Low-Rank and Row-Sparse Decomposition for Joint DOA Estimation and Distorted Sensor Detection

Huiping Huang, Student Member, IEEE
Darmstadt University of Technology, Germany

Qi Liu, Member, IEEE
South China University of Technology, China;
Pazhou Lab, Guangzhou 510330, China

Hing C. So, Fellow, IEEE
City University of Hong Kong, China

Abdelhak M. Zoubir, Fellow, IEEE
Darmstadt University of Technology, Germany

Abstract—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.

Index Terms—Alternating direction method of multipliers, distorted sensor, DOA estimation, iteratively reweighted least squares, low-rank and row-sparse decomposition

I. INTRODUCTION

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 [1]–[3]. Many classical approaches have been proposed, including multiple signal classification (MUSIC) [4], estimation of signal parameters via rotational invariance techniques (ESPRIT) [5], and maximum likelihood methods [6], [7]. 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 [8]–[13], such as unknown sensor gain and phase uncertainties, which is the focus of this work. More recently, techniques based on low-rank and sparse matrix decomposition have been applied to DOA estimation or tracking, see e.g. [14]–[17]. 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 [18]–[36]. In [18], the genetic algorithm [37] was applied for array failure correction. A minimal resource allocation network was used for DOA estimation under array sensor failure [19], which requires a training procedure with no failed sensors. A Bayesian compressive sensing approach was proposed in [21], which needs a noise-free array as a reference. Methods using difference co-array were developed in [22]–[24]. The idea of [22] 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, [23] and [24] 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 [28]–[33]. 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) [38]–[40]. 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 [41]–[45]. Another special structure

arXiv:2202.01140v3 [eess.SP] 25 Aug 2022

of LRSD is low-rank and column-sparse decomposition (LRCSD) [46]–[50], also known as RPCA-outlier pursuit [51]–[54], 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 [48], [50].

Several algorithms have been contributed to solve the LRSD and LRCSD problems, such as singular value thresholding (SVT) [55], accelerated proximal gradient (APG) [56], alternating direction method of multipliers (ADMM) [44], [50], and iteratively reweighted least squares (IRLS) [49], [50], [53], [54]. The SVT, APG, and ADMM methods will be reviewed in Section III 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 [57]–[61]. In this sense, the IRLS is more efficient in solving the corresponding problems.

Therefore, in the present work, 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.

The remainder of the paper is organized as follows. The signal model and problem statement are established in Section II. A review of state-of-the-art works is provided in Section III. Section IV derives an IRLS algorithm for joint DOA estimation and distorted sensor detection. Numerical results are given in Section V, while Section VI concludes this paper.

Notation: In this paper, 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

Fig. 1. Illustration of array structure of interest.

$\mathbf{V}_{i,:}$ is the ith row of $\mathbf{V}$. $\mathrm{rank}(\cdot)$ is the rank operator, defined as $\mathrm{rank}(\mathbf{Z}) \triangleq \mathrm{card}(\{\sigma_i(\mathbf{Z})\})$, with $\sigma_i(\mathbf{Z})$ being the ith 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.

II. 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. Fig. 1 illustrates the array model, where the black circles stand for *perfect* sensors and the boxes refer to distorted ones. The 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 θ_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 mth sensor, respectively.

Collecting all the snapshots into a matrix, we have

$$\mathbf{Y} = (\mathbf{I} + \boldsymbol{\Gamma})\mathbf{A}\mathbf{S} + \mathbf{N}, \tag{1}$$

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 $\mathbf{V} \triangleq \boldsymbol{\Gamma}\mathbf{A}\mathbf{S}$, (1) becomes:

$$\mathbf{Y} = \mathbf{Z} + \mathbf{V} + \mathbf{N}, \tag{2}$$

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 $\boldsymbol{\Gamma}$.

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.

III. 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 [55]. By adapting the SVT algorithm to our problem, we need to solve

$$\min_{\mathbf{Z},\mathbf{V},\mathbf{W}} \ \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1} + \frac{1}{2\tau}\|\mathbf{Z}\|_F^2$$
$$+ \frac{1}{2\tau}\|\mathbf{V}\|_F^2 + \frac{1}{\tau}\langle\mathbf{W},\mathbf{Y}-\mathbf{Z}-\mathbf{V}\rangle, \quad (3)$$

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 $\mathbf{V}$ [55]:

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

where $\mathbf{LSR}^H$ 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 [44].

