the upper bound, as represented by (4.16b) and (4.25c), is necessary for the purpose of selecting λ. Further, if uses the true prior information, the selector would not be sensitive to the value of G and can approximately locate λs with the largest probabilities of giving correct DOA estimates.

4.4.4 Automatic λ-Selector for Uncorrelated Sources

In this subsection, we compare the DOA estimate Root-Mean-Squared-Error (RMSE) of SpSF, using both fixed λ and the automatic selector with those of MUSIC, MVDR, L_1-SVD with its fixed regularizer, and the Cramer-Rao Lower Bound. Following that, the robustness of the automatic λ-selector with respect to errors in L and G is illustrated.

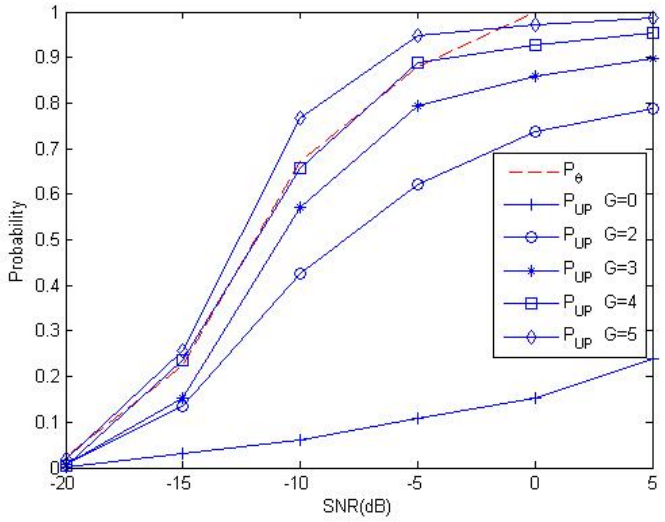


Figure 4.5: P_θ and P_{UP} for different G

Once again, $N = 300$ and 1000 independent trials and uncorrelated sources are used. We change the grid of candidate directions to $\{-80°, -79.9°, ..., 80°\}$, however.

First, in Figure 4.7, we compare the estimation performance of all these methods using the exact value of L. "SpSF (Manual)" is the RMSE of SpSF with the fixed λ manually selected. In this case, for each SNR, we evaluate the RMSE of SpSF for many candidate λs and select the best one. "SpSF (AS)" is the RMSE of SpSF combined with the automatic selector, using $G = 4$, the true DOAs and the exact distributions. "SpSF (AS+MVDR)" is the same as "SpSF (AS)" except that **the selector only knows that the sources and the noise are uncorrelated zero-mean Gaussian random variables**. In each trial, it uses the estimated DOAs ($\hat{\theta}_1$ and $\hat{\theta}_2$) by MVDR, the estimated noise power $\hat{\sigma}^2$ by the smallest eigenvalue of $\mathbf{R}$ and the estimated signal powers $\hat{p}_i = p_i^{mvdr} - \hat{\sigma}^2/M,\ i = 1, 2$ where p_i^{mvdr} is the value of the MVDR spectrum at $\hat{\theta}_i$. "MUSIC", "MVDR" and "L_1-SVD" are the RMSEs of the corresponding methods, and the regularization parameter of L_1-SVD is also manually selected in the same way as "SpSF (Manual)". "CRB" stands for the stochastic Cramer-Rao Lower Bound [12][53], which uses the knowledge that the sources are uncorrelated. As shown by Figure 4.7, the λs suggested by the selector using either the true or the initial rough estimates of DOAs

and powers can help SpSF achieve almost the same DOA estimation performance as an exhaustively searched and fixed regularization parameter ("SpSF (Manual)"). Further, the large error threshold of "SpSF (Manual)", "SpSF (AS)", "SpSF (AS+MVDR)" and L_1-SVD are $5dB$ lower than that of MUSIC.

Now, we examine the robustness of the automatic selector with respect to the errors in L and G. In Figure 4.8, we keep $G = 4$ and use different values of L for "SpSF (AS+MVDR)" (true value $L = 2$). As shown by the figure, SpSF is not sensitive to the errors in L. Especially when SNR is relatively large, $L = 1$ or 3 causes no significant performance degradation. In Figure 4.9, we assume availability of the exact knowledge of L but use different values of G for "SpSF (AS+MVDR)". As shown by this figure, SpSF is not sensitive to the values of G, and small perturbations in G do not decrease the DOA estimation performance of SpSF.

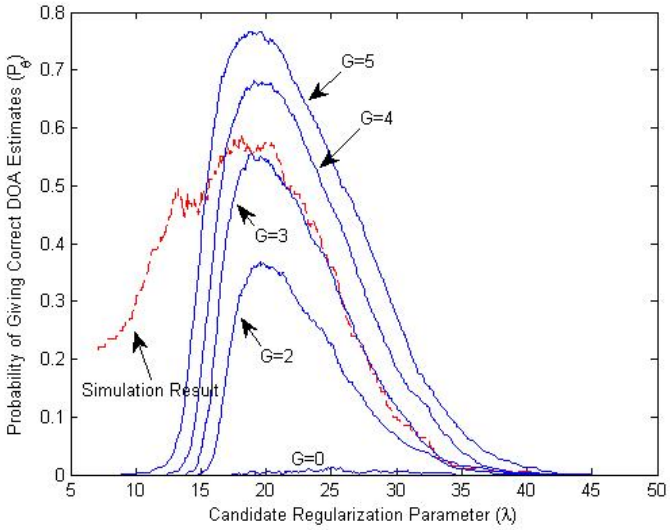
4.4.5 Automatic λ-Selector for Correlated Sources

In this subsection, we evaluate the DOA estimation performance of SpSF using the automatic λ-selector for the cases of correlated sources. The scenario, the Weighted Spatial Smoothing (WSS) and the Forward-Backward Smoothing (FBS) of Subsection 2.4.3 are used, and the difference between these two subsections is that, in this subsection, we apply SpSF to the covariance matrix after WSS and FBS, $\mathbf{R}_{SS}$, by using the regularization parameter provided by the automatic λ-selector. (Scenario: two sources with correlation coefficient 0.99, $N = 3000$, sub-array size equal 5. For more details, please refer to Subsection 2.4.3.)

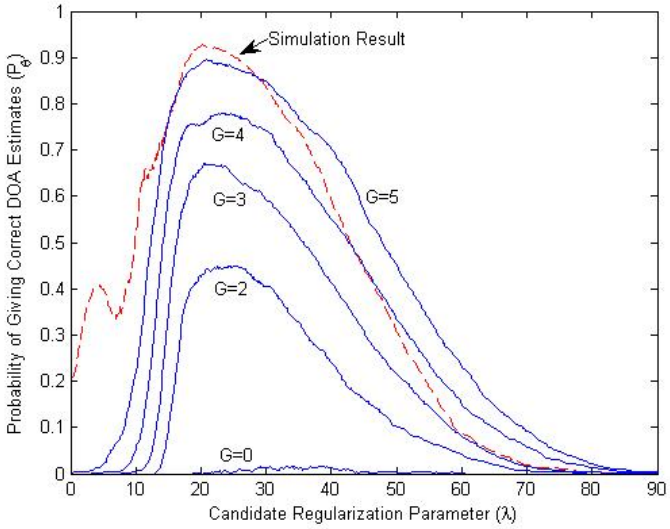
In Figure 4.10, "SpSF (AS+MVDR)" is obtained by applying SpSF to $\mathbf{R}_{SS}$ with the λs chosen by the selector using the rough estimates provided by MVDR. "SpSF (Clairvoyant λ)" is the performance of SpSF with λ adaptively tuned in such way that, in each trial, we select the λ which achieves the minimum estimation error among all the candidate regularization parameters. Note that, since it requires knowledge of the true DOAs, "SpSF (Clairvoyant λ)" is not achievable in practice. However, we show it here to illustrate the potential benefits of adaptively tuning λ and as a benchmark for our automatic selector, which also adaptively chooses λ. In these simulations, the grid of candidate directions of SpSF-C is set to be $\{-80°, -79°, ..., 80°\}$ when $SNR < 0dB$ and $\{-10°, -9.9°, ..., 10°\}$ when $SNR \geqslant 0dB$ in order to reduce its computational

complexity.

As shown by the figure, SpSF, using the automatic λ-selector, can perform better than MUSIC (using either $\mathbf{R}$ or $\mathbf{R}_{SS}$) and exhibits about $5dB$ gain over L_1-SVD. Further, in some ranges of SNR, the automatic selector can make improvement even over the best fixed λ ("SpSF (Manual)") and achieve almost the same performance as SpSF-C with much smaller computational complexity.



(a) SNR=-5dB



(b) SNR=0dB

Figure 4.6: Probability of Giving Correct DOA Estimates Versus Candidate Regularization Parameters

