TECHNISCHE
UNIVERSITÄT
DARMSTADT

Sparse Array Signal Processing

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

Huiping Huang, M.Sc.

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Hing Cheung So

Darmstadt, 2023

Huang, Huiping --- *Sparse Array Signal Processing*
Darmstadt, Technischen Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2023
URN: urn:nbn:de:tuda-tuprints-231214
Tag der mündlichen Prüfung: 19.01.2023

Acknowledgments

I sincerely thank Prof. Abdelhak M. Zoubir for opening my journey to the field of sparse array signal processing, and for giving me sufficient freedom on my research topics. I thank him for his helpful suggestions on not only my current research but also my academic career. I believe his passionate attitude towards teaching, research, scientific community service, etc. will have a profound and lasting impact on my future.

I thank Prof. Hing Cheung So for introducing me to Prof. Zoubir in 2018, and for being my co-supervisor and providing valuable feedback on my work. I have learnt a lot from his useful and detailed answers to my every question about various aspects. Thank Prof. So for always offering his guidance and support, both professionally and personally, for making my doctoral studies as pleasurable as possible.

I thank Prof. Bhavani Shankar, Dr. Linlong Wu, and Tong Wei at the University of Luxembourg for the pleasant collaboration, and thank Prof. Bhavani Shankar for supporting my research stays and visits in his group in Luxembourg. My thanks go also to Dr. Markus Lazanowski, Carina Schuster, and Christian Schmitt from the Graduate School of Excellence Computational Engineering, and the staff of the German Academic Exchange Service. I further thank Prof. Marius Pesavento, Prof. Björn Scheuermann and Prof. Stefan Niessen for being on my committee.

I wish to express my appreciation to all of my current and former Signal Processing Group colleagues and the members of the Communication Systems Group: Adrian Šošić, Tim Schäck, Mark Ryan Leonard, Freweyni Teklehaymanot, Michael Fauss, Ann-Kathrin Seifert, Di Jin, Sergey Sukhanov, Afief Dias Pambudi, Michael Muma, Christian Debes, Mahmoud El-Hindi, Martin Gölz, Pertami Kunz, Jasin Machkour, Dominik Reinhard, Christian Schroth, Aylin Tastan, Amare Kassaw, Ziliang Qiao, Julian Weick, Toufik Mouchini, Ibrahim Kakouche, Hauke Fath, Prof. Dr.-Ing. Henning Puder, Prof. Roy Howard, Tianyi Liu, Yufan Fan, David Schenck, Prof. Dr.-Ing. Marius Pesavento. Thank you all for providing a friendly environment, for offering me reliable support at work, for sharing so many scientific or funny talks. Special thanks to Renate Koschella for her constant support and comfort throughout all these years, and to Rojin Ulusoy for her constant support in the last year of my PhD.

Last but not least, I thank my families and friends in China: Jinshui Huang, Yuefen Hong, Luanyin Huang, Shuqing Huang, Yuzhu Huang, Jiahui Huang, Jiaming Huang, Yiwang Sun, and Zhuomin Zhao, for their unconditional love and support.

Darmstadt, January 2023

Kurzfassung

Diese Dissertation beschreibt drei Ansätze zur Richtungsschätzung (DOA) oder Beamforming in der Array-Signalverarbeitung aus der Perspektive der Sparsity. Im ersten Teil dieser Dissertation betrachten wir das Design von Sparse-Array-Beamformern basierend auf der Alternating Direction Method of Multipliers (ADMM); im zweiten Teil dieser Dissertation wird das Problem der gemeinsamen DOA-Schätzung und der Detektion von gestörten Sensoren untersucht; und Gitter unabhängige DOA-Schätzung wird im letzten Teil dieser Dissertation untersucht.

Im ersten Teil dieser Arbeit entwickeln wir einen Sparse-Array-Designalgorithmus für adaptives Beamforming. Unsere Strategie basiert darauf, ein sparses Beamformer-Gewicht zu finden, um das Ausgangssignal-zu-Interferenz-plus-Rausch-Verhältnis (SINR) zu maximieren. Das vorgeschlagene Verfahren verwendet das ADMM und lässt Lösungen in geschlossener Form bei jeder ADMM-Iteration zu. Die Konvergenzeigenschaften des Algorithmus werden analysiert, indem die Monotonie und Beschränktheit der erweiterten Lagrange-Funktion gezeigt werden. Außerdem beweisen wir, dass der vorgeschlagene Algorithmus gegen die Menge der stationären Punkte von Karush-Kuhn-Tucker konvergiert. Die numerischen Ergebnisse zeigen seine hervorragende Leistung, die mit der des erschöpfenden Suchansatzes vergleichbar ist, etwas besser als die der State-of-the-Art-Löser und mehrere andere Sparse-Array-Designstrategien in Bezug auf das Ausgabe-SINR deutlich übertrifft. Darüber hinaus übertrifft der vorgeschlagene ADMM-Algorithmus seine Konkurrenten hinsichtlich der Rechenkosten.

Gestörte Sensoren könnten zufällig auftreten und zum Ausfall eines Sensor-Array-Systems führen. Im zweiten Teil dieser Arbeit betrachten wir ein Array-Modell, in dem eine kleine Anzahl von Sensoren durch unbekannte Sensorverstärkung und Phasenfehler gestört wird. Mit einem solchen Array-Modell wird das Problem der gemeinsamen DOA-Schätzung und der Detektion von gestörten Sensoren im Rahmen der Low-Rank- und Row-Sparse-Zerlegung formuliert. Wir leiten ein Verfahren der iterativ-neugewichtete kleinste Quadrate (IRLS) her um das resultierende Problem zu lösen. Die Konvergenzeigenschaft des IRLS-Algorithmus wird anhand der Monotonie und Beschränktheit der Zielfunktion analysiert. Es werden umfangreiche Simulationen hinsichtlich Parameterauswahl, Konvergenzgeschwindigkeit, Rechenkomplexität und Leistung der DOA-Schätzung, sowie der Erkennung von gestörten Sensoren durchgeführt. Obwohl der IRLS-Algorithmus bei der Erkennung der gestörten Sensoren etwas schlechter als der ADMM ist, zeigen die Ergebnisse, dass unserer Ansatz mehrere hochmoderne Techniken in Bezug auf Konvergenzgeschwindigkeit, Rechenaufwand und DOA-Schätzleistung übertrifft.

Im letzten Teil dieser Arbeit wird das Problem der Gitter unabhängigen DOA-Schätzung untersucht. Wir entwickeln eine Methode, um gemeinsam die nächsten *Ortsfrequenz*-Gitter (den Sinus der DOA) und die Lücken zwischen den geschätzten Gittern und den entsprechenden Frequenzen zu schätzen. Unter Verwendung einer Taylor-Approximation zweiter Ordnung wird das Datenmodell im Rahmen der Joint-Sparse-Darstellung formuliert. Wir weisen auf eine wichtige Eigenschaft der interessierenden Signale im Modell hin, nämlich die Proportionalitätsbeziehung. Es hat sich empirisch gezeigt, dass die Proportionalitätsbeziehung insofern nützlich ist, da sie die Wahrscheinlichkeit erhöht, dass die Mischmatrix die blockbeschränkte Isometrieeigenschaft erfüllt. Simulationsbeispiele demonstrieren die Effektivität und Überlegenheit der vorgeschlagenen Methode gegenüber mehreren hochmodernen gitterbasierten Ansätzen.

Abstract

This dissertation details three approaches for direction-of-arrival (DOA) estimation or beamforming in array signal processing from the perspective of sparsity. In the first part of this dissertation, we consider sparse array beamformer design based on the alternating direction method of multipliers (ADMM); in the second part of this dissertation, the problem of joint DOA estimation and distorted sensor detection is investigated; and off-grid DOA estimation is studied in the last part of this dissertation.

In the first part of this thesis, we devise a sparse array design algorithm for adaptive beamforming. Our strategy is based on finding a sparse beamformer weight to maximize the output signal-to-interference-plus-noise ratio (SINR). The proposed method utilizes ADMM, and admits closed-form solutions at each ADMM iteration. The algorithm convergence properties are analyzed by showing the monotonicity and boundedness of the augmented Lagrangian function. In addition, we prove that the proposed algorithm converges to the set of Karush-Kuhn-Tucker stationary points. Numerical results exhibit its excellent performance, which is comparable to that of the exhaustive search approach, slightly better than those of the state-of-the-art solvers, and significantly outperforms several other sparse array design strategies, in terms of output SINR. Moreover, the proposed ADMM algorithm outperforms its competitors, in terms of computational cost.

Distorted sensors could occur randomly and may lead to the breakdown of a sensor array system. In the second part of this thesis, we consider an array model in which a small number of sensors are distorted by unknown sensor gain and phase errors. With such an array model, the problem of joint DOA estimation and distorted sensor detection is formulated under the framework of low-rank and row-sparse decomposition. We derive an iteratively reweighted least squares (IRLS) algorithm to solve the resulting problem. The convergence property of the IRLS algorithm is analyzed by means of the monotonicity and boundedness of the objective function. Extensive simulations are conducted in view of parameter selection, convergence speed, computational complexity, and performance of DOA estimation as well as distorted sensor detection. Even though the IRLS algorithm is slightly worse than the ADMM in detecting the distorted sensors, the results show that our approach outperforms several state-of-the-art techniques in terms of convergence speed, computational cost, and DOA estimation performance.

In the last part of this thesis, the problem of off-grid DOA estimation is investigated. We develop a method to jointly estimate the closest *spatial frequency* (the sine of DOA) grids, and the gaps between the estimated grids and the corresponding frequencies.

By using a second-order Taylor approximation, the data model under the framework of joint-sparse representation is formulated. We point out an important property of the signals of interest in the model, namely the proportionality relationship. The proportionality relationship is empirically demonstrated to be useful in the sense that it increases the probability of the mixing matrix satisfying the block restricted isometry property. Simulation examples demonstrate the effectiveness and superiority of the proposed method against several state-of-the-art grid-based approaches.

Contents

Chapter 1
Introduction

In various real-world applications, such as wireless communications, multiple input multiple output (MIMO) radar, radio astronomy, sonar, and bioinformatics, engineers are facing the problem that huge data are obtained and needed to be dealt with. It is observed that many practical signals are compressible in the sense that they can be well approximated by sparse signals [FR13a]. Note that exploiting the sparsity of the data is helpful in storing and processing them. Therefore, sparsity has become a more and more important property in many research fields, see for example [SFH22, CPP13, FGG+14, TYN14, AGM18, FHM+18, SZL14a, ZFWS20, LLWL18, DBXC17, DT15, MSG12, WSP17, YXZ13, LSG17, JYMZS21, TMZ21].

In array signal processing, the techniques of sparsity have been applied for different purposes, mainly including direction-of-arrival (DOA) estimation [BAPW08, LSZN09, MBZJ09, HM10, YXZ13, SZL14b, LAAZ15, YLSX18, SCL+21], beamforming [SZL14a, WATA14, HMP18, DXP21, XLD+21, SFH22, FHM+18, LLWL18, NSYC10, GK14, DT15, MSG13, HA19, ZFWS20, DGB22, ZCS+17, ATSMR+21, ZFW21, CZW+21], source detection [BZP02, PZBL07, DRZA10, SBL11b, CP11, RHE12, YDZ12, LZ13, LZ15, MMKBZ+16], and sensor diagnosis [YL99, MCR+12, ZWCS15, LV19, NLEF09, PGW02, SPPZ18, AA17, HZ21]. In this dissertation, we focus on sparse array beamformer design, DOA estimation using a sparsely distorted sensor array, and distorted sensor detection. Specifically speaking, this dissertation consists of the following three parts:

- We study the problem of sparse array beamformer design by using alternating direction method of multipliers (ADMM). Our goal is twofold. On the one hand, we select less sensors in order to reduce the processing cost. On the other hand, we wish to keep the beamformer output signal-to-interference-plus-noise ratio (SINR) as large as possible.
- We investigate the problem of joint DOA estimation and distorted sensor detection. We consider an array model within which a small number of sensors are distorted by sensor gain and phase uncertainties. The positions of the distorted sensors are random and unknown.
- We study off-grid DOA estimation problem. Since the number of sources is always much less than the number of possible pre-set angular grids, we have sparsity in the whole angular region. By using a second-order Taylor approximation, the data model under the framework of joint-sparse representation is formulated.

1.1 Research Contributions

In this thesis, several advances with respect to sparse array signal processing, are detailed. Our proposed sparse array design strategy admits closed-form solutions at each ADMM iteration. Besides, a convergence analysis of the proposed algorithm is provided by showing the monotonicity and boundedness of the augmented Lagrangian function. Additionally, it is proved that the proposed algorithm converges to the set of Karush-Kuhn-Tucker (KKT) stationary points. Simulation results demonstrate excellent behavior of our scheme, as it outperforms several existing methods, and is comparable to the exhaustive search approach. Moreover, the proposed method consumes much less computing power than the other tested approaches.

Our proposed iteratively reweighted least squares (IRLS) method has good performance in both DOA estimation and distorted sensor detection. We analyzed the convergence property of the algorithm, and show that the solution converges to a KKT point. The limit point is proved to be globally optimal. Moreover, the computational complexity of the IRLS algorithm as well as the singular value thresholding, accelerated proximal gradient, and ADMM methods are theoretically analyzed. Extensive simulations are conducted in view of parameter selection, convergence speed, computational time, and performance of DOA estimation and distorted sensor detection.

Our proposed off-grid DOA estimation algorithm is based on a second-order Taylor approximation. We point out an important property of the signals of interest in the model, namely the proportionality relationship, which is empirically demonstrated to be useful in the sense that it increases the probability of the mixing matrix satisfying the block restricted isometry property. We compare the proposed method with several state-of-the-art grid-based approaches, in terms of DOA estimation accuracy and computational cost. Simulation examples demonstrate the effectiveness and superiority of the proposed method against these competing approaches.

1.2 Thesis Structure

In Chapter 2, necessary background knowledge is presented. First, DOA estimation and beamforming are introduced. Second, standard ADMM iteration steps are reviewed. In Chapter 3, the problem of sparse array beamformer design is studied. We develop a method based on ADMM, which results in closed-form solution at each iteration. In Chapter 4, we consider the problem of joint DOA estimation and distorted sensor detection. It is formulated under the framework of low-rank and row-sparse decomposition. We derive an IRLS algorithm to solve the resulting problem. In Chapter 5, off-grid DOA estimation problem is investigated. We propose a second-order Taylor approximation method. Conclusions and potential for further research are detailed in Chapter 6.

1.3 Publications

The following publications have been produced during the period of doctoral candidacy.

Internationally Refereed Journal Articles

- **H. Huang**, H. C. So, and A. M. Zoubir, "Convergence Analysis of Consensus-ADMM for General QCQP," to be submitted.
- **H. Huang**, H. C. So, and A. M. Zoubir, "Sparse Array Beamformer Design via ADMM," submitted to *IEEE Transactions on Signal Processing*, August 2022.
- **H. Huang**, Q. Liu, H. C. So, and A. M. Zoubir, "Low-Rank and Row-Sparse Decomposition for Joint DOA Estimation and Distorted Sensor Detection," submitted to *IEEE Transactions on Aerospace and Electronic Systems*, August 2022.
- **H. Huang**, H. C. So, and A. M. Zoubir, "Off-Grid Direction-of-Arrival Estimation Using Second-Order Taylor Approximation," *Signal Processing*, vol. 196, pp. 108513, July 2022.

Internationally Refereed Conference Papers

- **H. Huang**, H. C. So, and A. M. Zoubir, "Sparse Array Beamformer Design via ADMM," In *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2022)*, Trondheim, Norway, June 2022.
- **H. Huang** and A. M. Zoubir, "Low-Rank and Sparse Decomposition for Joint DOA Estimation and Contaminated Sensors Detection with Sparsely Contaminated Arrays," In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2021)*, Toronto, Canada, June 2021.

- **H. Huang**, A. M. Zoubir, and H. C. So, "A Compressive Sensing Approach for Single-Snapshot Adaptive Beamforming," In *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2020)*, Hangzhou, China, June 2020.

- **H. Huang**, H. C. So, and A. M. Zoubir, "Extended Cyclic Coordinate Descent for Robust Row-Sparse Signal Reconstruction in the Presence of Outliers," In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020)*, Barcelona, Spain, May 2020.

- **H. Huang**, M. Fauß, and A. M. Zoubir, "Block Sparsity-Based DOA Estimation with Sensor Gain and Phase Uncertainties," In *Proceedings of European Signal Processing Conference (EUSIPCO 2019)*, A Coruña, Spain, September 2019.

Chapter 2

Fundamentals

In this chapter, some basic concepts that are relevant for the thesis are introduced. First, two main tasks in array signal processing, i.e., direction-of-arrival (DOA) estimation and beamforming, are detailed. Then, the alternating direction method of multipliers (ADMM) is briefly introduced, focussing on its most important steps.

2.1 DOA Estimation

The problem of retrieving information conveyed in propagating waves occurs in a wide range of applications including radar, sonar, wireless communications, geophysics and biomedical engineering. Methods for processing data measured by sensor arrays have attracted much attention over last four decades [KV96, CVY14].

Early space-time processing techniques represent DOA in terms of a spatial spectrum. The resulting Fourier transform based conventional beamformer, however, is subject to resolution limitation due to finite array aperture. Similar to its temporal counterpart, the spatial periodogram can not benefit from increasing signal-to-noise ratio (SNR) or number of samples. Better estimates can be achieved by applying a windowing function to reduce spectral leakage effects. The minimum variance distortionless response (MVDR) beamformer [Cap69] overcomes the resolution limitation of Fourier based techniques by formulating the spectrum estimation as a constrained optimization problem. Also, its performance can be enhanced by high SNR.

The multiple signal classification (MUSIC) algorithm [Sch86] is representative of subspace methods based on eigenstructure of the spatial correlation matrix. In addition to high resolution, MUSIC takes advantage of SNR, number of sensors and number of samples. It improves estimation accuracy with respect to all dimensions and is statistically efficient. However, in the presence of correlated source signals, subspace methods degrade dramatically as the signal subspace suffers from rank deficiency.

On the other hand, parametric methods such as the maximum likelihood (ML) approach [Bö86, ZW88] fully exploit the data model, leading to statistically efficient estimators. More importantly, they maintain their good performance in critical scenarios involving signal coherence, closely located signals and low SNRs. The optimal properties come at the price of an increased computational complexity. Hence, efficient implementation is crucial for parametric methods.

More recently, the methods proposed in [Fuc01, MCW05] view DOA estimation as sparse signal recovery and assign DOA estimates to signals with nonzero entry. In this approach, the first step is to find a sparse representation of the array output data. For example, the beamforming output in the frequency domain [Fuc01] or the array observation [MCW05] can be used to construct a sparse data representation. Then, the underlying optimization problem (typically convex) will be solved to find nonzero components. DOA estimates are finally obtained from angles associated with nonzero components.

2.2 Beamforming

Adaptive beamforming is a versatile approach to detect and estimate the signal-of-interest (SOI) at the output of sensor array using data adaptive spatial or spatio-temporal filtering and interference cancellation [VT02]. Being a very central problem of array processing, adaptive beamforming has found numerous application to radar, sonar, radio astronomy, biomedicine, wireless communications, among others. The connection of adaptive beamforming to adaptive filtering is emphasized in [Vor14]. The major differences, however, come from the fact that adaptive filtering is based on temporal processing of a signal, while adaptive beamforming is concerned with spatial processing. The latter also indicates that the signal is sampled in space, i.e., the signal is measured by an array of spatially distributed antenna sensors.

The traditional approach to the design of adaptive beamforming is to maximize the beamformer output signal-to-interference-plus-noise ratio (SINR) assuming that there is no SOI component in the beamforming training data [Vor14]. The data model can be given as follows. Consider a compact uniform linear array (ULA) consisting of M antenna sensors. Denote $\mathbf{x}(t) \in \mathbb{C}^M$ as the observation vector of the array, which can be modeled as:

$$\mathbf{x}(t) = \mathbf{a}(\theta_0)s_0(t) + \sum_{k=1}^{K} \mathbf{a}(\theta_k)s_k(t) + \mathbf{n}(t), \tag{2.1}$$

where $t = 1, 2, \cdots, T$ denotes the time index, with T being the total number of available snapshots, θ_0 and $s_0(t)$ are the DOA and waveform of the SOI, respectively, while θ_k

and $s_k(t)$ denote those of the k-th interference signal. We consider one SOI and K unknown interferers in the data model. In addition, $\mathbf{n}(t) \in \mathbb{C}^M$ stands for a spatially and temporally white zero-mean Gaussian noise vector, and the array steering vector $\mathbf{a}(\theta) \in \mathbb{C}^M$ takes the following form

$$\mathbf{a}(\theta) = [1, e^{-\jmath\pi \sin(\theta)}, \cdots, e^{-\jmath\pi(M-1)\sin(\theta)}]^{\mathrm{T}}. \tag{2.2}$$

For the simplicity of notation, we denote $\mathbf{a}(\theta_k)$ as $\mathbf{a}_k$, for all $k = 0, 1, \cdots, K$, here and subsequently.

The beamformer output is calculated as

$$y(t) = \mathbf{w}^{\mathrm{H}}\mathbf{x}(t), \tag{2.3}$$

in which $\mathbf{w} \in \mathbb{C}^M$ is the beamformer weight vector to be designed. The beamformer output SINR is defined as [Vor14]

$$\mathrm{SINR} = \frac{\sigma_{\mathrm{s}}^2|\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2}{\mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{i+n}}\mathbf{w}}, \tag{2.4}$$

where $\sigma_{\mathrm{s}}^2 = \mathrm{E}\{|s_0(t)|^2\}$ is the power of the SOI, and $\mathbf{R}_{\mathrm{i+n}}$ is the interference-plus-noise covariance matrix, which can be written as

$$\mathbf{R}_{\mathrm{i+n}} = \sum_{k=1}^{K} \sigma_k^2 \mathbf{a}_k\mathbf{a}_k^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}, \tag{2.5}$$

assuming that the interference signals are uncorrelated with the noise. In (2.5), $\sigma_k^2 = \mathrm{E}\{|s_k(t)|^2\}$ is the power of the k-th interference signal, and σ_{n}^2 is the noise power.

Maximizing the output SINR leads to the MVDR beamformer design [Cap69]:

$$\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{i+n}}\mathbf{w} \quad \text{subject to } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 = 1. \tag{2.6}$$

Since $\mathbf{R}_{\mathrm{i+n}}$ is not available in practice, a common approach is, to replace it by the received data covariance matrix $\mathbf{R}_{\mathrm{x}} = \sigma_{\mathrm{s}}^2\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}} + \mathbf{R}_{\mathrm{i+n}}$ which can be easily estimated [SGLW03], as:

$$\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{w} \quad \text{subject to } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1. \tag{2.7}$$

It is worth noting that the equality constraint is relaxed to an inequality one, since the output power of the SOI is included as part of the objective function [HA19, HA21].

The above problems have a closed-form solution as $\mathbf{w}_{\mathrm{opt}} = \mathcal{P}\{\sigma_{\mathrm{s}}^2\mathbf{R}_{\mathrm{i+n}}^{-1}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\} = \mathcal{P}\{\sigma_{\mathrm{s}}^2\mathbf{R}_{\mathrm{x}}^{-1}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\}$, where $\mathcal{P}\{\cdot\}$ denotes the principal eigenvector of its input matrix. Substituting $\mathbf{w}_{\mathrm{opt}}$ into (2.4) yields the corresponding optimum output SINR as

$$\mathrm{SINR}_{\mathrm{opt}} = \frac{\sigma_{\mathrm{s}}^2|\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{a}_0|^2}{\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{R}_{\mathrm{i+n}}\mathbf{w}_{\mathrm{opt}}} = \lambda_{\max}(\sigma_{\mathrm{s}}^2\mathbf{R}_{\mathrm{i+n}}^{-1}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}), \tag{2.8}$$

where $\lambda_{\max}(\cdot)$ is the maximal eigenvalue of the input matrix.

2.3 Alternating Direction Method of Multipliers

ADMM is an algorithm that is intended to blend the decomposability of dual ascent with the superior convergence properties of the method of multipliers [BPC+11]. This section describes the basic steps of ADMM, which are relevant for the main contributions in Chapter 3. ADMM solves problems in the form:

$$\min_{\mathbf{x}\in\mathbb{C}^n} \; f(\mathbf{x}) + g(\mathbf{x}) \qquad \text{s.t. } \mathbf{x} \in \mathcal{X}, \tag{2.9}$$

with $\mathcal{X}$ denoting the feasible set which is assumed to be nonempty.

The standard steps for scaled-form ADMM iterations are as follows.

Step i). We formulate Problem (2.9) by introducing an auxiliary variable $\mathbf{z} \in \mathbb{C}^n$ and settling the original variable $\mathbf{x}$ and the auxiliary variable $\mathbf{z}$ in a separable manner, as

$$\min \; f(\mathbf{x}) + g(\mathbf{z}) \qquad \text{s.t. } \mathbf{x} \in \mathcal{X} \text{ and } \mathbf{z} = \mathbf{x}, \tag{2.10}$$

with variables $\mathbf{x}$ and $\mathbf{z}$.

Step ii). We form the scaled-form augmented Lagrangian function according to Problem (2.10), by dealing with the equality constraints therein, i.e., $\mathbf{z} = \mathbf{x}$, as

$$\mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \triangleq f(\mathbf{x}) + g(\mathbf{z}) + \frac{\rho}{2}(\|\mathbf{z} - \mathbf{x} + \mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2), \tag{2.11}$$

where $\mathbf{u} \in \mathbb{C}^n$ is the scaled dual variable, and $\rho > 0$ is the augmented Lagrangian parameter.

Step iii). The scaled-form ADMM updating equations can be written down by separately solving for the variables, as follows

$$\mathbf{x}_{(k+1)} = \arg\min_{\mathbf{x}\in\mathcal{X}} \; f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{z}_{(k)} - \mathbf{x} + \mathbf{u}_{(k)}\|_2^2 \tag{2.12a}$$

$$\mathbf{z}_{(k+1)} = \arg\min_{\mathbf{z}} \; g(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{z} - \mathbf{x}_{(k+1)} + \mathbf{u}_{(k)}\|_2^2 \tag{2.12b}$$

$$\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)} + \mathbf{z}_{(k+1)} - \mathbf{x}_{(k+1)} \tag{2.12c}$$

where superscript $\cdot_{(k)}$ denotes the corresponding variable at the k-th ADMM iteration.

2.4 Summary

In this chapter, some basic concepts that are relevant for the thesis were introduced. Firstly, two main tasks in array signal processing, including direction-of-arrival (DOA) estimation and beamforming, were introduced in detail. Then, the alternating direction method of multipliers (ADMM) was briefly introduced, focussing on its significant steps.

Chapter 3

Beamforming with a Sparse Array

In this chapter, we devise a sparse array design algorithm for adaptive beamforming. Our strategy is based on finding a sparse beamformer weight to maximize the output signal-to-interference-plus-noise ratio (SINR). The proposed method utilizes the alternating direction method of multipliers (ADMM), and admits closed-form solutions at each ADMM iteration. The algorithms convergence properties are analyzed by showing the monotonicity and boundedness of the augmented Lagrangian function. In addition, we prove that the proposed algorithm converges to a set of Karush-Kuhn-Tucker stationary points. Numerical results exhibit its excellent performance, which is comparable to that of the exhaustive search approach. The proposed algorithm is slightly better than those of the state-of-the-art solvers, including the semidefinite relaxation (SDR), its variant (SDR-V), and the successive convex approximation (SCA) approaches, and significantly outperforms several other sparse array design strategies, in terms of output SINR. Moreover, the proposed ADMM algorithm outperforms the SDR, SDR-V, and SCA methods in terms of computational cost.

The key contributions presented in this chapter originate from [HSZ22b] and [HSZ22c]. The remainder of this chapter is organized as follows: The motivation is given in Section 3.1. The signal model is established in Section 3.2. The proposed approach is presented in Section 3.3 and the convergence analyses are given in Section 3.4. ADMM with re-weighted ℓ_1-norm regularization is proposed in Section 3.5. Section 3.6 shows the simulation results, and Section 3.7 summarizes the chapter.

3.1 Motivation

Adaptive arrays have been widely applied in diverse practical applications, such as radar, sonar, wireless communications, to name just a few [Gab76]. One of their uses is beamforming, which is to extract the signal-of-interest (SOI) while suppressing interference and noise [Vor14]. It has been reported that the performance of beamforming is affected by not only the beamformer weight, but also the array configuration [Lin82]. In this sense, conventional uniform arrays may not be the optimal choices for adaptive beamformer design. On the other hand, sparse arrays achieve increased array aperture and degrees of freedom while reducing the hardware complexity, as compared to conventional uniform arrays. Thus, sparse arrays could be a better option for adaptive beamformer design.

To this end, in recent years, several strategies for designing sparse array beamformers have been proposed [SZL14a, WATA14, HMP18, EM20, DXP21, XLD+21, VCN+21, SFH22, CPP13, FHM+18, LLWL18, NSYC10, GK14, HL14, DT15, MSG12, MSG13, HA19, HA21, ZFWS20, DGB22, ZCS+17, ATSMR+21, ZFW21, YQF+19, WGG21, CZW+21]. These methods can be roughly divided into three categories: Greedy based [SZL14a, WATA14, HMP18], machine learning based [EM20, DXP21, XLD+21, VCN+21, SFH22], and optimization based [SFH22, CPP13, FHM+18, LLWL18, NSYC10, MSG12, MSG13, GK14, HL14, DT15, ZFWS20, HA19, HA21, ZCS+17, ATSMR+21, ZFW21, YQF+19, WGG21, CZW+21, DGB22] approaches. The greedy procedure in [SZL14a, WATA14, HMP18] largely reduces the combinatorial exploration space, but can result in a highly suboptimal solution. Machine learning techniques [EM20, DXP21, XLD+21, VCN+21, SFH22] require prior data for their training step, which might be unavailable in some practical scenarios.

On the other hand, optimization based methods include branch and bound (B&B) [SFH22], mixed-integer programming (MIP) [CPP13, FHM+18, LLWL18], semidefinite relaxation (SDR) [MSG12, MSG13, HA19, HA21, ZFWS20, DGB22], and successive convex approximation (SCA) [ZCS+17, ATSMR+21, ZFW21]. B&B and MIP are capable of finding the global optimum at the cost of a large computational burden. SDR and SCA based methods are computationally expensive when the dimension of the resulting matrix is high [MHG+15, HS16]. Besides, the relaxation nature of SDR usually leads to a solution with a rank not being one, in which case extra post-processing based on randomization is needed [SDL06]. Moreover, note that the SDR methods in [MSG12, MSG13, HA19, HA21] utilize the ℓ_1-norm square instead of the ℓ_1-norm for sparsity promotion. Although the simulation results in these papers and also in Section 3.6.3 of the present chapter demonstrate the usefulness of the SDR-type methods, no theoretical support is available due to the convexity of the Pareto boundary that is not guaranteed (which has also been mentioned in [MSG12] and [MSG13]). On the other hand, another downside of the SCA approach lies in the fact that it requires a feasible starting point, which could be a difficult task on its own [MHG+15].

In this chapter, a sparse array design algorithm based on the alternating direction method of multipliers (ADMM) is devised for adaptive beamforming. The proposed technique admits closed-form solutions at each ADMM iteration. Convergence analyses of the proposed algorithm are provided by showing the monotonicity and boundedness of the augmented Lagrangian function. Additionally, it is proved that the proposed algorithm converges to the set of Karush-Kuhn-Tucker (KKT) stationary points. Simulation results demonstrate excellent behavior of our scheme, as it outperforms several existing methods, and is comparable to the exhaustive search approach.

Our algorithms and theoretical results are developed primarily on the basis of the ideas presented in [HS16] and [HLR16]. Several differences are highlighted as follows.

- Different from [HS16] which used ADMM to solve general quadratically constrained quadratic programming (QCQP) problems, we focus on a specific QCQP problem that arises in sparse array beamformer design. Our problem involves an ℓ_1-norm regularization and thus our solution is sparse, which is not the case in [HS16].

- Another important difference between our work and [HS16] lies in the fact that the latter only provides a weaker convergence result, i.e., *if ADMM converges for their problem, then* it converges to a KKT stationary point, see Theorem 1 in [HS16]. On the other hand, we show stronger convergence results for our algorithm. That is, we first prove the convergence of the proposed algorithm under a mild condition, and then we prove that it converges to a KKT stationary point.

- Note that [HLR16] needed extra assumptions on the Lipschitz gradient continuity as well as the boundedness of their objective function. In our work, we require neither such assumptions nor any other assumptions.

- Since [HLR16] considered general non-convex problems, no explicit expressions for their parameters were derived. On the contrary, as we consider a specific non-convex problem of sparse array beamformer design, the properties of our objective function have been investigated and thus several parameters are given in an explicit manner. See for example the augmented Lagrangian parameter ρ and the strongly convex parameter $\gamma_{\mathbf{v}}$ in Lemma 3.1.

- The results in [HLR16] were based on the augmented Lagrangian function, while we exploit the scaled-form augmented Lagrangian function. This results in significant differences in the following three aspects: i) the proof of the monotonicity of the augmented Lagrangian function, ii) the proof of the property of the point sequence, and iii) the proof that the algorithm converges to the set of KKT stationary points; see the proofs of Lemma 3.1, Theorem 3.2, and Theorem 3.3, respectively.

Notations: In this chapter, bold-faced lower-case and upper-case letters stand for vectors and matrices, respectively. $\mathbf{I}$ denotes the identity matrix of appropriate dimensions,

$\mathbf{1}_M$ is the $M \times M$ all-one matrix, $\mathbf{1}$ and $\mathbf{0}$ are the all-one and all-zero vectors of appropriate lengths, respectively. Superscripts $\cdot^{\mathrm{T}}$, $\cdot^{\mathrm{H}}$ and $\cdot^{-1}$ stand for transpose, Hermitian transpose and inverse operators, respectively. $\mathbb{C}$ is the set of complex numbers, $\Re\{\cdot\}$ returns the real part of its input variable, and $\jmath = \sqrt{-1}$. $\mathrm{E}\{\cdot\}$ denotes expectation. $\mathcal{P}\{\cdot\}$ returns the principal eigenvector of the input matrix, while $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the largest and smallest eigenvalues of the input matrix, respectively. $\|\cdot\|_0$ denotes the ℓ_0-quasi-norm counting the non-zero entries of the input vector, $\|\cdot\|_1$ and $\|\cdot\|_2$ represent the ℓ_1-norm and the ℓ_2-norm of a vector, respectively. Besides, $\mathrm{sign}(\cdot)$, $\odot$, $\oslash$, and $|\cdot|$ stand for the sign function, the multiplication, the division, and the absolute operators, respectively, all in an element-wise fashion. $(\cdot)_+$ stands for the element-wise plus function defined as $(x)_+ \triangleq \max\{x, 0\}$, where $\max\{a, b\}$ returns the maximum value between a and b. $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathrm{H}}\mathbf{y}$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$. $\mathrm{trace}\{\cdot\}$ denotes the trace of a matrix. $\mathbf{X} \succeq 0$ and $\mathbf{X} \succeq \mathbf{Y}$ indicate that $\mathbf{X}$ and $\mathbf{X} - \mathbf{Y}$ are positive semidefinite, respectively. The symbol $\geq$ between two matrices is element-wise larger than or equal to.

3.2 Signal Model

We consider a compact uniform linear array (ULA) consisting of M antenna sensors, where the term *compact* means that the element-spacing of two adjacent antennas is equal to half-wavelength of the incident signals. We refer to the ULA with element-spacing larger than half-wavelength as *sparse ULA*. Denote $\mathbf{x}(t) \in \mathbb{C}^M$ as the observation vector of the compact ULA, which can be modeled as:

$$\mathbf{x}(t) = \mathbf{a}(\theta_0)s_0(t) + \sum_{k=1}^{K} \mathbf{a}(\theta_k)s_k(t) + \mathbf{n}(t), \tag{3.1}$$

where $t = 1, 2, \cdots, T$ denotes the time index, with T being the total number of available snapshots, θ_0 and $s_0(t)$ are the direction-of-arrival (DOA) and waveform of the SOI, respectively, while θ_k and $s_k(t)$ denote those of the k-th interference signal. We consider one SOI and K unknown interferers in the data model. In addition, $\mathbf{n}(t) \in \mathbb{C}^M$ stands for a spatially and temporally white zero-mean Gaussian noise vector, and the array steering vector $\mathbf{a}(\theta) \in \mathbb{C}^M$ takes the form as

$$\mathbf{a}(\theta) = \left[1, e^{-\jmath\pi \sin(\theta)}, \cdots, e^{-\jmath\pi(M-1)\sin(\theta)}\right]^{\mathrm{T}}. \tag{3.2}$$

For the simplicity of notation, we denote $\mathbf{a}(\theta_k)$ as $\mathbf{a}_k$, for all $k = 0, 1, \cdots, K$, here and subsequently.

The beamformer output is calculated as

$$y(t) = \mathbf{w}^{\mathrm{H}}\mathbf{x}(t), \tag{3.3}$$

in which $\mathbf{w} \in \mathbb{C}^M$ is the beamformer weight vector to be designed. The beamformer output signal-to-interference-plus-noise ratio (SINR) is defined as [Vor14]

$$\mathrm{SINR} = \frac{\sigma_{\mathrm{s}}^2|\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2}{\mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{i+n}}\mathbf{w}}, \tag{3.4}$$

where $\sigma_{\mathrm{s}}^2 = \mathrm{E}\{|s_0(t)|^2\}$ is the power of the SOI, and $\mathbf{R}_{\mathrm{i+n}}$ is the interference-plus-noise covariance matrix, which can be written as

$$\mathbf{R}_{\mathrm{i+n}} = \sum_{k=1}^{K} \sigma_k^2 \mathbf{a}_k \mathbf{a}_k^{\mathrm{H}} + \sigma_{\mathrm{n}}^2 \mathbf{I}, \tag{3.5}$$

assuming that the interference signals are uncorrelated with the noise. In (3.5), $\sigma_k^2 = \mathrm{E}\{|s_k(t)|^2\}$ is the power of the k-th interference signal, and σ_{n}^2 is the noise power.

One of the most prevailing strategies for beamformer design is to maximize the output SINR, which leads to the minimum variance distortionless response (MVDR) beamformer design [Cap69]:

$$\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{i+n}}\mathbf{w} \quad \text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 = 1. \tag{3.6}$$

The above problem can be reformulated equivalently by replacing the in practice unattainable $\mathbf{R}_{\mathrm{i+n}}$ by the received data covariance matrix $\mathbf{R}_{\mathrm{x}} = \sigma_{\mathrm{s}}^2\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}} + \mathbf{R}_{\mathrm{i+n}}$ which can be easily estimated [SGLW03], as:

$$\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{w} \quad \text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1. \tag{3.7}$$

It is worth noting that the equality constraint is relaxed to an inequality one, since the output power of the SOI is included as part of the objective function in (3.7) [HA19, HA21].

The above problems have the closed-form solution $\mathbf{w}_{\mathrm{opt}} = \mathcal{P}\{\sigma_{\mathrm{s}}^2\mathbf{R}_{\mathrm{i+n}}^{-1}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\} = \mathcal{P}\{\sigma_{\mathrm{s}}^2\mathbf{R}_{\mathrm{x}}^{-1}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\}$. Substituting $\mathbf{w}_{\mathrm{opt}}$ into (3.4) yields the corresponding optimum output SINR

$$\mathrm{SINR}_{\mathrm{opt}} = \frac{\sigma_{\mathrm{s}}^2|\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{a}_0|^2}{\mathbf{w}_{\mathrm{opt}}^{\mathrm{H}}\mathbf{R}_{\mathrm{i+n}}\mathbf{w}_{\mathrm{opt}}} = \lambda_{\max}(\sigma_{\mathrm{s}}^2\mathbf{R}_{\mathrm{i+n}}^{-1}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}). \tag{3.8}$$

3.3 Sparse Array Beamformer Design

3.3.1 Sparse Beamforming Problem

In this subchapter, we consider the situation where only $L \leq M$ radio-frequency (RF) chains are available [MSG12], and thus only L antennas can be simultaneously utilized for beamformer design. The problem can be formulated as [MSG12, MSG13, HA19, HA21]:

$$\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{w} \ \text{ s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1 \text{ and } \|\mathbf{w}\|_0 = L. \tag{3.9}$$

This is a combinatorial problem, and there are $\binom{M}{L}$ possible options. It could be an extremely huge number when M is large and L is moderate. For instance, if $M = 100$ and $L = 20$, there are totally $\binom{100}{20} > 5 \times 10^{20}$ subproblems [DXP21]. Even if a modern machine (as fast as 10^{-10} seconds per subproblem) is used to solve this problem, it still needs more than 1.5 thousand years in total. Such computation times seem infeasible for the problem at hand and thus more computationally efficient approaches are required.