The second method is APG, whose updating equation can be given as [56]

$$(\mathbf{Z}_{k+1}, \mathbf{V}_{k+1}) = \arg\min_{\mathbf{Z},\mathbf{V}} \ h(\mathbf{Z},\mathbf{V}) \quad (4)$$

where subscript $\cdot_k$ denotes the variable in the kth iteration, $h(\mathbf{Z},\mathbf{V}) \triangleq p(\mathbf{Z}_k,\mathbf{V}_k) + \langle\nabla_{\mathbf{Z}_k}p(\mathbf{Z},\mathbf{V}_k),\mathbf{Z}-\mathbf{Z}_k\rangle + \langle\nabla_{\mathbf{V}_k}p(\mathbf{Z}_k,\mathbf{V}),\mathbf{V}-\mathbf{V}_k\rangle + \mu M\|\mathbf{Z}+\mathbf{V}-\mathbf{Z}_k-\mathbf{V}_k\|_F^2 + q(\mathbf{Z},\mathbf{V})$, with $p(\mathbf{Z},\mathbf{V}) \triangleq \frac{1}{\mu}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_F^2$, $q(\mathbf{Z},\mathbf{V}) \triangleq \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1}$, and μ being a small positive scalar. The detailed algorithm can be found in [56] and also [44].

As for ADMM, we consider the following problem

$$\min_{\mathbf{Z},\mathbf{V}} \ \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1} \quad \text{s.t. } \mathbf{Y} = \mathbf{Z} + \mathbf{V}, \quad (5)$$

and its augmented Lagrangian function is $\mathcal{L}_\mu(\mathbf{Z},\mathbf{V},\mathbf{W}) = \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1} + \langle\mathbf{W},\mathbf{Y}-\mathbf{Z}-\mathbf{V}\rangle + \frac{\mu}{2}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_F^2$, 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 $\mathbf{Z}$ (resp. $\mathbf{V}$) while keeping $\mathbf{V}$ (resp. $\mathbf{Z}$) and $\mathbf{W}$ unchanged; $\mathbf{W}$ is updated as $\mathbf{W} = \mathbf{W} + \mu(\mathbf{Y} - \mathbf{Z} - \mathbf{V})$ [62].

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 Sections V.B and V.C, respectively.

IV. 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.

A. NOISELESS CASE

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

$$\min_{\mathbf{Z},\mathbf{V}} \ \text{rank}(\mathbf{Z}) + \lambda\|\mathbf{V}\|_{2,0} \quad \text{s.t. } \mathbf{Y} = \mathbf{Z} + \mathbf{V}, \quad (6)$$

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

$$\min_{\mathbf{Z}} \ \|\mathbf{Z}\|_* + \lambda\|\mathbf{Y}-\mathbf{Z}\|_{2,1}. \quad (7)$$

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

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

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

$$\mathbf{Q} \triangleq \begin{bmatrix} \frac{1}{\sqrt{\|(\mathbf{Y}-\mathbf{Z})_{1,:}\|_2^2 + \mu^2}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\|(\mathbf{Y}-\mathbf{Z})_{M,:}\|_2^2 + \mu^2}} \end{bmatrix}. \quad (8)$$

The problem to be solved now turns to be

$$\min_{\mathbf{Z}} \ f(\mathbf{Z}) \triangleq \|[\mathbf{Z},\mu\mathbf{I}]\|_* + \lambda\|[\mathbf{Y}-\mathbf{Z},\mu\mathbf{1}]\|_{2,1}, \quad (9)$$

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

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

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

$$\mathbf{Z}_{k+1} = \lambda(\mathbf{P}_k + \lambda\mathbf{Q}_k)^{-1}\mathbf{Q}_k\mathbf{Y}, \quad (10)$$

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 1, where ϵ is a small scalar and $k_{\max}$ is a large scalar, used to terminate the algorithm.