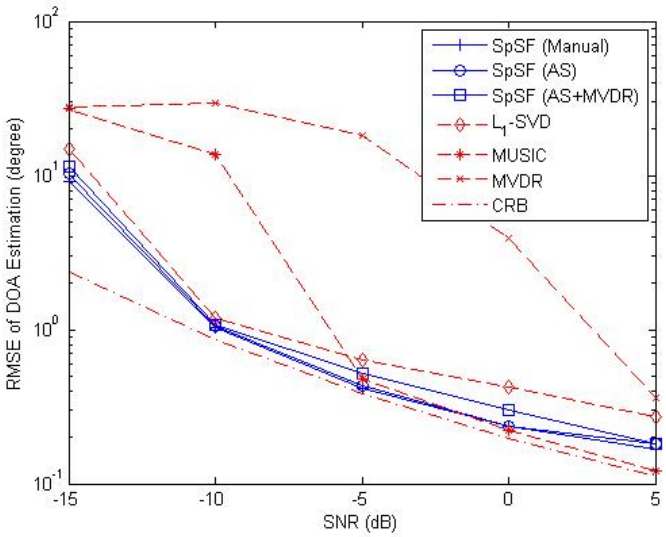


Figure 4.7: RMSE of DOA Estimation Using $L = 2$ and $G = 4$

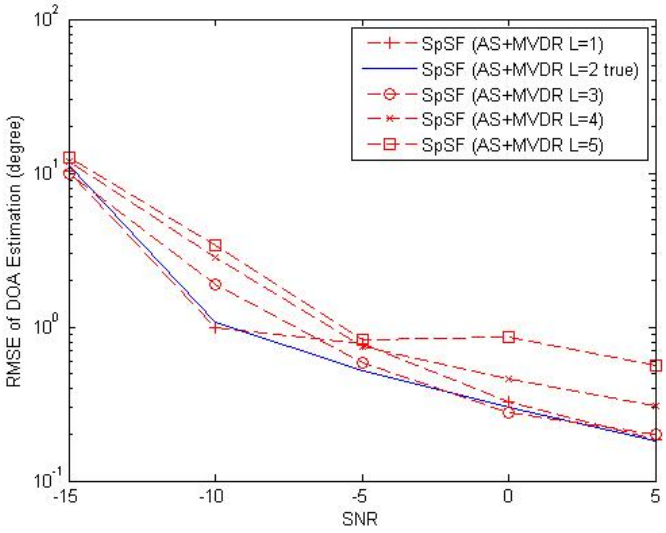


Figure 4.8: RMSE of SpSF DOA Estimation With Incorrect L

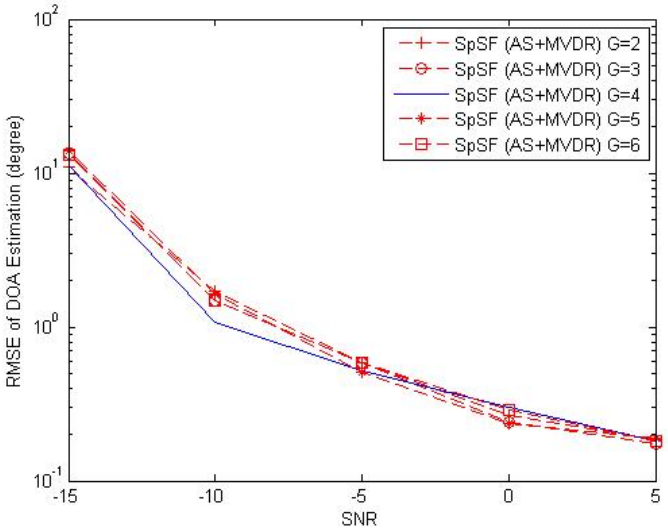


Figure 4.9: RMSE of SpSF DOA Estimation With Different G

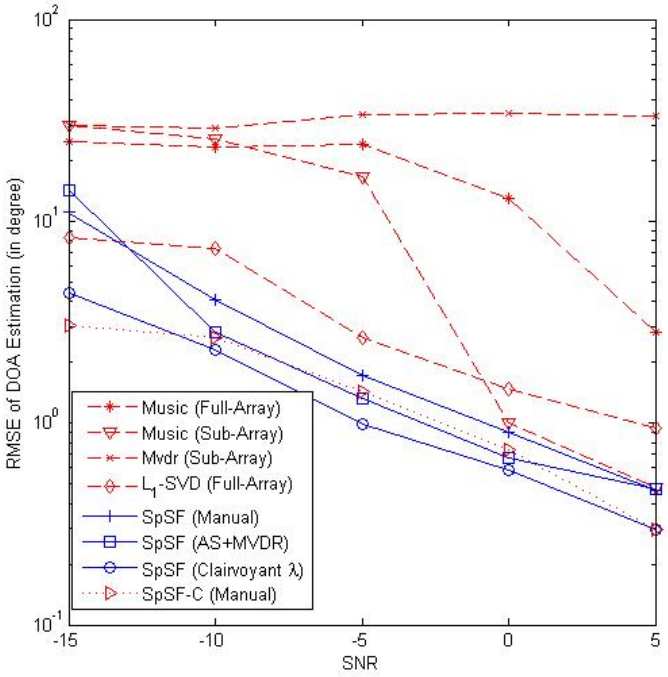


Figure 4.10: RMSE for Correlated Sources

Chapter 5

Extensions of SpSF

As presented in previous chapters, the SpSF method shows its advantages in DOA estimation performance, computational complexity and favorable theoretical properties. However, these results are based on several strict assumptions, including on-grid DOAs $\{\theta_1, \theta_2, ..., \theta_L\} \subset \{\phi_1, \phi_2, ..., \phi_K\}$ and spatially stationary sources. These assumptions may raise concern about the robustness and practical applications of not only SpSF but also many other sparsity-exploiting DOA estimation algorithms, including L_1-SVD. The modeling errors introduced by the failure of the assumptions are usually proportional to the signal powers, e.g. the error terms caused by off-grid DOAs, and hence can be much larger than the noise contributions, especially when SNR is large. Since the sparsity-exploiting DOA estimation methods intensively rely on the accuracy of their sparse representation models (as ordinary sparse signal reconstruction algorithms), such modeling errors can significantly degrade their estimation performances. This problem has been addressed for general sparse signal reconstruction algorithms in several works, including [45] and [46]. In [45], the performance degradation of the sparse signal reconstruction algorithm [14] under the presence of general modeling errors is analyzed and a lower bound on the corresponding estimation performance is provided. In contrast, the Sparse Total Least Square (STLS) [46] is proposed to reconstruct sparse signals with its formulation taking the general modeling errors into account; unfortunately, STLS is non-convex and thus unfavorable for DOA estimation applications.

In this chapter, we consider the errors in the sparse representation model of SpSF, especially for the DOA estimation problem, and propose convex extensions of SpSF to

improve its robustness with respect to various modeling errors. In Section 5.1, the SpSF with Modeling Uncertainty (SpSFMU) is proposed to address off-grid DOAs, the most common source of such errors. As an extension of SpSF, SpSFMU is also developed on the basis of the sparse representation model of the spatial covariance matrix, but utilizes the group sparsity by linearly approximating the modeling errors in the neighborhood of the candidate directions. Consequently, SpSFMU can provide continuous estimates of the DOAs and be used to either reduce the computational complexity of SpSF or improve its estimation performance when the off-grid DOAs are present. The diagonal loading technique, proposed for SpSF in Section 4.1, is also applied to SpSFMU to help select its regularization parameter.

Moving sources present the problem of time-varying DOA estimation. When this model is cast into a framework, one is faced with the problem of constructively combining multiple observations with inconsistent supports. By following the ideas of [2], in Section 5.2, we first define the concept of Block-level Sparsity and its differences with the widely used Element-level and Group-Level sparsity are discussed. Then, with the assumption of persistent activity, we propose a new objective function embedding penalization of Element-Level Sparsity into that of Block-Level Sparsity, based on which Persistently Active Block Sparsity (PABS) is proposed for the reconstruction of multiple inconsistent sparse signals. Following that, a Singular-Value-Decomposition (SVD)-based method is developed to reduce the computational complexity of PABS. In Section 5.3, the application of PABS to the DOA estimation problem is also discussed, and simulation results comparing its estimation performance with L_1-SVD [40] and SPICE [3] are presented.

5.1 SpSF with Modeling Uncertainty

In this section, the cases of off-grid DOAs are considered and the specialized extension of SpSF, SpSF with Modeling Uncertainty (SpSFMU), is proposed. As discussed before, in reality, the true DOAs, θ_i, can hardly be on the grid of candidate directions, and hence the modeling error

$$\mathbf{U}_v = \sum_{i=1}^{L} p_i(a_v(\theta_i) - a_v(\phi_{k_i})) \tag{5.1}$$

is introduced, where p_i is the power of the signal from θ_i and ϕ_{k_i} is the candidate direction closest to θ_i. Obviously, the size of $\mathbf{U}_v$ is proportional to the signal powers and dependent on the distance between the candidate directions (since the true DOAs vary continuously). Therefore, from the perspective of estimation performance, one would be interested if such an error term can be incorporated into the underlying sparse representation model so that it can enhance the performance rather than degrade it. However, since the relationship between the DOAs and the steering vectors is usually complicated, incorporating such error terms into the underlying model will significantly increase the complexity of the model and thus make it less favorable for, especially, real-time DOA estimation applications. In this section, by following some ideas presented in [46], we propose SpSFMU, which by linearly approximating the error term $\mathbf{U}_v$ of (5.1) can use very few optimization variables to parameterize the modeling error. Although the development of SpSFMU in Subsection 5.1.1 is based on formulation (2.16) of SpSF, it is straightforward to apply the same extension to other formulations of SpSF (as shown in Subsection 5.1.2).

5.1.1 Formulation of SpSFMU

Consider the following formulation of SpSF:

$$\min_{p\in\mathbb{R}^{K\times 1}} \quad \|p\|_1, \quad s.t. \quad \|\mathbf{R}_v - \mathbf{A}_K p\|_2 \leqslant \beta \; and \;\; p_k \geqslant 0, \;\; k = 1, ..., K. \tag{5.2}$$

Suppose there is a signal from DOA θ_i which belongs to the DOA bin centered at ϕ_k. We denote the true steering vector of this signal by $f(\theta_i)$ and $f_v(\theta_i) = vec(f(\theta_i)f^H(\theta_i))$. Then, by defining a "rotation matrix" $\mathbf{T}^{(k)}$ for the k^{th} candidate direction, we can write:

$$f_v(\theta_i) = \mathbf{T}^{(k)} a_v(\phi_k), \tag{5.3}$$

where $\mathbf{T}^{(k)}$, $M^2 \times M^2$, is *diagonal* and complex. By using (5.3), the sparse representation model of the spatial covariance matrix can be rewritten as:

$$\bar{\mathbf{R}}_v = \sum_{k=1}^{K} p_k \mathbf{T}^{(k)} a_v(\phi_k) + \mathbf{E}_v, \tag{5.4}$$

where $\mathbf{E}_v$ now only contains noise contribution. Denote t_k as the diagonal of the matrix $\mathbf{T}^{(k)}$, $t_k = diag(\mathbf{T}^{(k)}) \in \mathbb{C}^{M^2 \times 1}$, and define a $K \times M^2$ matrix

$$\mathbf{T} = [t_1, t_2, ..., t_K]^T.$$

Observe that if there is no signal from the DOA bin centered at ϕ_k, the corresponding spatial spectrum element $p_k = 0$ and there is no need to have non-zero $\mathbf{T}^{(k)}$. Thus, the corresponding row of $\mathbf{T}$ is null, which suggests that the spatial spectrum p and $\mathbf{T}$ have the same sparsity pattern (support) in row and thus *Group Sparsity*. Based on this observation, we can use the group sparsity of $[p, \mathbf{T}]$ as the objective function and formulate an estimator as

$$\begin{aligned} \min_{p,\mathbf{T}} \|[p, \mathbf{T}]\|_{2,1} \quad s.t. \quad \|\mathbf{R}_v - \sum_{k=1}^{K} p_k \mathbf{T}^{(k)} a_v(\phi_k)\|_2 \leqslant \beta \\ \& \; p_k \geq 0, \quad k = 1, ..., K, \end{aligned} \tag{5.5}$$

where

$$\|[p, \mathbf{T}]\|_{2,1} = \sum_{k=1}^{K} \|[p_k, t_k^T]\|_2.$$

Although the formulation (5.5) seems very generally applicable (for various types of modeling errors in $a_v(\phi)$), it is non-convex and hence not favorable to us.