One widespread method is to replace the non-convex constraint $\|\mathbf{w}\|_0 = L$ with its convex surrogates, such as the ℓ_1-norm. By doing so and writing the ℓ_1-norm in the objective function as a penalty, we relax Problem (3.9) to [MSG12,ZFWS20,HA19,HA21]

$$\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{w} + \lambda\|\mathbf{w}\|_1 \quad \text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1, \tag{3.10}$$

where $\lambda > 0$ is a tuning parameter controlling the sparsity of the solution (i.e., the number of selected sensors). The above problem is QCQP [PB17] with ℓ_1-regularization, and it is still non-convex because of its constraint. State-of-the-art solvers include SDR, SCA, ADMM, and their variants, see [PB17,MHG+15,BV04,HS16,BPC+11,LMS+10] for general QCQP problems and [MSG12, MSG13, ZFWS20, HA19, HA21, ZCS+17, DGB22, CHL21, CL22, LML+18, WWS21, CT17] for specific QCQP problems with applications in MIMO radar, wireless communications, and so on. In the following subsection, we develop a method based on ADMM for solving Problem (3.10), which will be shown to have closed-form solutions at each iteration.

3.3.2 Proposed ADMM

To solve Problem (3.10) using ADMM, we first introduce an auxiliary variable $\mathbf{v} \in \mathbb{C}^M$ and reformulate (3.10) into

$$\min_{\mathbf{w},\mathbf{v}} \ \mathbf{v}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{v} + \lambda\|\mathbf{w}\|_1 \tag{3.11a}$$

$$\text{s.t.} \begin{cases} |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1 \\ \mathbf{w} = \mathbf{v}. \end{cases} \tag{3.11b}$$

Then we can write down the scaled-form ADMM iterations for Problem (3.10) as [BPC+11]

$$\mathbf{w} \leftarrow \begin{cases} \arg\min\limits_{\mathbf{w}} \ \lambda\|\mathbf{w}\|_1 + \frac{\rho}{2}\|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \\ \text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1 \end{cases} \tag{3.12a}$$

$$\mathbf{v} \leftarrow \arg\min_{\mathbf{v}} \ \mathbf{v}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{v} + \frac{\rho}{2}\|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \tag{3.12b}$$

$$\mathbf{u} \leftarrow \mathbf{u} + \mathbf{w} - \mathbf{v} \tag{3.12c}$$

where the original variable $\mathbf{w}$ and the auxiliary variable $\mathbf{v}$ are separately treated in (3.12a) and (3.12b), respectively, $\mathbf{u}$ is the scaled dual variable corresponding to the equality constraint in (3.11b), i.e., $\mathbf{w} = \mathbf{v}$, and $\rho > 0$ is the augmented Lagrangian parameter.

In what follows, we show that (3.12) has closed-form solutions at each ADMM iteration, by deducing $\mathbf{w}$ and $\mathbf{v}$ from (3.12a) and (3.12b), respectively. First of all, from (3.12b), it is simple to arrive at the closed-form solution of $\mathbf{w}$ in a least-squares form, as

$$\mathbf{v} = \rho(2\mathbf{R}_{\mathrm{x}} + \rho\mathbf{I})^{-1}(\mathbf{w} + \mathbf{u}). \tag{3.13}$$

Now we turn to (3.12a). We solve (3.12a) in two steps: i) we consider the unconstrained minimization problem by directly removing its constraint; ii) we check whether the solution obtained from Step i) satisfies the constraint, and update the final solution accordingly. Details are provided as follows.

Step i): We consider the following unconstrained minimization problem

$$\min_{\mathbf{w}} \ \lambda\|\mathbf{w}\|_1 + \frac{\rho}{2}\|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2. \tag{3.14}$$

By calculating the subgradient of the objective function in (3.14) with respect to (w.r.t.) $\mathbf{w}$, and setting the resultant expression equal to zero, we obtain its solution, denoted by $\bar{\mathbf{w}}$, as

$$\bar{\mathbf{w}} = \text{sign}(\mathbf{v} - \mathbf{u}) \odot \left(|\mathbf{v} - \mathbf{u}| - \frac{\lambda}{\rho} \right)_+ . \tag{3.15}$$

The detailed derivation of (3.15) from Problem (3.14) is omitted here, and the interested readers are referred to the similar result in Lemma 1 in [ZKOM18].

Step ii): We check whether or not $\bar{\mathbf{w}}$ obtained from (3.15) statisfies $|\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0| \geq 1$. If it is, then the solution to (3.12a), referred to as $\widehat{\mathbf{w}}$, is $\widehat{\mathbf{w}} = \bar{\mathbf{w}}$. If it is not, then $\widehat{\mathbf{w}}$ can be found via the following theorem.

Theorem 3.1 *Denote $\widehat{\mathbf{w}}$ and $\bar{\mathbf{w}}$ as the solutions to Problems (3.12a) and (3.14), respectively. If $\bar{\mathbf{w}}$ does not satisfy $|\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0| \geq 1$, then $\widehat{\mathbf{w}}$ equals the one in $\{\mathbf{w} : |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1\}$, such that it is closest (in an ℓ_2-norm sense) to $\bar{\mathbf{w}}$.*

Proof: See Appendix A.1. □

Besides the above mathematical proof, we give an illustrative example for Theorem 3.1, by showing the *near-symmetric* structure of the objective function around its stationary point[1]. For simplicity, the variable is set to be real-valued and the dimension $M = 1$. The parameters are $\lambda = 1$, $\rho = 4$, $\mathbf{a}_0 = 1/2$, and $-\mathbf{v} + \mathbf{u} = -1$. Hence Problem (3.12a) becomes

$$\min_{w} \; f(w) \triangleq |w| + 2(w-1)^2 \quad \text{s.t. } |w| \geq 2, \tag{3.16}$$

with its stationary point w_0 falling outside its feasible region $|w| \geq 2$, as in Figure 3.1. Thanks to the convexity and near-symmetric structure of the objective function $f(w)$, finding its minimum is equivalent to determining the point (inside the feasible region) closest to its stationary point w_0. In Figure 3.1, it is easy to see that $w = 2$ is such a point, and thus it is the solution to Problem (3.16).

Consequently, if $\bar{\mathbf{w}}$ obtained from (3.15) does not satisfy $|\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0| \geq 1$, then according to Theorem 3.1, $\widehat{\mathbf{w}}$ can be found by solving the following minimization problem

$$\widehat{\mathbf{w}} \;\leftarrow\; \arg\min_{\mathbf{w}} \; \|\mathbf{w} - \bar{\mathbf{w}}\|_2^2 \quad \text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1, \tag{3.17}$$

[1] We say a function $f(\mathbf{x})$ has a *near-symmetric* structure around a point $\mathbf{x}_0$ if and only if $f(\mathbf{x}_0 + \mathbf{x}) \approx f(\mathbf{x}_0 - \mathbf{x})$ holds for any $\mathbf{x}$.

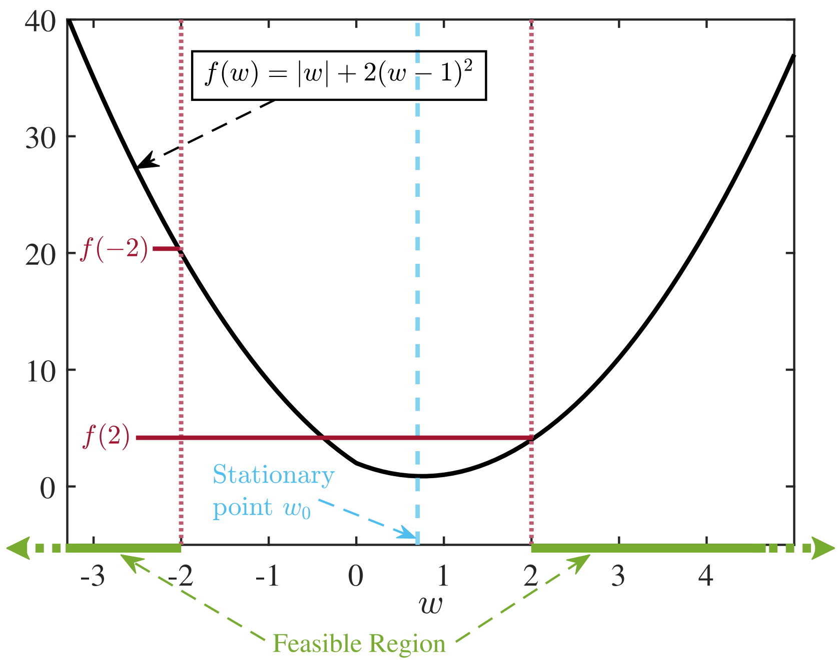


Figure 3.1: Illustration of the near-symmetric structure of $f(w)$.

which is equivalent to

$$\widehat{\mathbf{w}} \leftarrow \arg\min_{\mathbf{w}} \|\mathbf{w} - \bar{\mathbf{w}}\|_2^2 \quad \text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 = 1. \tag{3.18}$$

The equivalence between Problems (3.17) and (3.18) is straightforward, and it indicates that the solutions to Problem (3.17) always fall on the *boundary* of its feasible region. Problem (3.18) has the following closed-form solution as [HS16]

$$\widehat{\mathbf{w}} = \bar{\mathbf{w}} + \frac{1 - |\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0|}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0|}\mathbf{a}_0\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0. \tag{3.19}$$

Eventually, by considering Steps i) and ii) simultaneously and making use of the plus function, the solution to (3.12a) can be written in a single formula as

$$\widehat{\mathbf{w}} = \bar{\mathbf{w}} + \frac{\left(1 - |\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0|\right)_+}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0|}\mathbf{a}_0\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_0, \tag{3.20}$$

where $\bar{\mathbf{w}}$ is given in (3.15).

So far, we have derived closed-form solutions for $\mathbf{w}$ and $\mathbf{v}$ at each ADMM iteration. The complete ADMM for solving Problem (3.10) is summarized in Algorithm 3.1, where $k_{\max}$ denotes a large scalar and η a small one, used to terminate the iteration, and

Algorithm 3.1 ADMM for solving Problem (3.10)

Input : $\mathbf{R}_\mathrm{x} \in \mathbb{C}^{M\times M}$, $\mathbf{a}_0 \in \mathbb{C}^M$, λ, ρ, $k_{\max}$, η
Output : $\widehat{\mathbf{w}} \in \mathbb{C}^M$
Initialize: $\mathbf{v}_{(0)} \leftarrow \mathbf{v}_\mathrm{init}$, $\mathbf{u}_{(0)} \leftarrow \mathbf{u}_\mathrm{init}$, $k \leftarrow 0$

while not converged **do**
  $\bar{\mathbf{w}}_{(k+1)} \leftarrow \mathrm{sign}(\mathbf{v}_{(k)} - \mathbf{u}_{(k)}) \odot \left(|\mathbf{v}_{(k)} - \mathbf{u}_{(k)}| - \frac{\lambda}{\rho}\right)_+$
  $\mathbf{w}_{(k+1)} \leftarrow \bar{\mathbf{w}}_{(k+1)} + \frac{\left(1-|\bar{\mathbf{w}}^\mathrm{H}_{(k+1)}\mathbf{a}_0|\right)_+}{\|\mathbf{a}_0\|_2^2|\bar{\mathbf{w}}^\mathrm{H}_{(k+1)}\mathbf{a}_0|}\mathbf{a}_0\bar{\mathbf{w}}^\mathrm{H}_{(k+1)}\mathbf{a}_0$
  $\mathbf{v}_{(k+1)} \leftarrow \rho(2\mathbf{R}_\mathrm{x} + \rho\mathbf{I})^{-1}(\mathbf{w}_{(k+1)} + \mathbf{u}_{(k)})$
  $\mathbf{u}_{(k+1)} \leftarrow \mathbf{u}_{(k)} + \mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}$
  converged $\leftarrow k{+}1 \geq k_{\max}$ or $\|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}\|_2 \leq \eta$
  $k \leftarrow k+1$
end while
$\widehat{\mathbf{w}} \leftarrow \mathbf{w}_{(k)}$

subscript $\cdot_{(k)}$ denotes the variable at the k-th iteration. The convergence property of the proposed ADMM algorithm will be discussed in Section 3.4.

We have the following remarks:

- Note that we can attain any level of sparsity (i.e., any number L out of M sensors in sparse array design), by carefully tuning the value of λ. This shall be verified in the simulation section, see Figure 3.3 in Section 3.6.
- To ensure selection of L sensors, an appropriate value of λ is typically found by carrying out a binary search over a probable interval of λ, say $[\lambda_L, \lambda_U]$. To be precise, we begin by solving a sparse $\widehat{\mathbf{w}}$ using Algorithm 3.1, with $\lambda = (\lambda_L + \lambda_U)/2$. If $\|\widehat{\mathbf{w}}\|_0 > L$ (resp. $\|\widehat{\mathbf{w}}\|_0 < L$), then we update $\lambda_L = \lambda$ (resp. $\lambda_U = \lambda$), and solve another sparse $\widehat{\mathbf{w}}$ with $\lambda = (\lambda_L + \lambda_U)/2$. We repeat the above step until $\|\widehat{\mathbf{w}}\|_0 = L$.
- Note that the solution of (3.10) is not exactly equal to the one of (3.9). Therefore, after the solution of desired sparsity of (3.10) is obtained, one should solve a reduced-size minimization problem similar to (3.7) as a last step, omitting the sensors corresponding to the zero entries of the solution.

3.4 Convergence Analysis

The convergence properties of Algorithm 3.1 are presented in this section. We start with two lemmata, which show the monotonicity and boundedness of the augmented

Lagrangian function of Problem (3.11). Two theorems are then provided to show that the proposed algorithm converges and that it converges to a stationary point.

The augmented Lagrangian function regarding Problem (3.11) can be written as $\mathcal{L}(\mathbf{w},\mathbf{v},\mathbf{u}) \triangleq \lambda\|\mathbf{w}\|_1 + \mathbf{v}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{v} + \frac{\rho}{2}(\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2)$. As stated in Section 3.3.2, $\mathbf{w}$, $\mathbf{v}$, and $\mathbf{u}$ are the original, auxiliary, and dual variables, respectively, and $\rho > 0$ is the augmented Lagrangian parameter. In what follows, Lemma 3.1 shows that the function value of $\mathcal{L}(\mathbf{w},\mathbf{v},\mathbf{u})$ is non-increasing, and Lemma 3.2 shows that the function value $\mathcal{L}(\mathbf{w},\mathbf{v},\mathbf{u})$ is bounded from below, on the condition that ρ is larger than or equal to a certain value.

Lemma 3.1 *As long as the parameter $\rho \geq 2\sqrt{2}\lambda_{\max}(\mathbf{R}_{\mathrm{x}})$, the point sequence produces a monotonically non-increasing objective function value sequence $\{\mathcal{L}(\mathbf{w}_{(k)},\mathbf{v}_{(k)},\mathbf{u}_{(k)})\}$. That is, $\mathcal{L}(\mathbf{w}_{(k+1)},\mathbf{v}_{(k+1)},\mathbf{u}_{(k+1)}) \leq \mathcal{L}(\mathbf{w}_{(k)},\mathbf{v}_{(k)},\mathbf{u}_{(k)})$ holds for all $k = 0, 1, 2, \cdots$.*

Proof: See Appendix A.2. □

Lemma 3.2 *The function value of $\mathcal{L}(\mathbf{w},\mathbf{v},\mathbf{u})$ is bounded from below by 0, as long as*[2]

$$\rho \geq \frac{2\lambda_{\max}^2(\mathbf{R}_{\mathrm{x}})}{\lambda_{\min}(\mathbf{R}_{\mathrm{x}})}. \tag{3.21}$$

Proof: See Appendix A.3. □

With Lemmata 3.1 and 3.2, we have the following theorem.

Theorem 3.2 *As long as the augmented Lagrangian parameter*

$$\rho \geq \max\left\{2\sqrt{2}\lambda_{\max}(\mathbf{R}_{\mathrm{x}}), \frac{2\lambda_{\max}^2(\mathbf{R}_{\mathrm{x}})}{\lambda_{\min}(\mathbf{R}_{\mathrm{x}})}\right\}, \tag{3.22}$$

the objective function value sequence $\{\mathcal{L}(\mathbf{w}_{(k)},\mathbf{v}_{(k)},\mathbf{u}_{(k)})\}$ generated by Algorithm 3.1 converges. Furthermore, as $k \to \infty$, we have $\mathbf{w}_{(k+1)} = \mathbf{w}_{(k)}$, $\mathbf{v}_{(k+1)} = \mathbf{v}_{(k)}$, $\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)}$, and $\mathbf{w}_{(k)} = \mathbf{v}_{(k)}$.

[2] It is worth noting that the lower bound here is not tight, see Appendix A.3.

Proof: See Appendix A.4. □

Moreover, the following theorem shows that the limit point of Algorithm 3.1 is a stationary point.

Theorem 3.3 *Denote the limit point obtained by the proposed algorithm stated in Algorithm 3.1 as* $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$. *Then, it satisfies the KKT conditions of Problem (3.11), as*

$$\mathbf{0} = 2\mathbf{R}_{\mathrm{x}}\mathbf{v}_{(k+1)} - \mathbf{y}_{(k+1)}, \tag{3.23a}$$

$$\mathbf{w}_{(k+1)} \in \arg\min_{\mathbf{w}} \left\{ \begin{array}{l} \lambda\|\mathbf{w}\|_1 + \mu^\star(|\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 - 1) \\ + \Re\{\langle \mathbf{y}_{(k+1)}, \mathbf{w} - \mathbf{v}_{(k+1)} \rangle\} \end{array} \right\}, \tag{3.23b}$$

$$\mathbf{w}_{(k+1)} = \mathbf{v}_{(k+1)}, \tag{3.23c}$$

where $\mathbf{y}_{(k+1)} = \rho\mathbf{u}_{(k+1)}$ *is the dual variable corresponding to the equality constraint in (3.11b), and* $\mu^\star$ *denotes the optimal dual variable corresponding to the inequality constraint in (3.11b). In words, any limit point of Algorithm 3.1 is a stationary solution to Problem (3.11).*

Proof: See Appendix A.5. □

3.5 ADMM with Re-weighted ℓ_1-norm

As has been well-documented in the literature, see e.g. [CWB08], the iteratively re-weighted ℓ_1-norm penalty has remarkable advantages over the conventional ℓ_1-norm. Therefore, in this section, we propose an improved approach on the basis of Algorithm 3.1, by replacing the ℓ_1-norm regularization in (3.10) with the re-weighted ℓ_1-norm. That is, Problem (3.10) is modified as

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{w} + \lambda\|\mathbf{1} \oslash (|\mathbf{g}| + \epsilon) \odot \mathbf{w}\|_1 \tag{3.24a}$$

$$\text{s.t. } |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1, \tag{3.24b}$$

where $\mathbf{g}$ equals $\mathbf{w}$ obtained from the previous iteration, and $\epsilon > 0$ is a small scalar providing stability and ensuring that a zero-valued component in $\mathbf{w}$ does not strictly prohibit a non-zero estimate at the next iteration. Note that once the non-zero entries of the solution of Problem (3.24) are identified, their influence is down-weighted in order to

Algorithm 3.2 ADMM for solving Problem (3.24)

Input : $\mathbf{R}_{\mathrm{x}} \in \mathbb{C}^{M \times M}$, $\mathbf{a}_0 \in \mathbb{C}^M$, λ, ρ, ϵ, $k_{\max}$, η
Output : $\widehat{\mathbf{w}} \in \mathbb{C}^M$
Initialize: $\mathbf{v}_{(0)} \leftarrow \mathbf{v}_{\mathrm{init}}$, $\mathbf{u}_{(0)} \leftarrow \mathbf{u}_{\mathrm{init}}$, $k \leftarrow 0$

1: **while** not converged **do**
2: $\quad \bar{\mathbf{w}}_{(k+1)} \leftarrow \mathrm{sign}(\mathbf{v}_{(k)} - \mathbf{u}_{(k)}) \odot \left(|\mathbf{v}_{(k)} - \mathbf{u}_{(k)}| - (\lambda \mathbf{1}) \oslash [\rho(|\mathbf{v}_{(k)}| + \epsilon)]\right)_+$
3: $\quad \mathbf{w}_{(k+1)} \leftarrow \bar{\mathbf{w}}_{(k+1)} + \frac{\left(1 - |\bar{\mathbf{w}}_{(k+1)}^{\mathrm{H}} \mathbf{a}_0|\right)_+}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}_{(k+1)}^{\mathrm{H}} \mathbf{a}_0|} \mathbf{a}_0 \bar{\mathbf{w}}_{(k+1)}^{\mathrm{H}} \mathbf{a}_0$
4: $\quad \mathbf{v}_{(k+1)} \leftarrow \rho(2\mathbf{R}_{\mathrm{x}} + \rho \mathbf{I})^{-1}(\mathbf{w}_{(k+1)} + \mathbf{u}_{(k)})$
5: $\quad \mathbf{u}_{(k+1)} \leftarrow \mathbf{u}_{(k)} + \mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}$
6: $\quad$ converged $\leftarrow k+1 \geq k_{\max}$ or $\|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}\|_2 \leq \eta$
7: $\quad k \leftarrow k + 1$
8: **end while**
9: $\widehat{\mathbf{w}} \leftarrow \mathbf{w}_{(k)}$

allow more sensitivity for identifying the remaining small but non-zero entries [CWB08]. This results in a better behavior of (3.24) than (3.10), which will be corroborated in Section 3.6.1.

The ADMM iteration for Problem (3.24) is the same as (3.12) except for (3.12a) which should be replaced by

$$\mathbf{w} \leftarrow \begin{cases} \arg\min\limits_{\mathbf{w}} \ \lambda \|\mathbf{1} \oslash (|\mathbf{g}| + \epsilon) \odot \mathbf{w}\|_1 + \dfrac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \\ \text{s.t. } |\mathbf{w}^{\mathrm{H}} \mathbf{a}_0|^2 \geq 1. \end{cases} \tag{3.25}$$

Accordingly, the result of $\bar{\mathbf{w}}$ in (3.15) now becomes

$$\bar{\mathbf{w}} = \mathrm{sign}(\mathbf{v} - \mathbf{u}) \odot (|\mathbf{v} - \mathbf{u}| - (\lambda \mathbf{1}) \oslash [\rho(|\mathbf{g}| + \epsilon)])_+, \tag{3.26}$$

and the complete ADMM for solving Problem (3.24) is summarized in Algorithm 3.2. The convergence property of Algorithm 3.2, and the comparison between Algorithms 3.1 and 3.2, will be presented using simulations in Section 3.6.1.

3.6 Simulation Results

In this section, we present numerical examples to demonstrate the effectiveness of the proposed algorithms, i.e., Algorithms 3.1 and 3.2. We first examine behavior of the algorithms in Section 3.6.1, in terms of convergence property and beamformer weight sparsity control. Then, in Section 3.6.2, we compare the computational complexity of the proposed algorithms to those of other state-of-the-art approaches, including SDR, an SDR variant, and SCA, presented in [MSG12], [HA19], and [PB17], respectively. In Section 3.6.3, we finally test the performance of sparse array beamformers designed by using different strategies, in terms of array beampattern and output SINR.

3.6.1 Convergence and Sparsity Control

First example: A compact ULA consisting of $M = 12$ antenna sensors is utilized, while $T = 100$ snapshots, one SOI from $\theta_0 = 0°$ and $K = 2$ interference signals from $\theta_1 = -10°$ and $\theta_2 = 10°$, respectively, are considered. The signal-to-noise ratio (SNR) and interference-to-noise ratio (INR) are SNR = 10 dB and INR = 20 dB, respectively. The two proposed algorithms, i.e., Algorithms 3.1 and 3.2 are examined, where the parameters are $\epsilon = 10^{-10}$, $k_{\max} = 10^3$, $\eta = 10^{-12}$, $\mathbf{u}_{\text{init}} = \mathbf{0}$, and $\mathbf{v}_{\text{init}}$ is drawn from the complex standard normal distribution. Three scenarios with different values of the tuning parameter λ and the augmented Lagrangian parameter ρ are considered. The results of $\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$ versus the ADMM iteration index k are given in Figure 3.2. It is seen that the function value sequence $\{\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})\}$ of Algorithm 3.1 is monotonically non-increasing and bounded from below, which is consistent with the theoretical analyses in Section 3.4. Additionally, we also observe that although the function value sequence $\{\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})\}$ of Algorithm 3.2 is not monotonically non-increasing, it converges eventually. Moreover, Algorithm 3.2 converges faster than Algorithm 3.1.

Second example: A compact ULA of $M = 12$ antenna sensors is used, and we wish to select L sensors for the beamformer. We consider 8 situations with different SNR, decreasing from 20 dB to -15 dB with a stepsize of 5 dB. The interference-to-noise ratio is INR = 10 dB, the augmented Lagrangian parameter is $\rho = 2 \times 10^4$, and the other parameters are the same as those in the first example. We test the sparsity of the beamformer weight obtained via Algorithms 3.1 and 3.2 w.r.t. the tuning parameter λ. The curves are averaged over 1000 Monte Carlo runs, and they are displayed in Figure 3.3. It can be seen that, for all 8 situations, any level of sparsity (from 1 to 11) could be attained by both algorithms, and that a larger λ produces a smaller sparsity of $\widehat{\mathbf{w}}$, as expected. In addition, the curves by using Algorithm 3.1 decrease far more rapidly than those by using Algorithm 3.2, which implies that it is much easier to tune λ for a specific level of sparsity when Algorithm 3.2 is employed.

Because of the better behavior of Algorithm 3.2, rather than both Algorithms 3.1 and 3.2, we solely consider Algorithm 3.2 in the remaining simulations, and it will be labelled as "ADMM".

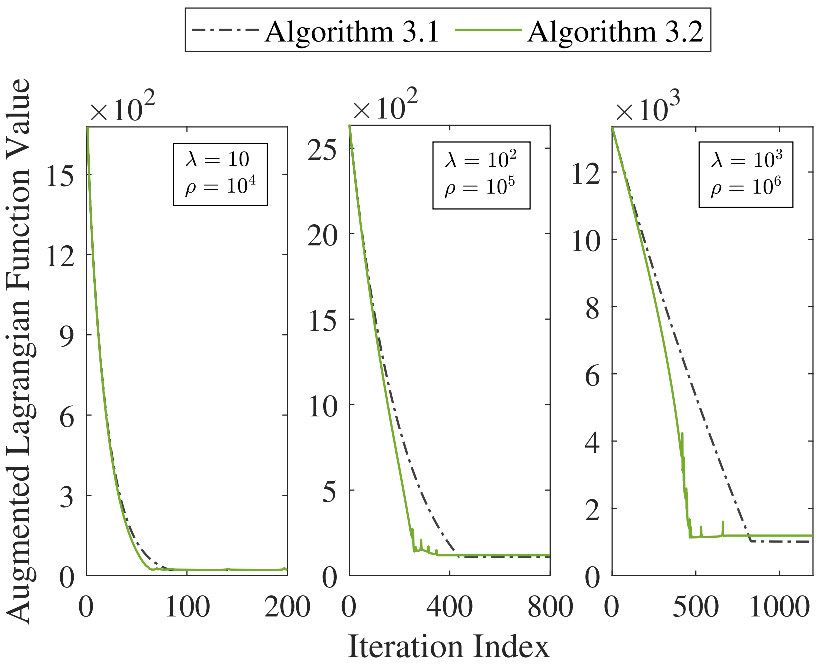


Figure 3.2: Function value of $\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$ versus iteration index k. (1st example)

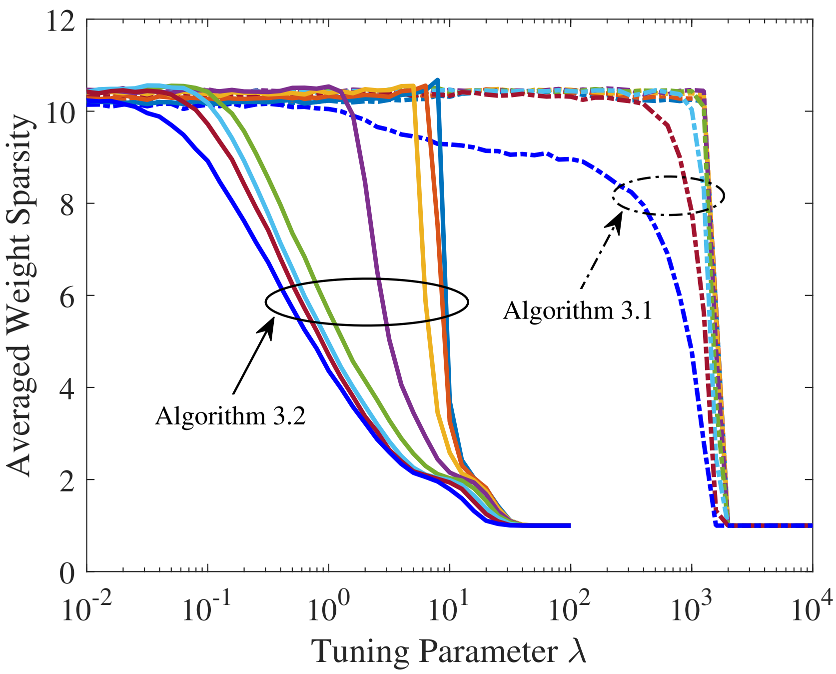


Figure 3.3: Sparsity of beamformer weight versus tuning parameter λ. (2nd example)

3.6.2 Computational Complexity

We start by rewriting the state-of-the-art methods to suit our problem, i.e., Problem (3.24), using our notations, as [MSG12]

$$\min_{\mathbf{W}} \operatorname{trace}\{\mathbf{R}_{\mathrm{x}}\mathbf{W}\} + \lambda \cdot \operatorname{trace}\{\mathbf{B}|\mathbf{W}|\} \tag{3.27a}$$

$$\text{s.t.} \begin{cases} \operatorname{trace}\{\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\mathbf{W}\} \geq 1, \\ \mathbf{W} \succeq 0, \end{cases} \tag{3.27b}$$

and [HA19]

$$\min_{\mathbf{W},\widetilde{\mathbf{W}}} \operatorname{trace}\{\mathbf{R}_{\mathrm{x}}\mathbf{W}\} + \lambda \cdot \operatorname{trace}\{\mathbf{B}\widetilde{\mathbf{W}}\} \tag{3.28a}$$

$$\text{s.t.} \begin{cases} \operatorname{trace}\{\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\mathbf{W}\} \geq 1, \\ \mathbf{W} \succeq 0, \\ \widetilde{\mathbf{W}} \geq |\mathbf{W}|, \end{cases} \tag{3.28b}$$

where $\mathbf{B} = \mathbf{1}_M \oslash (|\mathbf{G}| + \epsilon)$, with $\mathbf{G}$ being equal to $\mathbf{W}$ obtained from the previous iteration. Problems (3.27) and (3.28) are referred to as SDR and SDR-V (short for "SDR Variant"), respectively, in the remaining simulations. If the solution $\mathbf{W}$ to Problems (3.27) and (3.28) is of rank one, their beamformer weight can be calculated as the principal eigenvector of $\mathbf{W}$; otherwise, extra post-processing based on randomization is required [SDL06].

On the other hand, Problem (3.24) can be recast as [PB17]

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{w} + \lambda\|\mathbf{b} \odot \mathbf{w}\|_1 \tag{3.29a}$$

$$\text{s.t. } 2\Re\{\mathbf{g}^{\mathrm{H}}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\mathbf{w}\} - |\mathbf{g}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1, \tag{3.29b}$$

where $\mathbf{b} = \mathbf{1} \oslash (|\mathbf{g}| + \epsilon)$ with $\mathbf{g}$ being equal to $\mathbf{w}$ obtained from the previous iteration, the same as what has been introduced in Section 3.5. Problem (3.29) is denoted as SCA in the remaining simulations. The key of the success of SCA method lies in the fact that $|\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2$ is a concave function w.r.t. $\mathbf{w}$ and thus $|\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \leq 2\Re\{\mathbf{g}^{\mathrm{H}}\mathbf{a}_0\mathbf{a}_0^{\mathrm{H}}\mathbf{w}\} - |\mathbf{g}^{\mathrm{H}}\mathbf{a}_0|^2$ holds for all $\mathbf{w}$ and any given (known) $\mathbf{g}$.

If a general-purpose SDR solver, such as the interior point method, is adopted to solve Problems (3.27) and (3.28), the worst case complexity can be as high as $\mathcal{O}(M^{6.5})$ per iteration [HS16]. The cost of solving Problem (3.29) could be smaller, if further effort is made, for instance, by taking care of the structure of the problem. However, since this is out of the scope of this chapter, in our simulations, we simply utilize the

CVX toolbox [GB14] to solve the aforementioned three problems. As for the proposed algorithms, their computational cost primarily comes from the inverse operation, i.e., $(2\mathbf{R}_{\mathrm{x}} + \rho\mathbf{I})^{-1}$, which consumes $\mathcal{O}(M^3)$. Furthermore, we can cache the result of $(2\mathbf{R}_{\mathrm{x}} + \rho\mathbf{I})^{-1}$ to save computations in the subsequent iterations.

It is worth noting that when $\mathbf{B}$ in (3.27) and (3.28) and $\mathbf{b}$ in (3.29) are fixed as $\mathbf{B} = \mathbf{1}_M$ and $\mathbf{b} = \mathbf{1}$, Problems (3.27), (3.28), and (3.29) reduce to three approaches for solving Problem (3.10), which correspond to Algorithm 3.1. As has been confirmed by the numerical results in Section 3.6.1, algorithms with re-weighted ℓ_1-norm regularization are more efficient in the sense that they converge faster and are much easier to control the sparsity of solution, compared to their counterparts with conventional ℓ_1-norm regularization. Therefore, only the former group of approaches, i.e., the above-mentioned SDR (3.27), SDR-V (3.28), SCA (3.29), and Algorithm 3.2, are considered in the following simulations.

Third example: We wish to choose $L = 4$ out of $M = 12$ antenna sensors from a compact ULA. One SOI from $\theta_0 = 0°$ and $K = 2$ interferers from $\theta_1 = -10°$ and $\theta_2 = 10°$, respectively, are considered, while SNR = 0 dB and INR = 20 dB. The number of snapshots T varies uniformly from 10 to 150 with a stepsize of 10. The other parameters for Algorithm 3.2 are the same as those of the first example, except for ρ which is set to $\rho = 10^3$ in this example. The central processing unit (CPU) times of the examined approaches are averaged over 100 Monte Carlo runs, and they are plotted in Figure 3.4. It is seen that their CPU times are almost unchanged when T varies, and that of the ADMM method is around 10^{-1} seconds which is about 10^3 times less than those of the SDR, SDR-V, and SCA methods (which take around 10^2 seconds).

Fourth example: We wish to select $L = 4$ out of M sensors from a compact ULA. The number of snapshots is fixed as $T = 100$, and the number of sensors M changes from 10 to 20. The other parameters remain unchanged as those of the third example. The CPU times of the examined methods are shown in Figure 3.5, from which it is seen that the CPU times of the SDR, SDR-V, and ADMM methods increase as M increases, while the CPU time of the SCA method keeps almost unchanged when M varies. In addition, the CPU time of the proposed algorithm is much smaller than those of the other three tested approaches. Note that there is a jump of the ADMM curve at $M = 15$. This is caused by the increased number of iterations when $M \geq 16$.

Fifth example: We wish to select L out of $M = 12$ sensors from a compact ULA. The number of snapshots is fixed as $T = 100$, and the number of selected sensors L changes from 3 to 12. The other parameters remain unchanged as those of the third example. The CPU times of the examined methods are drawn in Figure 3.6, from which it is seen

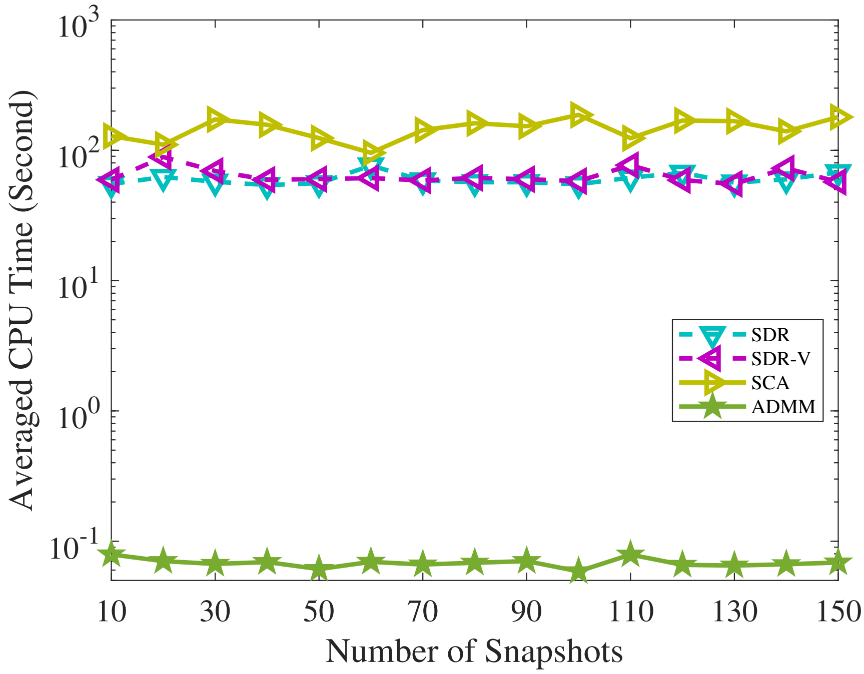


Figure 3.4: Averaged CPU time versus T. (3rd example)

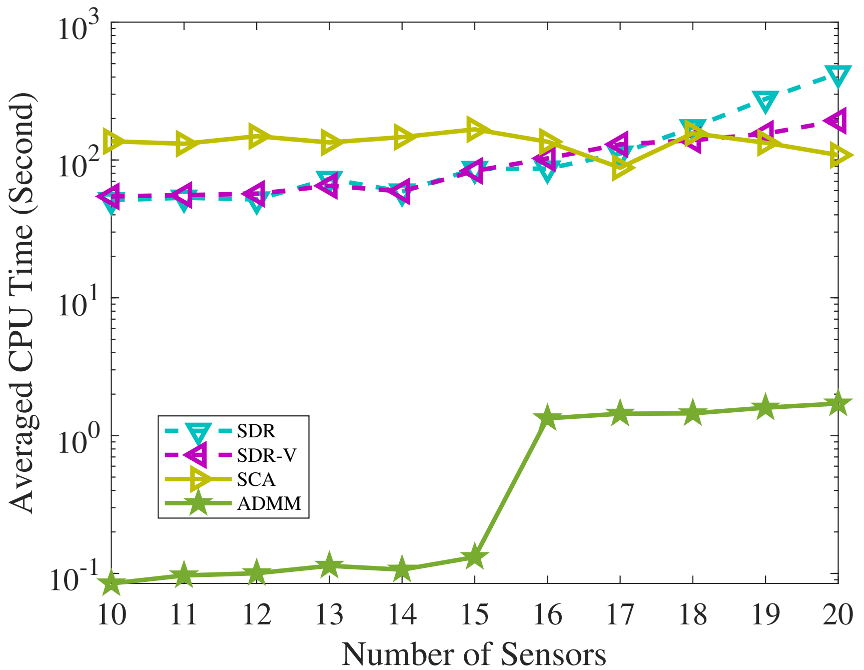


Figure 3.5: Averaged CPU time versus M. (4th example)

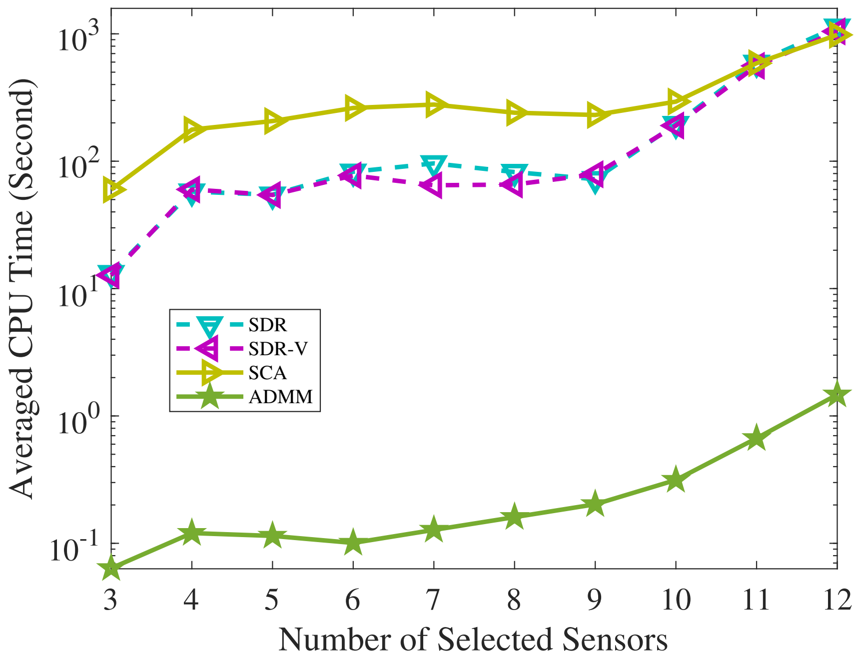


Figure 3.6: Averaged CPU time versus L. (5th example)

that the CPU times of all the four methods increase as L increases. Besides, the CPU time of the proposed algorithm is shown again much less than those of the other three tested methods.