Algorithm 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: $\mathbf{Z}_0 \leftarrow \mathbf{Z}_{\text{init}}$, $\mathbf{V}_0 \leftarrow \mathbf{V}_{\text{init}}$, $k \leftarrow 0$
  1: **while** not converged **do**
  2:     $k \leftarrow k + 1$
  3:     calculate $\mathbf{P}_k$ and $\mathbf{Q}_k$
  4:     update $\mathbf{Z}_k$ using $\mathbf{Z} = \lambda(\mathbf{P} + \lambda\mathbf{Q})^{-1}\mathbf{Q}\mathbf{Y}$
  5:     converged $\leftarrow k \geq k_{\max}$ or $\frac{|f(\mathbf{Z}_k) - f(\mathbf{Z}_{k-1})|}{|f(\mathbf{Z}_k)|} \leq \epsilon$
  6: **end while**
  7: $\widehat{\mathbf{Z}} \leftarrow \mathbf{Z}_k$, $\widehat{\mathbf{V}} \leftarrow \mathbf{Y} - \mathbf{Z}_k$

B. CONVERGENCE ANALYSIS FOR ALGORITHM 1

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 (9).

LEMMA 1 (Lemma 2 in [49]). *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 2. *For any matrices* $\mathbf{X}$ *and* $\mathbf{Y} \in \mathbb{C}^{M \times T}$*, we have* $\|\mathbf{Y}\|_{2,1} - \|\mathbf{X}\|_{2,1} \geq \frac{1}{2}\text{trace}\left(\mathbf{H}\left(\mathbf{Y}\mathbf{Y}^{\text{H}} - \mathbf{X}\mathbf{X}^{\text{H}}\right)\right)$*, where*

$$\mathbf{H} = \begin{bmatrix} \frac{1}{\|\mathbf{Y}_{1,:}\|_2} & & \\ & \ddots & \\ & & \frac{1}{\|\mathbf{Y}_{M,:}\|_2} \end{bmatrix}. \quad (11)$$

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,

$$\|\mathbf{Y}\|_{2,1} - \|\mathbf{X}\|_{2,1} = \sum_i^M \left[\sqrt{\|\mathbf{Y}_{i,:}\|_2^2} - \sqrt{\|\mathbf{X}_{i,:}\|_2^2}\right]$$

$$\geq \sum_i^M \left[\frac{1}{2\|\mathbf{Y}_{i,:}\|_2}\left(\|\mathbf{Y}_{i,:}\|_2^2 - \|\mathbf{X}_{i,:}\|_2^2\right)\right]$$

$$= \frac{1}{2}\text{trace}\left(\mathbf{H}\left(\mathbf{Y}\mathbf{Y}^{\text{H}} - \mathbf{X}\mathbf{X}^{\text{H}}\right)\right),$$

where $\mathbf{H}$ is given by (11). ∎

With Lemmata 1 and 2, we have the following theorem.

THEOREM 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 (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* $\lim_{k \to \infty} \|\mathbf{Z}_k - \mathbf{Z}_{k+1}\|_{\text{F}} = 0$.
Proof:
See Appendix VI.A. ∎

THEOREM 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 VI.B. ∎

THEOREM 3. *Any limit point of the sequence* $\{\mathbf{Z}_k\}$ *generated by (10) is a stationary point of Problem (9), and moreover, the stationary point is globally optimal.*
Proof:
See Appendix VI.C. ∎

C. NOISY CASE

In the noisy case, the data model is as (2), and the problem to be solved is given as

$$\min_{\mathbf{Z},\mathbf{V}} \frac{1}{2}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\text{F}}^2 + \lambda_1\|\mathbf{Z}\|_* + \lambda_2\|\mathbf{V}\|_{2,1}, \quad (12)$$

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 (12). Therefore, the problem to be addressed is transferred to

$$\min_{\mathbf{Z},\mathbf{V}} f(\mathbf{Z}, \mathbf{V}), \quad (13)$$

where the objective function is defined as $f(\mathbf{Z}, \mathbf{V}) \triangleq \frac{1}{2}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\text{F}}^2 + \lambda_1\|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda_2\|[\mathbf{V}, \mu\mathbf{1}]\|_{2,1}$. 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

$$\mathbf{Q} \triangleq \begin{bmatrix} \frac{1}{\sqrt{\|\mathbf{V}_{1,:}\|_2^2 + \mu^2}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\|\mathbf{V}_{M,:}\|_2^2 + \mu^2}} \end{bmatrix}. \quad (14)$$