Fortunately, if we only consider the modeling error introduced by off-grid DOAs, the estimator (5.5) can be significantly simplified by linear approximation of the entries in $\mathbf{T}$ and transformed to a convex formulation. For simplicity of development, we use Uniform Linear Array (ULA) as an example to present this idea, which can be applied to other array geometries. Since only the off-grid DOA is considered, we have $f_v(\theta_i) = a_v(\theta_i)$. For ULA, every entry in $a_v(\theta_i)$ has the form $e^{-j2\pi dm \sin\theta_i}$ where d is the element spacing normalized by signal wavelength, $j = \sqrt{-1}$ and m is an integer between $-(M-1)$ and $M-1$. We denote this entry by $a_v(\theta_i)_m$ and by (5.3):

$$a_v(\theta_i)_m = e^{-j2\pi dm\gamma_k} \times a_v(\phi_k)_m, \tag{5.6}$$

where $e^{-j2\pi dm\gamma_k}$ is the corresponding entry of $\mathbf{T}^{(k)}$ and $\gamma_k = \sin\theta_i - \sin\phi_k$ is called correction parameter. By using Taylor expansion, we have $e^{-j2\pi dm\gamma_k} \approx 1 + (-j2\pi dm\gamma_k)$, and (5.6) can be re-written as:

$$a_v(\theta_i)_m \approx [1 + (-j2\pi dm\gamma_k)] a_v(\phi_k)_m. \tag{5.7}$$

Replacing $\mathbf{T}^{(k)}a_v(\phi_k)$ by this approximation, (5.4) can be simplified to

$$\bar{\mathbf{R}}_v = \sum_{k=1}^{K} [p_k a_v(\phi_k) + q_k b_v(\phi_k)] + \mathbf{E}_v, \tag{5.8}$$

where $q_k = p_k\gamma_k$, $b_v(\phi_k) = \mathbf{P}a_v(\phi_k)$ and $\mathbf{P}$ is a diagonal matrix whose entry corresponding to $a_v(\phi)_m$ equals to $-j2\pi dm$. Since each p_k represents the signal power from the corresponding direction bin, we have the restriction: $l_k \leqslant \gamma_k \leqslant u_k$, $k = 1, ..., K$, where:

$$l_k = \sin\frac{\phi_{k-1} + \phi_k}{2} - \sin\phi_k,\ k = 2, ..., K,$$

with $l_1 = 0$ and

$$u_k = \sin\frac{\phi_{k+1} + \phi_k}{2} - \sin\phi_k,\ k = 1, ..., K-1,$$

with $u_K = 0$. Here, this restriction is posted since in the estimated spectrum, it is desirable to associate non-zero signal powers to the closest candidate directions. Further, since $p_k \geqslant 0$ for $k = 1, .., K$, we have

$$p_k l_k \leqslant q_k \leqslant p_k u_k, \quad k = 1, ..., K. \tag{5.9}$$

Define $q = [q_1, q_2, ..., q_K]^T$. Similar as $[p, \mathbf{T}]$, p and q have the same sparsity pattern. With this observation, the covariance matrix model (5.8) and the convex relaxation (5.9), we transform the non-convex optimization problem (5.5) into a convex one, Sparse Spectral Fitting with Modeling Uncertainty (SpSFMU):

$$\begin{aligned} \min_{p,q} \|[p, q]\|_{2,1} \quad s.t. \quad \|\mathbf{R}_v - \mathbf{A}_K p - \mathbf{B}_K q\|_2 \leqslant \beta \\ p_k l_k \leqslant q_k \leqslant p_k u_k, \quad p_k \geqslant 0, \quad k = 1, ..., K, \end{aligned} \tag{5.10}$$

where $\mathbf{B}_K = [b_v(\phi_1), b_v(\phi_2), ..., b_v(\phi_K)]$ and β is its regularization parameter. We denote the minimizer of (5.10) by p^* and q^*, and the indices of the largest L peaks in p^* by $k_1, ..., k_L$. Then the DOA estimates are:

$$\hat{\theta}_i = \arcsin\left[\sin\phi_{k_i} + q^*_{k_i}/p^*_{k_i}\right], \quad i = 1, 2, ..., L, \tag{5.11}$$

where we assume $p^*_{k_i} > 0, i = 1, 2, ..., L$.

Obviously, SpSFMU is a convex optimization problem which can be solved efficiently (just like for L_1-SVD) and provides estimates of the signal powers along with the DOAs.

Further, as shown by (5.11), its estimates of the DOAs are continuous rather than discrete, which makes it capable of improving the estimation accuracy for any given but fixed grid of candidate directions. Besides this improvement, SpSFMU can be used to reduce the computational complexity of SpSF by using a coarser grid of candidate directions while keeping its DOA estimation performance similar to or even better than that of SpSF. This advantage is obtained since, when off-grid DOAs are present, the estimation accuracy of SpSF will be bounded by the spacing of the candidate directions while the continuous estimates (or the corrections of (5.11)) provided by SpSFMU can compensate for such errors. However, as shown in the simulation examples of Section 5.3, the linear approximation utilized in SpSFMU still leaves error terms into the sparse representation model of spatial covariance. These errors becomes serious when the spacing of the candidate directions and SNR are large. This is as expected since when the spacing is large, first order approximation is very inaccurate and the error residual is still proportional to the signal powers.

Another concern of SpSFMU is its performance when there is spatial ambiguity, e.g. ULA with $d > 0.5$. In the analysis of Section 3.1, we proved that, under the assumption of on-grid DOAs, the condition for SpSF to work properly is that there is no repeated spatial frequency in the overcomplete dictionary of SpSF, $\mathbf{A}_K$. However, the same problem becomes more complicated for SpSFMU since, by using linear perturbation, there is going to be (approximately) repeated spatial frequencies in the effective representation model of SpSFMU even if the frequencies of the columns of $\mathbf{A}_K$ are distinctive. A simple method to avoid repetition of the spatial frequencies would be, first, to carefully select the candidate directions, $\phi_1, \phi_2, ..., \phi_K$ such that their corresponding spatial frequencies are as different as possible; and second, to setup the values of l_k and u_k, $k = 1, ..., K$ to restrict the ranges (regions) of spatial frequencies (candidate directions) which can be represented by the continuous DOA estimates. Specifically, the ranges of the DOAs which can be an estimate of SpSFMU are formulated by $[\arcsin(\sin\phi_k + l_k), \arcsin(\sin\phi_k + u_k)]$ (assuming that arcsin is monotonic in this range), and hence the θ_ks, l_ks and u_ks should be selected so that there is no overlap between any of two such ranges.

5.1.2 Diagonal Loading for SpSFMU

Similar to SpSF, the regularization parameter selection of SpSFMU is of concern especially for its practical applications. In this subsection, we use the Diagonal Loading technique, proposed in Section 4.1, to alleviate the sensitivity of SpSFMU to the selection of its regularization parameter.

Consider an equivalent formulation of (5.10):

$$\begin{aligned} \min_{p,q} & \|\mathbf{R}_v - \mathbf{A}_K p - \mathbf{B}_K q\|_2 + \delta \|[p,q]\|_{2,1} \\ & p_k l_k \leqslant q_k \leqslant p_k u_k, \quad p_k \geqslant 0, \quad k = 1, ..., K. \end{aligned} \tag{5.12}$$

Denote the optimizer of (5.12) by p_δ^* and q_δ^*. Our simulations show that if $\delta = \sqrt{M}$, the fitting residual of (5.12), which is $\mathbf{R}_v - \mathbf{A}_K p_\delta^* - \mathbf{B}_K q_\delta^*$, can be well approximated by $\alpha \mathbf{I}_v$, where α is the size of the identity matrix component in the error term $\mathbf{E}_v$ of (5.8) and $\mathbf{I}_v = vec(\mathbf{I})$.

Further, if we set $\delta = \sqrt{M} + \epsilon$ with ϵ being positive and sufficiently small, we have observed that the corresponding residual is approximately a linear function of α:

$$\|\mathbf{R}_v - \mathbf{A}_K p_\delta^* - \mathbf{B}_K q_\delta^*\|_2 \approx (\sqrt{M} + c_1)\alpha + c_2, \tag{5.13}$$

where $c_1 > 0$ and c_2 are the coefficients. Importantly, c_1 is very small and gets even smaller when ϵ becomes smaller. Note that for any given ϵ, the equation (5.13) only holds for a finite range of α. If α is too large, the optimizer p_δ^* and q_δ^* in (5.13) will be 0. This is because

$$\|\mathbf{R}_v\|_2 \leqslant \|\mathbf{R}_v - \alpha \mathbf{I}_v\|_2 + \|\alpha \mathbf{I}_v\|_2 = \|\mathbf{R}_v - \alpha \mathbf{I}_v\|_2 + \alpha\sqrt{M}.$$

If α is so large that $c_1\alpha + c_2 \geqslant \|\mathbf{R}_v - \alpha \mathbf{I}_v\|_2$, which gives:

$$\|\mathbf{R}_v - \mathbf{A}_K p_\delta^* - \mathbf{B}_K q_\delta^*\|_2 \geqslant \|\mathbf{R}_v\|_2,$$

then by minimizing the objective function of (5.12), we get $p_\delta^* = q_\delta^* = 0$. Based on the above observations and analysis, we propose the following alternative formulation of (5.10):

$$\begin{aligned} \min_{p,q} & \|\mathbf{R}_v + \rho \mathbf{I}_v - \mathbf{A}_K p - \mathbf{B}_K q\|_2 + \delta \|[p,q]\|_{2,1} \\ & p_k l_k \leqslant q_k \leqslant p_k u_k, \quad p_k \geqslant 0, \quad k = 1, ..., K, \end{aligned} \tag{5.14}$$

where $\delta = \sqrt{M} + \epsilon$ with ϵ being positive and sufficiently small and ρ is the parameter to be adjusted.

According to (5.13), the 2-norm of the residual of (5.14) is approximately $(\sqrt{M} + c_1)(\rho + \alpha) + c_2$, where $(\rho + \alpha)\sqrt{M}$ accounts for $(\rho + \alpha)\mathbf{I}_v$ and $c_1(\rho + \alpha) + c_2$ for the other error terms in $\mathbf{E}_v$. Similar to the Diagonal Loading method for SpSF, (5.14) will be much less sensitive to ρ than (5.10) to β since c_1 is very small. Furthermore, this sensitivity can be controlled by adjusting ϵ. Although using (5.14) may cause estimation performance loss, such loss will not be significant compared to the difference between the performances of SpSFMU and SpSF. Further, the performance of this diagonal loading formulation (5.14) is still better than that of SpSF if off-grid DOAs are present and the two methods use the same grid of candidate directions.

5.2 Persistently Active Block Sparsity for Moving Sources

Some basic/famous sparse signal reconstruction algorithms (SSRA), e.g. LASSO, use only one-dimensional observation data and seek the sparse decomposition of it on the pre-given dictionary. Some other SSRAs further assume and utilize the co-existence relationship of the active elements (atoms) of the sparse representations of multiple one-dimensional observation vectors, e.g. Group Lasso [29], L_1-SVD and SpSFMU. However, one common assumption of these methods has been that these multiple observations have the same sparsity pattern (support) on the same dictionary, which sometimes is not the case. In this section, the cases of moving sources are considered and, inspired by the ideas presented in [2], we develop a new method, Persistently Active Block Sparsity (PABS), for a more general framework for reconstruction of Block-sparse signals by exploiting persistent activity in the block-level sparsity. PABS will then be applied to the DOA estimation problem of moving sources.