3.6.3 Beamforming Performance

In the following simulations, we examine the beamforming performance of the proposed algorithm compared with several other sparse array design strategies, in terms of beampattern and output SINR.

Sixth example: We choose $L = 4$ out of $M = 12$ sensors. One SOI from $\theta_0 = 0°$ and $K = 3$ interferers from $\theta_1 = -40°$, $\theta_2 = 30°$, and $\theta_3 = 50°$, are considered, while SNR $= 0$ dB and INR $= 20$ dB. Different sparse array design strategies are examined, including enumeration (i.e., exhaustive search), compact ULA, sparse ULA, random array, nested array [PV10], coprime array [VP11, QZA15], SDR [MSG12], SDR-V [HA19], SCA [PB17], and the proposed ADMM. The result of using the whole ULA is also included. Their beampatterns are depicted in Figure 3.7, where we separate them into two subfigures and the ADMM is drawn in both, for a better comparison. From Figure 3.7 we observe that the proposed ADMM method provides lower sidelobe

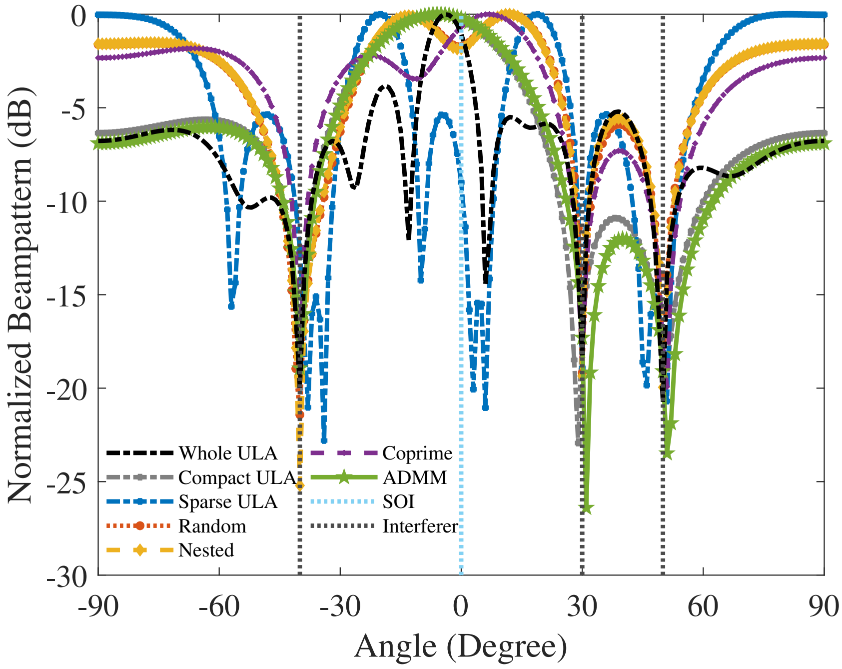
Table 3.1: Output SINR of sixth example.

Method	**SINR (dB)**	**Method**	**SINR (dB)**
Whole ULA	6.0913	Compact ULA	4.1546
Best via Enum.	4.7157	Coprime	3.0643
ADMM	4.3854	Random	0.5537
SDR	4.3589	Nested	0.4340
SDR-V	4.3589	Sparse ULA	−15.8124
SCA	4.2961	Worst via Enum.	−16.8793

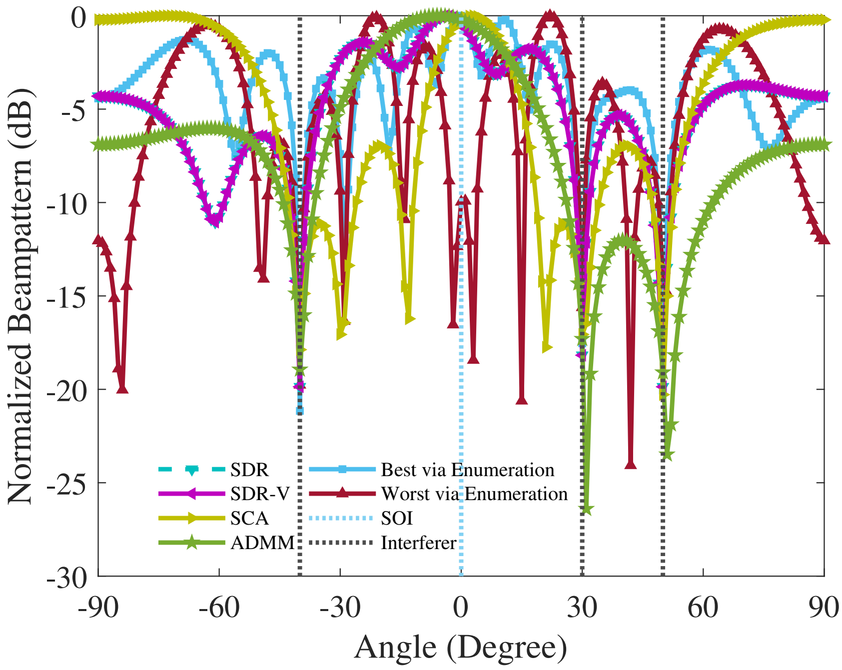
and deeper nulls towards the interferences, compared to the others. Their output SINRs in this example are given in TABLE 3.1.

Seventh example: We consider the scenario with one SOI whose DOA θ_0 changes from $-60°$ to $60°$, and $K = 2$ interferers from $\theta_1 = \theta_0 - 10°$ and $\theta_2 = \theta_0 + 10°$, respectively. The remaining parameters are unchanged as those in the third example. The output SINR versus DOA of the SOI is plotted in Figure 3.8. It is seen that the proposed method has excellent performance, whose output SINR is less than 0.4 dB lower than that of the best case via enumeration, very close (within 0.6 dB) to the optimal SINR, slightly larger than those of the SDR, SDR-V, and SCA methods, and at least about 2 dB larger than those of the other approaches. Two exceptions occur at $\theta_0 = -55°$ and $\theta_0 = 55°$, in which cases the output SINR of ADMM is about 2.5 dB lower than those of the best case via enumeration, SDR, SDR-V, and SCA methods, and is still significantly higher than those of the other approaches. Another interesting result is that the performance of the nested array and the coprime array is even worse than that of the random array. This is because the goal of nested array and coprime array is to make sure more continuous virtual sensors exist in their difference coarray, such that they can estimate more sources than physical sensors. In other words, nested array and coprime are designed to obtain better performance in DOA estimation, but not necessary to have good performance in beamforming in terms of SINR.

Eighth example: The SNR varies from -20 dB to 12 dB with a stepsize of 2 dB. The DOA of the SOI is $\theta_0 = 0°$ and $K = 2$ interference signals come from $\theta_1 = -10°$ and $10°$. The other parameters are unchanged compared with those of the previous example. The output SINR versus SNR is depicted in Figure 3.9a. To provide a clearer vision of the results, we calculate the SINR departure of the corresponding methods from the optimal SINR, and draw them in Figure 3.9b. The figures demonstrate better performance of the proposed scheme than the other sparse array design techniques (except for the best



(a)



(b)

Figure 3.7: Beampattern comparison with 1 SOI and 3 interferers. (6th example)

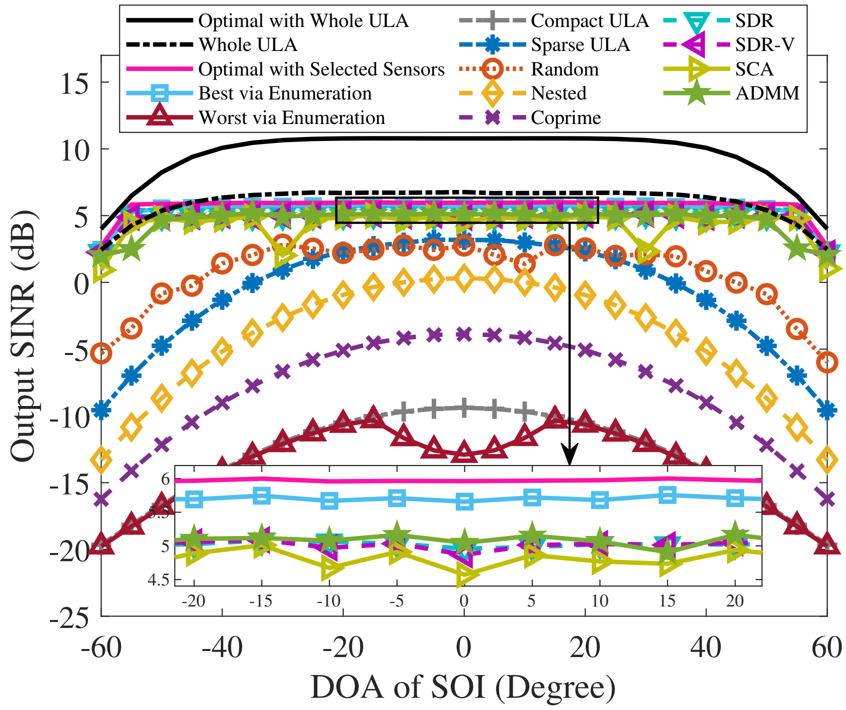
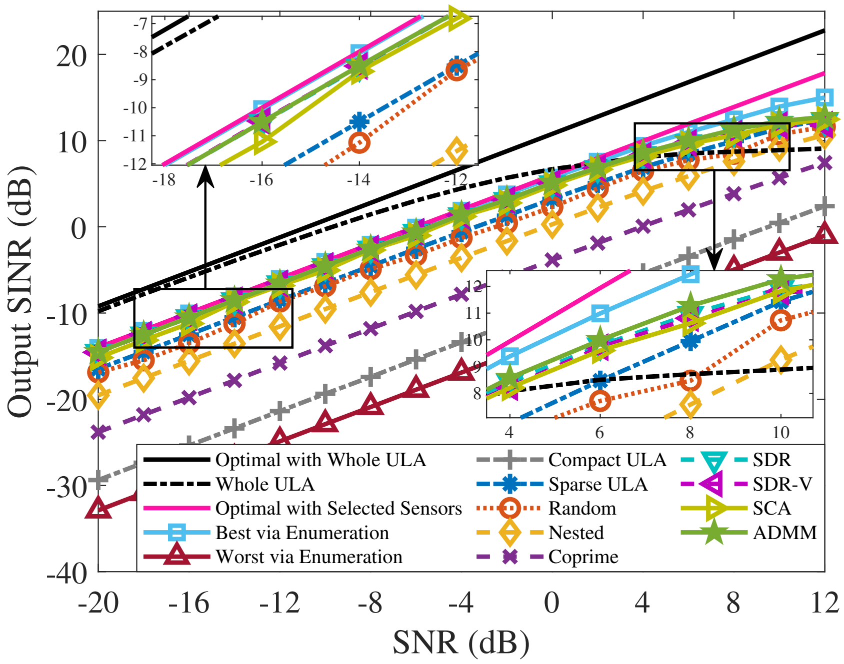


Figure 3.8: Output SINR versus DOA of SOI. (7th example)

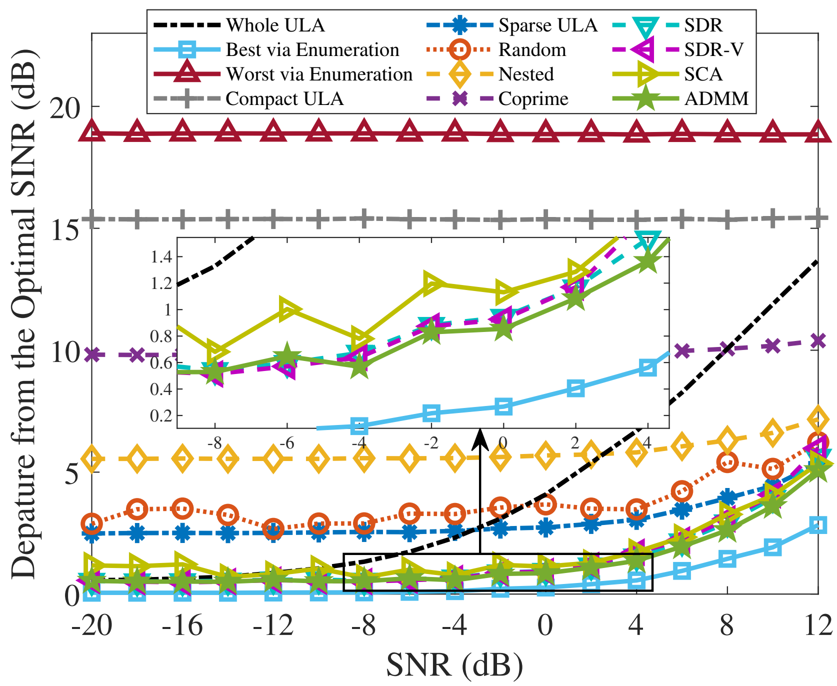
case via enumeration) in terms of output SINR. It is also observed that in the large SNR region, the SINR of the whole array is even smaller than those of sparse arrays. This verifies the statement that the performance of beamforming is affected by not only the beamformer weight, but also the array configuration [Lin82]. The reason comes from two aspects: On one hand, the performance of MVDR beamformer degrades when SNR is high, since the SOI presents in the training data [Vor14]. On the other hand, the calculation of SINR for the whole array is different from that for the sparse arrays, that is, the length of the beamformer weight and the size of $\mathbf{R}_{\mathrm{i+n}}$, used for calculating SINR, are different.

Ninth example: We examine the output SINR versus the number of snapshots. The simulation setup is the same as that of the third example. The results are shown in Figure 3.10, from which we observe that the output SINR of the proposed ADMM is close to that of the best case via enumeration, slightly larger than those of the SDR, SDR-V, and SCA approaches, and significantly larger than those of the others.

Tenth example: We examine the output SINR versus the number of sensors. The simulation setup is unchanged as that of the fourth example. The results are plotted in Figure 3.11, from which we again observe that the output SINR of the ADMM is close to that of the best case via enumeration, slightly larger than those of the SDR, SDR-V, and SCA methods, and significantly larger than those of the others.



(a)



(b)

Figure 3.9: Output SINR versus input SNR. (8th example)

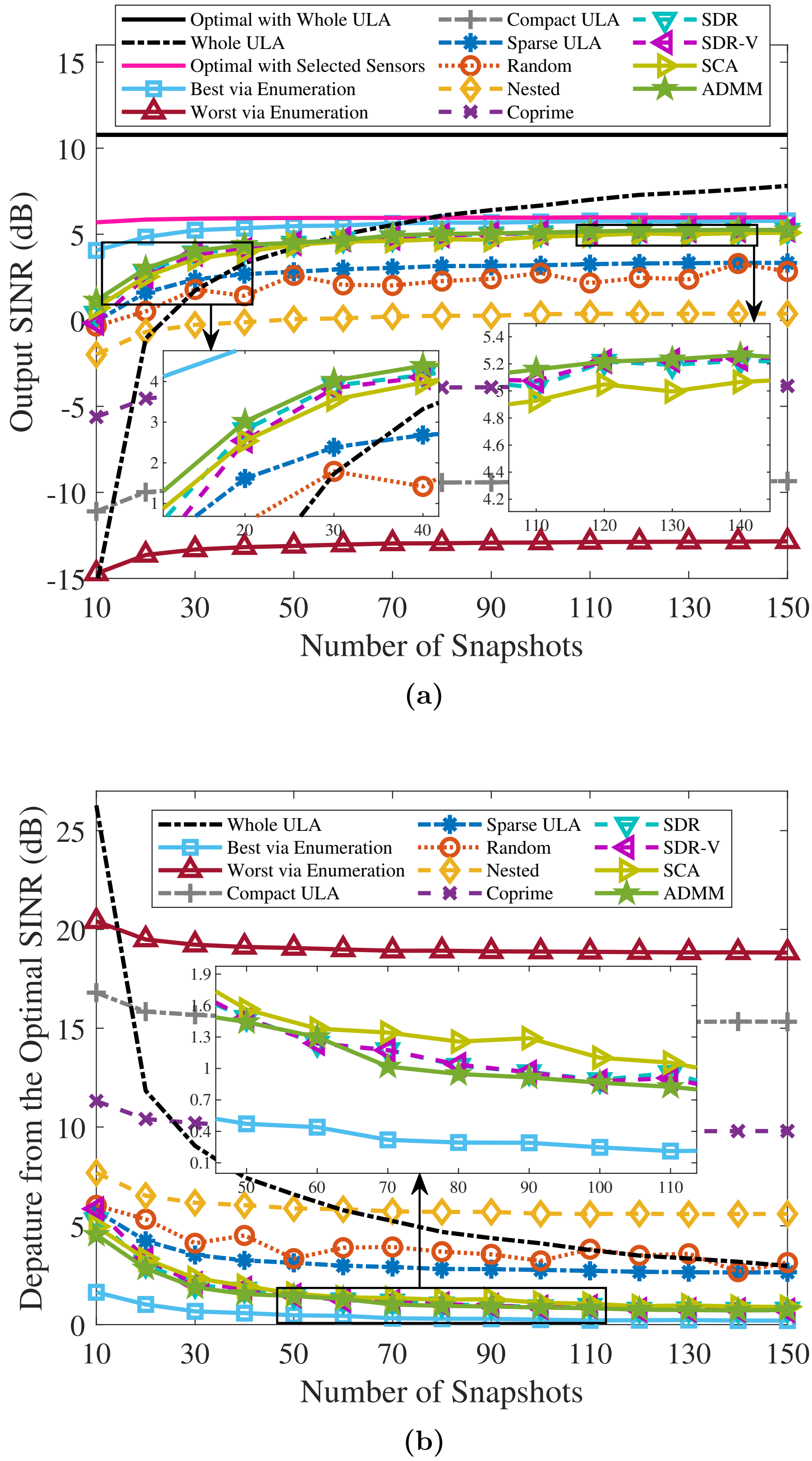


Figure 3.10: Output SINR versus T. (9th example)

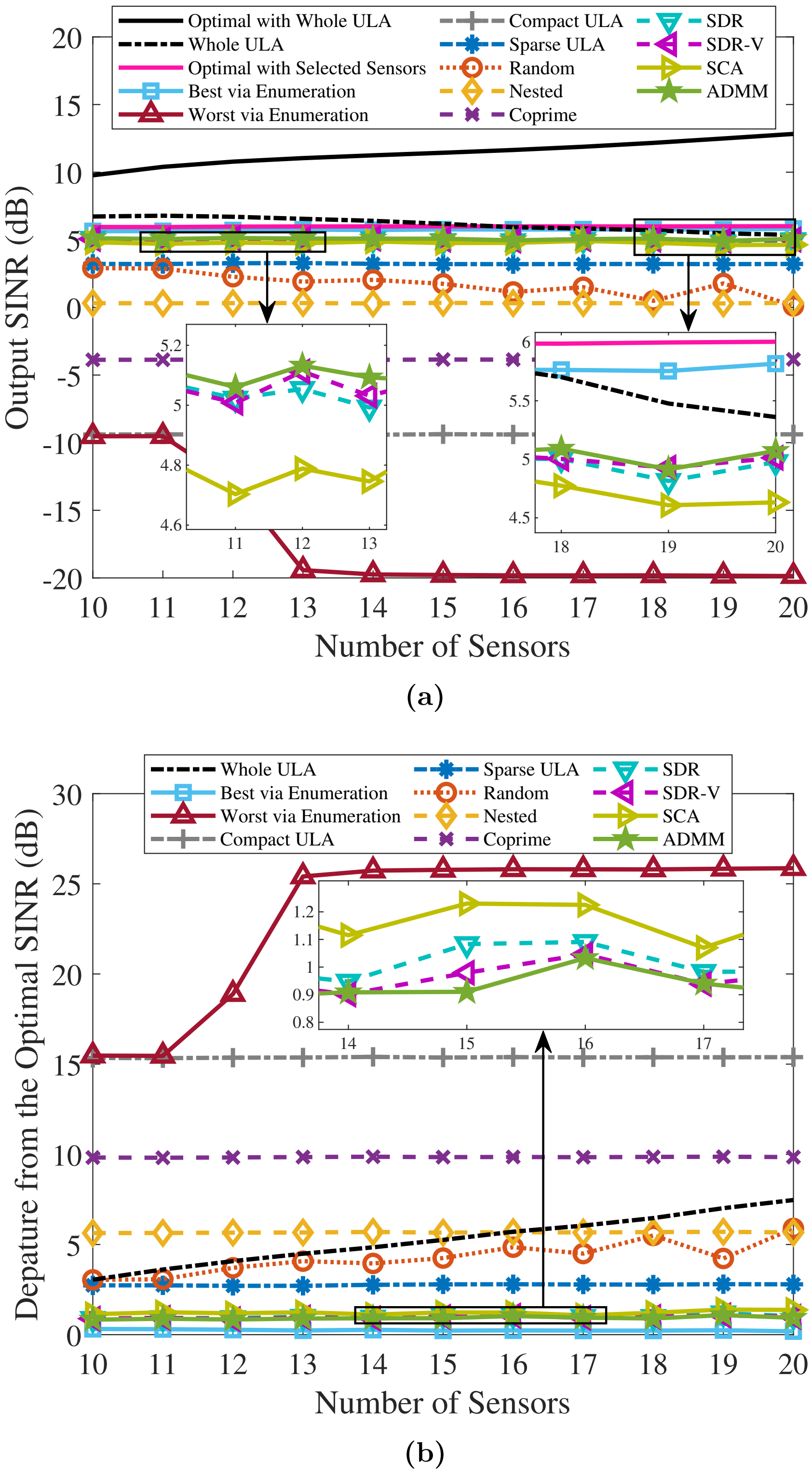
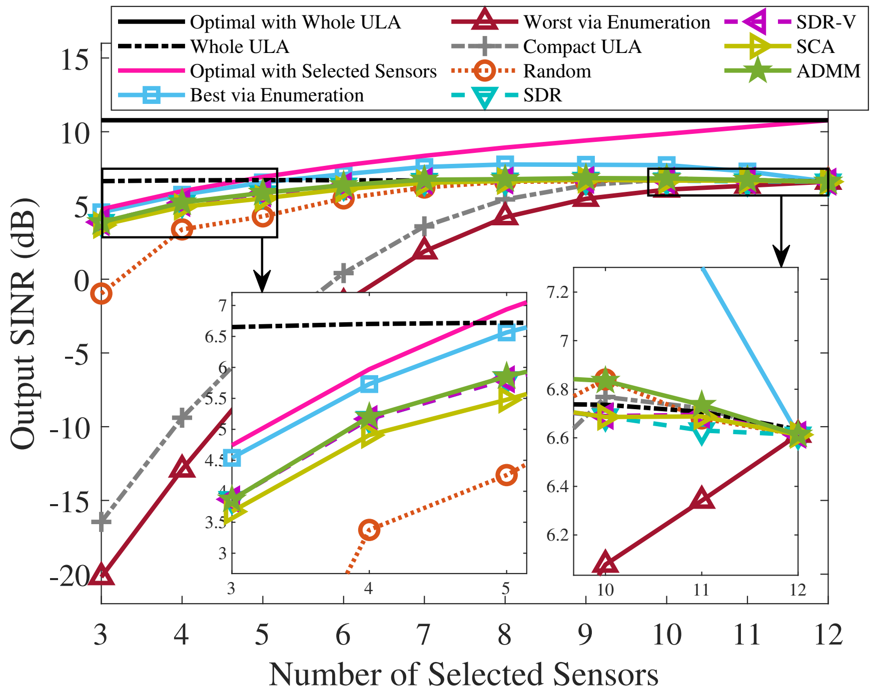


Figure 3.11: Output SINR versus M. (10th example)

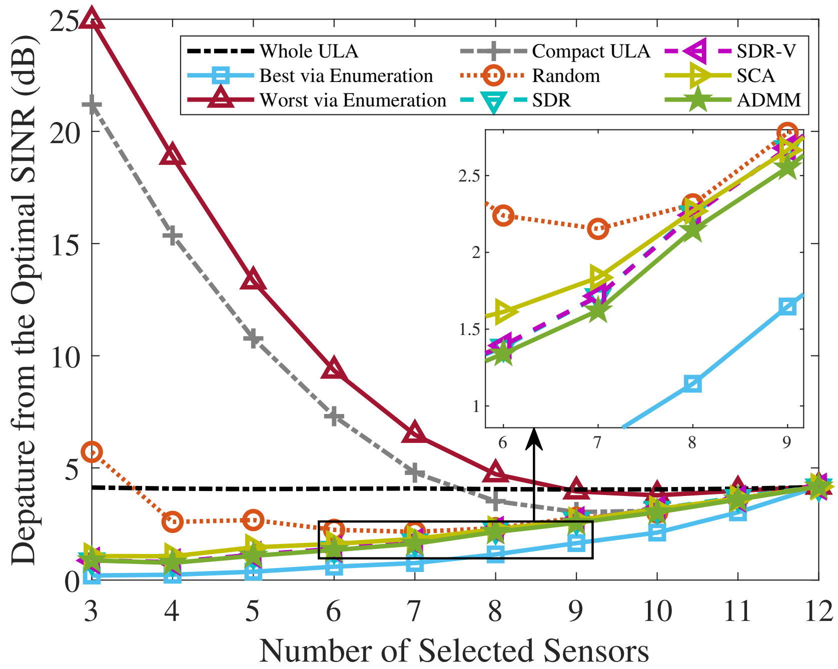
Eleventh example: We examine the output SINR versus the number of selected sensors. The simulation setup is the same as that of the fifth example. Note that, some previous methods are not included in this example since they do not apply to the entire range of L. The results are displayed in Figure 3.12. It is seen that when L approaches $M = 12$, the output SINR of all the tested methods converge to one point, because in this case ($L = M$) all the methods select the same sensors, i.e., the whole ULA. On the other hand, when $L < 12$, the proposed ADMM has higher output SINR than the other tested approaches (except for the best case via enumeration).

3.7 Summary

An algorithm based on alternating direction method of multipliers (ADMM) for sparse array beamformer design was proposed. Our approach provides closed-form solutions at each ADMM iteration. Theoretical analyses and numerical simulations were provided to show the convergence of the proposed algorithm. In addition, the algorithm was proven to converge to the set of stationary points. The ADMM algorithm was shown comparable to the exhaustive search method, and slightly better than the state-of-the-art solvers, including the semidefinite relaxation (SDR), an SDR variant (SDR-V), and the successive convex approximation (SCA) methods, and significantly better than several other sparse array design strategies in terms of output signal-to-interference-plus-noise ratio. Moreover, the proposed ADMM algorithm outperformed the SDR, SDR-V, and SCA approaches in terms of computational cost.



(a)



(b)

Figure 3.12: Output SINR versus L. (11th example)

Chapter 4

DOA Estimation with a Sparsely Distorted Array

Distorted sensors could occur randomly and may lead to the breakdown of a sensor array system. We consider an array model within which a small number of sensors are distorted by unknown sensor gain and phase errors. With such an array model, the problem of joint direction-of-arrival (DOA) estimation and distorted sensor detection is formulated under the framework of low-rank and row-sparse decomposition. We derive an iteratively reweighted least squares (IRLS) algorithm to solve the resulting problem in both noiseless and noisy cases. The convergence property of the IRLS algorithm is analyzed by means of the monotonicity and boundedness of the objective function. Extensive simulations are conducted regarding parameter selection, convergence speed, computational complexity, and performances of DOA estimation as well as distorted sensor detection. Even though the IRLS algorithm is slightly worse than the alternating direction method of multipliers in detecting the distorted sensors, the results show that our approach outperforms several state-of-the-art techniques in terms of convergence speed, computational cost, and DOA estimation performance.

The key contributions presented in this chapter originate from [HZ21] and [HLSZ22]. The structure of this chapter is as follows: Motivation is presented in 4.1. The signal model and problem statement are established in Section 4.2. A review of state-of-the-art works is provided in Section 4.3. Section 4.4 derives an IRLS algorithm for joint DOA estimation and distorted sensor detection. Numerical results are given in Section 4.5, while Section 4.6 summarizes this chapter.

4.1 Motivation

Direction-of-arrival (DOA) estimation is one of the most important topics in array signal processing, which has found numerous applications in radar, sonar, wireless communications, to name just a few [KV96, VT02, Vib14]. Many classical approaches have been proposed, including multiple signal classification (MUSIC) [Sch86], estimation of signal parameters via rotational invariance techniques (ESPRIT) [RK89], and maximum likelihood methods [Bö86, ZW88]. However, it is known that most of these high-resolution algorithms rely heavily on the exact knowledge of the array manifold, and hence their performance may greatly suffer when the sensor array encounters distortions [VGLM03, WZW17, WDC+17, YdL17, RCC+18, HFZ19], such as unknown sensor gain and phase uncertainties, which is the focus of this chapter. More recently, techniques based on low-rank and sparse matrix decomposition have been applied to DOA estimation or tracking, see e.g. [LLXZ15, Das17, MTPK14, MTPK19]. However, these works merely consider the well-calibrated array, and they are not straightforwardly applicable to an array with sensor errors.

There is a large number of works devoted to handle distorted or completely failed sensors [YL99, VSS07, MCR+12, ORM12, HZR12, ZWCS15, WZN17, LV19, SVWW96, NLEF09, JDC+13, PGW02, SG04, LC12, SPP14, LC14, SPPZ18, AA17, LRW19, HZ21]. In [YL99], the genetic algorithm [Hol92] was applied for array failure correction. A minimal resource allocation network was used for DOA estimation under array sensor failure [VSS07], which requires a training procedure with no failed sensors. A Bayesian compressive sensing approach was proposed in [ORM12], which needs a noise-free array as a reference. Methods using difference co-array were developed in [ZWCS15, WZN17, LV19]. The idea of [ZWCS15] was based on the fact that positions corresponding to damaged sensors may be occupied by virtual sensors and thus the impact of sensor failure could be avoided. However, this is not applicable when the failed sensors are located on the first or last position of the array, or when the malfunctioned sensors occur on symmetrical positions of the array, in which situations there exist *holes* in the difference co-array. On the other hand, [WZN17] and [LV19] restricted the array to some special sparse structures, such as co-prime and nested arrays. Approaches based on pre-calibrated sensors have been well-documented in the past decades [PGW02, SG04, LC12, SPP14, LC14, SPPZ18]. These methods require the knowledge of the calibrated sensors and they are time- and energy-consuming.

To circumvent the above-mentioned shortcomings, and to tackle the DOA estimation problem with an array in which a few sensors are distorted by unknown sensor gain and phase uncertainties, we formulate the problem under the framework of low-rank and row-sparse decomposition (LR^{2}SD), which can be regarded as a special structure of low-rank and sparse decomposition (LRSD). Note that LRSD is also known as robust principal component analysis (RPCA) [VBJN18, VCB18, BJZ+18]. The LRSD technique has become a popular tool in finding a low-dimensional subspace from sparsely and arbitrarily corrupted observations, and it has wide applications in science and engineering, ranging from bioinformatics, web search, to imaging, audio and video processing [Jol86, WPM+09, ZLT+11, LCM10, BIK+18]. Another special structure of LRSD is low-rank and column-sparse decomposition (LRCSD) [XCS12, LLY10, LLY+13, LLY15, LGS19], also known as RPCA-outlier pursuit [ZLZC15, LRR+18, GBZ12, RW14], which has been recently proposed to handle the scenarios where corruptions take place column-sparsely, meaning that the corruption matrix is column-wise sparse. Such situations occur for example when a fraction of the data vectors are grossly corrupted by outliers [LLY+13, LGS19].

Several algorithms have been contributed to solve the LRSD and LRCSD problems, such as singular value thresholding (SVT) [CCS10], accelerated proximal gradient (APG) [BT09], alternating direction method of multipliers (ADMM) [LCM10, LGS19], and iteratively reweighted least squares (IRLS) [LLY15, LGS19, GBZ12, RW14]. The SVT, APG, and ADMM methods will be reviewed in Section 4.3 in the context of joint DOA estimation and distorted sensor detection. The above three methods require one singular value decomposition (SVD) in each iteration, which may be unbearable for large scale problems. Instead, IRLS relies on simple linear algebra, and it generally has a linear convergence rate [DDFG09, BBPB14, EV19, SV21, KVS21]. In this sense, the IRLS is more efficient in solving the corresponding problems.

Therefore, in the present chapter, we develop an IRLS algorithm for joint DOA estimation and distorted sensor detection. The main contributions include:

- Both noiseless and noisy cases are considered. The convergence property of the algorithm is analyzed, via the monotonicity and boundedness of the objective function.
- The computational complexities of the IRLS algorithm as well as the SVT, APG, and ADMM methods are theoretically analyzed.
- Extensive simulations are conducted in view of parameter selection, convergence speed, computational time, and performance of DOA estimation and distorted sensor detection.

Notations: In this chapter, bold-faced lower-case and upper-case letters stand for vectors and matrices, respectively. Superscripts $\cdot^{\mathrm{T}}$ and $\cdot^{\mathrm{H}}$ denote transpose and Hermitian transpose, respectively. $\mathbb{C}$ is the set of complex numbers, and $\jmath = \sqrt{-1}$. For a real-valued scalar a, $|a|$ denotes its absolute value. The minimum value of two scalars a and b is denoted as $\min\{a, b\}$. $\|\cdot\|_2$ is the ℓ_2 norm of a vector. $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_*$ represent the Frobenius norm and the nuclear norm (sum of singular values) of a matrix, respectively. $\|\cdot\|_{2,0}$ and $\|\cdot\|_{2,1}$ denote the $\ell_{2,0}$ mixed-norm and $\ell_{2,1}$ mixed-norm of a matrix, respectively, whose definitions are given as $\|\mathbf{V}\|_{2,0} \triangleq \mathrm{card}(\{\|\mathbf{V}_{i,:}\|_2\})$ and $\|\mathbf{V}\|_{2,1} \triangleq \sum_{i=1}^{M} \|\mathbf{V}_{i,:}\|_2$, for $\mathbf{V} \in \mathbb{C}^{M\times T}$, where $\mathrm{card}(\cdot)$ is the cardinality of a set, $\{\|\mathbf{V}_{i,:}\|_2\} = \{\|\mathbf{V}_{1,:}\|_2, \|\mathbf{V}_{2,:}\|_2, \cdots, \|\mathbf{V}_{M,:}\|_2\}$, and $\mathbf{V}_{i,:}$ is the i-th row of $\mathbf{V}$. $\mathrm{rank}(\cdot)$ is

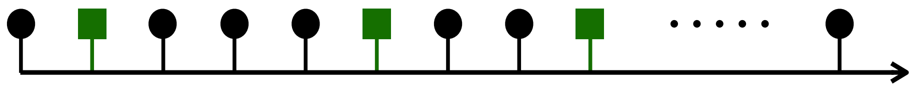


Figure 4.1: Illustration of array structure of interest.

the rank operator, defined as $\mathrm{rank}(\mathbf{Z}) \triangleq \mathrm{card}(\{\sigma_i(\mathbf{Z})\})$, with $\sigma_i(\mathbf{Z})$ being the i-th largest singular value of $\mathbf{Z}$ and $\{\sigma_i(\mathbf{Z})\}$ denoting the set containing all singular values of $\mathbf{Z}$. For two matrices $\mathbf{X}$ and $\mathbf{Y}$ of the same dimensions, we define their Frobenius inner product as $\langle \mathbf{X}, \mathbf{Y} \rangle \triangleq \mathrm{trace}(\mathbf{X}^{\mathrm{H}}\mathbf{Y})$, where $\mathrm{trace}(\cdot)$ denotes the trace of a square matrix.

4.2 Signal Model and Problem Statement

Suppose that a linear antenna array of M sensors receives K far-field narrowband signals from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_{(k)}]^{\mathrm{T}}$. The antenna array of interest is assumed to be randomly and *sparsely* distorted by sensor gain and phase uncertainty (the number of distorted sensors is far smaller than M). Further, we assume that the number of distorted sensors and their positions are unknown. Figure 4.1 illustrates the array model, where the black circles stand for *perfect* sensors and the green boxes refer to distorted ones. The green boxes appear randomly and sparsely within the whole linear array.

The array observation can be written as

$$\mathbf{y}(t) = \breve{\boldsymbol{\Gamma}}\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \triangleq (\mathbf{I} + \boldsymbol{\Gamma})\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t),$$

where $t = 1, 2, \cdots, T$ denotes the time index, T is the total number of available snapshots, $\mathbf{s}(t) \in \mathbb{C}^K$ and $\mathbf{n}(t) \in \mathbb{C}^M$ are signal and noise vectors, respectively. The steering matrix $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_{(k)})] \in \mathbb{C}^{M \times K}$ has steering vectors as columns, where the steering vector $\mathbf{a}(\theta_{(k)})$ is a function of $\theta_{(k)}$, for $k = 1, 2, \cdots, K$. In addition, $\breve{\boldsymbol{\Gamma}} \triangleq \mathbf{I} + \boldsymbol{\Gamma}$ indicates the electronic sensor status (either perfect or distorted), where $\mathbf{I}$ is the $M \times M$ identity matrix, and $\boldsymbol{\Gamma}$ is a diagonal matrix with its main diagonal, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_M]^{\mathrm{T}}$, being a sparse vector. Specifically, for $m = 1, 2, \cdots, M$

$$\gamma_m \begin{cases} = 0, & \text{if the } m\text{-th sensor is perfect,} \\ \neq 0, & \text{if the } m\text{-th sensor is distorted.} \end{cases}$$

The non-zero γ_m denotes sensor gain and phase error, namely, $\gamma_m = \rho_m e^{\jmath \phi_m}$, where ρ_m and ϕ_m are the gain and phase errors of the m-th sensor, respectively.