Note that $\mathbf{Q}$ given in (8) is exactly the same as the one in (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

$$\begin{cases} \mathbf{Z}_{k+1} = (\mathbf{I} + \lambda_1\mathbf{P}_k)^{-1}(\mathbf{Y} - \mathbf{V}_k) \\ \mathbf{V}_{k+1} = (\mathbf{I} + \lambda_2\mathbf{Q}_k)^{-1}(\mathbf{Y} - \mathbf{Z}_{k+1}), \end{cases} \quad (15)$$

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 2.

D. CONVERGENCE ANALYSIS FOR ALGORITHM 2

In this part, the monotonicity and boundedness of the objective function $f(\mathbf{Z}, \mathbf{V})$ in (13) are proved in Theorems 4 and 5, respectively.

THEOREM 4. *The sequence* $\{(\mathbf{Z}_k, \mathbf{V}_k)\}$ *generated by (15) produces a non-increasing objective function defined in (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* $\{(\mathbf{Z}_k, \mathbf{V}_k)\}$

Algorithm 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:** $\mathbf{Z}_0 \leftarrow \mathbf{Z}_{\text{init}}$, $\mathbf{V}_0 \leftarrow \mathbf{V}_{\text{init}}$, $k \leftarrow 0$

1: **while** not converged **do**
2:     $k \leftarrow k + 1$
3:     calculate $\mathbf{P}_k$ and $\mathbf{Q}_k$
4:     update $\mathbf{Z}_k$ using $\mathbf{Z} = (\mathbf{I} + \lambda_1 \mathbf{P})^{-1}(\mathbf{Y} - \mathbf{V})$
5:     update $\mathbf{V}_k$ using $\mathbf{V} = (\mathbf{I} + \lambda_2 \mathbf{Q})^{-1}(\mathbf{Y} - \mathbf{Z})$
6:     converged $\leftarrow \begin{cases} k \geq k_{\max} \text{ or} \\ \frac{|f(\mathbf{Z}_k,\mathbf{V}_k) - f(\mathbf{Z}_{k-1},\mathbf{V}_{k-1})|}{|f(\mathbf{Z}_k,\mathbf{V}_k)|} \leq \epsilon \end{cases}$
7: **end while**
8: $\widehat{\mathbf{Z}} \leftarrow \mathbf{Z}_k$, $\widehat{\mathbf{V}} \leftarrow \mathbf{V}_k$

Algorithm 3 Detection of distorted sensors

> **Input** : $\widehat{\mathbf{V}} \in \mathbb{C}^{M \times T}$, h
> **Output:** M_{fail}
> calculate $\mathbf{v} = [\|\widehat{\mathbf{V}}_{1,:}\|_2, \|\widehat{\mathbf{V}}_{2,:}\|_2, \cdots, \|\widehat{\mathbf{V}}_{M,:}\|_2]^T$
> calculate $\tilde{\mathbf{v}} = \text{sort}(\mathbf{v}, \text{'ascend'})$
> calculate $d = \tilde{\mathbf{v}}(2) - \tilde{\mathbf{v}}(1)$ and assign $i_{\text{fail}} = M + 1$

1: **for** $i = 3, 4, \cdots, M$ **do**
2:     **if** $\tilde{\mathbf{v}}(i) - \tilde{\mathbf{v}}(i-1) \geq h$ **then**
3:       $i_{\text{fail}} = i$ and break the **for** loop
4:     **end if**
5: **end for**
6: $M_{\text{fail}} \leftarrow M - i_{\text{fail}} + 1$

---

is bounded, and $\lim_{k \to \infty} \|\mathbf{Z}_k - \mathbf{Z}_{k+1}\|_F = 0$ and $\lim_{k \to \infty} \|\mathbf{V}_k - \mathbf{V}_{k+1}\|_F = 0$.