In Subsection 5.2.1, we first define the Block-Level Sparsity and compare it to the Element-Level and Group-Level sparsities. Then, the algorithm PABS is proposed in Subsection 5.2.2 on the basis of a novel objective function encouraging the persistently active block sparsity. In Subsection 5.2.3, an SVD-based method is presented to reduce the computational complexity of PABS.

5.2.1 Block-Level Sparsity

Suppose that the observation vectors are $y_t \in \mathbb{C}^{M\times 1}$ and that they are generated from the following model:

$$y_t = \mathbf{A}x_t + e_t,\ t = 1, 2, ..., T, \tag{5.15}$$

where $\mathbf{A} = [a_1, a_2, ..., a_K] \in \mathbb{C}^{M\times K}$ is constant with respect to t, $\{a_i,\ i = 1, ..., K\}$ are called the atoms, $x_t \in \mathbb{C}^{K\times 1}$ and $e_t \in \mathbb{C}^{M\times 1}$ are the underlying model and the observation error at index (or time) t respectively. Every x_t is assumed to be sparse and has very few non-zero (active) elements. This kind of sparsity is referred to as *Element-Level Sparsity*. If $T = 1$, the problem of estimating x_1 is an ordinary sparse recovery problem and can be efficiently solved by many algorithms, e.g. Lasso[1] and C-P_1 [14].

When multiple observations are available ($T > 1$), one can, of course, use Lasso or C-P_1 to estimate each x_t separately. However, if the sparsity patterns of $\{x_t,\ t = 1, ..., T\}$ are known to be the same, an algorithm taking advantage of this property (which is referred to as *Consistent*) will naturally result in improved performance. To benefit from Consistent models, we first rewrite (5.15) in the compact form

$$\mathbf{Y} = \mathbf{AX} + \mathbf{E}, \tag{5.16}$$

where $\mathbf{Y} = [y_1, y_2, ..., y_T]$, $\mathbf{X} = [x_1, x_2, ..., x_T]$ and $\mathbf{E} = [e_1, e_2, ..., e_T]$, and only a few rows of $\mathbf{X}$ are active. By viewing each row of $\mathbf{X}$ as a *Group*, we define such sparsity as *Group-Level Sparsity*, with the elements within each group either all non-zero (active) or zero (inactive). Group-Level Sparsity is exploited by Group Lasso, L_1-SVD, SpSFMU and etc. These algorithms share similar objective functions, which, with respect to the observation model (5.16), can be formulated as

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|_F + \lambda \sum_{k=1}^{K} \|\mathbf{X}_{k,:}\|_2, \tag{5.17}$$

where $\mathbf{X}_{k,:}$ is the k^{th} row of $\mathbf{X}$ and λ is the regularization parameter to be tuned. The penalization term in (5.17), $\lambda \sum_{k=1}^{K} \|\mathbf{X}_{k,:}\|_2$, encourages very few groups to be active (Group Sparsity) while smoothing the elements within the active groups. If the true models, x_ts, have their active elements within the same group in similar strengths, we call such model *Persistently-Active Model* and the smoothing effect of the objective

function of Group Lasso is favorable to it. (An example of consistent and persistently-active model is shown in Figure 5.1, where gray squares represent inactive elements and colored squares represent active elements.) However, if these elements have very different strengths (some of them may even be inactive), we refer to it as a *Non-Persistently-Active Model*, an example of which is presented in Figure 5.2. Comparing Figure 5.2 to Figure 5.1, it is obvious that the smoothing effect is not favorable to such non-persistent models since it will cause lots of "false active" elements. In order to encourage the Non-Persistently-Active property in the objective function, Hierarchical-Lasso (Hi-Lasso) [2] is proposed, which adds a L_1-Norm measurement of all the elements in $\mathbf{X}$ to the objective function of (5.17) to encourage, in addition to Group-Level Sparsity, Element-Level Sparsity. This additional penalization term introduces an extra regularization parameter to be tuned.

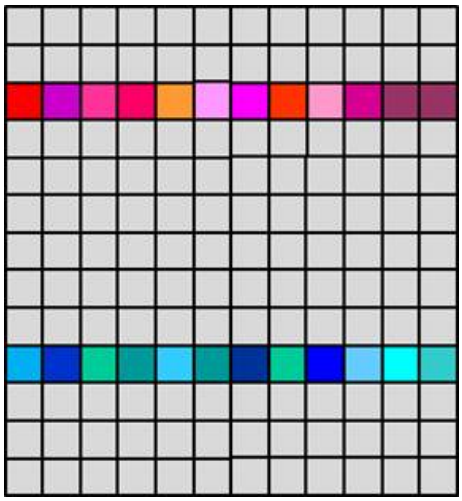


Figure 5.1: Consistent and Persistently-Active Model

Further, in some applications, even the sparsity patterns (support) of x_t can be different, which precludes the Group-Level Sparsity and degrades the estimation performance of the associated algorithms. We call such models as *Inconsistent*. In [2], C-HiLasso and its several applications are presented to deal with this issue. In this work, the groups are further divided into G blocks, and it is further assumed that, very few such blocks are active (containing non-zero elements). This structure is designated as one with *Block-Level Sparsity*. Figure 5.3 is an example of the model with

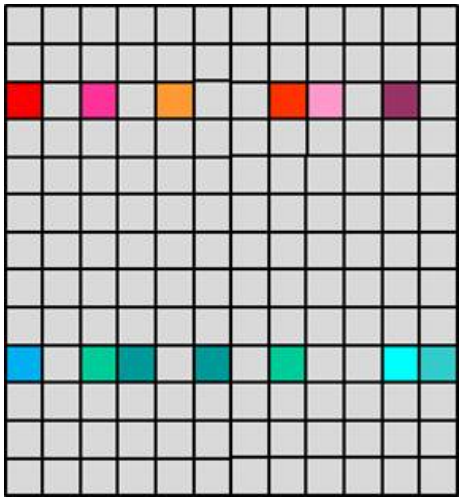


Figure 5.2: Consistent but Non-Persistently-Active Model

Block-Level (Non-Persistent) Sparsity, where the Blocks are defined by the red or blue rectangular. Unfortunately, the penalization term for the Block-Level Sparsity in C-HiLasso also tends to smooth the elements within the active blocks and hence usually leads to non-sparse estimates of x_t. Similar to Hi-Lasso, C-HiLasso uses an extra term of the L_1-Norm of all the elements in $\mathbf{X}$ to further promote Element-Level Sparsity. The formulation of C-HiLasso [2] is recited here:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F + \lambda_1 \sum_{t=1}^{T} \|x_t\|_1 + \lambda_2 \sum_{g=1}^{G} \|\mathbf{X}^g\|_F, \tag{5.18}$$

where $\mathbf{X}^g$ is the submatrix (block) constituted by the rows of $\mathbf{X}$ which belongs to the g^{th} block, and λ_1 and λ_2 are the regularization parameters to be tuned. As shown by (5.18), C-HiLasso separates the penalization for the Block-Level and the Element-Level sparsities, and this introduces an extra regularization parameter and the difficulty in balancing between these two types of sparsity. Further, some cases of interest present properties that are not exploited by C-HiLasso. For example, in many applications, an active block (submatrix) physically implies a target or objective existing through the whole observation duration. In addition, each x_t often has at most one non-zero element within each block since the vector is sparse. Therefore, the models similar to the one presented in Figure 5.4 are of interest, and the capability of the algorithm to distinguish among the elements of each x_t inside the same active block is desirable.

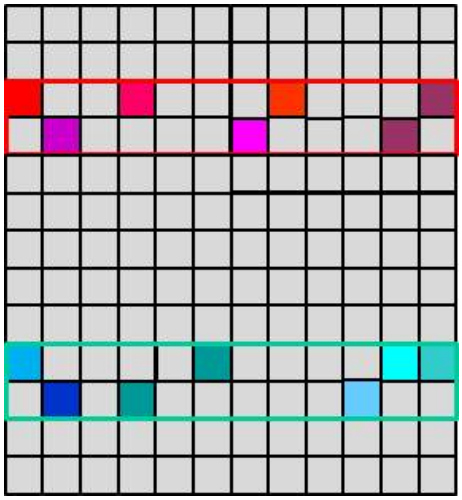


Figure 5.3: Inconsistent and Non-Persistently-Active Model

In next subsection, we propose PABS, which exploits these properties based on a new objective function embedding the penalization for the Element-Level Sparsity into that of the Block-Level Sparsity.

5.2.2 Persistent Active Block Sparsity

Suppose $\mathbf{X}$ is of size $K \times T$ and we divide the K rows into G blocks. The indices of the rows in the g^{th} block are denoted by $\{i_{g,n}\ n = 1, 2, ..., L_g\}$ where $g = 1, ..., G$ and $L_g > 0$ is the number of rows in this block. Without loss of generality, we can further assume that $i_{g,n} + 1 = i_{g,n+1}$ and $i_{g,L_g} + 1 = i_{g+1,1}$, which means that each block is constituted of contiguous rows and the blocks are indexed in the increasing order. Then, we define the subvector containing the elements of x_t inside the g^{th} block as x_t^g

$$x_t^g = [x_t(i_{g,1}), x_t(i_{g,2}), ..., x_t(i_{g,L_g})]^T,$$

where $x_t(i)$ is the i^{th} element of x_t and ^{T} stands for the transpose operation. As discussed in the last subsection, each block usually represents only one physical target or objective; hence, x_t^g is either a null vector or also very sparse (i.e. containing only one active element). Thus, to encourage sparsity within each x_t^g, we define its L_1-norm as

$$p_t^g = \|x_t^g\|_1.$$

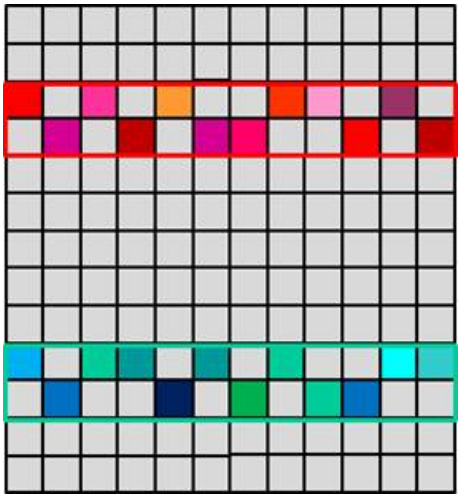


Figure 5.4: Inconsistent and Persistently-Active Model

Considering the property of persistent activity, if the g^{th} block is active, we expect all x_t^g, $t = 1, 2, ..., T$ to be active and sparse. Further, in many applications, e.g. slow fading wireless communication and some DOA estimation, the magnitude of the non-zero elements are roughly constant with respect to t, resulting in the persistently-active property. These observations suggest the following L_2-norm measurement for the g^{th} block

$$q^g = \|[p_1^g, p_2^g, ..., p_T^g]\|_2,$$

and the objective function or penalization term

$$f(\mathbf{X}) = \sum_{g=1}^{G} q^g,$$

which is expected to promote Block-Level as well as Element-Level sparsities of $\mathbf{X}$. Using this measurement function, Persistent Active Block Sparsity (PABS) can be formulated as

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F + \alpha \sum_{g=1}^{G} q^g. \tag{5.19}$$

Note that (5.19) is a convex optimization problem and can be efficiently solved by many off-the-shelf optimization softwares. By using the Lagrangian multiplier, PABS can be

reformulated as

$$\min_{\mathbf{X}} \sum_{g=1}^{G} q^g, \quad s.t. \ \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F \leqslant \beta, \tag{5.20}$$

which is equivalent to (5.19) in the sense that for any $\alpha > 0$, we can find a $\beta > 0$ leading to the same solution (set). When the distribution of $\mathbf{E}$ is known, (5.20) may be more favorable than (5.19) because, according to the Discrepancy Principle, we can select β to be larger than the mean value of $\|\mathbf{E}\|_F$ plus several times of its variance, while the equivalent α is hard to find.