Collecting all the snapshots into a matrix, we have

$$\mathbf{Y} = (\mathbf{I} + \mathbf{\Gamma})\mathbf{A}\mathbf{S} + \mathbf{N}, \tag{4.1}$$

where $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \cdots, \mathbf{y}(T)] \in \mathbb{C}^{M\times T}$ contains the measurements, $\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), \cdots, \mathbf{s}(T)] \in \mathbb{C}^{K\times T}$ denotes the signal matrix, and $\mathbf{N} = [\mathbf{n}(1), \mathbf{n}(2), \cdots, \mathbf{n}(T)] \in \mathbb{C}^{M\times T}$ is the noise matrix. Defining $\mathbf{Z} \triangleq \mathbf{A}\mathbf{S}$ and $\mathbf{V} \triangleq \mathbf{\Gamma}\mathbf{A}\mathbf{S}$, (4.1) becomes:

$$\mathbf{Y} = \mathbf{Z} + \mathbf{V} + \mathbf{N}, \tag{4.2}$$

where $\mathbf{Z} \in \mathbb{C}^{M\times T}$ is a low-rank matrix of rank K (in general $K < \min\{M, T\}$), and $\mathbf{V} \in \mathbb{C}^{M\times T}$ is a row-sparse (meaning that only a few rows are non-zero) matrix due to the sparsity of the main diagonal of $\mathbf{\Gamma}$.

Given the array measurements $\mathbf{Y}$, our task is to simultaneously estimate the incoming directions of signals and detect the distorted sensors within the array. Note that the number of distorted sensors is small, but unknown, and their positions are unknown as well.

4.3 Related Works

Related works for solving the joint DOA estimation and distorted sensor detection include SVT, APG, and ADMM. The SVT method was first proposed for matrix completion, see for example [CCS10]. By adapting the SVT algorithm to our problem, we need to solve

$$\min_{\mathbf{Z},\mathbf{V},\mathbf{W}} \ \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1} + \frac{1}{2\tau}\|\mathbf{Z}\|_{\mathrm{F}}^2 + \frac{1}{2\tau}\|\mathbf{V}\|_{\mathrm{F}}^2 + \frac{1}{\tau}\langle \mathbf{W}, \mathbf{Y}-\mathbf{Z}-\mathbf{V}\rangle, \tag{4.3}$$

where λ is a tuning parameter, τ is a large positive scalar such that the objective function is perturbed slightly. The SVT approach iteratively updates $\mathbf{Z}$, $\mathbf{V}$, and $\mathbf{W}$. $\mathbf{Z}$ and $\mathbf{V}$ are updated by solving the above problem with $\mathbf{W}$ fixed. Then $\mathbf{W}$ is updated as $\mathbf{W} = \mathbf{Y} - \mathbf{Z} - \mathbf{V}$. The following well-known results are used when updating $\mathbf{Z}$ and $\mathbf{V}$ [CCS10]:

$$\begin{aligned}
\mathbf{L}\mathcal{S}_\kappa(\mathbf{S})\mathbf{R}^{\mathrm{H}} &= \arg\min_{\mathbf{X}} \ \kappa\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{C}\|_{\mathrm{F}}^2, \\
\mathcal{S}_\kappa(\mathbf{C}) &= \arg\min_{\mathbf{X}} \ \kappa\|\mathbf{X}\|_{2,1} + \frac{1}{2}\|\mathbf{X} - \mathbf{C}\|_{\mathrm{F}}^2,
\end{aligned}$$

where $\mathbf{LSR}^{\mathrm{H}}$ is the SVD of $\mathbf{C}$ and the element-wise soft-thresholding operator is defined as:

$$\mathcal{S}_{\kappa}(x) = \begin{cases} x - \kappa, \text{ if } x > \kappa, \\ x + \kappa, \text{ if } x < \kappa, \\ 0, \text{ otherwise,} \end{cases}$$

with parameter $\kappa > 0$. The applicability of SVT is limited since it is difficult to select the step size for speedup [LCM10].

The second method is APG, whose updating equation can be given as [BT09]

$$(\mathbf{Z}_{(k+1)}, \mathbf{V}_{(k+1)}) = \arg\min_{\mathbf{Z},\mathbf{V}} \; h(\mathbf{Z}, \mathbf{V}) \tag{4.4}$$

where subscript $\cdot_{(k)}$ denotes the variable at the k-th iteration, $h(\mathbf{Z}, \mathbf{V}) \triangleq p(\mathbf{Z}_{(k)}, \mathbf{V}_{(k)}) + \langle \nabla_{\mathbf{Z}_{(k)}} p(\mathbf{Z}, \mathbf{V}_{(k)}), \mathbf{Z} - \mathbf{Z}_{(k)} \rangle + \langle \nabla_{\mathbf{V}_{(k)}} p(\mathbf{Z}_{(k)}, \mathbf{V}), \mathbf{V} - \mathbf{V}_{(k)} \rangle + \mu M \|\mathbf{Z} + \mathbf{V} - \mathbf{Z}_{(k)} - \mathbf{V}_{(k)}\|_{\mathrm{F}}^2 + q(\mathbf{Z}, \mathbf{V})$, with $p(\mathbf{Z}, \mathbf{V}) \triangleq \frac{1}{\mu} \|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\mathrm{F}}^2$, $q(\mathbf{Z}, \mathbf{V}) \triangleq \|\mathbf{Z}\|_* + \lambda \|\mathbf{V}\|_{2,1}$, and μ being a small positive scalar. The detailed algorithm can be found in [BT09] and also [LCM10].

As for ADMM, we consider the following problem

$$\min_{\mathbf{Z},\mathbf{V}} \; \|\mathbf{Z}\|_* + \lambda \|\mathbf{V}\|_{2,1} \quad \text{s.t. } \mathbf{Y} = \mathbf{Z} + \mathbf{V}, \tag{4.5}$$

and its augmented Lagrangian function is $\mathcal{L}_{\mu}(\mathbf{Z}, \mathbf{V}, \mathbf{W}) = \|\mathbf{Z}\|_* + \lambda \|\mathbf{V}\|_{2,1} + \langle \mathbf{W}, \mathbf{Y} - \mathbf{Z} - \mathbf{V} \rangle + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\mathrm{F}}^2$, where $\mathbf{W}$ denotes the dual variable and μ is the augmented Lagrangian parameter. Then ADMM updates $\mathbf{Z}$, $\mathbf{V}$, and $\mathbf{W}$, in a sequential manner. $\mathbf{Z}$ and $\mathbf{V}$ are solved by minimizing $\mathcal{L}_{\mu}(\mathbf{Z}, \mathbf{V}, \mathbf{W})$ with respect to (w.r.t.) $\mathbf{Z}$ (resp. $\mathbf{V}$) while keeping $\mathbf{V}$ (resp. $\mathbf{Z}$) and $\mathbf{W}$ unchanged; $\mathbf{W}$ is updated as $\mathbf{W} = \mathbf{W} + \mu(\mathbf{Y} - \mathbf{Z} - \mathbf{V})$ [BPC+11].

All the aforementioned three algorithms require performing one SVD per iteration. Therefore, their computational complexity is extremely high, especially when the problem size is large. Their convergence speed and computational cost will be compared in simulations.

4.4 Proposed Method

In this section, we develop an IRLS algorithm for the task of jointly estimating DOAs of sources and detecting distorted sensors. We start by considering the noiseless case, and then focus on the noisy case.

4.4.1 Noiseless Case

In the noiseless case, the data model (4.2) is simplified as $\mathbf{Y} = \mathbf{Z} + \mathbf{V}$. Therefore, we formulate the following LR^{2}SD problem, as

$$\min_{\mathbf{Z},\mathbf{V}} \ \text{rank}(\mathbf{Z}) + \lambda \|\mathbf{V}\|_{2,0} \quad \text{s.t. } \mathbf{Y} = \mathbf{Z} + \mathbf{V}, \tag{4.6}$$

where λ is a tuning parameter. By substituting the equality constraint into the objective, and replacing the rank and $\ell_{2,0}$ mixed-norm with the nuclear norm and $\ell_{2,1}$ mixed-norm, respectively, we have its convex counterpart, as

$$\min_{\mathbf{Z}} \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{Y} - \mathbf{Z}\|_{2,1}. \tag{4.7}$$

The nuclear norm and the $\ell_{2,1}$ mixed-norm are non-smooth, and thus they are not differentiable at some points. To deal with this issue, we introduce a smoothing parameter μ, and obtain the gradients as

$$\begin{aligned} \frac{\partial \|[\mathbf{Z}, \mu \mathbf{I}]\|_*}{\partial \mathbf{Z}} &= \mathbf{P}\mathbf{Z} \\ \frac{\partial \|[\mathbf{Y} - \mathbf{Z}, \mu \mathbf{1}]\|_{2,1}}{\partial \mathbf{Z}} &= \mathbf{Q}(\mathbf{Z} - \mathbf{Y}), \end{aligned}$$

where $\mathbf{1}$ is an all-ones vector of appropriate length, $\mathbf{P} \triangleq \left(\mathbf{Z}\mathbf{Z}^{\text{H}} + \mu^2 \mathbf{I}\right)^{-\frac{1}{2}}$ and

$$\mathbf{Q} \triangleq \begin{bmatrix} \frac{1}{\sqrt{\|(\mathbf{Y}-\mathbf{Z})_{1,:}\|_2^2 + \mu^2}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\|(\mathbf{Y}-\mathbf{Z})_{M,:}\|_2^2 + \mu^2}} \end{bmatrix}. \tag{4.8}$$

The problem to be solved now turns to be

$$\min_{\mathbf{Z}} \ f(\mathbf{Z}) \triangleq \|[\mathbf{Z}, \mu \mathbf{I}]\|_* + \lambda \|[\mathbf{Y} - \mathbf{Z}, \mu \mathbf{1}]\|_{2,1}, \tag{4.9}$$

where the objective function $f(\mathbf{Z})$ is differentiable everywhere w.r.t. $\mathbf{Z}$, as long as $\mu \neq 0$. The derivative of $f(\mathbf{Z})$ w.r.t. $\mathbf{Z}$ is

$$\frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}} = \mathbf{P}\mathbf{Z} + \lambda \mathbf{Q}(\mathbf{Z} - \mathbf{Y}).$$

According to the Karush-Kuhn-Tucker (KKT) condition, we have $\mathbf{P}\mathbf{Z} + \lambda \mathbf{Q}(\mathbf{Z} - \mathbf{Y}) = \mathbf{0}$, indicating that $\mathbf{Z} = \lambda(\mathbf{P} + \lambda \mathbf{Q})^{-1}\mathbf{Q}\mathbf{Y}$. This leads to the IRLS iterative process as

$$\mathbf{Z}_{(k+1)} = \lambda(\mathbf{P}_{(k)} + \lambda \mathbf{Q}_{(k)})^{-1} \mathbf{Q}_{(k)} \mathbf{Y}, \tag{4.10}$$

where both $\mathbf{P}_{(k)}$ and $\mathbf{Q}_{(k)}$ are dependent on $\mathbf{Z}_{(k)}$. The IRLS algorithm for the noiseless case is summarized in Algorithm 4.1, where ϵ is a small scalar and $k_{\max}$ is a large scalar, used to terminate the algorithm.

Algorithm 4.1 IRLS algorithm for noiseless case

Input : $\mathbf{Y} \in \mathbb{C}^{M\times T}$, λ, μ, ϵ, $k_{\max}$
Output : $\widehat{\mathbf{Z}} \in \mathbb{C}^{M\times T}$, $\widehat{\mathbf{V}} \in \mathbb{C}^{M\times T}$
Initialize: $\mathbf{Z}_0 \leftarrow \mathbf{Z}_{\text{init}}$, $\mathbf{V}_0 \leftarrow \mathbf{V}_{\text{init}}$, $k \leftarrow 0$

```
while not converged do
    k ← k + 1
    calculate P_(k) and Q_(k)
    update Z_(k) using Z = λ(P + λQ)^{-1} Q Y
    converged ← k ≥ k_max or |f(Z_(k)) − f(Z_(k−1))| / |f(Z_(k))| ≤ ε
end while
Ẑ ← Z_(k), V̂ ← Y − Z_(k)
```

4.4.2 Convergence Analysis for Noiseless Case

We first provide two lemmata giving two important inequalities regarding the trace function and the $\ell_{2,1}$ mixed-norm. Then, we prove the monotonicity and the boundedness of the objective function in Problem (4.9).

Lemma 4.1 (Lemma 2 in [LLY15]) *For any two symmetric positive definite matrices* $\mathbf{X}$ *and* $\mathbf{Y}$*, it holds that* $\text{trace}\left(\mathbf{Y}^{\frac{1}{2}}\right) - \text{trace}\left(\mathbf{X}^{\frac{1}{2}}\right) \geq \text{trace}\left(\frac{1}{2}(\mathbf{Y}-\mathbf{X})^{\text{H}}\mathbf{Y}^{-\frac{1}{2}}\right)$.

Lemma 4.2 *For any matrices* $\mathbf{X}$ *and* $\mathbf{Y} \in \mathbb{C}^{M\times T}$*, we have* $\|\mathbf{Y}\|_{2,1} - \|\mathbf{X}\|_{2,1} \geq \frac{1}{2}\text{trace}\big(\mathbf{H}\big(\mathbf{Y}\mathbf{Y}^{\text{H}} - \mathbf{X}\mathbf{X}^{\text{H}}\big)\big)$*, where*

$$\mathbf{H} = \begin{bmatrix} \frac{1}{\|\mathbf{Y}_{1,:}\|_2} & & \\ & \ddots & \\ & & \frac{1}{\|\mathbf{Y}_{M,:}\|_2} \end{bmatrix}. \tag{4.11}$$

Proof: Due to the concavity of function $\sqrt{x}$ $(x \geq 0)$, we have $\sqrt{y} - \sqrt{x} \geq \frac{1}{2\sqrt{y}}(y-x)$ for all $x \geq 0$ and $y \geq 0$. Therefore,

$$\begin{aligned} \|\mathbf{Y}\|_{2,1} - \|\mathbf{X}\|_{2,1} &= \sum_i^M \left[\sqrt{\|\mathbf{Y}_{i,:}\|_2^2} - \sqrt{\|\mathbf{X}_{i,:}\|_2^2}\right] \\ &\geq \sum_i^M \left[\frac{1}{2\|\mathbf{Y}_{i,:}\|_2}\left(\|\mathbf{Y}_{i,:}\|_2^2 - \|\mathbf{X}_{i,:}\|_2^2\right)\right] \\ &= \frac{1}{2}\text{trace}\big(\mathbf{H}\big(\mathbf{Y}\mathbf{Y}^{\text{H}} - \mathbf{X}\mathbf{X}^{\text{H}}\big)\big), \end{aligned}$$

where $\mathbf{H}$ is given by (4.11). $\square$

With Lemmata 4.1 and 4.2, we have the following theorem.

Theorem 4.1 *The sequence $\{\mathbf{Z}_{(k)}\}$ generated by $\mathbf{Z}_{(k+1)} = \lambda(\mathbf{P}_{(k)} + \lambda\mathbf{Q}_{(k)})^{-1}\mathbf{Q}_{(k)}\mathbf{Y}$ produces a non-increasing objective function defined in (4.9), i.e., $f(\mathbf{Z}_{(k)}) \geq f(\mathbf{Z}_{(k+1)})$ for $k = 0, 1, 2, \cdots$. Moreover, the sequence $\{\mathbf{Z}_{(k)}\}$ is bounded, and $\lim_{k\to\infty} \|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}} = 0$.*

Proof: See Appendix B.1. $\square$

Theorem 4.2 *The objective function $f(\mathbf{Z}) = \|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda\|[\mathbf{Y} - \mathbf{Z}, \mu\mathbf{1}]\|_{2,1}$ is bounded below by $|\mu|(\sqrt{M} + \lambda M)$.*

Proof: See Appendix B.2. $\square$

Theorem 4.3 *Any limit point of the sequence $\{\mathbf{Z}_{(k)}\}$ generated by (4.10) is a stationary point of Problem (4.9), and moreover, the stationary point is globally optimal.*

Proof: See Appendix B.3. $\square$

4.4.3 Noisy Case

In the noisy case, the data model is as (4.2), and the problem to be solved is given as

$$\min_{\mathbf{Z},\mathbf{V}} \frac{1}{2}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\mathrm{F}}^2 + \lambda_1\|\mathbf{Z}\|_* + \lambda_2\|\mathbf{V}\|_{2,1}, \tag{4.12}$$

where λ_1 and λ_2 are two tuning parameters. Different from the noiseless case, we have to optimize the problem with two variables, i.e., $\mathbf{Z}$ and $\mathbf{V}$. To proceed, we also introduce a smoothing parameter μ into the nuclear norm and the $\ell_{2,1}$ mixed-norm in Problem (4.12). Therefore, the problem to be addressed is transferred to

$$\min_{\mathbf{Z},\mathbf{V}} f(\mathbf{Z}, \mathbf{V}), \tag{4.13}$$

Algorithm 4.2 IRLS algorithm for noisy case

Input : $\mathbf{Y} \in \mathbb{C}^{M \times T}$, λ_1, λ_2, μ, ϵ, $k_{\max}$
Output : $\widehat{\mathbf{Z}} \in \mathbb{C}^{M \times T}$, $\widehat{\mathbf{V}} \in \mathbb{C}^{M \times T}$
Initialize: $\mathbf{Z}_0 \leftarrow \mathbf{Z}_{\text{init}}$, $\mathbf{V}_0 \leftarrow \mathbf{V}_{\text{init}}$, $k \leftarrow 0$

```
while not converged do
    k ← k + 1
    calculate P_(k) and Q_(k)
    update Z_(k) using Z = (I + λ1 P)^{-1} (Y − V)
    update V_(k) using V = (I + λ2 Q)^{-1} (Y − Z)
    converged ← k ≥ k_max or |f(Z_(k),V_(k)) − f(Z_(k−1),V_(k−1))| / |f(Z_(k),V_(k))| ≤ ε
end while
Ẑ ← Z_(k), V̂ ← V_(k)
```

where the objective function is defined as $f(\mathbf{Z}, \mathbf{V}) \triangleq \frac{1}{2}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_{\mathrm{F}}^2 + \lambda_1 \|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda_2 \|[\mathbf{V}, \mu\mathbf{1}]\|_{2,1}$. The derivatives of $f(\mathbf{Z}, \mathbf{V})$ w.r.t. $\mathbf{Z}$ and $\mathbf{V}$ are

$$\frac{\partial f(\mathbf{Z}, \mathbf{V})}{\partial \mathbf{Z}} = (-\mathbf{Y} + \mathbf{Z} + \mathbf{V}) + \lambda_1 \mathbf{P}\mathbf{Z},$$
$$\frac{\partial f(\mathbf{Z}, \mathbf{V})}{\partial \mathbf{V}} = (-\mathbf{Y} + \mathbf{Z} + \mathbf{V}) + \lambda_2 \mathbf{Q}\mathbf{V},$$

respectively, where $\mathbf{P}$ is defined the same as that in the noiseless case, and

$$\mathbf{Q} \triangleq \begin{bmatrix} \frac{1}{\sqrt{\|\mathbf{V}_{1,:}\|_2^2 + \mu^2}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\|\mathbf{V}_{M,:}\|_2^2 + \mu^2}} \end{bmatrix}. \tag{4.14}$$

Note that $\mathbf{Q}$ given in (4.8) is exactly the same as the one in (4.14) since $\mathbf{V} = \mathbf{Y} - \mathbf{Z}$ in the noiseless case.

According to the KKT condition, we have

$$\begin{cases} (\mathbf{I} + \lambda_1 \mathbf{P})\mathbf{Z} - \mathbf{Y} + \mathbf{V} = \mathbf{0} \\ (\mathbf{I} + \lambda_2 \mathbf{Q})\mathbf{V} - \mathbf{Y} + \mathbf{Z} = \mathbf{0} \end{cases}$$

which leads to the IRLS procedure as

$$\begin{cases} \mathbf{Z}_{(k+1)} = (\mathbf{I} + \lambda_1 \mathbf{P}_{(k)})^{-1}(\mathbf{Y} - \mathbf{V}_{(k)}) \\ \mathbf{V}_{(k+1)} = (\mathbf{I} + \lambda_2 \mathbf{Q}_{(k)})^{-1}(\mathbf{Y} - \mathbf{Z}_{(k+1)}), \end{cases} \tag{4.15}$$

where $\mathbf{P}_{(k)}$ and $\mathbf{Q}_{(k)}$ are dependent on $\mathbf{Z}_{(k)}$ and $\mathbf{V}_{(k)}$, respectively. The IRLS algorithm for the noisy case is summarized in Algorithm 4.2.

4.4.4 Convergence Analysis for Noisy Case

In this part, the monotonicity and boundedness of the objective function $f(\mathbf{Z}, \mathbf{V})$ in (4.13) are proved in Theorems 4.4 and 4.5, respectively.

Theorem 4.4 *The sequence* $\{(\mathbf{Z}_{(k)}, \mathbf{V}_{(k)})\}$ *generated by (4.15) produces a non-increasing objective function defined in (4.13), i.e.,* $f(\mathbf{Z}_{(k)}, \mathbf{V}_{(k)}) \geq f(\mathbf{Z}_{(k+1)}, \mathbf{V}_{(k+1)})$ *for* $k = 0, 1, 2, \cdots$. *Moreover, the sequence* $\{(\mathbf{Z}_{(k)}, \mathbf{V}_{(k)})\}$ *is bounded, and* $\lim_{k\to\infty} \|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}} = 0$ *and* $\lim_{k\to\infty} \|\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\|_{\mathrm{F}} = 0$.

Proof: See Appendix B.4. □

Theorem 4.5 *The objective function* $f(\mathbf{Z},\mathbf{V}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\mathrm{F}}^2 + \lambda_1\|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda_2\|[\mathbf{V}, \mu\mathbf{1}]\|_{2,1}$ *is bounded below by* $|\mu|(\lambda_1\sqrt{M} + \lambda_2 M)$.

Proof: See Appendix B.5. □

Theorem 4.6 *Any limit point of the sequence* $\{(\mathbf{Z}_{(k)}, \mathbf{V}_{(k)})\}$ *generated by (4.15) is a stationary point of Problem (4.13), and moreover, the stationary point is globally optimal.*

Proof: See Appendix B.6. □

The differences between our work and [LLY15] are stated as follows.

- The problem formulation in [LLY15] is column-sparse, while we have row-sparsity of $\mathbf{V}$. This leads to differences in matrix multiplication and matrix derivative.
- [LLY15] considers the noiseless case only, while we consider both noiseless and noisy cases.
- To update $\mathbf{Z}$ using matrices $\mathbf{P}$ and $\mathbf{Q}$, the approach in [LLY15] involves a Sylvester equation and utilizes the Matlab command `lyap`. However, our method admits a closed-form formula, see (4.10) and (4.15).
- The proofs of convergence are not exactly the same. [LLY15] proves the monotonicity of the objective and the boundedness of the sequence $\{\mathbf{Z}_{(k)}\}$. We prove the monotonicity and the boundedness of the objective in both noiseless and noisy cases.

Algorithm 4.3 Detection of distorted sensors

Input : $\widehat{\mathbf{V}} \in \mathbb{C}^{M\times T}$, h
Output: M_{fail}
calculate $\mathbf{v} = [\|\widehat{\mathbf{V}}_{1,:}\|_2, \|\widehat{\mathbf{V}}_{2,:}\|_2, \cdots, \|\widehat{\mathbf{V}}_{M,:}\|_2]^{\mathrm{T}}$
calculate $\tilde{\mathbf{v}} = \mathrm{sort}(\mathbf{v}, \text{`ascend'})$
calculate $d = \tilde{\mathbf{v}}(2) - \tilde{\mathbf{v}}(1)$ and assign $i_{\text{fail}} = M + 1$
for $i = 3, 4, \cdots, M$ **do**
    if $\tilde{\mathbf{v}}(i) - \tilde{\mathbf{v}}(i-1) \geq h$ **then**
        $i_{\text{fail}} = i$ and break the **for** loop
    end if
end for
$M_{\text{fail}} \leftarrow M - i_{\text{fail}} + 1$

4.4.5 DOA Estimation and Distorted Sensor Detection

Once $\widehat{\mathbf{Z}}$ and $\widehat{\mathbf{V}}$ are resolved, they can be adopted to estimate the DOAs and detect the distorted sensors, respectively. Note that $\mathbf{Z} = \mathbf{AS}$ can be viewed as a noise-free data model. DOAs can be found via subspace-based methods, such as MUSIC, whose spatial spectrum is

$$P(\theta) = \frac{1}{\mathbf{a}^{\mathrm{H}}(\theta)(\mathbf{I} - \mathbf{LL}^{\mathrm{H}})\mathbf{a}(\theta)}.$$

The SVD of $\widehat{\mathbf{Z}}$ is $\widehat{\mathbf{Z}} = \mathbf{L\Sigma R}^{\mathrm{H}}$, where the columns of $\mathbf{L}$ and $\mathbf{R}$ contain the left and right orthogonal base vectors of $\widehat{\mathbf{Z}}$, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix whose diagonal elements are the singular values of $\widehat{\mathbf{Z}}$ arranged in descending order. Under the assumption that the number of sources, i.e., K, is known, the DOAs are determined by searching for the K largest peaks of $P(\theta)$.

On the other hand, the number of distorted sensors and their positions can be determined by $\|\widehat{\mathbf{V}}_{i,:}\|_2$, $i = 1, 2, \cdots, M$. Algorithm 4.3 shows a strategy for detecting the distorted sensors. In words, we first calculate the ℓ_2 norm of each row of $\widehat{\mathbf{V}}$ and form a vector, say $\mathbf{v}$, and then we sort these ℓ_2 norms in ascending order and obtain $\tilde{\mathbf{v}}$. We define the difference of the first two entries of $\tilde{\mathbf{v}}$ as $d = \tilde{\mathbf{v}}(2) - \tilde{\mathbf{v}}(1)$. Next, for $i = 3, 4, \cdots, M$, we compute $\tilde{\mathbf{v}}(i) - \tilde{\mathbf{v}}(i-1)$ and compare it with a threshold, say h, of large value: if it is larger than or equal to h, we set $i_{\text{fail}} = i$ and break the for loop; if it is less than h, we have $i_{\text{fail}} = M + 1$. Finally, the number of distorted sensors is obtained as $M_{\text{fail}} = M - i_{\text{fail}} + 1$.

4.5 Simulation Results

4.5.1 Parameter Selection

In this subsection, we discuss the problem of choosing appropriate values for μ, λ_1, and λ_2 in Problem (4.13) used in Algorithm 4.2. We set $\epsilon = 10^{-16}$, $k_{\max} = 1000$, and $\mathbf{Z}_{\text{init}} = \mathbf{V}_{\text{init}} = \mathbf{O}$ in Algorithm 4.2, where $\mathbf{O}$ denotes the $M \times T$ all-zeros matrix. We define the root-mean squared error (RMSE) of DOA estimates as:

$$\text{RMSE} = \sqrt{\frac{1}{QK} \sum_{q=1}^{Q} \sum_{k=1}^{K} (\hat{\theta}_{k,q} - \theta_{(k)})^2},$$

where $\hat{\theta}_{k,q}$ is the estimate of the k-th signal in the q-th Monte Carlo trial, and Q is the total number of Monte Carlo trials. The RMSE is used as a metric to select appropriate values for μ, λ_1, and λ_2. The plots in this subsection are averaged over $Q = 1000$ trials.

Consider a uniform linear array (ULA) of $M = 10$ sensors, 4 of which at random positions are distorted by gain and phase errors, receiving $K = 2$ signals with DOAs $\boldsymbol{\theta} = [-10°, 10°]^{\mathrm{T}}$. The sensor gain and phase errors are randomly generated by drawing from uniform distributions on $[0, 10]$ and $[-15°, 15°]$, respectively. In the first example, we test 6 scenarios with different signal-to-noise ratios (SNRs) and different numbers of snapshots. In Figure 4.2, we fix $\lambda_1 = 2$ and $\lambda_2 = 0.2$, and plot RMSE versus μ. In the second example, we examine RMSE versus the tuning parameters λ_1 and λ_2 with $\mu = 0.01$, SNR $= 0$ dB, and $T = 100$ snapshots. The result is drawn in Figure 4.3.

We observe from Figure 4.2 that the RMSE remains unchanged and stays minimal when μ lies within the interval $[10^{-13}, 10^{0}]$ for all 6 tested scenarios. Hence, we can choose any value for μ within this interval. Since the interval covers such a large range, the IRLS algorithm is insensitive to the smoothing parameter μ. Note that in Figure 4.3, our goal is to find a pair of (λ_1, λ_2) such that the RMSE is minimized. This demonstrates that there are many pairs of (λ_1, λ_2) meeting such a condition, such as $(\lambda_1, \lambda_2) = (2\,, 0.2)$, which is used for Algorithm 4.2 in the following simulations.

4.5.2 Convergence Speed

We compare the convergence speed of the IRLS with several existing methods, i.e., SVT, APG, and ADMM. Considering again a ULA of $M = 10$ sensors, 4 of which

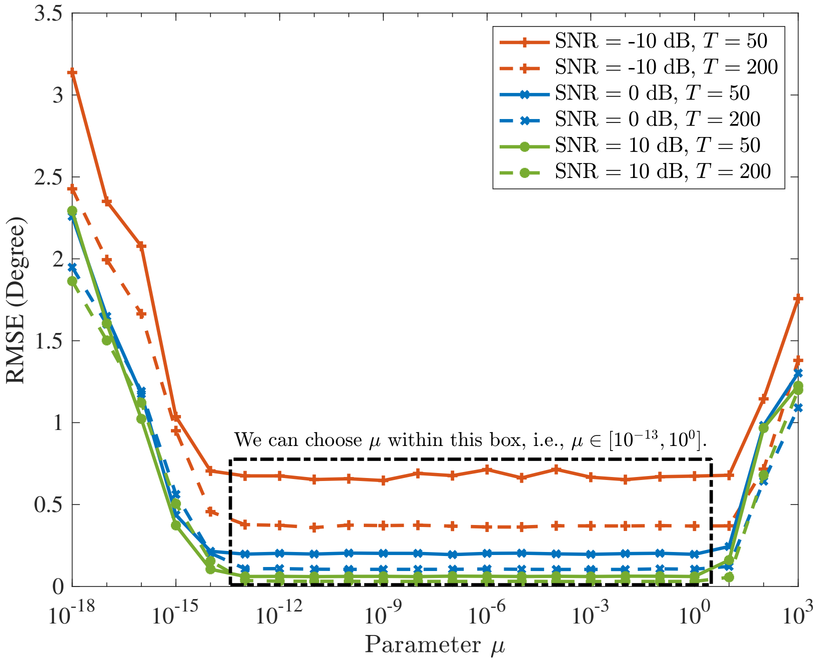


Figure 4.2: RMSE versus μ, with $M = 10$ sensors (4 of which fail), $K = 2$ sources, $\lambda_1 = 2$, and $\lambda_2 = 0.2$.

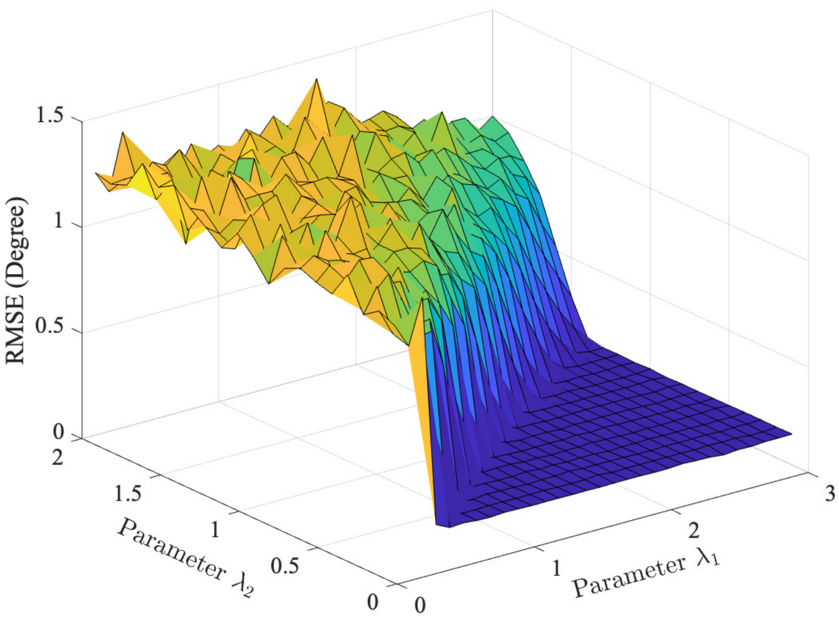


Figure 4.3: RMSE versus λ_1 and λ_2, with $M = 10$ sensors (4 of which fail), $K = 2$ sources, $T = 100$ snapshots, SNR = 0 dB, and $\mu = 0.01$.

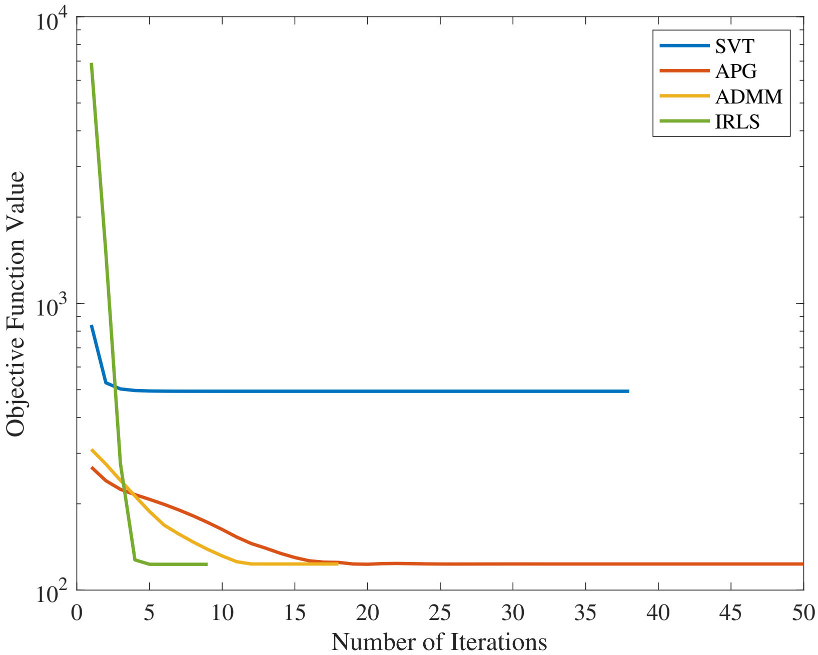


Figure 4.4: Objective function value versus number of iterations at SNR = 0 dB and $T = 100$ snapshots.

at random positions are distorted, receives $K = 2$ signals from $-10°$ and $10°$. The objective function values of the algorithms versus the number of iterations are depicted in Figure 4.4 with SNR = 0 dB and $T = 100$ snapshots. We see that the IRLS algorithm converges fastest in the sense that its objective function value decreases most rapidly, and it requires the least number of iterations to terminate, compared with the other three competitors.

The objective function value, CPU time and number of iterations are tabulated in Table 4.1 (upper) for SNR = 0 dB and $T = 100$ snapshots, and Table 4.1 (lower) for SNR = 0 dB and $T = 500$ snapshots. In both settings, the IRLS algorithm has the smallest objective function value, the least CPU time, and the least number of iterations, among all the examined algorithms.

4.5.3 Computational Complexity

We compare the computational complexity in this subsection. Note that the SVT, APG, and ADMM algorithms require one SVD of an $M \times T$ matrix per iteration, and the SVD consumes the most CPU time. As for the IRLS algorithm, the main calculation is

Table 4.1: Comparison of objective function value, CPU time and number of iterations in two different settings.

$T = 100$ snapshots, SNR = 0 dB

Algorithm	$f(\mathbf{Z}, \mathbf{V})$	Time (sec)	No. Iter.
SVT [CCS10]	493.8653	0.1205	**5**
APG [BT09]	123.3227	0.8403	22
ADMM [LCM10]	123.3227	0.1029	13
IRLS	**123.0227**	**0.0890**	**5**

$T = 500$ snapshots, SNR = 0 dB

Algorithm	$f(\mathbf{Z}, \mathbf{V})$	Time (sec)	No. Iter.
SVT [CCS10]	1965.5994	0.6166	25
APG [BT09]	282.5776	4.3472	32
ADMM [LCM10]	282.5776	5.8497	78
IRLS	**282.2776**	**0.1063**	**10**

Table 4.2: Computational complexity.

Algorithm	Complexity
SVT [CCS10]	$K_{\text{svt}}\mathcal{O}(TM^2)$
APG [BT09]	$K_{\text{apg}}\mathcal{O}(TM^2)$
ADMM [LCM10]	$K_{\text{admm}}\mathcal{O}(TM^2)$
IRLS	$K_{\text{irls}}\mathcal{O}(M^3)$

to find the inverse of an $M \times M$ matrix per iteration. Their main computational cost is summarized in Table 5.1, where K_{svt}, K_{apg}, K_{admm}, and K_{irls} denote the numbers of iterations for the SVT, APG, ADMM, and IRLS algorithms, respectively.

Figure 4.5 plots the averaged CPU time against the number of snapshots at $M = 10$ sensors (4 of which distorted), $K = 2$ sources, SNR = 0 dB, and $Q = 1000$ Monte Carlo runs. It is seen that the CPU times of the SVT, APG, and ADMM[1] algorithms are nearly linearly increasing with T. This is consistent with the theoretical analysis in Table 5.1. Figure 4.6 displays the CPU time versus the number of sensors with $T = 100$ snapshots and the other parameters are the same as those in Figure 4.5. We see that the curves of the CPU time of the SVT, APG, and ADMM algorithms are approximately

[1] Note that there is a jump of ADMM at $T = 250$. This is caused by the rapid increment of its number of iterations K_{admm}.

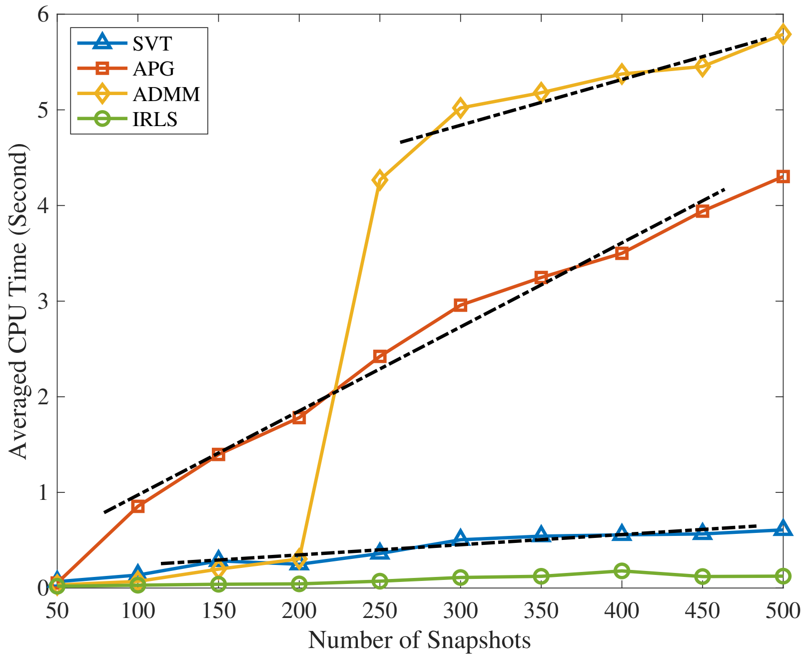


Figure 4.5: Computational complexity versus number of snapshots.

linearly correlated to M in a log scale, which again matches the theoretical calculations.

4.5.4 DOA Estimation Performance

We use the RMSE and resolution probability as DOA estimation performance measures. The resolution probability is calculated by N_{succ}/Q, where Q is the number of Monte Carlo runs, and N_{succ} denotes the number of trials where all the DOAs are successfully estimated. The trial is counted as a successful one if the following inequality is satisfied: $\max_{(k)}\{|\hat{\theta}_{(k)} - \theta_{(k)}|\} \leq 0.5°$.