The definition process of the sparsity encouraging term, $f(\mathbf{X})$, is complicated and hence it (and also the objective function of 5.19) may appear confusing to the reader. In Figure 5.5, we use an example of Inconsistent and Persistently-Active Model to better illustrate the definition of $f(\mathbf{X})$. In this example, the rows (groups) of $\mathbf{X}$ are divided into G blocks, each of which contains two successive rows, and the second and the second to last blocks are the only active blocks. For the active blocks, each of the subvectors x_t^i, $t = 1, ..., T$ contains only one active element and their L_1-Norms, p_t^i are fed into the "L_2-Norm Calculator". The "L_2-Norm" of each block, q^i, $i = 1, ..., G$, are summed together, which gives $f(\mathbf{X})$. By minimizing this objective function, PABS tends to prefer the solution with sparse $q = [q^1, q^2, ..., q^G]$ and thus encourages Block-Level Sparsity. However, at the same time, if g^{th} block is active ($q^g > 0$), the algorithm will try to smooth these p_t^g $t = 1, ..., T$ since q^g is the L_2-Norm of p_t^g $t = 1, ..., T$. This effect leads to the Persistent Activity. Further, the minimization process will have to suppress the magnitude value of each p_t^g, which in turn will promotes the Element-Level Sparsity within x_t^g if g^{th} block is active. This effect is actually the aforementioned embedded (nested) penalization of Element-Level Sparsity into that of Block-Level Sparsity.

Compared to Group Lasso, PABS incorporates the inconsistent sparse models into its formulation and, if we set $G = K$ and $L_g = 1$ for $\forall g$, PABS reduces to Group Lasso. In contrast to C-HiLasso, PABS has only one regularization parameter to be tuned because of the nesting of the penalization for the sparsities. Further, the embedded penalization term, $f(\mathbf{X})$, prefers that each x_t^g has exactly one non-zero element with similar absolute values. However, this property is not clear for C-HiLasso, which only suggests the non-zero elements in the active block to be few and of similar magnitudes without requiring them to be uniformly distributed to each x_t^g. Unfortunately, the computational

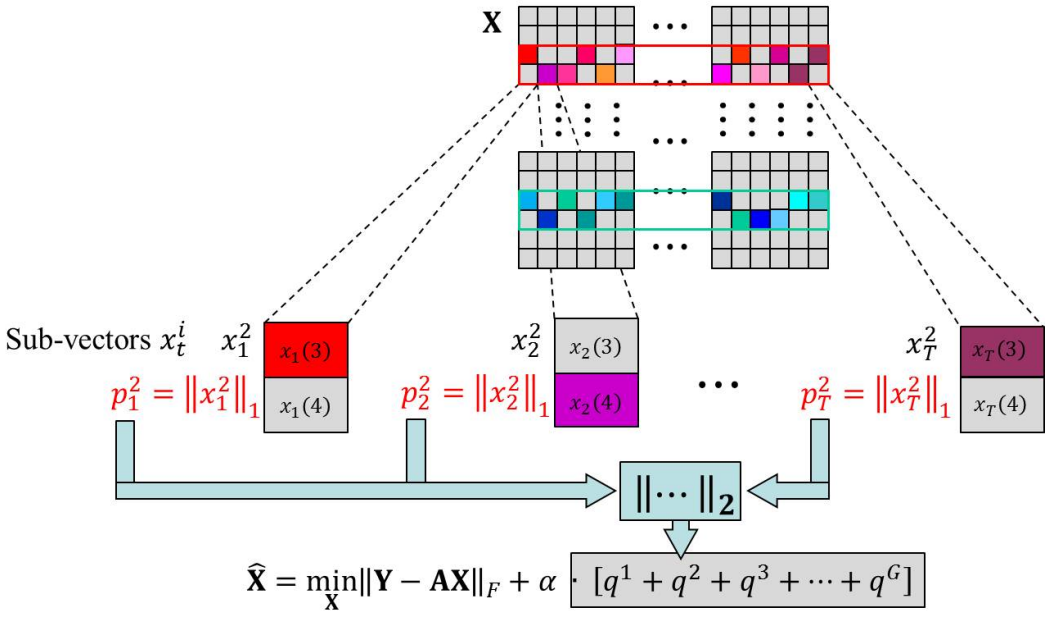


Figure 5.5: Illustration of the Objective Function of PABS

complexities of Group Lasso, C-HiLasso and PABS are significant concerns, especially for real-time applications, because the number of the optimization variables (NOV) in these algorithms is KT and either K or T is usually very large. Therefore, in next section, we present a method to reduce the computational complexity of PABS and C-HiLasso (see [59] regarding the computation of C-HiLasso).

5.2.3 Reduction of Computational Complexity

Similar to the idea of blocks, we can divide the T observation indexes into N observation sets and denote them by $S_i = \{t_{i,n}\ n = 1, 2, ..., T_i\}$ with $i = 1, 2, ..., N$, and T_i being the number of entries in S_i. Assuming that, for $\forall i$, the corresponding sparse models $\{x_t | t \in S_i\}$ share the same sparsity pattern (which can be the case when the sparsity patterns of x_t change very slowly), the Singular-Value-Decomposition (SVD) technique used in [40] can be utilized to reduce NOV of PABS and C-HiLasso which, in turn, leads to lower computational complexity.

Defining $\mathbf{Y}_i = [y_{t_{i,1}}, ..., y_{t_{i,T_i}}]$ and $\mathbf{X}_i$ and $\mathbf{E}_i$ in similar ways, we have

$$\mathbf{Y}_i = \mathbf{A}\mathbf{X}_i + \mathbf{E}_i, \tag{5.21}$$

where, by the assumption in the last paragraph, $\mathbf{X}_i$ enjoys Group-Level sparsity. If we denote the SVD of $\mathbf{Y}_i = \mathbf{U}_i\mathbf{\Omega}_i\mathbf{V}_i^H$ with the singular values in $\mathbf{\Omega}_i$ arranged in decreasing order, it follows from (5.21) that

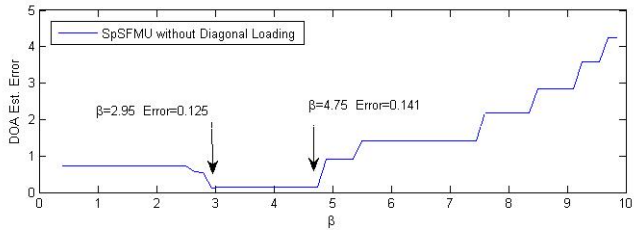
$$y^i = \mathbf{A}x^i + e^i,$$

where $y^i = \mathbf{Y}_i\mathbf{V}_i u_1$, $x^i = \mathbf{X}_i\mathbf{V}_i u_1$, $e^i = \mathbf{E}_i\mathbf{V}_i u_1$ and $u_1 = [1, 0, ..., 0]^T$. By the orthogonality of $\mathbf{V}_i$, y^i is the first column of $\mathbf{U}_i\mathbf{\Omega}_i$. Here, we drop all the columns except for the first one because they will increase NOV proportionally while not providing much more information than keeping only the first column. Further, since the Element-Level Sparsity pattern of $x^i \in \mathbb{C}^{K\times 1}$ is the same as the Group-Level Sparsity pattern of $\mathbf{X}_i$, the non-zero elements of $\{x^i,\ i = 1, 2, ..., N\}$ also lie in the same active blocks of $\mathbf{X}$. This fact keeps the Block-Level Sparsity. Then, by applying PABS, we get an estimator

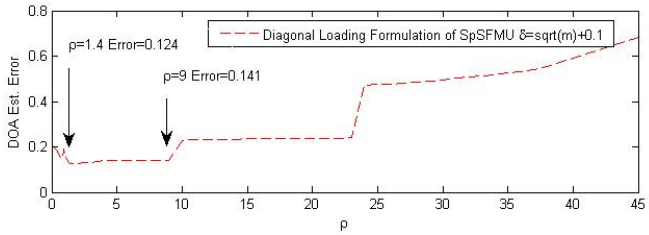
$$\min_{\mathbf{X}^N} \sum_{g=1}^{G} q_N^g, \quad s.t.\ \|\mathbf{Y}^N - \mathbf{A}\mathbf{X}^N\|_F \leqslant \beta, \tag{5.22}$$

where $\mathbf{Y}^N = [y^1, y^2, ..., y^N]$, $\mathbf{X}^N = [x^1, x^2, ..., x^N]$ and $\{q_N^g,\ g = 1, ..., G\}$ are defined on the basis of $\mathbf{X}^N$ in a similar manner as in Subsection 5.2.2.

Note that NOV of (5.22) is KN, which can be much smaller than KT of (5.19). This advantage is obtained at the price of the lack of the capability to estimate the exact value of the non-zero elements in $\mathbf{X}$ and the additional SVD calculation (the computational complexity of this SVD process is usually dominated by the decrease of that of the optimization process). If we can further assume that the non-zero elements from different active blocks are zero-mean and mutually uncorrelated, the vectorized covariance matrix technique of SpSF can be employed to even avoid the SVD calculation. Since this is a straightforward calculation from the discussions of Section 2.2, we omit the details of this additional formulation.



(a) DOA estimation error of SpSFMU vs β (5.10)



(b) DOA estimation error of SpSFMU vs ρ (5.14)

Figure 5.6: DOA Estimation Error vs Regularization Parameters

5.3 Simulation Results

In this section, the simulation results are presented to illustrate the effectiveness and performance of the methods proposed in this chapter. First, in Subsection 5.3.1, we give an example illustrating the effectiveness of the diagonal loading approach of Subsection 5.1.2 and compare the RMSEs of the DOA estimation of SpSF, L_1-SVD and SpSFMU. Then, in Subsection 5.3.2, the application of PABS to the DOA estimation of moving sources is discussed and its estimation performance is compared to that of SPICE [3] and C-HiLasso [2].