In the first example, we consider a ULA of $M = 10$ sensors, 3 of which at random positions are distorted, $K = 2$ signals from $-10°$ and $10°$, $T = 100$ snapshots, and $Q = 5000$ Monte Carlo trials. The RMSE and resolution probability are depicted in Figures 4.7 and 4.8, respectively. The traditional Cramér-Rao bound (CRB) with known sensor errors [Del14] is plotted as a benchmark. Note that the curve labelled as "MUSIC-Known" denotes the MUSIC method with exact knowledge of the distorted sensors. It is seen that the SVT and MUSIC have bad performance even when the SNR becomes large. The APG, ADMM, and IRLS algorithms perform well when the SNR increases, their RMSEs decrease and their resolution probabilities increase up to

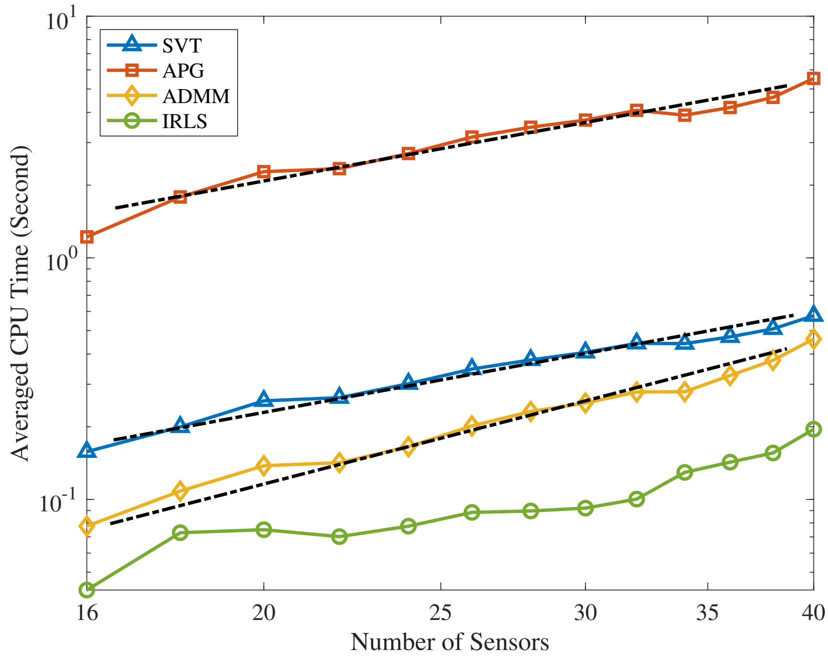


Figure 4.6: Computational complexity versus number of sensors.

1. The IRLS algorithm outperforms the other two state-of-the-art methods, i.e., APG and ADMM.

In the next example, we examine the DOA estimation performance for different numbers of snapshots. The SNR is set to be 0 dB, and the remaining parameters are the same as those of the former example. The RMSE and resolution probability of the methods are plotted in Figures 4.9 and 4.10, respectively. The results demonstrate a better performance of the IRLS algorithm compared with the SVT, APG, and ADMM methods.

In the last example of this subsection, we evaluate the DOA estimation performance in view of the source separation angle. The settings of SNR = 0 dB, $K = 2$ sources, and $T = 100$ snapshots are employed. The first signal is from $0°$, while the DOA of the second signal changes from $1°$ to $20°$ with a stepsize of $1°$. The other parameters are unchanged as those in the first example of this subsection. The RMSE and resolution probability versus angular separation are displayed in Figures 4.11 and 4.12, respectively. These again indicate that the IRLS algorithm outperforms the SVT, APG, and ADMM algorithms in terms of RMSE and resolution probability.

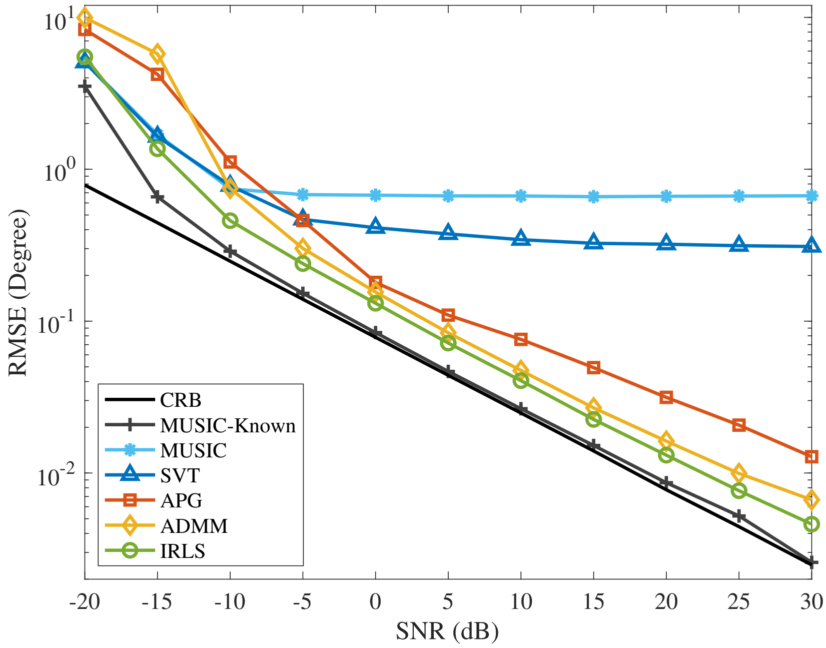


Figure 4.7: RMSE versus SNR.

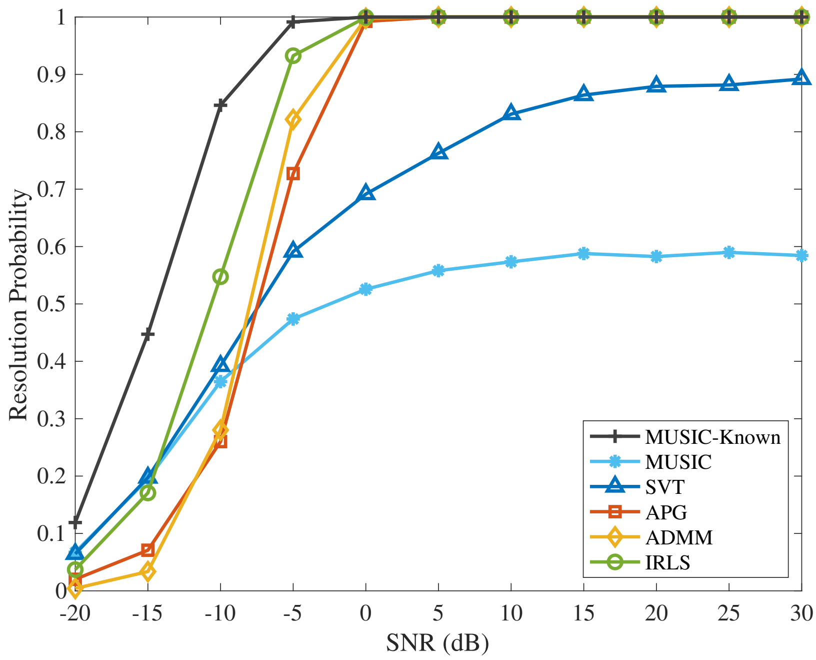


Figure 4.8: Resolution probability versus SNR.

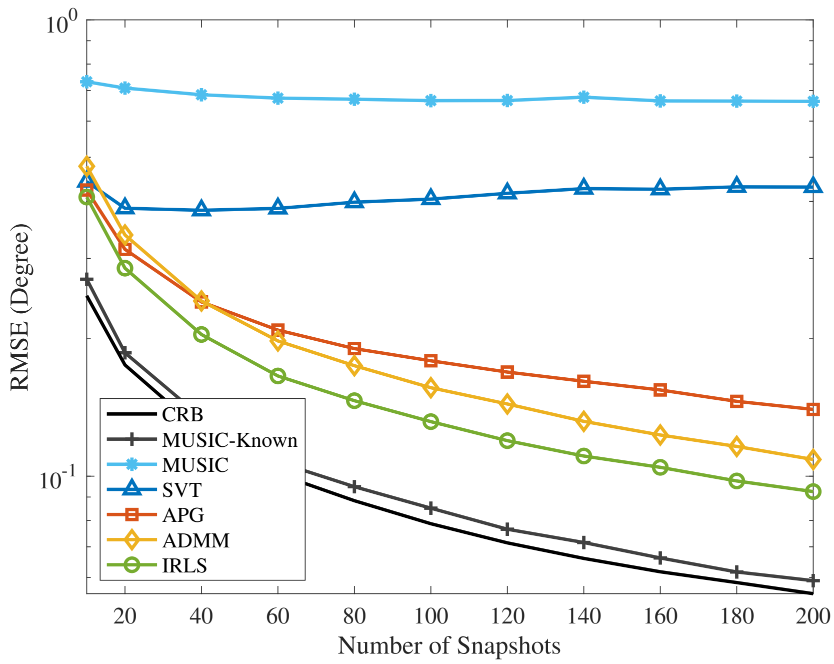


Figure 4.9: RMSE versus number of snapshots.

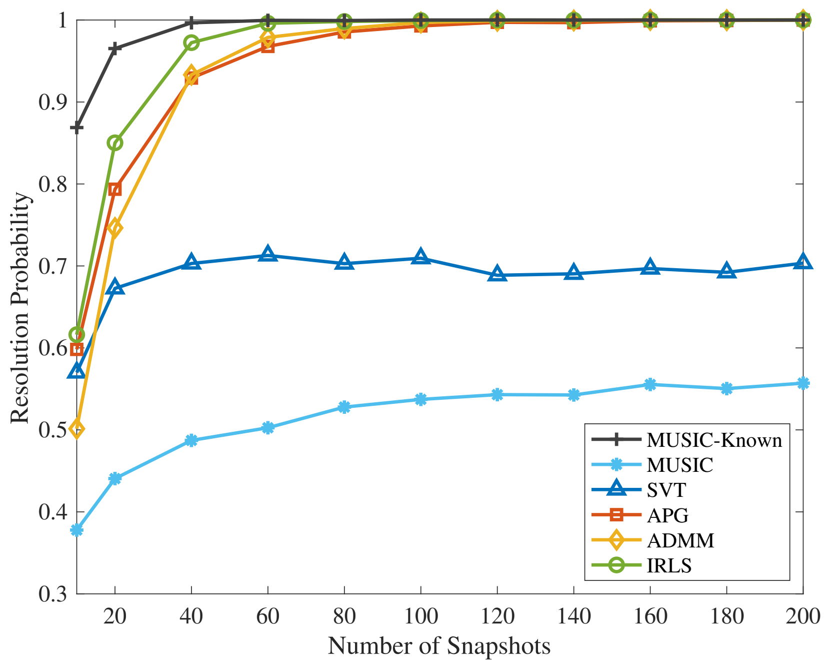


Figure 4.10: Resolution probability versus number of snapshots.

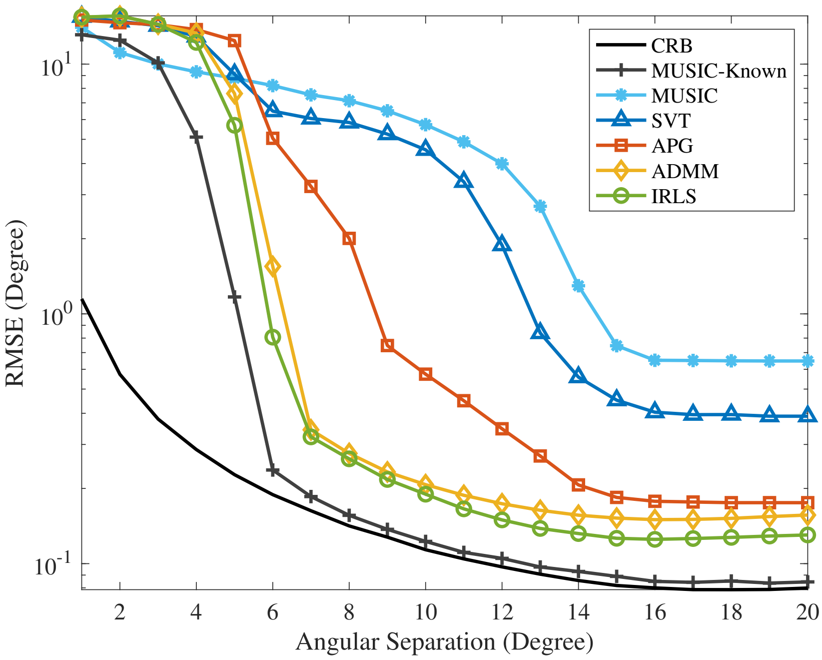


Figure 4.11: RMSE versus source separation angle.

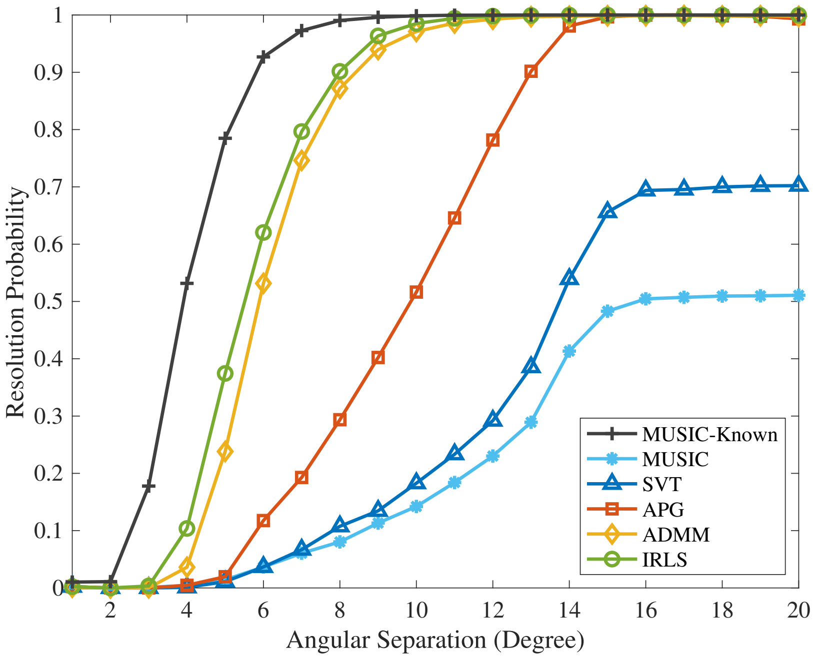


Figure 4.12: Resolution probability versus source separation angle.

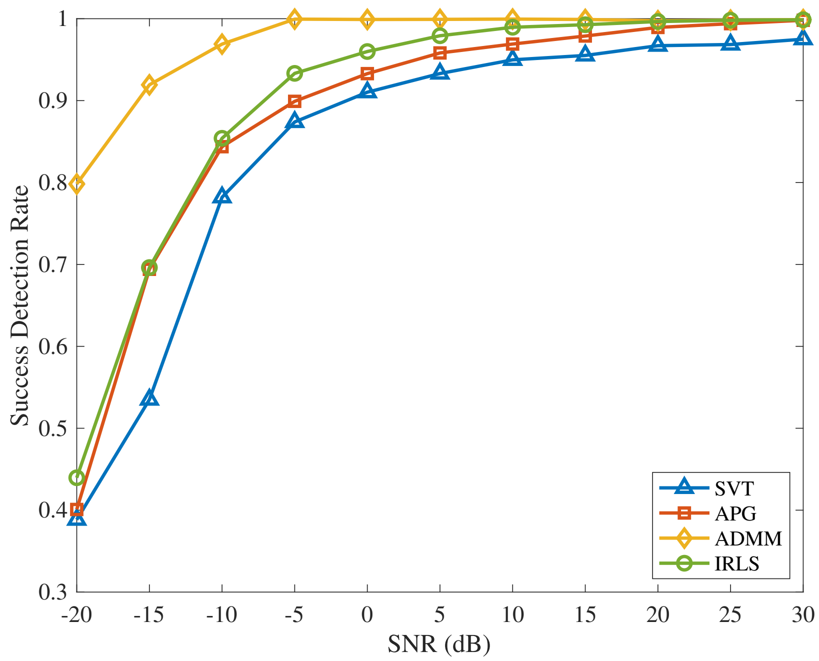


Figure 4.13: Success detection rate versus SNR.

4.5.5 Distorted Sensor Detection Performance

Parallel to the three examples in Section 4.5.4, we now examine the performance of the detection of distorted sensors of the SVT, APG, ADMM, and IRLS algorithms. The threshold in Algorithm 4.3 is set as $h = 10d$. We utilize the success detection rate as a metric, which is defined as N_{detec}/Q. N_{detec} is the number of trials where the number of distorted sensors is correctly estimated, and meanwhile their positions are exactly found. The results are given in Figures 4.13, 4.14, and 4.15, which show that the ADMM is the best amongst all tested methods in terms of identifying the distorted sensors, followed by the IRLS algorithm.

4.6 Summary

We studied the problem of simultaneously estimating direction-of-arrival (DOA) of signals and detecting distorted sensors. It is assumed that the distorted sensors occur randomly, and the number of distorted sensors is much smaller than the total number of sensors. The problem was formulated via low-rank and row-sparse decomposition, and solved by iteratively reweighted least squares (IRLS). Both noiseless and noisy

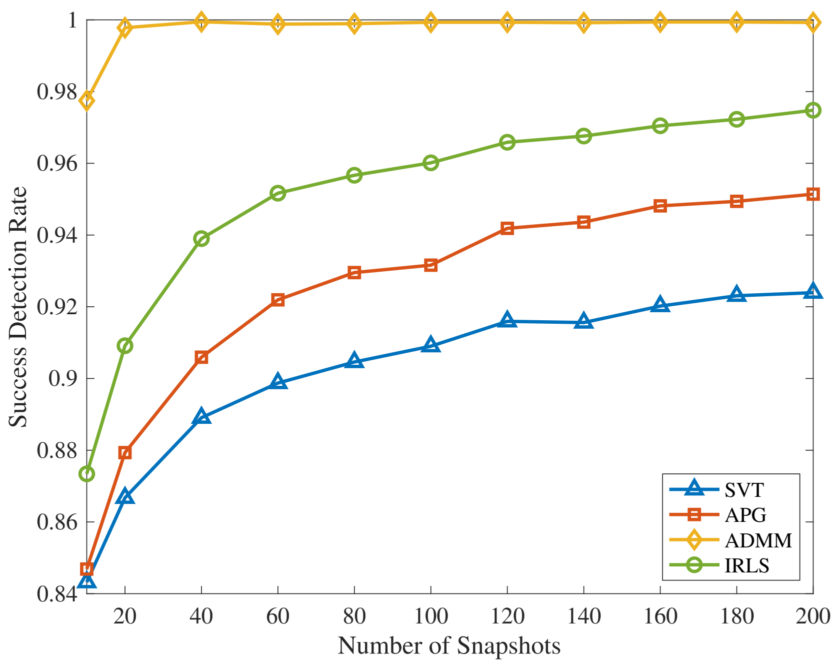


Figure 4.14: Success detection rate versus number of snapshots.

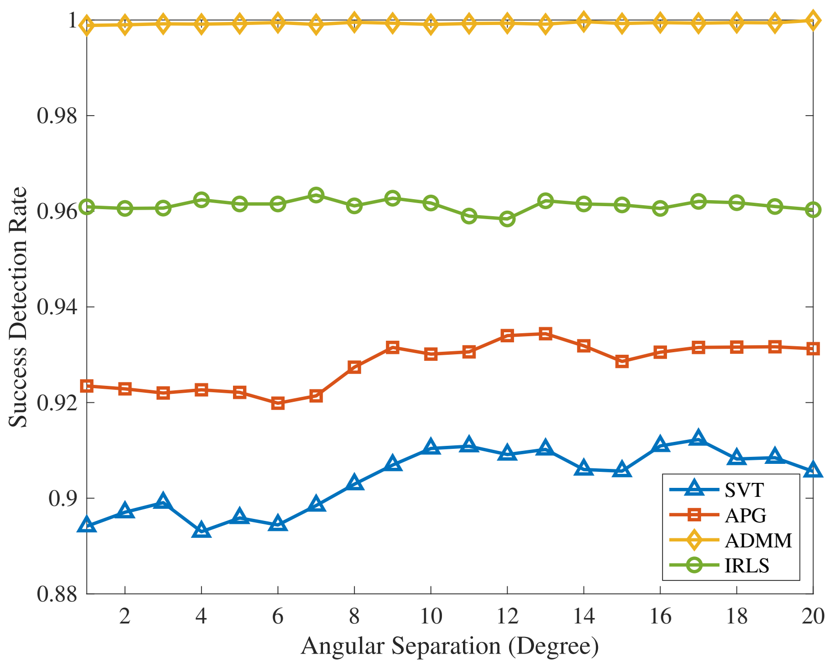


Figure 4.15: Success detection rate versus source separation angle.

cases were considered. Theoretical analyses of algorithm convergence were provided. Computational cost of the IRLS algorithm was compared with that of several existing methods. Simulation results were conducted for parameter selection, convergence speed, computational time, and performance of DOA estimation as well as distorted sensor detection. The IRLS method was demonstrated to have higher DOA estimate accuracy and lower computational cost than other methods, and the alternating direction method of multipliers was shown to be slightly better than the IRLS algorithm in distorted sensor detection.

Chapter 5

Off-grid DOA Estimation Using a Sparse Representation

In this chapter, the problem of off-grid direction-of-arrival (DOA) estimation is investigated. We develop a grid-based method to jointly estimate the closest *spatial frequency* (the sine of DOA) grids, and the gaps between the estimated grids and the corresponding frequencies. By using a second-order Taylor approximation, the data model under the framework of joint-sparse representation is formulated. We point out an important property of the signals of interest in the model, namely the proportionality relationship, which is empirically demonstrated to be useful in the sense that it increases the probability of the mixing matrix satisfying the block restricted isometry property. Simulation examples demonstrate the effectiveness and superiority of the proposed method against several state-of-the-art grid-based approaches.

The key contributions presented in this chapter originate from [HSZ22a]. The structure of this chapter is as follows: Motivation is presented in Section 5.1. Section 5.2 shows the signal model. The proposed method and uniqueness property of the proposed solution are detailed in Sections 5.3 and 5.4, respectively. Simulation results are given in Section 5.5, while Section 5.6 summarizes this chapter.

5.1 Motivation

Grid-based methods have gained interest in direction-of-arrival (DOA) estimation in recent years. Such approaches include least absolute shrinkage and selection operator (LASSO) [Tib96, MCW05, WZ14] and sparse iterative covariance-based estimation [SBL11a, SBL11b, SB12, BS14], among others. See [YLSX18] for a comprehensive review of grid-based sparse methods for DOA estimation. The advantage of grid-based methods is that they have super-high resolution even in the case when only one single snapshot is available, provided that all the source spatial frequencies align exactly with the preset grid. However, this condition may not be satisfied in practice, since the region of interest (ROI) contains infinite candidates and hence grid mismatch almost always exists when we split the ROI into a finite number of grids. This is known as the off-grid issue and has attracted a lot of research interest in array signal processing during the past decade, see for example [ZLG11, YZX12, YXZ13, DB13, JH13, FGG+14, TYN14, DBXC17, WSP17, LSG17, AGM18, WZCN18, WZYZ18, ZGSZ18, WGW19, WL21, MCW+21].

Existing solutions to tackle the off-grid problem can be categorized into three groups. The first group uses denser grids or the coarse-to-fine strategy such as [MCW05]. The drawbacks of these methods are twofold. On one hand, denser grids lead to extremely expensive computational complexity; on the other hand, too dense grids may result in weak incoherence among the steering vectors. The second group consists of the so-called *gridless* approach [TBSR13, YX15, SRS+16, SSSP17, WGP19, ZRH21]. Its weakness is that most of these methods are restricted to regularly sampled measurements that can only be taken from a uniform linear array (ULA) [WPG21]. The last group of methods estimates the off-grid bias together with the grids closest to the true spatial frequencies. Representative works include the first-order Taylor approximation [JH13, TYN14] and the neighbor-grid based method [AGM18], denoted in this chapter as 1st Taylor G-LASSO and Neighbor G-LASSO, respectively.

It is known that in general the first-order Taylor approximation is accurate enough, especially when the grid size is small. However, when the grid size is set not small enough so as to save computational cost, there still exists a large bias. In such a situation, a high-order Taylor approximation decreases the approximation error. To this end, we introduce a second-order Taylor approximation in off-grid DOA estimation. We observe in this case the proportionality relationship of the signals of interest. With this, we propose a novel optimization approach which is shown by simulation to produce more accurate frequency estimates in off-grid scenarios. Moreover, the uniqueness issue of the proposed method is discussed by means of the restricted isometry property (RIP), which is one of the most important tools in compressive sensing [EM09].

Notations: In this chapter, bold-faced lower-case and upper-case letters stand for vectors and matrices, respectively. Superscripts $\cdot^{\mathrm{T}}$, $\cdot^{\mathrm{H}}$, and $\cdot^{*}$ denote transpose, Hermitian transpose, and complex conjugate operators, respectively. $\mathrm{vec}\{\cdot\}$ denotes the vectorization operator, $\mathrm{diag}\{\cdot\}$ returns a diagonal matrix whose main diagonal is given in the curly bracket, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are real and imaginary parts of a complex-valued variable, respectively. $\diamond$ symbolizes the Khatri-Rao product. $\mathbb{C}$ and $\mathbb{R}$ are the sets of complex and real numbers, respectively. $\mathbf{I}$ is the identity matrix of appropriate dimension. $\mathbf{0}$ and $\mathbf{1}$ denote the all-zeros and the all-ones vectors of appropriate length, respectively. For a vector $\mathbf{x}$, $|\mathbf{x}|$ and $\|\mathbf{x}\|_2$ represent the element-wise absolute value and the L_2 norm of $\mathbf{x}$, respectively. The symbols $\geq$, $\leq$, $>$, and $<$ are element-wise greater than or equal to, less than or equal to, greater than, and less than operators, respectively.

5.2 Signal Model

Suppose that a linear array of M sensors whose positions are contained in $\mathbf{q} = [q_1, q_2, \cdots, q_M]^{\mathrm{T}}$, receives K far-field narrowband signals from directions $\boldsymbol{\phi} = [\phi_1, \phi_2, \cdots, \phi_K]^{\mathrm{T}}$ with $\phi_k \in [-\pi/2, \pi/2)$. For simplicity, we define the spatial frequencies as $\boldsymbol{u} = [u_1, u_2, \cdots, u_K]^{\mathrm{T}}$ with $u_k = \sin(\phi_k) \in [-1, 1)$. The array observation can be modeled as

$$\mathbf{y} = \sum_{k=1}^{K} s_k \mathbf{a}(u_k) + \mathbf{n} = \mathbf{A}(\boldsymbol{u})\mathbf{s} + \mathbf{n},$$

where s_k is the k-th signal waveform, $\mathbf{s} = [s_1, s_2, \cdots, s_K]^{\mathrm{T}}$ represents the signal vector, and $\mathbf{n} \in \mathbb{C}^M$ is the noise vector. The steering matrix $\mathbf{A}(\boldsymbol{u}) = [\mathbf{a}(u_1), \mathbf{a}(u_2), \cdots, \mathbf{a}(u_K)] \in \mathbb{C}^{M \times K}$ has the steering vectors as columns, where $\mathbf{a}(u_k) = [e^{\jmath \frac{2\pi q_1}{\lambda} u_k}, e^{\jmath \frac{2\pi q_2}{\lambda} u_k}, \cdots, e^{\jmath \frac{2\pi q_M}{\lambda} u_k}]^{\mathrm{T}}$, for $k = 1, 2, \cdots, K$, with λ being the signal wavelength and $\jmath = \sqrt{-1}$.

In grid-based methods, we formulate the signal model by means of a sparse representation, as

$$\mathbf{y} = \sum_{l=1}^{L} x_l \mathbf{a}(v_l) + \mathbf{n} = \mathbf{A}(\boldsymbol{v})\mathbf{x} + \mathbf{n},$$

where $\boldsymbol{v} = [v_1, v_2, \cdots, v_L]^{\mathrm{T}}$ denotes the frequency grid vector with L being the number of grids (in general $L \gg M > K$), $\mathbf{A}(\boldsymbol{v}) \in \mathbb{C}^{M \times L}$ stands for the overcomplete dictionary matrix, and $\mathbf{x} = [x_1, x_2, \cdots, x_L]^{\mathrm{T}}$ is a sparse vector whose elements $x_l = s_k$ if $v_l = u_k$, and $x_l = 0$ otherwise. When the true frequencies do not exactly lie in the preset grids, we encounter the off-grid issue. To handle this problem, we propose a method to simultaneously estimate the closest frequency grids, and the gaps between the closest grids and the true frequencies, using a second-order Taylor approximation.

5.3 Proposed Method

5.3.1 Second-order Taylor Approximation

We start by considering a second-order Taylor approximation of the steering vectors. For any u_l, we have

$$\mathbf{a}(u_l) \approx \mathbf{a}(v_l) + \mathbf{a}'(v_l)p_l + \frac{\mathbf{a}''(v_l)}{2}p_l^2,$$

where v_l is the grid closest to u_l, $\mathbf{a}'(v_l) = \frac{\mathrm{d}\mathbf{a}(v)}{\mathrm{d}v}\Big|_{v=v_l}$, $\mathbf{a}''(v_l) = \frac{\mathrm{d}^2\mathbf{a}(v)}{\mathrm{d}v^2}\Big|_{v=v_l}$, and $p_l = u_l - v_l \in [-\delta/2, \delta/2]$ with δ being the grid size. Collecting all the candidates, we have

$$[\mathbf{a}(u_1),\cdots,\mathbf{a}(u_L)] \approx \mathbf{A}(\boldsymbol{v})+\mathbf{A}'(\boldsymbol{v})\mathrm{diag}\{\mathbf{p}\}+\frac{1}{2}\mathbf{A}''(\boldsymbol{v})\mathrm{diag}\{\mathbf{p}\}^2,$$

where $\mathbf{A}'(\boldsymbol{v}) = [\mathbf{a}'(v_1), \cdots, \mathbf{a}'(v_L)] \in \mathbb{C}^{M\times L}$, $\mathbf{A}''(\boldsymbol{v}) = [\mathbf{a}''(v_1), \cdots, \mathbf{a}''(v_L)] \in \mathbb{C}^{M\times L}$, and $\mathbf{p} = [p_1, p_2, \cdots, p_L]^{\mathrm{T}}$. Hence, the signal model can be approximately written as:

$$\begin{aligned}\mathbf{y} &\approx \left[\mathbf{A}(\boldsymbol{v}) + \mathbf{A}'(\boldsymbol{v})\mathrm{diag}\{\mathbf{p}\} + \frac{1}{2}\mathbf{A}''(\boldsymbol{v})\mathrm{diag}\{\mathbf{p}\}^2\right]\mathbf{x} + \mathbf{n} \\ &= \left[\mathbf{A}(\boldsymbol{v}), \mathbf{A}'(\boldsymbol{v}), \frac{1}{2}\mathbf{A}''(\boldsymbol{v})\right]\begin{bmatrix}\mathbf{x} \\ \mathrm{diag}\{\mathbf{p}\}\mathbf{x} \\ \mathrm{diag}\{\mathbf{p}\}^2\mathbf{x}\end{bmatrix} + \mathbf{n},\end{aligned} \tag{5.1}$$

where the signals of interest $[\mathbf{x}^{\mathrm{T}}, (\mathrm{diag}\{\mathbf{p}\}\mathbf{x})^{\mathrm{T}}, (\mathrm{diag}\{\mathbf{p}\}^2\mathbf{x})^{\mathrm{T}}]^{\mathrm{T}}$ are referred to as block signal in the sequel.

5.3.2 Properties of the Block Signal

As shown in signal model (5.1), the unknown block signal is divided into three parts: (i) $\mathbf{x}_1 \triangleq \mathbf{x}$, (ii) $\mathbf{x}_2 \triangleq \mathrm{diag}\{\mathbf{p}\}\mathbf{x}$, and (iii) $\mathbf{x}_3 \triangleq \mathrm{diag}\{\mathbf{p}\}^2\mathbf{x}$. Without loss of generality, we assume $\mathbf{x}$ is a real-valued vector, see the remark below. Denote the l-th entries of $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as $x_{1,l}$, $x_{2,l}$, and $x_{3,l}$, respectively. We notice the following properties of the block signal $[\mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}, \mathbf{x}_3^{\mathrm{T}}]^{\mathrm{T}}$.

- Since $\mathbf{x}$ is a sparse vector as mentioned in Section 5.2, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are all sparse and share the same sparsity pattern. This property is known as block-sparsity [JH13] or joint-sparsity [TYN14].
- It holds that $x_{2,l} = p_l x_{1,l}$ and $x_{3,l} = p_l^2 x_{1,l}$. Due to $-\delta/2 \le p_l \le \delta/2$, $\forall l \in \{1, 2, \cdots, L\}$, it is easy to verify that the following inequalities hold:
$$-\frac{\delta}{2}|\mathbf{x}_1| \le \mathbf{x}_2 \le \frac{\delta}{2}|\mathbf{x}_1|, \quad -\left(\frac{\delta}{2}\right)^2|\mathbf{x}_1| \le \mathbf{x}_3 \le \left(\frac{\delta}{2}\right)^2|\mathbf{x}_1|. \tag{5.2}$$
- It can be seen that $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ satisfy the proportionality relationship, as
$$x_{2,l}^2 = x_{1,l}x_{3,l}, \ \forall l \in \{1, 2, \cdots, L\}. \tag{5.3}$$

Remark: For any complex-valued data model, say $\mathbf{y} = \mathbf{A}\mathbf{x}$, we have its real-valued counterpart as $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}$, where $\tilde{\mathbf{y}} = [\Re\{\mathbf{y}\}^{\mathrm{T}}, \Im\{\mathbf{y}\}^{\mathrm{T}}]^{\mathrm{T}}$, $\tilde{\mathbf{x}} = [\Re\{\mathbf{x}\}^{\mathrm{T}}, \Im\{\mathbf{x}\}^{\mathrm{T}}]^{\mathrm{T}}$, and $\tilde{\mathbf{A}} = \begin{bmatrix}\Re\{\mathbf{A}\} & -\Im\{\mathbf{A}\} \\ \Im\{\mathbf{A}\} & \Re\{\mathbf{A}\}\end{bmatrix}$.

5.3.3 Problem Formulation Development

Based on the aforementioned relationships among $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, we propose the following minimization problem:

$$\min_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \quad g(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3) \qquad \text{s.t.} \;\; (5.2) \text{ and } (5.3). \tag{5.4}$$

The cost function in (5.4) is given by

$$g(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3) \triangleq \frac{1}{2}\left\|\mathbf{y}-\mathbf{A}(\boldsymbol{v})\mathbf{x}_1-\mathbf{A}'(\boldsymbol{v})\mathbf{x}_2-\frac{1}{2}\mathbf{A}''(\boldsymbol{v})\mathbf{x}_3\right\|_2^2+\mu\left\|\left[\mathbf{x}_1^{\mathrm{T}},\mathbf{x}_2^{\mathrm{T}},\mathbf{x}_3^{\mathrm{T}}\right]^{\mathrm{T}}\right\|_{2,1}, \tag{5.5}$$

where μ is a regularization parameter balancing the data fitting and the model sparsity, and $\|\cdot\|_{2,1}$ is the mixed $L_{2,1}$ norm of a vector, defined as

$$\left\|\left[\mathbf{x}_1^{\mathrm{T}},\mathbf{x}_2^{\mathrm{T}},\mathbf{x}_3^{\mathrm{T}}\right]^{\mathrm{T}}\right\|_{2,1} = \sum_{l=1}^{L}\sqrt{|x_{1,l}|^2+|x_{2,l}|^2+|x_{3,l}|^2}.$$

Problem (5.4) is non-convex and hard to solve due to its constraints. We first consider the constraints of (5.2). The difficulty of dealing with (5.2) comes from the absolute value operator [TYN14]. However, when the signals are assumed to be real positive, i.e., $\mathbf{s} > \mathbf{0}$ (and $\mathbf{x}_1 = \mathbf{x} \geq \mathbf{0}$), the constraints of (5.2) in (5.4) become

$$-\frac{\delta}{2}\mathbf{x}_1 \leq \mathbf{x}_2 \leq \frac{\delta}{2}\mathbf{x}_1, \quad \mathbf{0} \leq \mathbf{x}_3 \leq \left(\frac{\delta}{2}\right)^2\mathbf{x}_1, \quad \mathbf{x}_1 \geq \mathbf{0}, \tag{5.6}$$

which are linear and thus convex. It is worth pointing out that $\mathbf{0} \leq \mathbf{x}_3$ in (5.6) is the result of $x_{3,l} = p_l^2 x_{1,l},\ \forall l \in \{1,2,\cdots,L\}$ and $\mathbf{x}_1 \geq \mathbf{0}$. Note that the assumption of real positive signals is valid in various situations. For instance, in multiple-snapshot scenarios, the signal vector denotes the signal powers which are naturally positive, see the remark below.

In the sequel, we consider the last constraint in (5.4), viz. (5.3). Firstly, we convert (5.3) to its equivalent form as in [PB17]:

$$\left\|\begin{bmatrix} 2x_{2,l} \\ x_{1,l}-x_{3,l}\end{bmatrix}\right\|_2 = x_{1,l}+x_{3,l},\ \forall l \in \{1,2,\cdots,L\}. \tag{5.7}$$

Then, we introduce an additional variable $\mathbf{z} \in \mathbb{R}^L$ with entries z_l satisfying

$$0 \leq z_l \leq \eta,\ \forall l \in \{1,2,\cdots,L\}, \tag{5.8}$$

where η is a small user-defined parameter, and rewrite (5.7) as

$$\left\| \begin{bmatrix} 2x_{2,l} \\ x_{1,l} - x_{3,l} \end{bmatrix} \right\|_2 \leq x_{1,l} + x_{3,l} + z_l, \ \forall l \in \{1, 2, \cdots, L\}, \tag{5.9}$$

which belongs to the set of standard second-order cone and hence is convex.

By replacing the constraint (5.2) with (5.6) and replacing (5.3) with (5.8) and (5.9), we finally relax the non-convex problem (5.4) into a convex one, as

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{z}} \ g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \qquad \text{s.t.} \ \ (5.6), \ (5.8), \ \text{and} \ (5.9). \tag{5.10}$$

Remark: Note that the proposed method is developed for the single-snapshot scenario. However, it can be easily extended to the case of multiple snapshots. To be precise, when multiple snapshots are available, we have the covariance matrix $\mathbf{R} = \mathbf{A}\mathbf{R}_s\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}$, where σ^2 is the noise power. Note that we assume the signals to be uncorrelated with the noise, and the noise components are independent and identically distributed. Vectoring $\mathbf{R}$ yields

$$\mathrm{vec}\{\mathbf{R}\} = (\mathbf{A}^* \diamond \mathbf{A})\mathbf{r}_s + \sigma^2 \mathrm{vec}\{\mathbf{I}\}, \tag{5.11}$$

where $\mathbf{r}_s$ is the main diagonal of $\mathbf{R}_s$, denoting the signal powers. The data model (5.11) is similar to the signal model introduced in Section 5.2, and therefore, we can develop our method on the basis of (5.11).