5.3.1 Performance of SpSFMU

We consider an ULA of M=8 sensors with d=0.5, receiving L=2 plane waves from $-5.4°$ and $4.6°$. These sources are assumed to be narrow-band, zero-mean and mutually uncorrelated. We set N=3000 and noise to be AWGN with unit variance. The candidate directional grids are assumed to be uniform and the DOA estimation error is defined as $\|[\theta_1 - \hat{\theta}_1, \theta_2 - \hat{\theta}_2]\|_2$.

In Figure 5.6, two sources are of equal SNR= $0dB$ and the grid is of $1°$ separation

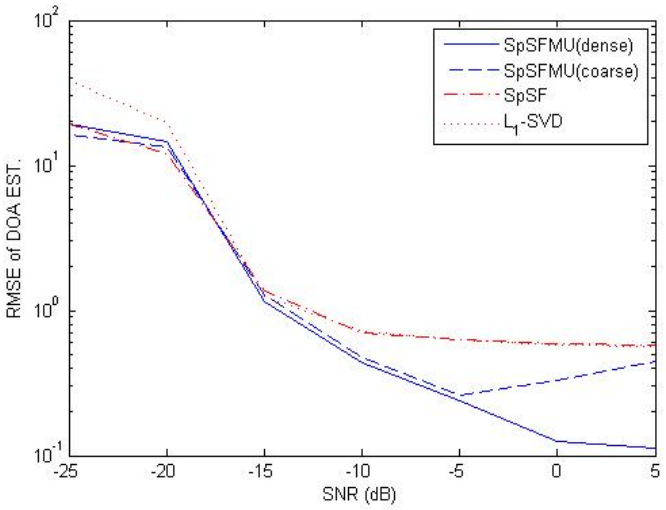


Figure 5.7: RMSE of DOA Estimation of SpSFMU

with the grid points constituting the set $\{-90°, -89°, ..., 90°\}$. For (5.14), we set $\delta = \sqrt{M} + 0.1$. Figure 5.6(a) and 5.6(b) show the curves of DOA estimation error versus β of (5.10) and ρ of (5.14), respectively. As shown in Figure 5.6, when $2.95 < \beta < 4.75$, SpSFMU (5.10) reaches its best performance (compared to the other values of β), and its DOA estimation error is between 0.12 and 0.14. But when β is outside this range, this error becomes larger than 0.57. The best performance of the diagonal loading formulation (5.14) is almost the same as that of SpSFMU and attained in a much wider range $1.4 < \rho < 9$. Furthermore, when ρ is outside this range but less than 24, its DOA estimation error is still less than 0.24. Through this example, it is illustrated that the diagonal loading approach can potentially keep the best performance of SpSFMU while being much less sensitive to its regularization parameters.

In Figure 5.7, we show a comparison of the RMSEs of the DOA estimations for SpSF, L_1-SVD and SpSFMU(5.10). We use 500 trials for each SNR and assume L known to all the methods. The grids of SpSF, L_1-SVD and SpSFMU(dense) are the same as the grid used in Figure 5.6, and the grid of SpSFMU(coarse) is of the same range but with 4° separation. Therefore, the problem dimensions of L_1-SVD and SpSFMU(dense) are the same, twice that of SpSF and four times that of SpSFMU(coarse). The regularization parameters of these methods are empirically and independently chosen. As

shown in Figure 5.7, all the four methods have the same large error threshold(=-15dB) while SpSFMU(coarse), with the lowest problem dimension, achieves similar or better performance compared to SpSF and L_1-SVD. Furthermore, SpSFMU(dense), with the same dimension of L_1-SVD, attains much better DOA estimation performance than the other three methods, especially when SNR is large. Note that when $L > 2$, the problem dimension of SpSFMU(dense) is smaller than that of L_1-SVD and, in this example, when SNR>5dB the linear approximation error of SpSFMU(coarse) becomes dominating. Thus its performance becomes very similar to that of SpSF and L_1-SVD.

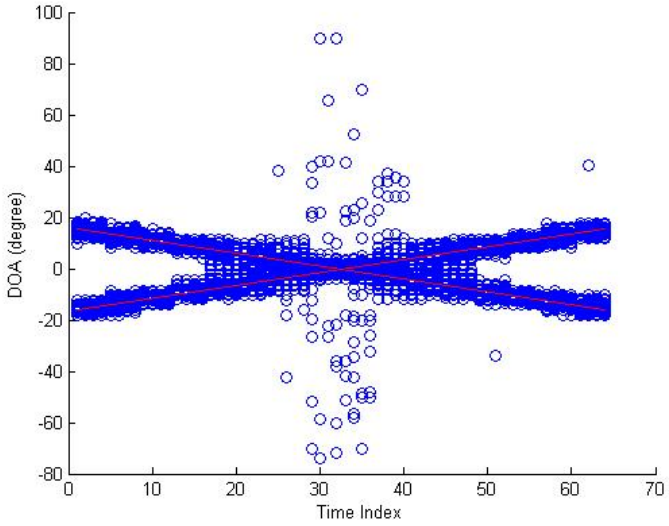


Figure 5.8: Estimated Trajectories by PABS

5.3.2 Application of PABS to DOA Estimation

In this section, we present an example discussing the application of PABS to the DOA estimation using sensor arrays with the presence of moving sources. Suppose, at time t, $L = 2$ uncorrelated far-field sources are from the directions $\Theta_t = \{\theta_t(1), \theta_t(2)\}$ impinging on a ULA of $M = 8$ sensors and $d = 0.5$. The two sources are further assumed to be persistent and moving, one from -16° to 15.5° and the other from 16° to -15.5°, with SNR=0dB and step size= 0.5°. The samples of the two signals are denoted by $s_t(k) = e^{j\psi_t(k)}$, $k = 1, 2$, and the phases $\psi_t(k)$ are independently and uniformly distributed in

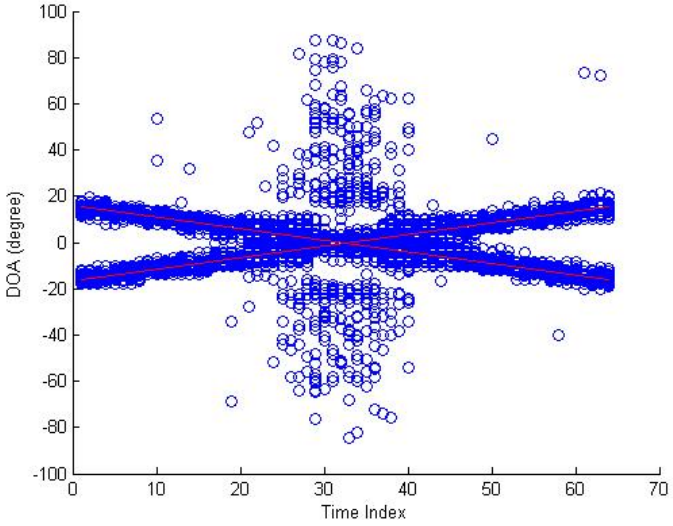


Figure 5.9: Estimated Trajectories by C-HiLasso

$[0, 2\pi]$. $K = 360$ candidate directions constitute the set $\{-90°, -89.5°, ..., 89.5°\}$ with a uniform separation of $0.5°$. These candidate directions are divided into $G = 90$ blocks with $L_g = 4$ for $\forall g$.

Then, following the idea of the sparse representation model of L_1-SVD and SpSF, the array snapshots y_t can be written as

$$y_t = \mathbf{A}x_t + e_t, \tag{5.23}$$

where $\mathbf{A} = [a(\phi_1), a(\phi_2), ..., a(\phi_K)]$ and ϕ_k $k = 1, ..., K$ are the candidate directions. Unlike the usual settings for sparsity-exploiting DOA estimation methods, the sparsity patterns of x_t are time dependent since we assume that the sources are moving in the DOA domain. In [3], SPICE is applied to the moving-source-case, where it uses the array snapshots within a short time window to calculate a sample covariance matrix and then obtains the DOA estimates for the corresponding time index based on this sample covariance matrix. Although other sparsity-exploiting DOA estimation methods, e.g. L_1-SVD and SpSF, can deal with the moving sources in a similar way, this short time window technique conflicts with the basic assumption of consistent sparsity patterns and therefore leads to non-sparse estimates of the models and even biased estimates of the DOAs, as shown in simulation examples of [3]. Since its formulation constructively

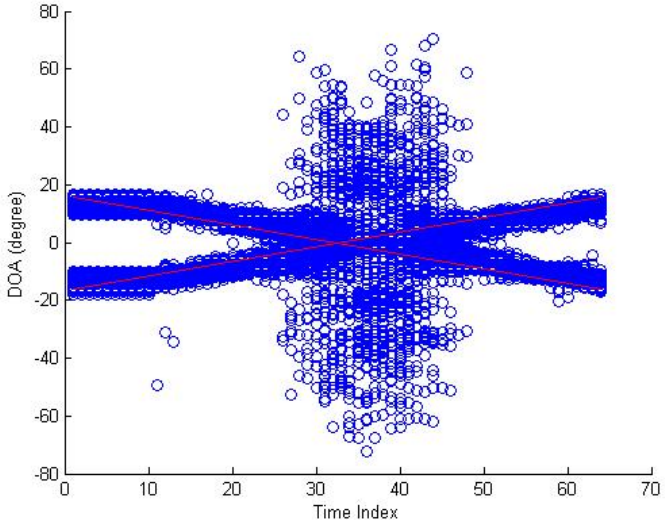


Figure 5.10: Estimated Trajectories by SPICE

utilizes inconsistent sparsity patterns of x_t, we expect PABS to have advantages over these methods in estimating the DOAs of the moving sources.

In the following, we use 100 independent trials (with 64 snapshots available in each trial) to compare the estimation performance of PABS to that of C-HiLasso and SPICE. The length of the time window used by SPICE is set to 10; thus, starting from the 10^{th} snapshot, it updates its DOA estimates every one snapshot with 20 iterations. The DOA estimates corresponding to the first 9 snapshots of SPICE are simply set to be that of the 10^{th} snapshot. Because the moving ranges of the sources cross many blocks of the candidate directions, we solve PABS and C-HiLasso every 4 snapshots to ensure the block-level sparsity of their models. Therefore, in each solving of these two methods, $T = 4$. Note that we can solve them for every $T > 4$ snapshots by using larger DOA blocks. This, however, leads to lower Block-Level Sparsity. The regularization parameters of PABS and C-HiLasso are independently and empirically chosen and kept constant through the simulations.