To analyze the computational cost, we formulate Problem (5.10) under the framework of standard second-order cone programming (SOCP) [LVBL98], as

$$\begin{aligned} \min_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{z}, \mathbf{t}} \ & \sum_{l=1}^{L} t_l = \mathbf{1}^{\mathrm{T}}\mathbf{t} \\ \text{s.t.} \quad & (5.6), \ (5.8), \ \text{and} \ (5.9), \\ & \sqrt{|x_{1,l}|^2 + |x_{2,l}|^2 + |x_{3,l}|^2} \leq t_l, \ \forall l \in \{1, 2, \cdots, L\}, \\ & \left\| \mathbf{y} - \mathbf{A}(\boldsymbol{v})\mathbf{x}_1 - \mathbf{A}'(\boldsymbol{v})\mathbf{x}_2 - \frac{1}{2}\mathbf{A}''(\boldsymbol{v})\mathbf{x}_3 \right\|_2 \leq \epsilon, \end{aligned}$$

where $\mathbf{t} = [t_1, t_2, \cdots, t_L]^{\mathrm{T}}$ is an auxiliary variable vector, and ϵ is a tuning parameter related to μ in (5.10). The computational cost of the above problem with implementation of SOCP is $\mathcal{O}(9(M+1)L^2 + 72L)$ per iteration, and the number of iterations is bounded above by $\mathcal{O}(\sqrt{L})$ [LVBL98]. The proposed second-order Taylor approximation method is referred to as 2nd Taylor G-LASSO. The computational complexity of the 2nd Taylor G-LASSO, as well as those of LASSO [Tib96, MCW05], Neighbor G-LASSO [AGM18], and 1st Taylor G-LASSO [JH13, TYN14], are summarized in Table 5.1.

Table 5.1: Computational cost using SOCP implementation.

Method	Cost per Iteration	No. of Iterations
LASSO	$\mathcal{O}((M+1)L^2)$	$\mathcal{O}(1)$
Neighbor G-LASSO	$\mathcal{O}(4(M+1)L^2+12L)$	$\mathcal{O}(\sqrt{L})$
1st Taylor G-LASSO	$\mathcal{O}(4(M+1)L^2+28L)$	$\mathcal{O}(\sqrt{L})$
2nd Taylor G-LASSO	$\mathcal{O}(9(M+1)L^2+72L)$	$\mathcal{O}(\sqrt{L})$

5.4 Uniqueness Property of the Proposed Solution

Note that, for underdetermined linear systems, uniqueness of a sparse solution is one of the fundamental problems in compressive sensing [FR13b]. In this section, we discuss this issue in view of the proposed signal model in (5.1). To this end, we first introduce the following definition and theorem [EM09]:

Definition 5.1 *An $M \times bL$ block matrix $\mathbf{D}$ is said to have the block RIP with parameter β_K, if for every K block-sparse vector $\mathbf{c}$ of length bL, it holds that*

$$(1-\beta_K)\|\mathbf{c}\|_2^2 \leq \|\mathbf{D}\mathbf{c}\|_2^2 \leq (1+\beta_K)\|\mathbf{c}\|_2^2.$$

Theorem 5.1 *Let $\mathbf{y} = \mathbf{D}\mathbf{c}_0$ be measurements of a K block-sparse vector $\mathbf{c}_0$. If $\mathbf{D}$ satisfies the block RIP with parameter $\beta_{2K} < 1$, then there exists a unique block-sparse vector $\mathbf{c}$ satisfying $\mathbf{y} = \mathbf{D}\mathbf{c}$; and further, if $\mathbf{D}$ satisfies the block RIP with $\beta_{2K} < \sqrt{2}-1$, then the convex optimization problem: $\min_{\mathbf{c}} \|\mathbf{c}\|_{2,1}$ s.t. $\mathbf{y} = \mathbf{D}\mathbf{c}$, has a unique solution and the solution is equal to $\mathbf{c}_0$.*

Define $\mathbf{c}_0 = [\mathbf{x}^{\mathrm{T}}, (\mathrm{diag}\{\mathbf{p}\}\mathbf{x})^{\mathrm{T}}, (\mathrm{diag}\{\mathbf{p}\}^2\mathbf{x})^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{D} = [\mathbf{A}(\boldsymbol{v}), \mathbf{A}'(\boldsymbol{v}), \frac{1}{2}\mathbf{A}''(\boldsymbol{v})]$. In the absence of noise, our proposed model in (5.1) can be rewritten as: $\mathbf{y} = \mathbf{D}\mathbf{c}_0$. Without loss of generality, we denote $\bar{\mathbf{D}}$ as the column-normalized matrix structured from $\mathbf{D}$. Our task is to check whether or not $\bar{\mathbf{D}}$ satisfies the block RIP with parameter $\beta_{2K} < 1$ and $\beta_{2K} < \sqrt{2}-1$. Note that determining the RIP parameter, i.e., β_{2K}, of a given matrix is in general an NP-hard problem [EKB10, TP14]. In what follows, we introduce a Monte Carlo test to check the condition of the block RIP of $\bar{\mathbf{D}}$.

According to the definition, if $\bar{\mathbf{D}}$ has the block RIP with parameter β_{2K}, then for any $2K$ block-sparse vector $\mathbf{c}$ of length bL, it holds that

$$(1-\beta_{2K})\|\mathbf{c}\|_2^2 \leq \|\bar{\mathbf{D}}\mathbf{c}\|_2^2 \leq (1+\beta_{2K})\|\mathbf{c}\|_2^2. \tag{5.12}$$

Note that, for any $2K$ block-sparse vector $\mathbf{c}$, we can write its unit-norm vector as $\bar{\mathbf{c}} = \mathbf{c}/\|\mathbf{c}\|_2$, such that $\|\bar{\mathbf{c}}\|_2 = 1$. As a result, (5.12) becomes:

$$(1-\beta_{2K}) \leq \frac{\|\bar{\mathbf{D}}\mathbf{c}\|_2^2}{\|\mathbf{c}\|_2^2} = \|\bar{\mathbf{D}}\bar{\mathbf{c}}\|_2^2 \leq (1+\beta_{2K}).$$

Based on the above inequalities, the parameter β_{2K} is calculated as

$$\beta_{2K} = \max\big\{\|\bar{\mathbf{D}}\bar{\mathbf{c}}\|_2^2 - 1\,, 1 - \|\bar{\mathbf{D}}\bar{\mathbf{c}}\|_2^2\big\}. \tag{5.13}$$

We randomly generate a unit-norm $2K$ block-sparse vector $\bar{\mathbf{c}}$, and calculate β_{2K} using (5.13). By repeatedly performing the above steps for 10^4 Monte Carlo runs, we estimate the empirical probabilities of $\{\beta_{2K} < 1\}$ and $\{\beta_{2K} < \sqrt{2}-1\}$. The empirical probabilities versus block-sparsity $2K$ are presented in Figure 5.1, with $M = 8$, $q_m = \frac{(m-1)\lambda}{2}$ $(m = 1, 2, \cdots, M)$, $L = 200$, and $b = 1$ for LASSO, $b = 2$ for Neighbor G-LASSO and 1st Taylor G-LASSO, and $b = 3$ for 2nd Taylor G-LASSO. It is seen that when the block-sparsity is small (less than 8), the probabilities of $\{\beta_{2K} < 1\}$ of all the tested methods are high (greater than 0.9), and their probabilities of $\{\beta_{2K} < \sqrt{2}-1\}$ are larger than 0.5. Note that in Figure 5.1, the plot of 2nd Taylor G-LASSO with proportional signals (abbreviated as "Prop. Sig." in the figure), i.e., (5.3), has the highest probability. This reveals that the proportionality relationship of the block signal contains useful information in the sense that it increases the probabilities of $\{\beta_{2K} < 1\}$ and $\{\beta_{2K} < \sqrt{2}-1\}$.

5.5 Simulation Results

We evaluate the frequency estimation performance of 2nd Taylor G-LASSO, compared with LASSO [Tib96, MCW05, WZ14], Neighbor G-LASSO [AGM18], and 1st Taylor G-LASSO [JH13, TYN14]. We adopt the root-mean squared error (RMSE) and the empirical probability of correct detection (PCD) as performance metrics, defined as in [SP18]:

$$\text{RMSE} = 10\log_{10}\left(\sqrt{\frac{1}{KQ}\sum_{k=1}^{K}\sum_{q=1}^{Q}(\widehat{u}_{k,q} - u_k)^2}\right)$$

and $\text{PCD} = Q_{\text{suc}}/Q$, respectively, where $\widehat{u}_{k,q}$ denotes the frequency estimates of the k-th signal in the q-th Monte Carlo run, Q is the total number of Monte Carlo trials, and Q_{suc} is the number of trials where the frequency estimates $\{\widehat{u}_k | k = 1, 2, \cdots, K\}$ fulfill: $\max_k\{|\widehat{u}_k - u_k|\} \leq \delta/2$. The Cramér–Rao bound (CRB) [MCW+21] is drawn as a benchmark for RMSE comparison. The results of single-snapshot situation are shown in Section 5.5.1, while the results of multiple-snapshot situation are drawn in Section 5.5.2.

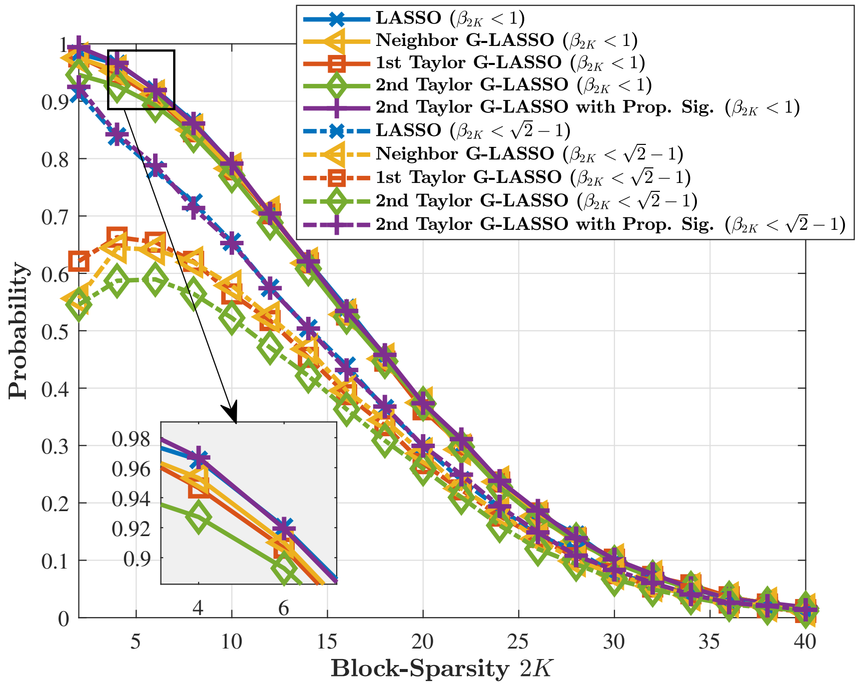


Figure 5.1: Empirical probabilities of $\{\beta_{2K} < 1\}$ and $\{\beta_{2K} < \sqrt{2} - 1\}$ versus block-sparsity $2K$ with 10^4 Monte Carlo runs, $M = 8$, and $L = 200$.

5.5.1 Single-snapshot Situation

In the first experiment, a linear array of $M = 16$ omnidirectional sensors is considered to receive $K = 2$ signals with spatial frequencies $\boldsymbol{u} = [0.1815, 0.7942]^{\mathrm{T}}$. The $M = 16$ sensors are randomly selected from a ULA of 20 sensors with half-wavelength inter-element spacing. The frequency grid size is set to be $\delta = 0.01$, and hence the number of grids is $L = 200$. That is, the preset frequency grids are $\{-1, -0.99, \cdots, 0.98, 0.99\}$. Two parameters utilized in (5.10) are given as $\eta = 10^{-5}$ and $\mu = \sigma\sqrt{M \ln(M)}$ [BTR13] with σ denoting the standard deviation of the noise vector, which is assumed to be known *a priori* in our simulations. $Q = 1000$ Monte Carlo trials are performed. The results of RMSE versus SNR and PCD versus SNR are plotted in Figures 5.2 and 5.3, respectively. It is seen that, in the large SNR region, 2nd Taylor G-LASSO has significantly lower RMSE compared with the other grid-based approaches, and the PCD of 2nd Taylor G-LASSO is higher than those of the other tested methods.

In the second experiment, we randomly select M sensors from a ULA of 20 sensors with half-wavelength inter-element spacing, and M varies from 4 to 20. SNR is fixed to 20 dB, while the remaining parameters are the same as those in the first experiment. The RMSE and PCD are depicted in Figures 5.4 and 5.5, respectively. The results

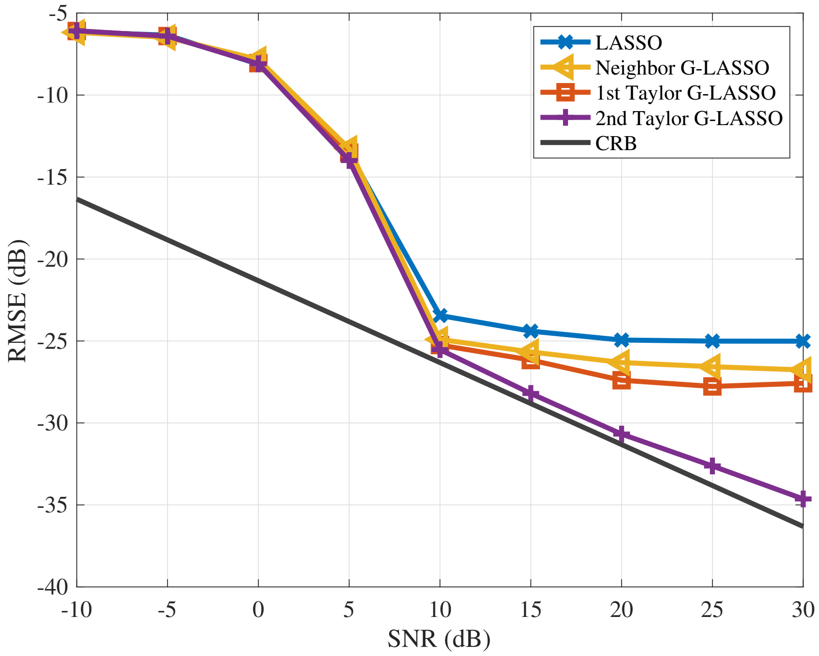


Figure 5.2: RMSE versus SNR with $M = 16$ sensors, $K = 2$ sources, $L = 200$ frequency grids, and grid size $\delta = 0.01$.

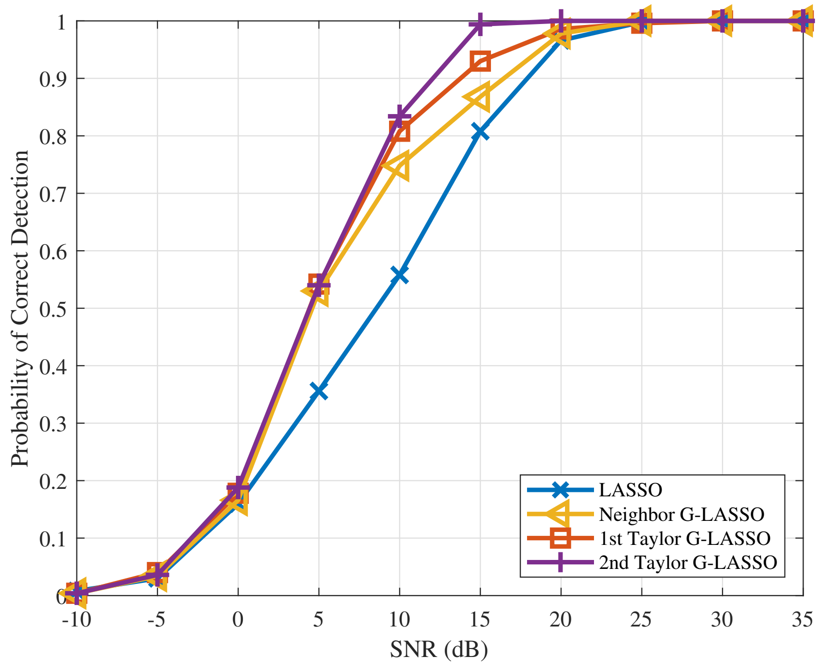


Figure 5.3: PCD versus SNR with $M = 16$ sensors, $K = 2$ sources, $L = 200$ frequency grids, and grid size $\delta = 0.01$.

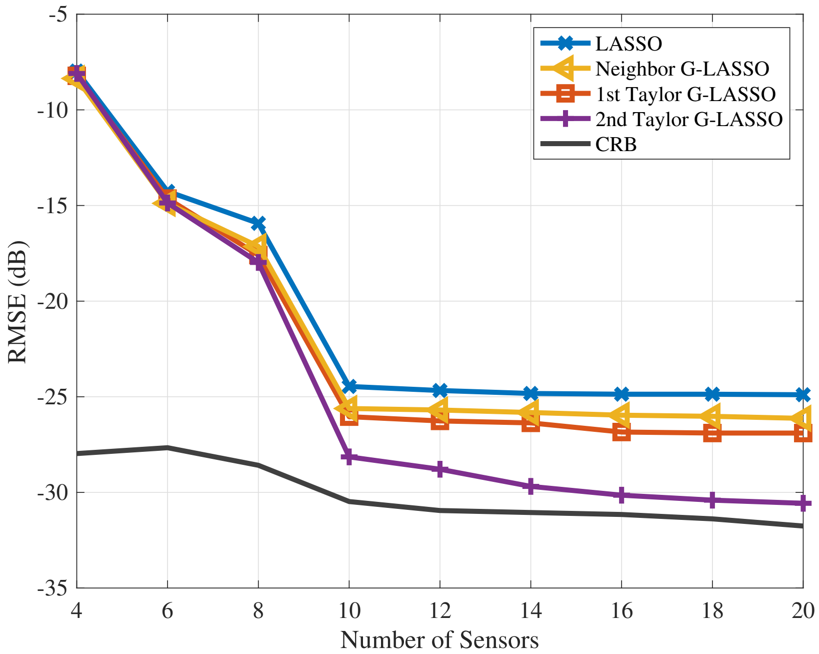


Figure 5.4: RMSE versus number of sensors with SNR = 20 dB, $K = 2$ sources, $L = 200$ frequency grids, and grid size $\delta = 0.01$.

exhibit again better performance of the proposed 2nd Taylor G-LASSO than the other competitors.

In the third experiment, the number of frequency grids, i.e., L, varies from 50 to 500 with a step size of 50, the SNR is fixed to 20 dB, while the other parameters are unchanged as those in the first experiment. The RMSE and PCD results are shown in Figures 5.6 and 5.7, respectively. It can be seen that (i) When the number of grids is $L < 400$ (equivalently grid size of $\delta > 1/200$), the RMSE of 2nd Taylor G-LASSO is evidently smaller than those of the other tested methods; and (ii) When $L \geq 400$ (that is $\delta \leq 1/200$), the RMSE of 1st Taylor G-LASSO is very close to that of 2nd Taylor G-LASSO. This verifies that 2nd Taylor G-LASSO works better than 1st Taylor G-LASSO in terms of DOA estimation accuracy, especially when the grid size is not sufficiently small.

5.5.2 Multiple-snapshot Situation

In this section, we test the performances of the proposed method and several other algorithms in multiple-snapshot scenarios. We utilize 100 snapshots, and the other

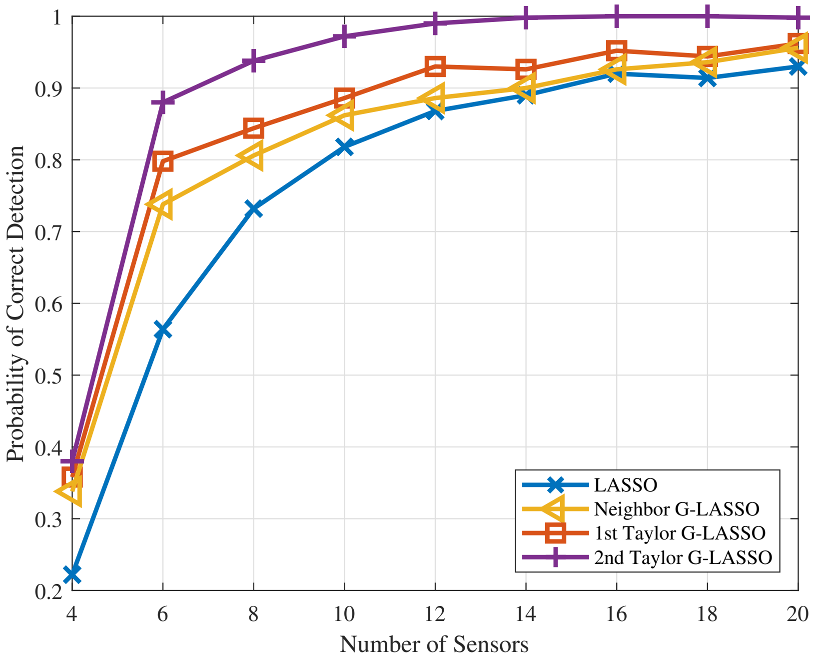


Figure 5.5: PCD versus number of sensors with SNR = 20 dB, $K = 2$ sources, $L = 200$ frequency grids, and grid size $\delta = 0.01$.

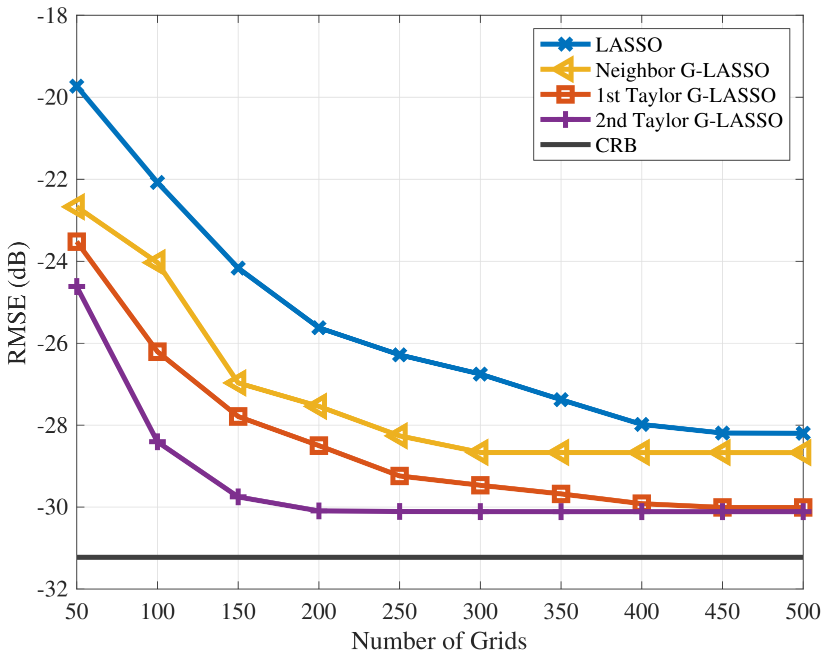


Figure 5.6: RMSE versus number of frequency grids with SNR = 20 dB, $M = 16$ sensors, and $K = 2$ sources.

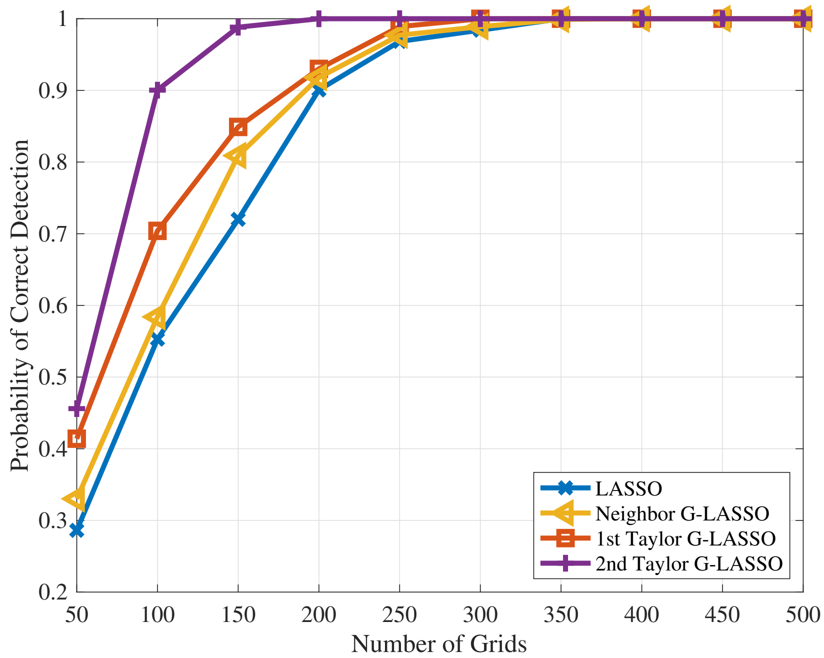


Figure 5.7: PCD versus number of frequency grids with SNR = 20 dB, $M = 16$ sensors, $K = 2$ sources.

parameters are set to be the same as those in the first experiment. The strategy of transforming the multiple-snapshot signal model into a single-snapshot one, which has been detailed in the remark in Section 5.3.3, is applied to LASSO, Neighbor G-LASSO, 1st Taylor G-LASSO, and 2nd Taylor G-LASSO. In addition, in this example, we also consider two classical methods, namely, the Capon beamforming and multiple signal classification (MUSIC) algorithms [CVY14]. For comparison, on-grid MUSIC with a much tinier grid size $\delta = 0.0001$ is also examined. The RMSE and PCD are plotted in Figures 5.8 and 5.9, respectively, from which it is seen that both Capon beamforming and MUSIC algorithms share similar performance with LASSO in the off-grid setup. On-grid MUSIC has the smallest RMSE and the largest PCD among all the tested approaches. The proposed 2nd Taylor G-LASSO outperforms LASSO, Neighbor G-LASSO, and 1st Taylor G-LASSO.

5.6 Summary

We have investigated the off-grid DOA estimation problem and have proposed a method using the second-order Taylor approximation. By exploring the properties of the block signal, we have added the proportionality relationship to our optimization problem. A

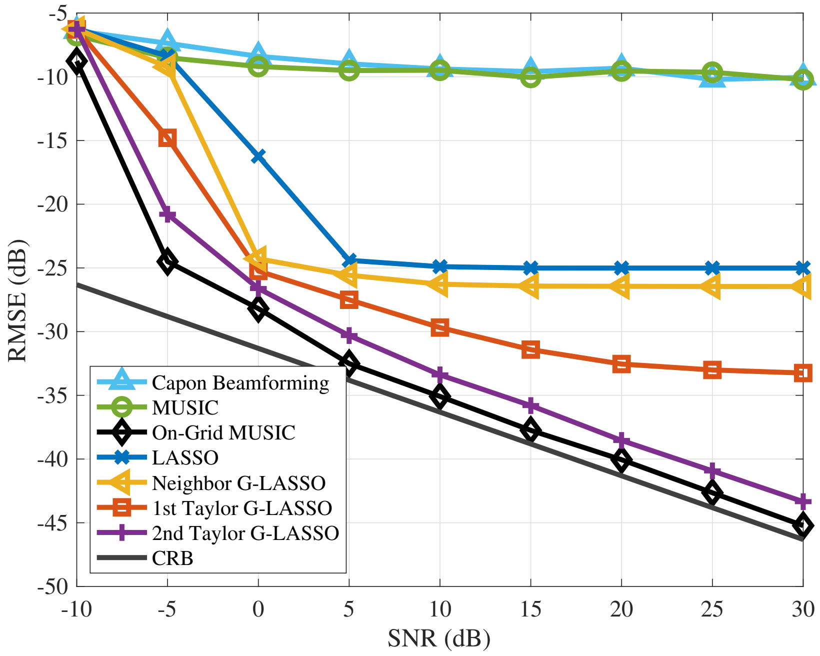


Figure 5.8: RMSE versus SNR in multiple-snapshot scenarios with 100 snapshots, $M = 16$ sensors, $K = 2$ sources.

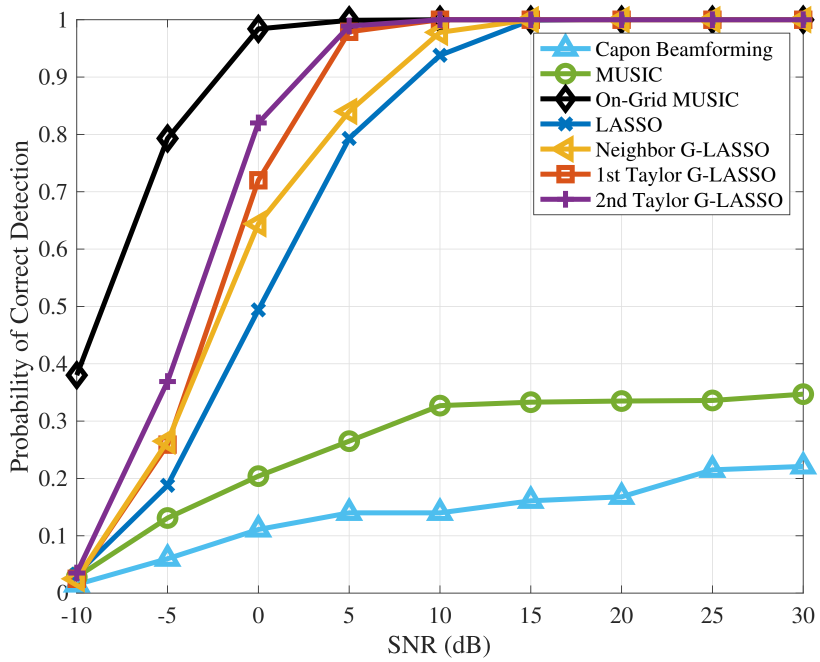


Figure 5.9: PCD versus SNR in multiple-snapshot scenarios with 100 snapshots, $M = 16$ sensors, $K = 2$ sources.

Monte Carlo test has shown the usefulness of such proportionality relationship in the sense that it increases the probabilities of $\{\beta_{2K} < 1\}$ and $\{\beta_{2K} < \sqrt{2} - 1\}$. Numerical results have demonstrated that the proposed method outperforms several existing grid-based DOA estimation approaches.

Chapter 6
Conclusions and Outlook

This dissertation has contributed new methods of sparse array signal processing. In particular, we have investigated two main problems in array signal processing, i.e., direction-of-arrival (DOA) estimation and beamforming, from the perspective of sparsity. A summary and the main conclusions of this thesis are given in Section 6.1, while Section 6.2 provides an outlook for possible future work.

6.1 Conclusions

In this dissertation, we focused on sparse array beamformer design, DOA estimation using a sparsely distorted sensor array, distorted sensor detection, and DOA estimation in an off-grid scenario. More specifically speaking, we conducted the following three main works.

In Chapter 3, an algorithm based on alternating direction method of multipliers (ADMM) for sparse array beamformer design was proposed. Our approach provides closed-form solutions at each ADMM iteration. Theoretical analyses and numerical simulations were provided to show the convergence of the proposed algorithm. In addition, the algorithm was proved to converge to the set of stationary points. The ADMM algorithm was shown to be comparable to the exhaustive search method, and slightly better than the state-of-the-art solvers, including the semidefinite relaxation (SDR), an SDR variant (SDR-V), and the successive convex approximation (SCA) methods, and significantly better than several other sparse array design strategies in terms of output signal-to-interference-plus-noise ratio. Moreover, the proposed ADMM algorithm outperformed the SDR, SDR-V, and SCA approaches in terms of computational cost.

In Chapter 4, we studied the problem of simultaneously estimating the DOAs of signals and detecting distorted sensors. It was assumed that the distorted sensors occur randomly, and the number of distorted sensors is much smaller than the total number of sensors. The problem was formulated via low-rank and row-sparse decomposition, and solved by iteratively reweighted least squares (IRLS). Both noiseless and noisy cases were considered. Theoretical analyses of algorithm convergence were provided. The computational complexity of the IRLS algorithm was compared with that of several existing methods. Simulation results were conducted for parameter selection, convergence speed, computational time, and performance of DOA estimation as well as distorted sensor detection. The IRLS method was demonstrated to have higher DOA estimation accuracy and lower computational cost than other methods, and the ADMM was shown to be slightly better than the IRLS algorithm in distorted sensor detection.

In Chapter 5, we investigated the off-grid DOA estimation problem and proposed a method using a second-order Taylor approximation. By exploring the properties of the block signal, we added the proportionality relationship to our optimization problem. A Monte Carlo test showed the usefulness of such proportionality relationship in the sense that it increases the probabilities of $\{\beta_{2K} < 1\}$ and $\{\beta_{2K} < \sqrt{2} - 1\}$. Numerical results demonstrated that the proposed method outperforms several existing grid-based DOA estimation approaches.

By developing these efficient methods, the property of sparsity has been demonstrated to be very important and useful in array signal processing. More and more works based on sparsity has appeared or will appear in the related research areas. Our work will lay the groundwork for fantastic research to follow.

6.2 Outlook

Possible extensions of this work and open problems are listed as below. Some of them are left open, some are the subject of ongoing research.

- In Chapter 3, we considered only one signal of interest. Along this direction, we can extend our work to more signals of interest. The extended setting results in a quadratically constrained quadratic program with more constraints, and consensus ADMM, rather than ADMM, can be applied.
- In Chapter 3, our focused was on narrowband signals. We can extend our method to wideband signals, which has found various real-world applications, see for example [YG04, Pan10, LW10, HA19]. The tapped delay line (TDL) filtering and discrete Fourier transform (DFT) are two typical schemes for wideband beamforming systems [LW10, HA19]. Different from the work in Chapter 3, which has a sparse solution, the resulting problem with wideband signals leads to group sparsity in its beamformer weights.

- In Chapter 4, we assumed that the number of distorted sensors is small. One of the future works is to investigate the maximal number of distorted sensors such that the proposed algorithm can still correctly detect all the distorted sensors, and at the same time accurately estimate the DOAs of sources.
- In Chapter 4, the distorted sensors were considered to be contaminated by unknown gain and phase uncertainties. Different types of sensor errors, such as sensor mutual coupling and sensor position errors, can be considered in future works.
- In Chapter 5, we adopted a Monte Carlo test to check whether or not the mixing matrix satisfies the block restricted isometry property (RIP) with proper parameter. Future work could be put on theoretically analyzing this result. Moreover, we can extend our algorithm from second-order Taylor approximation to higher-order Taylor approximation and explore the structure of the resulting signal to be estimated.

Appendix A

Appendix for Chapter 3

A.1 Proof of Theorem 3.1

Define the objective function as $f(\mathbf{w}) \triangleq \lambda\|\mathbf{w}\|_1 + \frac{\rho}{2}\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2$, and denote $\widetilde{\mathbf{w}}$ as the point within $\{\mathbf{w} : |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1\}$, such that it is closest to $\bar{\mathbf{w}}$. That is,

$$\|\widetilde{\mathbf{w}} - \bar{\mathbf{w}}\|_2 \leq \|\mathbf{w} - \bar{\mathbf{w}}\|_2 \tag{A.1}$$

holds for any $\mathbf{w} \in \{\mathbf{w} : |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1\}$. Our goal is to show $f(\widetilde{\mathbf{w}}) \leq f(\mathbf{w})$, for any $\mathbf{w} \in \{\mathbf{w} : |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1\}$.

As shall be shown later, the augmented Lagrangian parameter ρ is set to be large in order to make our algorithm converge. In such a case, we have $\rho \gg \lambda$, and thus the objective function

$$f(\mathbf{w}) \approx \frac{\rho}{2}\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2 \approx \frac{\rho}{2}\|\mathbf{w}-\bar{\mathbf{w}}\|_2^2. \tag{A.2}$$

Note that, in the second approximate equality above, we have used the fact that $\bar{\mathbf{w}} \approx \mathbf{v} - \mathbf{u}$ as $\rho \gg \lambda$.

Suppose that there exists a point $\mathbf{w}' \in \{\mathbf{w} : |\mathbf{w}^{\mathrm{H}}\mathbf{a}_0|^2 \geq 1\}$, such that $f(\mathbf{w}') < f(\widetilde{\mathbf{w}})$. Thus, by using (A.2), we obtain that $\frac{\rho}{2}\|\mathbf{w}'-\bar{\mathbf{w}}\|_2^2 < \frac{\rho}{2}\|\widetilde{\mathbf{w}}-\bar{\mathbf{w}}\|_2^2$, which contradicts (A.1). This implies that $f(\widetilde{\mathbf{w}}) \leq f(\mathbf{w})$ holds for all feasible $\mathbf{w}$, that is, $\widetilde{\mathbf{w}}$ is the solution to Problem (3.12a).

A.2 Proof of Lemma 3.1

In order to show that $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) \leq \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$ holds $\forall k = 0, 1, 2, \cdots$, where the objective function is defined as $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u}) \triangleq \lambda\|\mathbf{w}\|_1 + \mathbf{v}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{v} + \frac{\rho}{2}(\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2)$, we formulate their difference as

$$\begin{aligned}
&\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \\
&= [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)})] && \text{(A.3a)} \\
&\quad + [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})]. && \text{(A.3b)}
\end{aligned}$$

In what follows, we separately deal with (A.3a) and (A.3b). For (A.3a), it is calculated as

$$
\begin{aligned}
&\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) \\
\overset{(a)}{=} &\frac{\rho}{2}\big(\|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k+1)}\|_2^2 - \|\mathbf{u}_{(k+1)}\|_2^2 \\
&\qquad - \|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k)}\|_2^2 + \|\mathbf{u}_{(k)}\|_2^2\big) \\
\overset{(b)}{=} &\frac{\rho}{2}\big(\|2\mathbf{u}_{(k+1)} - \mathbf{u}_{(k)}\|_2^2 - 2\|\mathbf{u}_{(k+1)}\|_2^2 + \|\mathbf{u}_{(k)}\|_2^2\big) \\
= &\frac{\rho}{2}\big(2\|\mathbf{u}_{(k+1)} - \mathbf{u}_{(k)}\|_2^2\big) \\
\overset{(c)}{=} &\rho\left\|\frac{2}{\rho}\mathbf{R}_{\mathrm{x}}\mathbf{v}_{(k+1)} - \frac{2}{\rho}\mathbf{R}_{\mathrm{x}}\mathbf{v}_{(k)}\right\|_2^2 \\
\overset{(d)}{\leq} &\frac{4}{\rho}\lambda_{\max}^2(\mathbf{R}_{\mathrm{x}})\|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2,
\end{aligned}
$$

where the definition of $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$ is used in (a); $\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} = \mathbf{u}_{(k+1)} - \mathbf{u}_{(k)}$ (which is from Line 5 in Algorithm 3.1) has been utilized in (b); $\mathbf{u}_{(k+1)} = \frac{2}{\rho}\mathbf{R}_{\mathrm{x}}\mathbf{v}_{(k+1)}$ (which is the result by combining Lines 4 and 5 in Algorithm 3.1) has been employed in (c); and inequality $\|\mathbf{R}_{\mathrm{x}}\mathbf{v}\|_2^2 = \mathbf{v}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}^{\mathrm{H}}\mathbf{R}_{\mathrm{x}}\mathbf{v} \leq \mathbf{v}^{\mathrm{H}}[\lambda_{\max}^2(\mathbf{R}_{\mathrm{x}})\mathbf{I}]\mathbf{v} = \lambda_{\max}^2(\mathbf{R}_{\mathrm{x}})\|\mathbf{v}\|_2^2$ has been used in (d).