Figure 5.8, 5.9 and 5.10 illustrate the trajectories estimated by the three methods, with the red lines indicating the true trajectories of the two sources and the blue dots indicating the estimated DOAs at each time index. First, as discussed before, the DOA

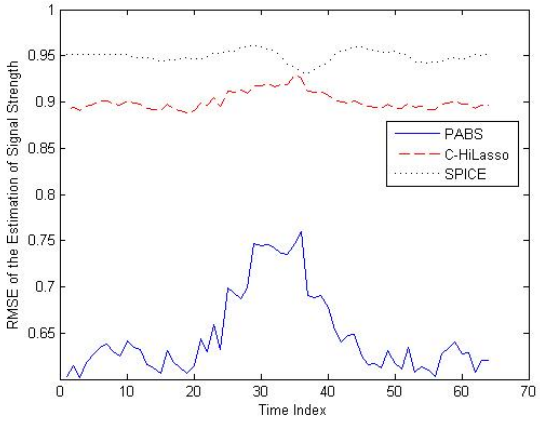


Figure 5.11: RMSE of the Estimation of Signal Strength

estimates of SPICE are biased, while that of the other two are not. Further, when the two sources are well separated ($t < 15$ or $t > 50$), the bands of the blue dots of the three methods have very similar widths, which are indicators of the variances of the DOA estimates. But when the two sources are very close to each other (around time index 32), PABS shows significantly better performance than the other two, in terms of the number of outlier DOA estimates. This improved performance is achieved by PABS because it encourages the persistent activity property, which becomes crucial especially when it becomes hard for the array to resolve the two sources. It should be noted that, although SPICE performs worse than the other methods in this case, it is free of regularization parameter selection and computationally less demanding.

Since sometimes the estimation of the signal strength is also a concern, we show in Figure 5.11 the RMSEs of the signal strength estimation of the three methods versus time index (or equivalently the separation in DOA domain). According to Figure 5.11, PABS has much better signal strength estimation performance than the other two and, not surprisingly, its performance degrades when the two sources are very close to each other (around time index 32).

Chapter 6

Conclusion

6.1 Conclusion

In this thesis, we considered the problem of Direction-Of-Arrival (DOA) estimation using sensor arrays. Following the idea of L_1-SVD [40], a sparse representation model is constructed for the spatial covariance matrix of array snapshots, based on which the DOA estimation target is fulfilled by utilizing sparse signal recovery algorithms. This new framework for DOA estimation is attractive because of its computational complexity, resolution capability and estimation performance, especially at relatively low SNRs. We focused on the algorithm development, performance analysis, regularization parameter selection and more general extensions of such sparsity-exploiting DOA estimation methods.

Assuming a grid of candidate directions and on-grid DOAs, a sparse representation model, based on the extended array snapshots, is established for the spatial covariance matrix (or its estimate, sample covariance matrix) of possibly correlated sources. By applying a L_1-Norm-based sparse signal reconstruction algorithm to this model, SpSF-C is obtained as an optimization problem, which can estimate DOAs, signal powers and correlation factors of arbitrarily correlated signals. After simplifying its sparse representation model, SpSF is proposed as a version of SpSF-C specialized for uncorrelated signals. SpSF retains the advantage of SpSF-C in estimating both DOAs and signal powers, but is much more computationally efficient than SpSF-C. Although SpSF is proposed for uncorrelated signals, its application to the cases of correlated signals is

discussed, where the weighted spatial smoothing and forward-backward smoothing are employed to enhance its performance. Through simulations, it is shown that, when signals are uncorrelated, SpSF can perform similarly to L_1-SVD and provide $5dB$ improvement of large error threshold over MUSIC. Further, when the sources are highly correlated, SpSF can achieve a DOA estimation performance comparable to that of SpSF-C at a much lower computational complexity. It is also worth noting that SpSF and SpSF-C both perform remarkably better than L_1-SVD and MUSIC.

Although the formulation of SpSF is very similar to the ordinary sparse signal reconstruction (SSR) algorithms, its additional positivity constraints make it able to perform much better than predicted by the theoretical analysis in the literature of SSR. In this thesis, analyses and conditions of the existence and uniqueness of Model Selection Consistent (MSC) solutions of SpSF are provided, which further lead to the discussion of the maximum number of sources that SpSF can reliably work with. Our results show that such a number is dependent on the observation matrix $\mathbf{A}_K$ of SpSF and is hard to be related to the number of sensors for arbitrary array. However, as an exception, we provided the explicit formula describing the relationship between the maximum number of sources and the number of sensors for the widely used Uniform-Linear-Array (ULA), and this formula shows that SpSF can resolve the same number of sources as MUSIC and L_1-SVD. Further, it is recognized that this maximum number is actually the degree of freedom of the co-array; hence, by using nonuniform linear array, SpSF can work with many more sources than the number of sensors. Based on this uniqueness result, the asymptotic estimation consistencies of SpSF with respect to infinitely many snapshots and sensors, respectively, are proved on the basis of the optimality conditions of MSC solutions.

Inspired by the previous theoretical analyses, Diagonal Loading is utilized to reduce the sensitivity of SpSF with respect to its regularization parameter, and this leads to a new formulation of SpSF. Although this new formulation has an extra parameter, it is very easy to select and enables the control of the sensitivity of SpSF with respect to the other parameter. An iterative version of SpSF, Iterative SpSF (I-SpSF), is proposed as a parameter-free sparsity-exploiting DOA estimation algorithm, which starts with a fixed initialization of the regularization parameter β of SpSF and alternates between solving SpSF and refining estimates of good β. Simulation results show that I-SpSF can

retain the performance of SpSF using an exhaustively searched but fixed regularization parameter and may perform even better than SpSF at high SNRs. Based on the analyses of the optimality conditions for SpSF, an upper bound on the probability of correct support recovery of SpSF, evaluated by Monte Carlo simulations, is proposed. As shown by simulation examples, this upper bound may well approximate the behavior of SpSF's probability of correct DOA estimate as a function of the regularization parameter. Using this upper bound, an automatic selector of the regularization parameter of SpSF is proposed for the purpose of estimating the DOAs when the number of snapshots is finite. The effectiveness and robustness (with respect to its parameters) of the automatic selector are illustrated through examples. It is shown that this automatic selector can help SpSF achieve almost the same performance as exhaustively searched and fixed regularization parameters. The application of this selector (used by SpSF) to the cases of correlated sources is considered, and its performance is illustrated by a numerical example.

The extension of SpSF to the more general cases of off-grid DOAs and moving sources is considered. By adding extra parameter (variables) describing the modeling uncertainty caused by the off-grid DOAs into the sparse representation model of SpSF, the group sparsity between the spatial spectrum and such additional variables is recognized. Further, using Group Lasso to exploit such group sparsity, SpSF with Modeling Uncertainty (SpSFMU) is formulated as a convex optimization problem which incorporates off-grid DOAs into its sparse representation model and can provide continuous DOA estimates. When off-grid DOAs are present, SpSFMU can be used either to improve the estimation performance of SpSF or reduce its computational complexity. The diagonal loading technique is also applied to SpSFMU to reduce its sensitivity with respect to its regularization parameter. In order to estimate the DOAs of moving sources, we start with the introduction of Block-Level Sparsity and compare it to Element-Level and Group-Level Sparsities. Then, a general sparse model estimator, Persistently Active Block Sparsity (PABS), is proposed, which exploits sparsity in multiple observations vectors and takes inconsistent models into consideration. Based on a new objective function, PABS is formulated as a convex optimization problem which has only one regularization parameter and promotes both Block-Level and Element-level Sparsities, as well as the persistent activity. A Singular-Value-Decomposition-based method to reduce

the computational complexity of this class of algorithms is also presented. PABS is then applied to the DOA estimation of moving sources, and simulation results are presented to illustrate its performance in DOA and signal power estimation.

6.2 Future Research

• **Computational Complexity Reduction of SpSF-C**

Although in Subsections 2.4.3 and 4.4.5 SpSF is shown to achieve similar DOA estimation performance as SpSF-C, the spatial smoothing techniques utilized by SpSF significantly reduces the number of (correlated) sources whose DOAs can be estimated by SpSF. Further, the capability of SpSF-C in estimating the correlation factors can be necessary in some applications. Therefore, a "simplified version" of SpSF-C, which can work with more correlated sources than SpSF but has much less computational complexity than SpSF-C, is highly desirable. The computational complexity of SpSF-C is very large because its dictionary contains the atoms corresponding to the source correlations between any two candidate directions. So, in order to obtain such a desired estimator, adaptive dictionary may be worth investigation which, similar as Capped-L_1[19], adaptively includes some candidate atoms into its dictionary during the iterations.

• **Estimation Consistency of SpSF with Respect to M**

In Subsection 3.2.2, we proved the asymptotic consistency of SpSF with respect to infinitely large M, under the assumption of fixed candidate directions and the number of sources. Also, it was shown that the existing theoretical results from sparse signal reconstruction literature cannot support such asymptotic consistency of SpSF if the number of candidate directions and the number of sources grow to infinity along with M. However, it is still possible that SpSF has such consistency because of its additional positivity constraints and the special structure of its atoms which are the vectorized version of $a(\theta)a^H(\theta)$. Thus, it is of theoretical interest to prove that SpSF does or does not have such asymptotic consistency and, if it has, the corresponding conditions is of importance since they may suggest "good" array configurations and the selection of candidate directions. Further, these conditions have to be different from those in the existing theoretical analyses and the ease of checking these conditions is important to practical applications of SpSF.

• **Application of Non-convex Regularization Methods**

As mentioned in Subsection 3.2.1, the true spatial spectrum of more than $N_A/2$ sources is still inside the solution set of SpSF if the number of snapshots is sufficiently large and the regularization parameter is carefully chosen. In order to distinguish such desired solution from the solution set, one can use non-convex approximations of L_1-Norm, e.g. L_g-norm with $0 < g < 1$ [18] and Capped-L_1 [19]. However, through simulations, we found that direct application of these methods cannot produce satisfactory results. For example, in order to reach a "good" solution, the magnitude threshold of Capped-L_1 usually has to be changed for each iteration, but a systematic (and automatic) way of deciding the amount of change is still lacking. Further, how to utilize these non-convex methods to estimate the DOAs of $L > N_A/2$ sources in the cases of finite snapshots is of more interest and some modifications to these methods may be necessary to improve their robustness with respect to off-grid DOAs and calibration errors etc.

• **Calculation of N_A**

Since $N_A/2$ is the maximum number of sources which SpSF can reliably work with, it is very important to decide N_A for the array of interest (and the pre-given selection of candidate directions). If an explicit formula of N_A or a computationally efficient method of determining N_A are not feasible, one should either find a method to roughly but computationally efficiently estimate N_A for any given observation matrix $\mathbf{A}_K$ or decide N_A through exhaustive search for the array of interest (or some widely used array geometries). Along with that, it is desirable to understand how the candidate directions should be selected so that the corresponding $\mathbf{A}_K$ can have its maximum N_A and hence SpSF is capable to resolve the maximum number of sources.

• **Parameter Selection for Regularization Methods**

Although the automatic parameter selector of Section 4.3 is proposed especially for SpSF, it is possible to apply those ideas to some other (L_1-Norm) regularization methods, e.g. LASSO. Specifically, by examining the differences between the optimality conditions of the Model Selection Consistent solutions and the solutions leading to correct feature estimates (referred to as CFE solution), a relaxation of the optimality conditions of MSC solutions can be obtained by using randomly generated indices to replace the unknown support of CFE solutions. Then, rough estimates of signal and

noise distributions can be used to perform the Monte Carlo evaluation process. For each candidate regularization parameter, such evaluation process provides an estimate of the probability that this parameter can lead to CFE solution. The candidate regularization parameter with the largest such probability should then be used by the corresponding regularization method. The possible difficulties of this application can lie in the analyses of the optimality conditions and the formulation of the aforementioned relaxation.