Now we focus on (A.3b), which can be written as

$$
\begin{aligned}
&\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \\
= &[\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})] \\
&+ [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})] \\
\overset{(a)}{\leq} &\Big[\Re\{\langle\nabla_{\mathbf{v}}\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}), \mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\rangle\} - \frac{\gamma_{\mathbf{v}}}{2}\|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2\Big] \\
&+ \Big[\Re\{\langle\zeta_{\mathbf{w}}, \mathbf{w}_{(k+1)} - \mathbf{w}_{(k)}\rangle\} - \frac{\gamma_{\mathbf{w}}}{2}\|\mathbf{w}_{(k+1)} - \mathbf{w}_{(k)}\|_2^2\Big] \\
\overset{(b)}{=} &-\frac{\gamma_{\mathbf{v}}}{2}\|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2 - \frac{\gamma_{\mathbf{w}}}{2}\|\mathbf{w}_{(k+1)} - \mathbf{w}_{(k)}\|_2^2 \\
\overset{(c)}{=} &-\Big[\lambda_{\min}(\mathbf{R}_{\mathrm{x}}) + \frac{\rho}{2}\Big]\|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2 - \frac{\gamma_{\mathbf{w}}}{2}\|\mathbf{w}_{(k+1)} - \mathbf{w}_{(k)}\|_2^2,
\end{aligned}
$$

where in (a) we have used the fact that $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$ is strongly convex w.r.t. $\mathbf{v}$ and $\mathbf{w}$, with parameters $\gamma_{\mathbf{v}} > 0$ and $\gamma_{\mathbf{w}} > 0$, respectively [RB16], $\nabla_{\mathbf{v}}\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) = 2\mathbf{R}_{\mathrm{x}}\mathbf{v}_{(k+1)} - \rho(\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k)})$ and $\zeta_{\mathbf{w}} \in \partial_{\mathbf{w}}\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$; in (b) we have used the optimality conditions of Problems (3.12b) and (3.12a); in (c) we have used $\gamma_{\mathbf{v}} = 2\lambda_{\min}(\mathbf{R}_{\mathrm{x}}) + \rho$, which is due to the facts that the objective function $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$ is twice continuously differentiable w.r.t. $\mathbf{v}$, and thus its strong convexity parameter $\gamma_{\mathbf{v}}$ satisfies $\nabla_{\mathbf{v}}^2\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u}) \succeq \gamma_{\mathbf{v}}\mathbf{I}$ for all $\mathbf{v}$ [RB16].

By substituting the above two inequalities back to (A.3), and denoting $\lambda_{\max}(\mathbf{R}_\mathrm{x})$ and $\lambda_{\min}(\mathbf{R}_\mathrm{x})$ as $\lambda_{\max}$ and $\lambda_{\min}$ for brevity, respectively, we have

$$\begin{aligned}&\mathcal{L}(\mathbf{w}_{(k+1)},\mathbf{v}_{(k+1)},\mathbf{u}_{(k+1)})-\mathcal{L}(\mathbf{w}_{(k)},\mathbf{v}_{(k)},\mathbf{u}_{(k)})\\ \le\;&\underbrace{\left[\frac{4}{\rho}\lambda_{\max}^2-\lambda_{\min}-\frac{\rho}{2}\right]\|\mathbf{v}_{(k+1)}-\mathbf{v}_{(k)}\|_2^2}_{\text{(i)}}-\underbrace{\frac{\gamma_\mathbf{w}}{2}\|\mathbf{w}_{(k+1)}-\mathbf{w}_{(k)}\|_2^2}_{\text{(ii)}}.\end{aligned}$$

We have the following discussions regarding the two terms in the above inequality, i.e., (i) and (ii).

- We observe that if $\rho<-\sqrt{\lambda_{\min}^2+8\lambda_{\max}^2}-\lambda_{\min}$, which should be deleted as $\rho>0$, or $\rho>\sqrt{\lambda_{\min}^2+8\lambda_{\max}^2}-\lambda_{\min}$, the coefficient $\frac{4}{\rho}\lambda_{\max}^2-\lambda_{\min}-\frac{\rho}{2}<0$ and thus (i) ≤ 0. Furthermore, because of $\sqrt{\lambda_{\min}^2+8\lambda_{\max}^2}-\lambda_{\min}<\lambda_{\min}+2\sqrt{2}\lambda_{\max}-\lambda_{\min}=2\sqrt{2}\lambda_{\max}$, we have the first conclusion that as long as $\rho\ge 2\sqrt{2}\lambda_{\max}$, (i) ≤ 0.
- Obviously, (ii) ≥ 0 thanks to $\gamma_\mathbf{w}>0$.

To sum up, $\mathcal{L}(\mathbf{w}_{(k+1)},\mathbf{v}_{(k+1)},\mathbf{u}_{(k+1)})-\mathcal{L}(\mathbf{w}_{(k)},\mathbf{v}_{(k)},\mathbf{u}_{(k)})\le$ (i) $-$ (ii) ≤ 0, as long as $\rho\ge 2\sqrt{2}\lambda_{\max}(\mathbf{R}_\mathrm{x})$.

A.3 Proof of Lemma 3.2

Note that the augmented Lagrangian function satisfies

$$\begin{aligned}&\mathcal{L}(\mathbf{w},\mathbf{v},\mathbf{u})\\ =\;&\lambda\|\mathbf{w}\|_1+\mathbf{v}^\mathrm{H}\mathbf{R}_\mathrm{x}\mathbf{v}+\frac{\rho}{2}\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2-\frac{\rho}{2}\|\mathbf{u}\|_2^2\\ \overset{\text{(a)}}{=}\;&\lambda\|\mathbf{w}\|_1+\mathbf{v}^\mathrm{H}\mathbf{R}_\mathrm{x}\mathbf{v}+\frac{\rho}{2}\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2-\frac{\rho}{2}\left\|\frac{2}{\rho}\mathbf{R}_\mathrm{x}\mathbf{v}\right\|_2^2\\ \overset{\text{(b)}}{\ge}\;&\lambda\|\mathbf{w}\|_1+\mathbf{v}^\mathrm{H}\mathbf{R}_\mathrm{x}\mathbf{v}+\frac{\rho}{2}\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2-\frac{2}{\rho}\mathbf{v}^\mathrm{H}[\lambda_{\max}^2(\mathbf{R}_\mathrm{x})\mathbf{I}]\mathbf{v}\\ =\;&\lambda\|\mathbf{w}\|_1+\mathbf{v}^\mathrm{H}\left(\mathbf{R}_\mathrm{x}-\frac{2}{\rho}\lambda_{\max}^2(\mathbf{R}_\mathrm{x})\mathbf{I}\right)\mathbf{v}+\frac{\rho}{2}\|\mathbf{w}-\mathbf{v}+\mathbf{u}\|_2^2\\ \overset{\text{(c)}}{\ge}\;&0,\end{aligned}$$

where in (a) we have used $\mathbf{u}=\frac{2}{\rho}\mathbf{R}_\mathrm{x}\mathbf{v}$ (which is the result by combining (3.12c) and (3.13)); in (b) we have used the inequality $\|\mathbf{R}_\mathrm{x}\mathbf{v}\|_2^2=\mathbf{v}^\mathrm{H}\mathbf{R}_\mathrm{x}^\mathrm{H}\mathbf{R}_\mathrm{x}\mathbf{v}\le\mathbf{v}^\mathrm{H}[\lambda_{\max}^2(\mathbf{R}_\mathrm{x})\mathbf{I}]\mathbf{v}$;

and inequality (c) holds if $\mathbf{R}_{\mathrm{x}} - \frac{2}{\rho}\lambda_{\max}^2(\mathbf{R}_{\mathrm{x}})\mathbf{I} \succeq 0$, which indicates (3.21). Note that inequality (c) is not tight, since the ℓ_1-norm term and the ℓ_2-norm term could be bounded from below by some large positive value. Therefore, the second term, i.e., the quadratic term w.r.t. $\mathbf{v}$, has much space to be tuned, which results in the fact that the lower bound for ρ in (3.21) is not tight.

A.4 Proof of Theorem 3.2

Denote $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$ and $\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$ by $\mathcal{L}_{(k+1)}$ and $\mathcal{L}_{(k)}$, respectively. According to Lemmata 3.1 and 3.2, the objective function value sequence $\{\mathcal{L}_{(k)}\}$ produced by Algorithm 3.1 converges. Further, since sequence $\{\mathcal{L}_{(k)}\}$ converges if ρ satisfies (3.22), we have $\mathcal{L}_{(k+1)} - \mathcal{L}_{(k)} \to 0$ as $k \to \infty$. On the other hand, we know from Appendix A.2 that $\mathcal{L}_{(k+1)} - \mathcal{L}_{(k)} \leq \text{(i)} - \text{(ii)} \leq 0$, as long as (3.22) holds, where (i) and (ii) are defined in Appendix A.2. Therefore, when (3.22) holds and $k \to \infty$, we have

$$0 = \mathcal{L}_{(k+1)} - \mathcal{L}_{(k)} \leq \text{(i)} - \text{(ii)} \leq 0, \tag{A.4}$$

meaning that (i) − (ii) = 0 or equivalently (i) = (ii). Moreover, we have (i) ≤ 0 when $\rho \geq 2\sqrt{2}\lambda_{\max}(\mathbf{R}_{\mathrm{x}})$ holds, and (ii) ≥ 0. Combining (i) = (ii), (i) ≤ 0, and (ii) ≥ 0, we have that (i) = (ii) = 0. This further indicates that

$$\mathbf{w}_{(k+1)} = \mathbf{w}_{(k)} \text{ and } \mathbf{v}_{(k+1)} = \mathbf{v}_{(k)}. \tag{A.5}$$

As already mentioned in Appendix A.2, $\mathbf{u} = \frac{2}{\rho}\mathbf{R}_{\mathrm{x}}\mathbf{v}$. By jointly considering $\mathbf{u} = \frac{2}{\rho}\mathbf{R}_{\mathrm{x}}\mathbf{v}$ and $\mathbf{v}_{(k+1)} = \mathbf{v}_{(k)}$ in (A.5), we obtain

$$\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)}. \tag{A.6}$$

Moreover, combining (A.6) and $\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)} + \mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}$ (i.e., Line 5 in Algorithm 3.1) yields

$$\mathbf{w}_{(k+1)} = \mathbf{v}_{(k+1)}, \tag{A.7}$$

as $k \to \infty$. Equivalently, $\mathbf{w}_{(k)} = \mathbf{v}_{(k)}$ as $k \to \infty$.

A.5 Proof of Theorem 3.3

The Lagrangian function of (3.11) is given by

$$\lambda\|\mathbf{w}\|_1 + \mathbf{v}\mathbf{R}_\mathrm{x}\mathbf{v} + \mu(|\mathbf{w}^\mathrm{H}\mathbf{a}_0| - 1) + \Re\{\langle\mathbf{y}, \mathbf{w}-\mathbf{v}\rangle\}, \tag{A.8}$$

where μ and $\mathbf{y}$ are Lagrangian dual variables corresponding to the inequality and equality constraints, respectively. Note that the (Lagrangian) dual variable $\mathbf{y}$ and the scaled dual variable $\mathbf{u}$ are related to each other as $\mathbf{y} = \rho\mathbf{u}$ [BPC+11]. A KKT point $(\mathbf{w}^\star, \mathbf{v}^\star)$ of Problem (3.11), together with the corresponding dual variables $\mu^\star$ and $\mathbf{y}^\star$, satisfies [HLR16]

$$\mathbf{0} = 2\mathbf{R}_\mathrm{x}\mathbf{v}^\star - \mathbf{y}^\star, \tag{A.9a}$$

$$\mathbf{w}^\star \in \arg\min_{\mathbf{w}} \left\{ \begin{array}{l} \lambda\|\mathbf{w}\|_1 + \mu^\star(|\mathbf{w}^\mathrm{H}\mathbf{a}_0|^2 - 1) \\ + \Re\{\langle\mathbf{y}^\star, \mathbf{w}-\mathbf{v}^\star\rangle\} \end{array} \right\}, \tag{A.9b}$$

$$\mathbf{w}^\star = \mathbf{v}^\star. \tag{A.9c}$$

Our aim is to show any limit point of Algorithm 3.1, referred to as $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$ or $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{y}_{(k+1)}/\rho)$, satisfies (A.9). Firstly, note that Line 4 in Algorithm 3.1 indicates

$$2\mathbf{R}_\mathrm{x}\mathbf{v}_{(k+1)} - \rho(\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k)}) = \mathbf{0}. \tag{A.10}$$

Jointly considering (A.6), (A.7), (A.10), and $\mathbf{y}_{(k+1)} = \rho\mathbf{u}_{(k+1)}$ yields $2\mathbf{R}_\mathrm{x}\mathbf{v}_{(k+1)} - \mathbf{y}_{(k+1)} = \mathbf{0}$, which is (A.9a). Additionally, (A.7) shows that (A.9c) is also achieved.

Now we turn to (A.9b). According to (3.12a), we have: $\mathbf{w}_{(k+1)}$

$$\begin{aligned}
&= \arg\min_{\mathbf{w}}\ \lambda\|\mathbf{w}\|_1 + \frac{\rho}{2}\|\mathbf{w}-\mathbf{v}_{(k)}+\mathbf{u}_{(k)}\|_2^2 \quad \text{s.t. } |\mathbf{w}^\mathrm{H}\mathbf{a}_0| \geq 1 \\
&\overset{\text{(a)}}{=} \arg\min_{\mathbf{w}}\ \lambda\|\mathbf{w}\|_1 + \mu^\star(|\mathbf{w}^\mathrm{H}\mathbf{a}_0|-1) + \frac{\rho}{2}\|\mathbf{w}-\mathbf{v}_{(k)}+\mathbf{u}_{(k)}\|_2^2 \\
&= \arg\min_{\mathbf{w}} \left\{ \begin{array}{l} \lambda\|\mathbf{w}\|_1 + \mu^\star(|\mathbf{w}^\mathrm{H}\mathbf{a}_0|-1) + \frac{\rho}{2}\|\mathbf{w}-\mathbf{v}_{(k)}\|_2^2 \\ + \frac{\rho}{2}\|\mathbf{u}_{(k)}\|_2^2 + \rho\Re\{\langle\mathbf{u}_{(k)}, \mathbf{w}-\mathbf{v}_{(k)}\rangle\} \end{array} \right\} \\
&\overset{\text{(b)}}{=} \arg\min_{\mathbf{w}}\ \lambda\|\mathbf{w}\|_1 + \mu^\star(|\mathbf{w}^\mathrm{H}\mathbf{a}_0|-1) + \rho\Re\{\langle\mathbf{u}_{(k)}, \mathbf{w}-\mathbf{v}_{(k)}\rangle\} \\
&\overset{\text{(c)}}{=} \arg\min_{\mathbf{w}}\ \lambda\|\mathbf{w}\|_1 + \mu^\star(|\mathbf{w}^\mathrm{H}\mathbf{a}_0|-1) + \Re\{\langle\mathbf{y}_{(k+1)}, \mathbf{w}-\mathbf{v}_{(k+1)}\rangle\},
\end{aligned}$$

where in (a) we have written the constraint into the objective function by involving its optimal dual variable $\mu^\star$; in (b) we have utilized the facts that $\mathbf{w}_{(k+1)} = \mathbf{w}_{(k)} = \mathbf{v}_{(k)}$ at any limit point, and $\frac{\rho}{2}\|\mathbf{u}_{(k)}\|_2^2$ is a scalar term unrelated to $\mathbf{w}$; in (c) we have used $\rho\mathbf{u}_{(k)} = \rho\mathbf{u}_{(k+1)} = \mathbf{y}_{(k+1)}$ and $\mathbf{v}_{(k)} = \mathbf{v}_{(k+1)}$. This completes the proof of Theorem 3.3.

Appendix B

Appendix for Chapter 4

B.1 Proof of Theorem 4.1

We calculate the difference between the objective function values in two successive iterations as

$$
\begin{aligned}
&f(\mathbf{Z}_{(k)}) - f(\mathbf{Z}_{(k+1)}) \\
&= \|[\mathbf{Z}_{(k)}, \mu\mathbf{I}]\|_* - \|[\mathbf{Z}_{(k+1)}, \mu\mathbf{I}]\|_* + \lambda\big(\|[\mathbf{Y} - \mathbf{Z}_{(k)}, \mu\mathbf{1}]\|_{2,1} - \|[\mathbf{Y} - \mathbf{Z}_{(k+1)}, \mu\mathbf{1}]\|_{2,1}\big) \\
&= \mathrm{trace}\Big(\big(\mathbf{Z}_{(k)}\mathbf{Z}_{(k)}^{\mathrm{H}} + \mu^2\mathbf{I}\big)^{\frac{1}{2}}\Big) - \mathrm{trace}\Big(\big(\mathbf{Z}_{(k+1)}\mathbf{Z}_{(k+1)}^{\mathrm{H}} + \mu^2\mathbf{I}\big)^{\frac{1}{2}}\Big) \\
&\quad + \lambda\big(\|[\mathbf{Y} - \mathbf{Z}_{(k)}, \mu\mathbf{1}]\|_{2,1} - \|[\mathbf{Y} - \mathbf{Z}_{(k+1)}, \mu\mathbf{1}]\|_{2,1}\big) \\
&\geq \mathrm{trace}\Big(\frac{1}{2}\big(\mathbf{Z}_{(k)}\mathbf{Z}_{(k)}^{\mathrm{H}} - \mathbf{Z}_{(k+1)}\mathbf{Z}_{(k+1)}^{\mathrm{H}}\big)\mathbf{P}_{(k)}\Big) \\
&\quad + \frac{\lambda}{2}\mathrm{trace}\Big(\mathbf{Q}_{(k)}\Big[\big(\mathbf{Y}-\mathbf{Z}_{(k)}\big)\big(\mathbf{Y}-\mathbf{Z}_{(k)}\big)^{\mathrm{H}} - \big(\mathbf{Y}-\mathbf{Z}_{(k+1)}\big)\big(\mathbf{Y}-\mathbf{Z}_{(k+1)}\big)^{\mathrm{H}}\Big]\Big) \qquad \text{(B.1)} \\
&= \mathrm{trace}\Big(\frac{1}{2}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\mathbf{P}_{(k)}\Big) + \mathrm{trace}\big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})\mathbf{Z}_{(k+1)}^{\mathrm{H}}\mathbf{P}_{(k)}\big) \\
&\quad + \frac{\lambda}{2}\mathrm{trace}\big(2\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k+1)} - \mathbf{Z}_{(k)})^{\mathrm{H}}\big) + \frac{\lambda}{2}\mathrm{trace}\big(\mathbf{Q}_{(k)}(\mathbf{Z}_{(k)}\mathbf{Z}_{(k)}^{\mathrm{H}} - \mathbf{Z}_{(k+1)}\mathbf{Z}_{(k+1)}^{\mathrm{H}})\big) \\
&= \mathrm{trace}\Big(\frac{1}{2}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\mathbf{P}_{(k)}\Big) + \mathrm{trace}\big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})\mathbf{Z}_{(k+1)}^{\mathrm{H}}\mathbf{P}_{(k)}\big) \\
&\quad + \frac{\lambda}{2}\mathrm{trace}\big(2\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k+1)} - \mathbf{Z}_{(k)})^{\mathrm{H}}\big) + \frac{\lambda}{2}\mathrm{trace}\big(\mathbf{Q}_{(k)}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\big) \\
&\quad + \lambda\,\mathrm{trace}\big(\mathbf{Q}_{(k)}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})\mathbf{Z}_{(k+1)}^{\mathrm{H}}\big) \\
&\geq \mathrm{trace}\big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})\mathbf{Z}_{(k+1)}^{\mathrm{H}}\mathbf{P}_{(k)}\big) + \mathrm{trace}\big(\lambda\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k+1)} - \mathbf{Z}_{(k)})^{\mathrm{H}}\big) \\
&\quad + \mathrm{trace}\big(\lambda\mathbf{Q}_{(k)}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})\mathbf{Z}_{(k+1)}^{\mathrm{H}}\big) \qquad \text{(B.2)} \\
&= \mathrm{trace}\big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})\mathbf{Z}_{(k+1)}^{\mathrm{H}}(\mathbf{P}_{(k)} + \lambda\mathbf{Q}_{(k)})\big) + \mathrm{trace}\big(\lambda\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k+1)} - \mathbf{Z}_{(k)})^{\mathrm{H}}\big) \\
&= \mathrm{trace}\big((\mathbf{P}_{(k)} + \lambda\mathbf{Q}_{(k)})\mathbf{Z}_{(k+1)}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\big) + \mathrm{trace}\big(\lambda\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k+1)} - \mathbf{Z}_{(k)})^{\mathrm{H}}\big) \\
&= \mathrm{trace}\big(\lambda\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\big) + \mathrm{trace}\big(\lambda\mathbf{Q}_{(k)}\mathbf{Y}(\mathbf{Z}_{(k+1)} - \mathbf{Z}_{(k)})^{\mathrm{H}}\big) \qquad \text{(B.3)} \\
&= 0,
\end{aligned}
$$

which indicates that $f(\mathbf{Z})$ is a non-increasing function with sequence $\{\mathbf{Z}_{(k)}\}$ generated by the IRLS procedure (4.10). The first equality is based on the definition of the objective function in (4.9), and the second equality uses $\|\mathbf{Z}\|_* = \mathrm{trace}\Big(\big(\mathbf{Z}\mathbf{Z}^{\mathrm{H}}\big)^{-\frac{1}{2}}\Big)$ when $M < T$. Inequality (B.1) holds thanks to Lemmata 4.1 and 4.2. Inequality (B.2) holds because of $\mathrm{trace}\big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\mathbf{P}\big) \geq 0$ and

$\mathrm{trace}\big(\mathbf{Q}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\big) \geq 0$, which result from the fact that $\mathbf{P}$ and $\mathbf{Q}$ are symmetric matrices and $\mathrm{trace}\big(\mathbf{X}\mathbf{X}^{\mathrm{H}}\big) = \|\mathbf{X}\|_{\mathrm{F}}^2 \geq 0$. Equality (B.3) holds true according to the KKT condition, i.e., $(\mathbf{P}_{(k)} + \lambda\mathbf{Q}_{(k)})\mathbf{Z}_{(k+1)} = \lambda\mathbf{Q}_{(k)}\mathbf{Y}$.

Since $f(\mathbf{Z})$ is a non-increasing, we have

$$\|\mathbf{Z}_{(k)}\|_* = \mathrm{trace}\Big(\big(\mathbf{Z}\mathbf{Z}^{\mathrm{H}}\big)^{-\frac{1}{2}}\Big) < \mathrm{trace}\Big(\big(\mathbf{Z}\mathbf{Z}^{\mathrm{H}} + \mu^2\mathbf{I}\big)^{-\frac{1}{2}}\Big) = \|[\mathbf{Z}_{(k)}, \mu\mathbf{I}]\|_* \leq f(\mathbf{Z}_{(k)}) \leq f(\mathbf{Z}_0),$$

which indicates that the sequence $\{\mathbf{Z}_{(k)}\}$ is bounded in terms of its nuclear norm.

Besides, combining (4.10) and (B.1) yields

$$\begin{aligned}
& f(\mathbf{Z}_{(k)}) - f(\mathbf{Z}_{(k+1)}) \\
\geq\ & \mathrm{trace}\Big(\frac{1}{2}\big(\mathbf{Z}_{(k)}\mathbf{Z}_{(k)}^{\mathrm{H}} - \mathbf{Z}_{(k+1)}\mathbf{Z}_{(k+1)}^{\mathrm{H}}\big)\mathbf{P}_{(k)}\Big) \\
& + \frac{\lambda}{2}\mathrm{trace}\Big(\mathbf{Q}_{(k)}\Big[\big(\mathbf{Y}-\mathbf{Z}_{(k)}\big)\big(\mathbf{Y}-\mathbf{Z}_{(k)}\big)^{\mathrm{H}} - \big(\mathbf{Y}-\mathbf{Z}_{(k+1)}\big)\big(\mathbf{Y}-\mathbf{Z}_{(k+1)}\big)^{\mathrm{H}}\Big]\Big) \\
=\ & \frac{1}{2}\mathrm{trace}\big((\mathbf{P}_{(k)} + \lambda\mathbf{Q}_{(k)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\big) \\
\geq\ & \frac{1}{2}\sum_{i=1}^{M}\zeta_i(\mathbf{P}_{(k)}+\lambda\mathbf{Q}_{(k)})\zeta_{M-i+1}\big((\mathbf{Z}_{(k)}-\mathbf{Z}_{k+1})(\mathbf{Z}_{(k)}-\mathbf{Z}_{k+1})^{\mathrm{H}}\big) \qquad \text{(B.4)} \\
\geq\ & \frac{1}{2}\zeta_M(\mathbf{P}_{(k)}+\lambda\mathbf{Q}_{(k)})\|\mathbf{Z}_{(k)}-\mathbf{Z}_{k+1}\|_{\mathrm{F}}^2 \geq \frac{1}{2}\zeta_{\min}\times\|\mathbf{Z}_{(k)}-\mathbf{Z}_{k+1}\|_{\mathrm{F}}^2,
\end{aligned}$$

where $\zeta_i(\cdot)$ denotes the ith largest eigenvalue of its input Hermitian matrix, inequality (B.4) follows from the fact that $\mathrm{trace}(\mathbf{X}\mathbf{Y}) \geq \sum_{i=1}^{M}\zeta_i(\mathbf{X})\zeta_{M-i+1}(\mathbf{Y})$ holds for any two positive semi-definite matrices $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{C}^{M\times M}$ [LLY15], and in the last inequality, we have defined $\zeta_{\min} > 0$ as the smallest eigenvalue of $\mathbf{P}_{(k)}+\lambda\mathbf{Q}_{(k)}$ over all k. Summing all the above inequalities for all $k \geq 0$, we have

$$f(\mathbf{Z}_0) \geq \frac{1}{2}\zeta_{\min}\sum_{k=0}^{\infty}\|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}}^2,$$

which implies that $\lim_{k\to\infty}\|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}} = 0$.

B.2 Proof of Theorem 4.2

For any matrices $\mathbf{Y}$ and $\mathbf{Z} \in \mathbb{C}^{M\times T}$, we have

$$\begin{aligned}
\|[\mathbf{Y}-\mathbf{Z},\mu\mathbf{1}]\|_{2,1} &= \sum_{i=1}^{M}\sqrt{\|(\mathbf{Y}-\mathbf{Z})_{i,:}\|_2^2+\mu^2} \;\geq\; \sum_{i=1}^{M}|\mu| \;=\; |\mu|M \\
\|[\mathbf{Z},\mu\mathbf{I}]\|_* &= \mathrm{trace}\Big(\big(\mathbf{Z}\mathbf{Z}^{\mathrm{H}}+\mu^2\mathbf{I}\big)^{\frac{1}{2}}\Big) \\
&\geq \big(\mathrm{trace}\big(\mathbf{Z}\mathbf{Z}^{\mathrm{H}}+\mu^2\mathbf{I}\big)\big)^{\frac{1}{2}} \\
&= \big(\mathrm{trace}\big(\mathbf{Z}\mathbf{Z}^{\mathrm{H}}\big)+M\mu^2\big)^{\frac{1}{2}} \\
&\geq \big(M\mu^2\big)^{\frac{1}{2}} \;=\; |\mu|\sqrt{M}.
\end{aligned} \tag{B.5}$$

Inequality (B.5) holds because $\mathrm{trace}\Big(\mathbf{X}^{\frac{1}{2}}\Big)=\sum_i\sqrt{\zeta_i} \geq \sqrt{\sum_i \zeta_i} = (\mathrm{trace}(\mathbf{X}))^{\frac{1}{2}}$ for any symmetric matrix $\mathbf{X}$, with ζ_i being the eigenvalue of $\mathbf{X}$. Therefore, the objective function in (4.9) is bounded below as $f(\mathbf{Z}) = \|[\mathbf{Z},\mu\mathbf{I}]\|_* + \lambda\|[\mathbf{Y}-\mathbf{Z},\mu\mathbf{1}]\|_{2,1} \geq |\mu|(\sqrt{M}+\lambda M)$.

B.3 Proof of Theorem 4.3

Denote the limit point of sequence $\{\mathbf{Z}_{(k)}\}$ as $\mathbf{Z}_{(k+1)}$. Then, according to $\lim_{k\to\infty}\|\mathbf{Z}_{(k)}-\mathbf{Z}_{(k+1)}\|_{\mathrm{F}}=0$ in Theorem 4.1 and (4.10), we have $\mathbf{Z}_{(k+1)}=\lambda(\mathbf{P}_{(k+1)}+\lambda\mathbf{Q}_{(k+1)})^{-1}\mathbf{Q}_{(k+1)}\mathbf{Y}$, that is, $\mathbf{P}_{(k+1)}\mathbf{Z}_{(k+1)}+\lambda\mathbf{Q}_{(k+1)}(\mathbf{Z}_{(k+1)}-\mathbf{Y})=\mathbf{0}$. This indicates that $\mathbf{Z}_{(k+1)}$ satisfies the KKT condition. Since Problem (4.9) is convex w.r.t. $\mathbf{Z}$, the stationary point is globally optimal.

B.4 Proof of Theorem 4.4

Similar to the proof of Theorem 4.1, we calculate the difference between the objective function values in two successive iterations as

$$\begin{aligned}
&f(\mathbf{Z}_{(k)},\mathbf{V}_{(k)})-f(\mathbf{Z}_{(k+1)},\mathbf{V}_{(k+1)}) \\
=\;&\frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{(k)}-\mathbf{V}_{(k)}\|_{\mathrm{F}}^2-\frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{(k+1)}-\mathbf{V}_{(k+1)}\|_{\mathrm{F}}^2+\lambda_1\|[\mathbf{Z}_{(k)},\mu\mathbf{I}]\|_*-\lambda_1\|[\mathbf{Z}_{(k+1)},\mu\mathbf{I}]\|_* \\
&+\lambda_2\|[\mathbf{V}_{(k)},\mu\mathbf{1}]\|_{2,1}-\lambda_2\|[\mathbf{V}_{(k+1)},\mu\mathbf{1}]\|_{2,1} \\
\geq\;&\frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{(k)}-\mathbf{V}_{(k)}\|_{\mathrm{F}}^2-\frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{(k+1)}-\mathbf{V}_{(k+1)}\|_{\mathrm{F}}^2+\lambda_1\,\mathrm{trace}\Big(\frac{1}{2}\big(\mathbf{Z}_{(k)}\mathbf{Z}_{(k)}^{\mathrm{H}}-\mathbf{Z}_{(k+1)}\mathbf{Z}_{(k+1)}^{\mathrm{H}}\big)\mathbf{P}_{(k)}\Big)
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{\lambda_2}{2}\mathrm{trace}\Big(\mathbf{Q}_{(k)}\big(\mathbf{V}_{(k)}\mathbf{V}_{(k)}^{\mathrm{H}} - \mathbf{V}_{(k+1)}\mathbf{V}_{(k+1)}^{\mathrm{H}}\big)\Big) \\
&= \frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{(k)}-\mathbf{V}_{(k)}\|_{\mathrm{F}}^2 - \frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{(k+1)}-\mathbf{V}_{(k+1)}\|_{\mathrm{F}}^2 + \lambda_1\,\mathrm{trace}\bigg(\frac{1}{2}\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)^{\mathrm{H}}\mathbf{P}_{(k)}\bigg) \\
&\quad + \lambda_1\,\mathrm{trace}\Big(\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)\mathbf{Z}_{(k+1)}^{\mathrm{H}}\mathbf{P}_{(k)}\Big) + \frac{\lambda_2}{2}\mathrm{trace}\Big(\mathbf{Q}_{(k)}\big(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\big)\big(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\big)^{\mathrm{H}}\Big) \\
&\quad + \lambda_2\,\mathrm{trace}\Big(\mathbf{Q}_{(k)}(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})\mathbf{V}_{(k+1)}^{\mathrm{H}}\Big) \qquad \text{(B.6)} \\
&\geq \frac{1}{2}\mathrm{trace}\Big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}} + (\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})^{\mathrm{H}}\Big) \\
&\quad + \mathrm{trace}\Big((\mathbf{Z}_{(k+1)}-\mathbf{Y})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}} + (\mathbf{V}_{(k+1)}-\mathbf{Y})(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})^{\mathrm{H}} + \mathbf{Z}_{(k)}\mathbf{V}_{(k)}^{\mathrm{H}} - \mathbf{Z}_{(k+1)}\mathbf{V}_{(k+1)}^{\mathrm{H}}\Big) \\
&\quad + \lambda_1\,\mathrm{trace}\Big(\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)\mathbf{Z}_{(k+1)}^{\mathrm{H}}\mathbf{P}_{(k)}\Big) + \lambda_2\,\mathrm{trace}\Big(\mathbf{Q}_{(k)}(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})\mathbf{V}_{(k+1)}^{\mathrm{H}}\Big) \\
&= \frac{1}{2}\mathrm{trace}\Big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}} + (\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})^{\mathrm{H}}\Big) \\
&\quad + \mathrm{trace}\Big(\mathbf{Z}_{(k)}\mathbf{V}_{(k)}^{\mathrm{H}} - \mathbf{Z}_{(k+1)}\mathbf{V}_{(k+1)}^{\mathrm{H}}\Big) + \mathrm{trace}\Big(-\mathbf{V}_{(k)}(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\Big) + \mathrm{trace}\Big(-\mathbf{Z}_{(k)}(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})^{\mathrm{H}}\Big) \\
&= \frac{1}{2}\mathrm{trace}\Big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}} + (\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)})^{\mathrm{H}}\Big) \\
&\quad + \mathrm{trace}\Big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{V}_{(k+1)} - \mathbf{V}_{(k)})^{\mathrm{H}}\Big) \\
&= \frac{1}{2}\mathrm{trace}\Big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)} - \mathbf{V}_{(k)} + \mathbf{V}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)} - \mathbf{V}_{(k)} + \mathbf{V}_{(k+1)})^{\mathrm{H}}\Big) \\
&= \frac{1}{2}\|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)} - \mathbf{V}_{(k)} + \mathbf{V}_{(k+1)}\|_{\mathrm{F}}^2 \;\geq\; 0,
\end{aligned}
$$

which indicates that $f(\mathbf{Z},\mathbf{V})$ is a non-increasing function. Since $f(\mathbf{Z},\mathbf{V})$ is non-increasing, we have

$$
\begin{aligned}
\min\{\lambda_1,\lambda_2\}\big(\|[\mathbf{Z}_{(k)},\mu\mathbf{I}]\|_* + \|[\mathbf{V}_{(k)},\mu\mathbf{1}]\|_{2,1}\big) &\leq \lambda_1\|[\mathbf{Z}_{(k)},\mu\mathbf{I}]\|_* + \lambda_2\|[\mathbf{V}_{(k)},\mu\mathbf{1}]\|_{2,1} \\
&\leq f(\mathbf{Z}_{(k)},\mathbf{V}_{(k)}) \;\leq\; f(\mathbf{Z}_0,\mathbf{V}_0).
\end{aligned}
$$

Hence, $\|\mathbf{Z}_{(k)}\|_* + \|\mathbf{V}_{(k)}\|_{2,1} < \|[\mathbf{Z}_{(k)},\mu\mathbf{I}]\|_* + \|[\mathbf{V}_{(k)},\mu\mathbf{1}]\|_{2,1} \leq \frac{f(\mathbf{Z}_0,\mathbf{V}_0)}{\min\{\lambda_1,\lambda_2\}}$, which shows that $\{(\mathbf{Z}_{(k)},\mathbf{V}_{(k)})\}$ is bounded.

Besides, combining (4.15) and (B.6) yields

$$
\begin{aligned}
&f(\mathbf{Z}_{(k)},\mathbf{V}_{(k)}) - f(\mathbf{Z}_{(k+1)},\mathbf{V}_{(k+1)}) \\
&\geq \frac{\lambda_1}{2}\,\mathrm{trace}\Big(\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)^{\mathrm{H}}\mathbf{P}_{(k)}\Big) + \frac{\lambda_2}{2}\mathrm{trace}\Big(\mathbf{Q}_{(k)}\big(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\big)\big(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\big)^{\mathrm{H}}\Big) \\
&\quad + \frac{1}{2}\mathrm{trace}\Big((\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)})^{\mathrm{H}}\Big) \\
&\geq \frac{\lambda_1}{2}\,\mathrm{trace}\Big(\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)\big(\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\big)^{\mathrm{H}}\mathbf{P}_{(k)}\Big) + \frac{\lambda_2}{2}\mathrm{trace}\Big(\mathbf{Q}_{(k)}\big(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\big)\big(\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\big)^{\mathrm{H}}\Big) \\
&\geq \frac{\lambda_1}{2}\sum_{i}^{M}\zeta_i(\mathbf{P}_{(k)})\zeta_{M-i+1}\big((\mathbf{Z}_{(k)}-\mathbf{Z}_{(k+1)})(\mathbf{Z}_{(k)}-\mathbf{Z}_{(k+1)})^{\mathrm{H}}\big) \\
&\quad + \frac{\lambda_2}{2}\sum_{i}^{M}\zeta_i(\mathbf{Q}_{(k)})\zeta_{M-i+1}\big((\mathbf{V}_{(k)}-\mathbf{V}_{(k+1)})(\mathbf{V}_{(k)}-\mathbf{V}_{(k+1)})^{\mathrm{H}}\big)
\end{aligned}
$$

$$\geq \frac{\lambda_1}{2}\zeta_{\min}^{(\mathbf{P})} \times \|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}}^2 + \frac{\lambda_1}{2}\zeta_{\min}^{(\mathbf{Q})} \times \|\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\|_{\mathrm{F}}^2,$$

where $\zeta_{\min}^{(\mathbf{P})}$ and $\zeta_{\min}^{(\mathbf{Q})}$ are the smallest eigenvalues of $\mathbf{P}_{(k)}$ and $\mathbf{Q}_{(k)}$, respectively, over all k. Summing all the above inequalities for all $k \geq 0$, we have

$$f(\mathbf{Z}_0, \mathbf{V}_0) \geq \frac{\lambda_1}{2}\zeta_{\min}^{(\mathbf{P})} \sum_{k=0}^{\infty} \|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}}^2 + \frac{\lambda_2}{2}\zeta_{\min}^{(\mathbf{Q})} \sum_{k=0}^{\infty} \|\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\|_{\mathrm{F}}^2,$$

which implies that $\lim_{k\to\infty} \|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}} = 0$ and $\lim_{k\to\infty} \|\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\|_{\mathrm{F}} = 0$.

B.5 Proof of Theorem 4.5

Considering the inequalities in Appendix B.2 and $\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\mathrm{F}}^2 \geq 0$, we can prove that the objective function in (4.13) is bounded below as $f(\mathbf{Z}, \mathbf{V}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\mathrm{F}}^2 + \lambda_1\|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda_2\|[\mathbf{V}, \mu\mathbf{1}]\|_{2,1} \geq |\mu|(\lambda_1\sqrt{M} + \lambda_2 M)$.

B.6 Proof of Theorem 4.6

Denote the limit point of the sequence $\{(\mathbf{Z}_{(k)}, \mathbf{V}_{(k)})\}$ as $(\mathbf{Z}_{(k+1)}, \mathbf{V}_{(k+1)})$. Then, according to $\lim_{k\to\infty} \|\mathbf{Z}_{(k)} - \mathbf{Z}_{(k+1)}\|_{\mathrm{F}} = 0$ and $\lim_{k\to\infty} \|\mathbf{V}_{(k)} - \mathbf{V}_{(k+1)}\|_{\mathrm{F}} = 0$ in Theorem 4.4 and (4.15), we have

$$\begin{cases} \mathbf{Z}_{(k+1)} = (\mathbf{I} + \lambda_1 \mathbf{P}_{(k+1)})^{-1}(\mathbf{Y} - \mathbf{V}_{(k+1)}) \\ \mathbf{V}_{(k+1)} = (\mathbf{I} + \lambda_2 \mathbf{Q}_{(k+1)})^{-1}(\mathbf{Y} - \mathbf{Z}_{(k+1)}), \end{cases}$$

which is the KKT condition of Problem (4.13). Since Problem (4.13) is convex w.r.t. $\mathbf{Z}$ and $\mathbf{V}$, the stationary point is globally optimal.

List of Acronyms

ADMM	alternating direction method of multipliers
APG	accelerated proximal gradient
B&B	branch and bound
CPU	central processing unit
CRB	Cramér–Rao bound
DFT	discrete Fourier transform
DOA	direction-of-arrival
ESPRIT	estimation of signal parameters via rotational invariance techniques
G-LASSO	group least absolute shrinkage and selection operator
INR	interference-to-noise ratio
IRLS	iteratively reweighted least squares
KKT	Karush-Kuhn-Tucker
LASSO	least absolute shrinkage and selection operator
LR^2SD	low-rank and row-sparse decomposition
LRCSD	low-rank and column-sparse decomposition
LRSD	low-rank and sparse decomposition
MIMO	multiple input multiple output
MIP	mixed-integer programming
ML	maximum likelihood
MUSIC	multiple signal classification
MVDR	minimum variance distortionless response
PCD	probability of correct detection
RF	radio-frequency
RIP	restricted isometry property
RMSE	root-mean squared error
ROI	region of interest
RPCA	robust principal component analysis
SCA	successive convex approximation
SDR	semidefinite relaxation
SDR-V	semidefinite relaxation variant
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SOCP	second-order cone programming

SOI	signal-of-interest
SVD	singular value decomposition
SVT	singular value thresholding
TDL	tapped delay line
ULA	uniform linear array
w.r.t.	with respect to

List of Symbols

The most important symbols in the dissertation are listed in alphabetical order.