• **Further Extensions of SpSF**

In Chapter 5 the off-grid DOAs and moving sources are considered and the corresponding extensions of SpSF are presented. However, there are some other cases which can cause modeling error to SpSF and are very common in practical applications. For example, the calibration error can have similar effect as off-grid DOAs and its modeling error can be described by (5.4). But in contrast, such modeling error cannot be described by simple and explicit formulas and parameterized by very few variables. Thus, the idea of linear approximation of SpSFMU cannot be applied for calibration error. In reference [60], another extension of SpSF is proposed especially for arbitrary calibration error. Unfortunately, the computational complexity of this method is significantly larger than that of SpSFMU and therefore a similar but constitutionally efficient extension (for calibration error) of SpSF may be of interest. Non-ideal or directional selective sensor is another example causing modeling error to SpSF. This modeling error can also be described by (5.4) but a convex problem formulation is needed for the corresponding extension of SpSF (since (5.5) is non-convex and the linear approximation idea is also not applicable).

References

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

[2] P. Sprechmann, I. Ramirez, P. Cancela, and G. Sapiro. Collaborative sources identification in mixed signals via hierarchical sparse modeling. In *in Proc. IEEE Acoust. Speech and Signal Process.*, Prague, May 2011.

[3] P. Stoica, P. Babu, and J. Li. Spice: A sparse covariance-based estimation method for array processing. *IEEE Trans. on Signal Process.*, 59(2):629–638, Feb. 2011.

[4] H. Tsuji, J. Zheng, and M. Kaveh. Experiments in radio location estimation using an airbone array. pages 189–193, July 2008.

[5] H.L. Van Trees. *Optimum Array Processing.* John Wiely, 1st edition edition, 2003.

[6] H. Krim and M. Viberg. Two decades of array signal processing research. the parametric approach. *IEEE Trans. on Signal Process. Mag.*, 13:67–94, July 1996.

[7] J. Capon. High resolution frequency-wavenumber spectrum analysis. *Proc. IEEE.*, pages 1408–1418, Aug. 1969.

[8] R.O. Schmidt. Multiple emitter location and signal parameter estimation. In *Proc. RADC Spectrum Estimation Workshop*, pages 243–258, Rome, NY, Sept. 1979.

[9] H. Wang and M. Kaveh. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *IEEE Trans. ASSP.*, 33(4):823–831, Aug. 1985.

[10] K.D. Mauck. Wideband cyclic music. In *in Proc. IEEE Acoust. Speech and Signal Process.*, volume 4, pages 288–291, April 1993.

[11] A.J. Barabell. Improving the resolution performance of eigenstructurebased direction-finding algorithms. In *in Proc. IEEE Acoust. Speech and Signal Process.*, pages 336–339, April 1983.

[12] P. Stoica and A. Nehorai. Music, maximum likelihood and cramer-rao bound. *IEEE Trans. Acoust., Speech Signal Process.*, 33:720–741, May 1989.

[13] M. Kaveh and A.J. Barabell. The statistical performance of the music and the minimum-norm algorithms in resolving plane waves in noise. *IEEE Trans. ASSP*, 34(2):331–341, April 1986.

[14] D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable recovery of sparse overcomplete representation in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1):6–18, Jan. 2006.

[15] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Scientific Computing*, 20(1):33–61, 1998.

[16] A.N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Doklady Akademii Nauk SSSR*, 151:501–504, 1963.

[17] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, Nov. 1992.

[18] S. Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leqslant 1$. *Appl. Computat. Harmon. Anal.*, 26(3):395–407, 2009.

[19] T. Zhang. Multi-stage convex relaxation for learning with sparse regularization. In *In Proc. Adv. Neural Inf. Process. Syst.*, pages 1929–1936, 2008.

[20] H. Zou. The adaptive lasso and its oracle property. *J. Am. Statist. Ass.*, 101(476):1418–1429, 2006.

[21] E.J. Candes. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Paris, Series I*, 346:589–592, 2008.

[22] T.T. Wu and K. Lange. Coordinate descent algorithms for lasso penalized regression. *Annals of Applied Statistics*, 2(1):224–244, 2008.

[23] E. Candsand J. Romberg and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, Aug. 2006.

[24] R.G. Baraniuk. Compressive sensing. *IEEE Signal Processing Mag.*, 24(4):118–120, 124, July 2007.

[25] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. In *Proc. Natl, Acad. Sci.*, volume 100, pages 2197–2202, USA, 2003.

[26] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, Nov. 2001.

[27] M.R. Osbirne, B. Presnell, and B.A. Turlach. On the lasso and its dual. *Journal of Computational and Graphical Statistics*, 9(2):319–337, 2000.

[28] P. Zhao and B. Yu. On model selection consistency of lasso. *J. Machine Learning Research*, pages 2541–2563, Dec. 2006.

[29] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of Royal Statistical Society*, 68:49–67, 2006.

[30] J. Friedman, T. Hastie, and R. Tibshirani. A note on the group lasso and a sparse group lasso. Technical report, Department of Statistics, Stanford University, Jan. 2010. preprint arXiv:1001:0736v1.

[31] V. Roth and B. Fischer. The group-lasso for generalized linear models: Uniqueness of solutions and efficient algorithms. In *Proc. Int. Conf. Mach. Learn.*, pages 848–855, Helsinki, Finland, 2008.

[32] F.R. Bach. Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Research*, 9:1179–1225, June 2008.

[33] Y. Nardi and A. Rinaldo. On the asymptotic properties of the group lasso estimator for linear models. *Electronic Journal of Statistics*, 2:605–633, 2008.

[34] H. Wang and C. Leng. A note on adaptive group lasso. *Comput. Statist. Data Anal.*, 52:5277–5286, 2008.

[35] I.F. Gorodnitsky and B.D. Rao. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *IEEE Trans. ASSP.*, 45(3):600–616, Mar. 1997.

[36] B.D. Rao and K. Kreutz-Delgado. An affine scaling methodology for best basis selection. *IEEE Trans. Signal Process.*, 47(1):187–200, Jan. 1999.

[37] B.D. Jeffs. Sparse inverse solution methods for signal and image processing applications. In *in Proc. IEEE Acoust. Speech and Signal Process.*, volume 3, pages 1885–1888, 1998.

[38] M.D. Sacchi, T.J. Ulrych, and C.J. Walker. Interpolation and extrapolation using a high-resolution discrete fourier transform. *IEEE Trans. Signal Process.*, 46(1):31–38, Jan. 1998.

[39] J.J. Fuchs. On the application of global matched filter to doa estimation with uniform circular arrays. *IEEE Trans. Signal Process.*, 47(1):187–200, Jan. 1999.

[40] D.M. Malioutov, M. Cetin, and A.S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.*, 53(8):3010–3022, Aug. 2005.

[41] D. Model and M. Zibulevsky. Signal reconstruction in sensor arrays using sparse representations. *Signal Processing*, 86(3):624–638, Mar. 2006.

[42] T. Yardibi, J. Li, P. Stoica, and L.N. Cattafesta. Sparsity constrained deconvolution approaches for acoustic source mapping. *Journal of the Acoustical Society of America*, 123(5):2631–2642, Feb. 2008.

[43] J.S. Picard and A.J. Weiss. Direction finding of multiple emitters by spatial sparsity and linear programming. In *Communications and Information Technology (ISCIT)*, pages 1258–1262, Incheon, Korea, Dec. 2009.

[44] Z. Liu, Z. Huang, and Y. Zhou. Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation. *IEEE Trans. on Signal Process.*, 59(9):4256–4270, Sep. 2011.

[45] M.A. Herman and T. Strohmer. General deviants: An analysis of perturbations in compressed sensing. *IEEE Journal of Selected Topics in Signal Process.: Special Issue on Compressive Sensing*, April 2010.

[46] H. Zhu, G. Leus, and G.B. Giannakis. Sparsity-cognizant total least-squares for perturbed compressive sampling. *IEEE Trans. on Signal Process.*, 59(5):2002–2016, 2011.

[47] Z. Yang, L. Xie, and C. Zhang. Off-grid direction of arrival estimation using sparse bayesian inference. Available: http://arxiv.org/abs/1108.5838, 2011.

[48] S.U. Pillai and B.H. Kwon. Forward/backward spatial smoothing techniques for coherent signal identification. *IEEE Trans. Acoust. Speech Signal Process.*, 37(1):8–15, Jan. 1989.

[49] K. Takao and N. Kikuma. An adaptive array utilizing an adaptive spatial averaging technique for multipath environments. *IEEE Trans. Antennas Propagat.*, AP-35:1389–1396, Dec. 1987.

[50] A. Paulraj, V.U. Reddy, and T. Kailath. Analysis of signal cancellation due to multipath in optimum beamformers for moving arrays. *IEEE J. Ocean. Eng.*, OE-12:163–172, Jan. 1987.

[51] K.J. Raghunath and V.U. Reddy. A note on spatially weighted subarray covariance averaging schemes. *IEEE Trans. Antennas Propagat.*, 40(6):720–723, June 1992.

[52] J. Dai and Z. Ye. Spatial smoothing for direction of arrival estimation of coherent signals in the presence of unknown mutual coupling. *IET Signal Processing J.*, 5(4):418–425, July 2011.

[53] H.L. Van Trees. *Optimum Array Processing.* John Wiely, 1st edition edition, 2003.

[54] P. Pal and P.P. Vaidyanathan. Nest arrays: A novel approach to array processing with enhanced degrees of freedom. *IEEE Trans. on Signal Processing*, 58(8):4167–4184, Aug. 2010.

[55] Y.I. Abramovich, D.A. Gray, A.Y. Gorokhov, and N.K. Spencer. Positive-definite toeplitz completion in doa estimation for nonuniform linear antenna arrays. i. fully augmentable arrays. *IEEE Trans. on Signal Processing*, 46(9):2458–2471, Sep. 1998.

[56] A.B. Gershman. A note on most favorable array geometries for doa estimation and array interpolation. 4(8):232–235, Aug. 1997.

[57] S.U. Pillai, Y. Bar-Ness, and F. Haber. A new approach to array geometry for improved spatial spectrum estimation. *Proc. IEEE*, 73:1522–1524, Oct. 1985.

[58] J.S. Picard and A.J. Weiss. Error bounds for convex parameter estimation. *Signal Process.*, 92(5):1328–1337, Dec. 2011.

[59] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. C-hilasso: A collaborative hierachical sparse modeling framework. In *http://arxiv.org/abs/1003.0400*, June 2010.

[60] W. Shi, J. Zheng, M. Kaveh, and J. Huang. Robust sparse spectral fitting element and beam spaces for directions-of-arrival and power estimation. In *in Proc. IEEE Acoust. Speech and Signal Process. (ICASSP'12)*, Kyoto, Japan, April 2012.