$\mathbf{0}$	all-zero vector of appropriate length
$\mathbf{1}$	all-one vector of appropriate length
$\mathbf{1}_M$	$M \times M$ all-one matrix
$\mathbb{C}$	set of complex numbers
$\mathrm{card}(\cdot)$	cardinality of a set
$\mathrm{diag}\{\cdot\}$	returns a diagonal matrix whose main diagonal is given in the bracket
$\mathrm{E}\{\cdot\}$	expectation
$\mathbf{I}$	identity matrix of appropriate dimension
$\mathbf{I}_M$	$M \times M$ identity matrix
$\Im\{\cdot\}$	imaginary part of its input variable
$\jmath$	imaginary unit
$\max\{a, b\}$	maximum value between a and b
$\min\{a, b\}$	minimum value between a and b
$\mathcal{P}\{\cdot\}$	principal eigenvector
$\mathbb{R}$	set of real numbers
$\mathrm{rank}(\cdot)$	rank of a matrix defined as $\mathrm{rank}(\mathbf{Z}) \triangleq \mathrm{card}(\{\sigma_i(\mathbf{Z})\})$
$\Re\{\cdot\}$	real part of its input variable
$\mathrm{sign}(\cdot)$	sign function
s.t.	subject to
$\mathrm{trace}\{\cdot\}$	matrix trace
$\mathrm{vec}\{\cdot\}$	vectorization operator
$\mathbf{V}_{i,:}$	the ith row of $\mathbf{V}$
$\mathbf{X} \succeq 0$	$\mathbf{X}$ is positive semidefinite
$\mathbf{X} \succeq \mathbf{Y}$	$\mathbf{X} - \mathbf{Y}$ is positive semidefinite
$\lambda_{\max}(\cdot)$	the largest eigenvalue
$\lambda_{\min}(\cdot)$	the smallest eigenvalue
$\sigma_i(\cdot)$	the i-th largest singular value
$.^{-1}$	matrix inverse
$.^{\mathrm{H}}$	Hermitian transpose
$.^{\mathrm{T}}$	transpose
$.^{*}$	complex conjugate
$\diamond$	Khatri-Rao product

$\odot$	element-wise multiplication
$\oslash$	element-wise division
$\lvert \cdot \rvert$	element-wise absolute value
$\\|\cdot\\|_0$	ℓ_0-quasi-norm counting the non-zero entries of the input vector
$\\|\cdot\\|_1$	ℓ_1-norm of a vector
$\\|\cdot\\|_2$	ℓ_2-norm of a vector
$\\|\cdot\\|_{2,0}$	$\ell_{2,0}$ mixed-norm of a matrix defined as $\\|\mathbf{V}\\|_{2,0} \triangleq \mathrm{card}(\{\\|\mathbf{V}_{i,:}\\|_2\})$
$\\|\cdot\\|_{2,1}$	$\ell_{2,1}$ mixed-norm of a matrix defined as $\\|\mathbf{V}\\|_{2,1} \triangleq \sum_{i=1}^{M} \\|\mathbf{V}_{i,:}\\|_2$ with $\mathbf{V} \in \mathbb{C}^{M\times T}$
$\\|\cdot\\|_{\mathrm{F}}$	Frobenius norm of a matrix
$\\|\cdot\\|_*$	Nuclear norm (sum of singular values) of a matrix
$(\cdot)_+$	element-wise plus function
$\langle \mathbf{x}, \mathbf{y} \rangle$	inner product of $\mathbf{x}$ and $\mathbf{y}$
$\geq$	element-wise greater than or equal to
$\leq$	element-wise less than or equal to
$>$	element-wise greater than
$<$	element-wise less than
$\gg$	much larger than
$\ll$	much less than

List of Figures

List of Tables

References

[AA17] F. Afkhaminia and M. Azghani, ‘‘Sparsity-based direction of arrival estimation in the presence of gain/phase uncertainty,’’ in *Proceedings of European Signal Processing Conference (EUSIPCO)*, Kos, Greece, September 2017, pp. 2616--2619.

[AGM18] A. Abtahi, S. Gazor, and F. Marvasti, ‘‘Off-grid localization in MIMO radars using sparsity,’’ *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 313--317, February 2018.

[ATSMR+21] A. Arora, C. G. Tsinos, B. Shankar Mysore R, S. Chatzinotas, and B. Ottersten, ‘‘Analog beamforming with antenna selection for large-scale antenna arrays,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 4795--4799.

[BAPW08] C. R. Berger, J. Areta, K. Pattipati, and P. Willett, ‘‘Compressed sensing - A look beyond linear programming,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, USA, 2008, pp. 3857--3860.

[BBPB14] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, ‘‘Convergence and stability of iteratively re-weighted least squares algorithms,’’ *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183--195, January 2014.

[BIK+18] Y. Bando, K. Itoyama, M. Konyo, S. Tadokoro, K. Nakadai, K. Yoshii, T. Kawahara, and H. G. Okuno, ‘‘Speech enhancement based on Bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms,’’ *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 215--230, February 2018.

[BJZ+18] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, ‘‘On the applications of robust PCA in image and video processing,’’ *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1427--1457, August 2018.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, ‘‘Distributed optimization and statistical learning via the alternating direction method of multipliers,’’ *Foundation and Trends in Machine Learning*, vol. 3, no. 1, pp. 1--122, January 2011.

[BS14] P. Babu and P. Stoica, ‘‘Connection between SPICE and square-root LASSO for sparse parameter estimation,’’ *Signal Processing*, vol. 95, pp. 10--14, February 2014.

[BT09] A. Beck and M. Teboulle, ‘‘A fast iterative shrinkage-thresholding algorithm for linear inverse problems,’’ *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183--202, March 2009.

[BTR13] B. N. Bhaskar, G. Tang, and B. Recht, ‘‘Atomic norm denoising with applications to line spectral estimation,’’ *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987--5999, December 2013.

[BV04] S. Boyd and L. Vandenberghe, ‘‘Appendix B: Problems involving two quadratic functions,’’ in *Convex Optimization*. Cambridge university press, 2004, pp. 653--659.

[BZP02] R. Brcich, A. Zoubir, and P. Pelin, ‘‘Detection of sources using bootstrap techniques,’’ *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 206--215, February 2002.

[Bö86] J. F. Böhme, ‘‘Estimation of spectral parameters of correlated signals in wavefields,’’ *Signal Processing*, vol. 11, no. 4, pp. 329--337, December 1986.

[Cap69] J. Capon, ‘‘High-resolution frequency-wavenumber spectrum analysis,’’ *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408--1418, August 1969.

[CCS10] J.-F. Cai, E. J. Candès, and Z. Shen, ‘‘A singular value thresholding algorithm for matrix completion,’’ *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956--1982, March 2010.

[CHL21] Z. Cheng, Z. He, and B. Liao, ‘‘Hybrid beamforming design for OFDM dual-function radar-communication system,’’ *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1455--1467, November 2021.

[CL22] Z. Cheng and B. Liao, ‘‘QoS-aware hybrid beamforming and DOA estimation in multi-carrier dual-function radar-communication systems,’’ *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1890--1905, March 2022.

[CP11] E. J. Candes and Y. Plan, ‘‘A probabilistic and RIPless theory of compressed sensing,’’ *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235--7254, November 2011.

[CPP13] Y. Cheng, M. Pesavento, and A. Philipp, ‘‘Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming,’’ *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3972--3987, August 2013.

[CT17] E. Chen and M. Tao, ‘‘ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems,’’ *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2685--2698, June 2017.

[CVY14] P.-J. Chung, M. Viberg, and J. Yu, ‘‘Chapter 14 - DOA estimation methods and algorithms,’’ in *Academic Press Library in Signal Processing, Volume 3*, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds. Elsevier, 2014, pp. 599--650.

[CWB08] E. J. Candès, M. B. Wakin, and S. Boyd, ''Enhancing sparsity by reweighted ℓ_1 minimization,'' *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877--905, December 2008.

[CZW+21] Z. Chen, X. Zhang, S. Wang, Y. Xu, J. Xiong, and X. Wang, ''Enabling practical large-scale MIMO in WLANs with hybrid beamforming,'' *IEEE/ACM Transactions on Networking*, vol. 29, no. 4, pp. 1605--1619, August 2021.

[Das17] A. Das, ''A Bayesian sparse-plus-low-rank matrix decomposition method for direction-of-arrival tracking,'' *IEEE Sensors Journal*, vol. 17, no. 15, pp. 4894--4902, August 2017.

[DB13] M. F. Duarte and R. G. Baraniuk, ''Spectral compressive sensing,'' *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111--129, July 2013.

[DBXC17] J. Dai, X. Bao, W. Xu, and C. Chang, ''Root sparse Bayesian learning for off-grid DOA estimation,'' *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 46--50, January 2017.

[DDFG09] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, ''Iteratively reweighted least squares minimization for sparse recovery,'' *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1--38, October 2009.

[Del14] J. P. Delmas, ''Chapter 16 - Performance bounds and statistical analysis of DOA estimation,'' in *Academic Press Library in Signal Processing: Volume 3*, ser. Academic Press Library in Signal Processing, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds. Elsevier, 2014, vol. 3, pp. 719--764.

[DGB22] J. Dan, S. Geirnaert, and A. Bertrand, ''Grouped variable selection for generalized eigenvalue problems,'' *Signal Processing*, vol. 195, p. 108476, June 2022.

[DRZA10] C. Debes, J. Riedler, A. M. Zoubir, and M. G. Amin, ''Adaptive target detection with application to through-the-wall radar imaging,'' *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5572--5583, November 2010.

[DT15] O. T. Demir and T. E. Tuncer, ''Multicast beamforming with antenna selection using exact penalty approach,'' in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Australia, April 2015, pp. 2489--2493.

[DXP21] K. Diamantaras, Z. Xu, and A. Petropulu, ''Sparse antenna array design for MIMO radar using softmax selection,'' *arXiv*, February 2021. [Online]. Available: https://arxiv.org/abs/2102.05092

[EKB10] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, ''Block-sparse signals: Uncertainty relations and efficient recovery,'' *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042--3054, June 2010.

[EM09] Y. C. Eldar and M. Mishali, ''Robust recovery of signals from a structured union of subspaces,'' *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302--5316, November 2009.

[EM20] A. M. Elbir and K. V. Mishra, ''Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks,'' *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1677--1688, March 2020.

[EV19] A. Ene and A. Vladu, ''Improved convergence for ℓ_1 and ℓ_∞ regression via iteratively reweighted least squares,'' in *Proceedings of International Conference on Machine Learning (ICML)*, California, USA, June 2019, pp. 1794--1801.

[FGG$^+$14] S. Fortunati, R. Grasso, F. Gini, M. S. Greco, and K. LePage, ''Single-snapshot DOA estimation by using compressed sensing,'' *EURASIP Journal on Advances in Signal Processing*, vol. 120, pp. 1--17, July 2014.

[FHM$^+$18] T. Fischer, G. Hegde, F. Matter, M. Pesavento, M. E. Pfetsch, and A. M. Tillmann, ''Joint antenna selection and phase-only beamforming using mixed-integer nonlinear programming,'' in *Proceedings of International ITG Workshop on Smart Antennas (WSA)*, Bochum, Germany, March 2018, pp. 1--7.

[FR13a] S. Foucart and H. Rauhut, ''Chapter 1 - An invitation to compressive sensing,'' in *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013, pp. 1--39.

[FR13b] ------, ''Chapter 2 - Sparse solutions of underdetermined systems,'' in *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013, pp. 41--59.

[Fuc01] J.-J. Fuchs, ''On the application of the global matched filter to DOA estimation with uniform circular arrays,'' *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 702--709, Arpil 2001.

[Gab76] W. F. Gabriel, ''Adaptive arrays-An introduction,'' *Proceedings of the IEEE*, vol. 64, no. 2, pp. 239--272, February 1976.

[GB14] M. Grant and S. Boyd, ''CVX: Matlab software for disciplined convex programming, version 2.1,'' March 2014. [Online]. Available: http://cvxr.com/cvx

[GBZ12] C. Guyon, T. Bouwmans, and E.-H. Zahzah, ''Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression,'' in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, November 2012, pp. 2805--2808.

[GK14] M. Gkizeli and G. N. Karystinos, ‘‘Maximum-SNR antenna selection among a large number of transmit antennas,’’ *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 891--901, October 2014.

[HA19] S. A. Hamza and M. G. Amin, ‘‘Hybrid sparse array beamforming design for general rank signal models,’’ *IEEE Transactions on Signal Processing*, vol. 67, no. 24, pp. 6215--6226, December 2019.

[HA21] ------, ‘‘Sparse array beamforming design for wideband signal models,’’ *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 2, pp. 1211--1226, April 2021.

[HFZ19] H. Huang, M. Fauß, and A. M. Zoubir, ‘‘Block sparsity-based DOA estimation with sensor gain and phase uncertainties,’’ in *Proceedings of European Signal Processing Conference (EUSIPCO)*, A Coruna, Spain, September 2019, pp. 1--5.

[HL14] M. B. Hawes and W. Liu, ‘‘Sparse array design for wideband beamforming with reduced complexity in tapped delay-lines,’’ *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 8, pp. 1236--1247, August 2014.

[HLR16] M. Hong, Z.-Q. Luo, and M. Razaviyayn, ‘‘Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems,’’ *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337--364, 2016.

[HLSZ22] H. Huang, Q. Liu, H. C. So, and A. M. Zoubir, ‘‘Low-rank and row-sparse decomposition for joint DOA estimation and distorted sensor detection,’’ *arXiv*, August 2022. [Online]. Available: https://arxiv.org/abs/2202.01140

[HM10] M. M. Hyder and K. Mahata, ‘‘Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation,’’ *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646--4655, September 2010.

[HMP18] G. Hegde, C. Masouros, and M. Pesavento, ‘‘Analog beamformer design for interference exploitation based hybrid beamforming,’’ in *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield, UK, July 2018, pp. 109--113.

[Hol92] J. H. Holland, ‘‘Reproductive plans and genetic operators,’’ in *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* MIT Press, 1992, pp. 89--120.

[HS16] K. Huang and N. D. Sidiropoulos, ‘‘Consensus-ADMM for general quadratically constrained quadratic programming,’’ *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5297--5310, October 2016.

[HSZ22a] H. Huang, H. C. So, and A. M. Zoubir, ‘‘Off-grid direction-of-arrival estimation using second-order Taylor approximation,’’ *Signal Processing*, vol. 196, p. 108513, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168422000603

[HSZ22b] H. Huang, H. C. So, and A. M. Zoubir, ‘‘Sparse array beamformer design via ADMM,’’ in *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Trondheim, Norway, June 2022, pp. 336--340.

[HSZ22c] H. Huang, H. C. So, and A. M. Zoubir, ‘‘Sparse array beamformer design via ADMM,’’ *arXiv*, August 2022. [Online]. Available: https://arxiv.org/abs/2208.12313

[HZ21] H. Huang and A. M. Zoubir, ‘‘Low-rank and sparse decomposition for joint DOA estimation and contaminated sensors detection with sparsely contaminated arrays,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 4615--4619.

[HZR12] P. Heidenreich, A. M. Zoubir, and M. Rubsamen, ‘‘Joint 2-D DOA estimation and phase calibration for uniform rectangular arrays,’’ *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4683--4693, September 2012.

[JDC+13] J. Jiang, F. Duan, J. Chen, Z. Chao, Z. Chang, and X. Hua, ‘‘Two new estimation algorithms for sensor gain and phase errors based on different data models,’’ *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1921--1930, May 2013.

[JH13] R. Jagannath and K. Hari, ‘‘Block sparse estimator for grid matching in single snapshot DoA estimation,’’ *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1038--1041, November 2013.

[Jol86] I. T. Jolliffe, ‘‘Outlier detection, influential observations and robust estimation of principal components,’’ in *Principal Component Analysis*. Springer New York, 1986, pp. 173--198.

[JYMZS21] D. Jin, F. Yin, A. M. Zoubir, and H. C. So, ‘‘Exploiting sparsity of ranging biases for NLOS mitigation,’’ *IEEE Transactions on Signal Processing*, vol. 69, pp. 3782--3795, June 2021.

[KV96] H. Krim and M. Viberg, ‘‘Two decades of array signal processing research: The parametric approach,’’ *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67--94, July 1996.

[KVS21] C. Kümmerle, C. M. Verdun, and D. Stöger, ‘‘Iteratively reweighted least squares for basis pursuit with global linear convergence rate,’’ in *Proceedings of Conference on Neural Information Processing Systems (NeurIPS)*, Virtual Conference, December 2021, pp. 1--14.

[LAAZ15] M. Leigsnering, F. Ahmad, M. G. Amin, and A. M. Zoubir, ‘‘Compressive sensing-based multipath exploitation for stationary and moving indoor target localization,’’ *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1469--1483, December 2015.

[LC12] B. Liao and S. C. Chan, ‘‘Direction finding with partly calibrated uniform linear arrays,’’ *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 922--929, February 2012.

[LC14] ------, ‘‘A review on direction finding in partly calibrated arrays,’’ in *Proceedings of International Conference on Digital Signal Processing (DSP)*, Hong Kong, China, August 2014, pp. 812--816.

[LCM10] Z. Lin, M. Chen, and Y. Ma, ‘‘The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices,’’ 2010. [Online]. Available: https://arxiv.org/abs/1009.5055

[LGS19] Q. Liu, Y. Gu, and H. C. So, ‘‘DOA estimation in impulsive noise via low-rank matrix approximation and weakly convex optimization,’’ *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3603--3616, December 2019.

[Lin82] H.-C. Lin, ‘‘Spatial correlations in adaptive arrays,’’ *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 2, pp. 212--223, March 1982.

[LLWL18] H. Li, Q. Liu, Z. Wang, and M. Li, ‘‘Transmit antenna selection and analog beamforming with low-resolution phase shifters in mmWave MISO systems,’’ *IEEE Communications Letters*, vol. 22, no. 9, pp. 1878--1881, September 2018.

[LLXZ15] B. Lin, J. Liu, M. Xie, and J. Zhu, ‘‘Direction-of-arrival tracking via low-rank plus sparse matrix decomposition,’’ *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1302--1305, February 2015.

[LLY10] G. Liu, Z. Lin, and Y. Yu, ‘‘Robust subspace segmentation by low-rank representation,’’ in *Proceedings of International Conference on Machine Learning (ICML)*, Madison, USA, June 2010, pp. 663--670.

[LLY+13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, ‘‘Robust recovery of subspace structures by low-rank representation,’’ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171--184, January 2013.

[LLY15] C. Lu, Z. Lin, and S. Yan, ‘‘Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization,’’ *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 646--654, February 2015.

[LML+18] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, ‘‘MU-MIMO communications with MIMO radar: From co-existence to joint transmission,’’ *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2755--2770, April 2018.

[LMS+10] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, ‘‘Semidefinite relaxation of quadratic optimization problems,’’ *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20--34, May 2010.

[LRR+18] X. Li, J. Ren, S. Rambhatla, Y. Xu, and J. Haupt, ‘‘Robust PCA via dictionary based outlier pursuit,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, April 2018, pp. 4699--4703.

[LRW19] G. C. F. Lee, A. S. Rawat, and G. W. Wornell, ‘‘Robust direction of arrival estimation in the presence of array faults using snapshot diversity,’’ in *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Ottawa, Canada, November 2019, pp. 1--5.

[LSG17] Q. Liu, H. C. So, and Y. Gu, ‘‘Off-grid DOA estimation with nonconvex regularization via joint sparse representation,’’ *Signal Processing*, vol. 140, pp. 171--176, November 2017.

[LSZN09] C.-H. Lim, S. C.-M. See, A. M. Zoubir, and B. P. Ng, ‘‘Robust adaptive trimming for high-resolution direction finding,’’ *IEEE Signal Processing Letters*, vol. 16, no. 7, pp. 580--583, July 2009.

[LV19] C.-L. Liu and P. P. Vaidyanathan, ‘‘Robustness of difference coarrays of sparse arrays to sensor failures-Part I: A theory motivated by coarray MUSIC,’’ *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3213--3226, June 2019.

[LVBL98] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, ‘‘Applications of second-order cone programming,’’ *Linear Algebra and Its Applications*, vol. 284, no. 1, pp. 193--228, November 1998.

[LW10] W. Liu and S. Weiss, *Wideband Beamforming: Concepts and Techniques*. Wiley Publishing, 2010.

[LZ13] Z. Lu and A. M. Zoubir, ‘‘Flexible detection criterion for source enumeration in array processing,’’ *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1303--1314, March 2013.

[LZ15] ------, ‘‘Source enumeration in array processing using a two-step test,’’ *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2718--2727, May 2015.

[MBZJ09] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, ‘‘A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm,’’ *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289--301, January 2009.

[MCR+12] M. Muma, Y. Cheng, F. Roemer, M. Haardt, and A. M. Zoubir, ‘‘Robust source number enumeration for R-dimensional arrays in case of brief sensor failures,’’ in *Proceedings of IEEE International Conference* *on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 3709--3712.

[MCW05] D. Malioutov, M. Cetin, and A. S. Willsky, ‘‘A sparse signal reconstruction perspective for source localization with sensor arrays,’’ *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010--3022, August 2005.

[MCW+21] Y. Ma, X. Cao, X. Wang, M. S. Greco, and F. Gini, ‘‘Multi-source off-grid DOA estimation with single snapshot using non-uniform linear arrays,’’ *Signal Processing*, vol. 189, p. 108238, 2021.

[MHG+15] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, ‘‘Feasible point pursuit and successive approximation of non-convex QCQPs,’’ *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 804--808, July 2015.

[MMKBZ+16] M. Malek-Mohammadi, A. Koochakzadeh, M. Babaie-Zadeh, M. Jansson, and C. R. Rojas, ‘‘Successive concave sparsity approximation for compressed sensing,’’ *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5657--5671, November 2016.

[MSG12] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, ‘‘Multicast beamforming with antenna selection,’’ in *Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cesme, Turkey, September 2012, pp. 70--74.

[MSG13] ------, ‘‘Joint multicast beamforming and antenna selection,’’ *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2660--2674, May 2013.

[MTPK14] P. P. Markopoulos, N. Tsagkarakis, D. A. Pados, and G. N. Karystinos, ‘‘Direction finding with L1-norm subspaces,’’ in *Compressive Sensing III*, F. Ahmad, Ed., vol. 9109, International Society for Optics and Photonics. SPIE, May 2014, pp. 130--140.

[MTPK19] P. P. Markopoulos, N. Tsagkarakis, D. A. Pados, and G. N. Karystinos, ‘‘Realified L1-PCA for direction-of-arrival estimation: Theory and algorithms,’’ *EURASIP Journal on Advances in Signal Processing*, vol. 30, pp. 1--16, June 2019.

[NLEF09] B. Ng, J. P. Lie, M. Er, and A. Feng, ‘‘A practical simple geometry and gain/phase calibration technique for antenna array processing,’’ *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 7, pp. 1963--1972, July 2009.

[NSYC10] S. E. Nai, W. Ser, Z. L. Yu, and H. Chen, ‘‘Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization,’’ *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 12, pp. 3923--3930, 2010.

[ORM12] G. Oliveri, P. Rocca, and A. Massa, ‘‘Reliable diagnosis of large linear arrays-A Bayesian compressive sensing approach,’’ *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 10, pp. 4627--4636, October 2012.

[Pan10] E. Pancera, ‘‘Medical applications of the ultra wideband technology,’’ in *Proceedings of Loughborough Antennas & Propagation Conference*, Loughborough, UK, 2010, pp. 52--56.

[PB17] J. Park and S. Boyd, ‘‘General heuristics for nonconvex quadratically constrained quadratic programming,’’ *arXiv*, May 2017. [Online]. Available: https://arxiv.org/abs/1703.07870

[PGW02] M. Pesavento, A. B. Gershman, and K. M. Wong, ‘‘Direction finding in partly calibrated sensor arrays composed of multiple subarrays,’’ *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2103--2115, September 2002.

[PV10] P. Pal and P. P. Vaidyanathan, ‘‘Nested arrays: A novel approach to array processing with enhanced degrees of freedom,’’ *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167--4181, August 2010.

[PZBL07] D. S. Pham, A. M. Zoubir, R. F. Brcic, and Y. H. Leung, ‘‘A nonlinear M-estimation approach to robust asynchronous multiuser detection in non-gaussian noise,’’ *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1624--1633, May 2007.

[QZA15] S. Qin, Y. D. Zhang, and M. G. Amin, ‘‘Generalized coprime array configurations for direction-of-arrival estimation,’’ *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377--1390, March 2015.

[RB16] E. K. Ryu and S. P. Boyd, ‘‘A primer on monotone operator methods,’’ *Applied and Computational Mathematics*, vol. 15, no. 1, pp. 3--43, 2016.

[RCC+18] K. N. Ramamohan, S. P. Chepuri, D. F. Comesaña, G. C. Pousa, and G. Leus, ‘‘Blind calibration for acoustic vector sensor arrays,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, April 2018, pp. 3544--3548.

[RHE12] M. Rossi, A. M. Haimovich, and Y. C. Eldar, ‘‘Spatial compressive sensing in MIMO radar with random arrays,’’ in *Proceedings of 46th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, USA, 2012, pp. 1--6.

[RK89] R. Roy and T. Kailath, ‘‘ESPRIT-estimation of signal parameters via rotational invariance techniques,’’ *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984--995, July 1989.

[RW14] P. Rodrêguez and B. Wohlberg, ‘‘Performance comparison of iterative reweighting methods for total variation regularization,’’ in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Paris, France, October 2014, pp. 1758--1762.

[SB12] P. Stoica and P. Babu, ‘‘SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation,’’ *Signal Processing*, vol. 92, no. 7, pp. 1580--1590, July 2012.

[SBL11a] P. Stoica, P. Babu, and J. Li, ‘‘New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data,’’ *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35--47, January 2011.

[SBL11b] ------, ‘‘SPICE: A sparse covariance-based estimation method for array processing,’’ *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629--638, February 2011.

[Sch86] R. Schmidt, ‘‘Multiple emitter location and signal parameter estimation,’’ *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276--280, March 1986.

[SCL+21] E. Soubies, A. Chinatto, P. Larzabal, J. M. T. Romano, and L. Blanc-Féraud, ‘‘Direction-of-arrival estimation through exact continuous $\ell_{2,0}$-norm relaxation,’’ *IEEE Signal Processing Letters*, vol. 28, pp. 16--20, 2021.

[SDL06] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, ‘‘Transmit beamforming for physical-layer multicasting,’’ *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239--2251, June 2006.

[SFH22] S. Shrestha, X. Fu, and M. Hong, ‘‘Optimal solutions for joint beamforming and antenna selection: From branch and bound to machine learning,’’ *arXiv*, June 2022. [Online]. Available: https://arxiv.org/abs/2206.05576

[SG04] C. M. S. See and A. B. Gershman, ‘‘Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays,’’ *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 329--338, February 2004.

[SGLW03] S. Shahbazpanahi, A. Gershman, Z.-Q. Luo, and K. M. Wong, ‘‘Robust adaptive beamforming for general-rank signal models,’’ *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2257--2269, September 2003.

[SP18] C. Steffens and M. Pesavento, ‘‘Block- and rank-sparse recovery for direction finding in partly calibrated arrays,’’ *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 384--399, January 2018.

[SPP14] C. Steffens, P. Parvazi, and M. Pesavento, ‘‘Direction finding and array calibration based on sparse reconstruction in partly calibrated arrays,’’ in *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruna, Spain, June 2014, pp. 21--24.

[SPPZ18] W. Suleiman, P. Parvazi, M. Pesavento, and A. M. Zoubir, ‘‘Non-coherent direction-of-arrival estimation using partly calibrated arrays,’’ *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5776--5788, November 2018.

[SRS+16] J. Steinwandt, F. Roemer, C. Steffens, M. Haardt, and M. Pesavento, ‘‘Gridless super-resolution direction finding for strictly non-circular sources based on atomic norm minimization,’’ in *Proceedings of 50th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, USA, November 2016, pp. 1518--1522.

[SSSP17] C. Steffens, W. Suleiman, A. Sorg, and M. Pesavento, ‘‘Gridless compressed sensing under shift-invariant sampling,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, March 2017, pp. 4735--4739.

[SV21] D. Straszak and N. K. Vishnoi, ‘‘Iteratively reweighted least squares and slime mold dynamics: Connection and convergence,’’ *Mathematical Programming*, pp. 509--515, April 2021.

[SVWW96] P. Stoica, M. Viberg, K. M. Wong, and Q. Wu, ‘‘Maximum-likelihood bearing estimation with partly calibrated arrays in spatially correlated noise fields,’’ *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 888--899, April 1996.

[SZL14a] Y. Shi, J. Zhang, and K. B. Letaief, ‘‘Group sparse beamforming for green Cloud-RAN,’’ *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809--2823, May 2014.

[SZL14b] P. Stoica, D. Zachariah, and J. Li, ‘‘Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation,’’ *Digital Signal Processing*, vol. 33, pp. 1--12, October 2014.

[TBSR13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, ‘‘Compressed sensing off the grid,’’ *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465--7490, November 2013.

[Tib96] R. Tibshirani, ‘‘Regression shrinkage and selection via the LASSO,’’ *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267--288, 1996.

[TMZ21] A. Taştan, M. Muma, and A. M. Zoubir, ‘‘Sparsity-aware robust community detection (SPARCODE),’’ *Signal Processing*, vol. 187, p. 108147, October 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168421001857

[TP14] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248--1259, February 2014.

[TYN14] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4997--5008, October 2014.

[VBJN18] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32--55, July 2018.

[VCB18] N. Vaswani, Y. Chi, and T. Bouwmans, "Rethinking PCA for modern data sets: Theory, algorithms, and applications," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1274--1276, August 2018.

[VCN+21] T. X. Vu, S. Chatzinotas, V.-D. Nguyen, D. T. Hoang, D. N. Nguyen, M. D. Renzo, and B. Ottersten, "Machine learning-enabled joint antenna selection and precoding design: From offline complexity to online performance," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3710--3722, June 2021.

[VGLM03] S. Vorobyov, A. B. Gershman, Z.-Q. Luo, and N. Ma, "Adaptive beamforming with joint robustness against signal steering vector errors and interference nonstationarity," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, Hong Kong, China, April 2003, pp. 345--348.

[Vib14] M. Viberg, "Chapter 11 - Introduction to array processing," in *Academic Press Library in Signal Processing: Volume 3*, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds. Elsevier, 2014, vol. 3, pp. 463--502.

[Vor14] S. A. Vorobyov, "Chapter 12 - Adaptive and robust beamforming," in *Academic Press Library in Signal Processing: Volume 3*, ser. Academic Press Library in Signal Processing, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds. Elsevier, 2014, vol. 3, pp. 503--552.

[VP11] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573--586, February 2011.

[VSS07] S. Vigneshwaran, N. Sundararajan, and P. Saratchandran, "Direction of arrival (DoA) estimation under array sensor failures using a minimal resource allocation neural network," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 2, pp. 334--343, February 2007.

[VT02] H. L. Van Trees, "Chapter 1 - Introduction," in *Optimum Array Processing.* John Wiley & Sons, Ltd, 2002, pp. 1--16.

[WATA14] X. Wang, E. Aboutanios, M. Trinkle, and M. G. Amin, "Reconfigurable adaptive array beamforming by antenna selection," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2385--2396, May 2014.

[WDC+17] Q. Wang, T. Dou, H. Chen, W. Yan, and W. Liu, "Effective block sparse representation algorithm for DOA estimation with unknown mutual coupling," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2622--2625, December 2017.

[WGG21] X. Wang, M. S. Greco, and F. Gini, "Adaptive sparse array beamformer design by regularized complementary antenna switching," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2302--2315, March 2021.

[WGP19] M. Wagner, P. Gerstoft, and Y. Park, "Gridless DOA estimation via. alternating projections," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019, pp. 4215--4219.

[WGW19] B. Wang, Y. Gu, and W. Wang, "Off-grid direction-of-arrival estimation based on steering vector approximation," *Circuits, Systems, and Signal Processing*, vol. 38, no. 3, pp. 1287--1300, March 2019.

[WL21] Z. Wan and W. Liu, "Non-coherent DOA estimation of off-grid signals with uniform circular arrays," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 4370--4374.

[WPG21] M. Wagner, Y. Park, and P. Gerstoft, "Gridless DOA estimation and root-MUSIC for non-uniform linear arrays," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2144--2157, March 2021.

[WPM+09] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization," in *Proceedings of International Conference on Neural Information Processing Systems (NIPS)*, Red Hook, USA, December 2009, pp. 2080--2088.

[WSP17] A. C. Walewski, C. Steffens, and M. Pesavento, "Off-grid parameter estimation based on joint sparse regularization," in *Proceedings of International ITG Conference on Systems, Communications and Coding (SCC)*, Hamburg, Germany, February 2017, pp. 1--6.

[WWS21] T. Wei, L. Wu, and B. M. R. Shankar, "Sparse array beampattern synthesis via majorization-based ADMM," in *Proceedings of IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Norman, USA, September 2021, pp. 1--5.

[WZ14] C. Weiss and A. M. Zoubir, ‘‘Robust high-resolution DOA estimation with array pre-calibration,’’ in *Proceedings of European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September 2014, pp. 1049--1052.

[WZCN18] Q. Wang, Z. Zhao, Z. Chen, and Z. Nie, ‘‘Grid evolution method for DOA estimation,’’ *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2374--2383, May 2018.

[WZN17] M. Wang, Z. Zhang, and A. Nehorai, ‘‘Direction finding using sparse linear arrays with missing data,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, March 2017, pp. 3066--3070.

[WZW17] B. Wang, Y. D. Zhang, and W. Wang, ‘‘Robust DOA estimation in the presence of miscalibrated sensors,’’ *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1073--1077, July 2017.

[WZYZ18] X. Wu, W.-P. Zhu, J. Yan, and Z. Zhang, ‘‘Two sparse-based methods for off-grid direction-of-arrival estimation,’’ *Signal Processing*, vol. 142, pp. 87--95, January 2018.

[XCS12] H. Xu, C. Caramanis, and S. Sanghavi, ‘‘Robust PCA via outlier pursuit,’’ *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3047--3064, May 2012.

[XLD+21] Z. Xu, F. Liu, K. Diamantaras, C. Masouros, and A. Petropulu, ‘‘Learning to select for MIMO radar based on hybrid analog-digital beamforming,’’ in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 8228--8232.

[YdL17] Z. Yang, R. C. de Lamare, and W. Liu, ‘‘Sparsity-based STAP using alternating direction method with gain/phase errors,’’ *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 6, pp. 2756--2768, December 2017.

[YDZ12] F. Yin, C. Debes, and A. M. Zoubir, ‘‘Parametric waveform design using discrete prolate spheroidal sequences for enhanced detection of extended targets,’’ *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4525--4536, September 2012.

[YG04] L. Yang and G. Giannakis, ‘‘Ultra-wideband communications: An idea whose time has come,’’ *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 26--54, November 2004.

[YL99] B.-K. Yeo and Y. Lu, ‘‘Array failure correction with a genetic algorithm,’’ *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 5, pp. 823--828, May 1999.

[YLSX18] Z. Yang, J. Li, P. Stoica, and L. Xie, “Chapter 11 - Sparse methods for direction-of-arrival estimation,” in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, pp. 509--581.

[YQF+19] H. Yu, W. Qu, Y. Fu, C. Jiang, and Y. Zhao, “A novel two-stage beam selection algorithm in mmWave hybrid beamforming system,” *IEEE Communications Letters*, vol. 23, no. 6, pp. 1089--1092, June 2019.

[YX15] Z. Yang and L. Xie, “On gridless sparse methods for line spectral estimation from complete and incomplete data,” *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139--3153, June 2015.

[YXZ13] Z. Yang, L. Xie, and C. Zhang, “Off-grid direction of arrival estimation using sparse Bayesian inference,” *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38--43, January 2013.

[YZX12] Z. Yang, C. Zhang, and L. Xie, “Robustly stable signal recovery in compressed sensing with structured matrix perturbation,” *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658--4671, September 2012.

[ZCS+17] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, “Joint transceiver design with antenna selection for large-scale MU-MIMO mmWave systems,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2085--2096, 2017.

[ZFW21] Z. Zheng, Y. Fu, and W.-Q. Wang, “Sparse array beamforming design for coherently distributed sources,” *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 5, pp. 2628--2636, May 2021.

[ZFWS20] Z. Zheng, Y. Fu, W.-Q. Wang, and H. C. So, “Sparse array design for adaptive beamforming via semidefinite relaxation,” *IEEE Signal Processing Letters*, vol. 27, pp. 925--929, May 2020.

[ZGSZ18] C. Zhou, Y. Gu, Z. Shi, and Y. D. Zhang, “Off-grid direction-of-arrival estimation using coprime array interpolation,” *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1710--1714, November 2018.

[ZKOM18] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, “Chapter 3 - Robust penalized regression in the linear model,” in *Robust Statistics for Signal Processing*. Cambridge University Press, 2018, pp. 74--99.

[ZLG11] H. Zhu, G. Leus, and G. B. Giannakis, “Sparsity-cognizant total least-squares for perturbed compressive sampling,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002--2016, May 2011.

[ZLT+11] C. Zhang, J. Liu, Q. Tian, C. Xu, H. Lu, and S. Ma, “Image classification by non-negative sparse coding, low-rank and sparse decomposition,” in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Spring, USA, August 2011, pp. 1673--1680.

[ZLZC15] H. Zhang, Z. Lin, C. Zhang, and E. Chang, ''Exact recoverability of robust PCA via outlier pursuit with tight recovery bounds,'' in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, Austin Texas, USA, February 2015.

[ZRH21] J. Zhang, D. Rakhimov, and M. Haardt, ''Gridless channel estimation for hybrid mmWave MIMO systems via tensor-ESPRIT algorithms in DFT beamspace,'' *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 816--831, April 2021.

[ZW88] I. Ziskind and M. Wax, ''Maximum likelihood localization of multiple sources by alternating projection,'' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553--1560, October 1988.

[ZWCS15] C. Zhu, W.-Q. Wang, H. Chen, and H. C. So, ''Impaired sensor diagnosis, beamforming, and DOA estimation with difference co-array processing,'' *IEEE Sensors Journal*, vol. 15, no. 7, pp. 3773--3780, July 2015.

Erklärung laut Promotionsordnung

§8 Abs. 1 lit. c PromO Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§8 Abs. 1 lit. d PromO Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§9 Abs. 1 PromO Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§9 Abs. 2 PromO Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 14 October 2022

M.Sc. Huiping Huang

Curriculum Vitae

Name:	Huiping Huang
Date of birth:	04.10.1991
Place of birth:	Guangdong, China

Education

09/2015 - 07/2018	Shenzhen University, China Electronics and Communication Engineering Master of Science (M.Sc.)
09/2011 - 07/2015	Shenzhen University, China Electronic and Information Engineering Bachelor of Science (B.Sc.)
09/2008 - 06/2011	Shantou Jinshan Senior High School, China High School Degree

Work Experiences

12/2018 - Present	Research Associate Signal Processing Group Technische Universität Darmstadt, Germany
07/2022 - 08/2022	Visiting Ph.D. Student Signal Processing Applications in Radar and Communications Group University of Luxembourg, Luxembourg
07/2018 - 11/2018	Research Assistant Department of Electronic and Information Engineering Southern University of Science and Technology, China