Proof:
See Appendix VI.D. ∎

THEOREM 5. *The objective function $f(\mathbf{Z},\mathbf{V}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_F^2 + \lambda_1\|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda_2\|[\mathbf{V}, \mu\mathbf{1}]\|_{2,1}$ is bounded below by $|\mu|(\lambda_1\sqrt{M} + \lambda_2 M)$.*

Proof:
See Appendix VI.E. ∎

THEOREM 6. *Any limit point of the sequence $\{(\mathbf{Z}_k, \mathbf{V}_k)\}$ generated by (15) is a stationary point of Problem (13), and moreover, the stationary point is globally optimal.*

Proof:
See Appendix VI.F. ∎

REMARK 1. *The differences between our work and [49] are stated as follows.*

- *The problem formulation in [49] is column-sparse, while we have row-sparsity of $\mathbf{V}$. This leads to differences in matrix multiplication and matrix derivative.*
- *[49] 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 [49] involves a Sylvester equation and utilizes the Matlab command `lyap`. However, our method admits a closed-form formula, see (10) and (15).*
- *The proofs of convergence are not exactly the same. [49] 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.*

E. 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

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

The SVD of $\widehat{\mathbf{Z}}$ is $\widehat{\mathbf{Z}} = \mathbf{L\Sigma R}^H$, 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 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 $M_{\text{fail}} = M - i_{\text{fail}} + 1$.

V. SIMULATIONS

A. PARAMETER SELECTION

In this subsection, we discuss the problem of choosing appropriate values for μ, λ_1, and λ_2 in Problem (13) used in Algorithm 2. We set $\epsilon = 10^{-16}$, $k_{\max} = 1000$, and $\mathbf{Z}_{\text{init}} = \mathbf{V}_{\text{init}} = \mathbf{O}$ in Algorithm 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 kth signal in the qth 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.

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

Fig. 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 $\mu = 0.01$.

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 Fig. 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 Fig. 3.

We observe from Fig. 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 Fig. 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 2 in the following simulations.

B.   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 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 Fig. 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 I (upper) for SNR $= 0$

Fig. 4.   Objective function value versus number of iterations at SNR $= 0$ dB and $T = 100$ snapshots.

dB and $T = 100$ snapshots, and Table I (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.

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

In Fig. 5, we plot the averaged CPU time against the number of snapshots at $M = 10$ sensors (4 of which distorted), $K = 2$ sources, SNR $= 0$ dB, and

TABLE I

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 [55]	493.8653	0.1205	**5**
APG [56]	123.3227	0.8403	22
ADMM [44]	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 [55]	1965.5994	0.6166	25
APG [56]	282.5776	4.3472	32
ADMM [44]	282.5776	5.8497	78
IRLS	**282.2776**	**0.1063**	**10**

Fig. 5.   Computational complexity versus number of snapshots.

Fig. 6.   Computational complexity versus number of sensors.

TABLE II

Computational complexity.

Algorithm	Complexity
SVT [55]	$K_{\text{svt}}\mathcal{O}(TM^2)$
APG [56]	$K_{\text{apg}}\mathcal{O}(TM^2)$
ADMM [44]	$K_{\text{admm}}\mathcal{O}(TM^2)$
IRLS	$K_{\text{irls}}\mathcal{O}(M^3)$

$Q = 1000$ Monte Carlo runs. It is seen that the CPU times of the SVT, APG, and ADMM[1] algorithms are nearly linearly increasing with the number of snapshots. This is consistent with the theoretical analysis in Table II. Fig. 6 displays the CPU time versus the number of sensors with $T = 100$ snapshots and the other parameters are the same as those in Fig. 5. We see that the curves of the CPU time of the SVT, APG, and ADMM algorithms are approximately linearly correlated to the number of sensors in a log scale, which again matches the theoretical calculations in Table II.

D. DOA ESTIMATION PERFORMANCE

We use the RMSE and resolution probability as DOA estimation performance measures. The resolution probability is defined as

$$\text{ResProb} = N_{\text{succ}}/Q,$$

where as in the previous examples 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

---

[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}.

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 Figs. 7 and 8, respectively. The traditional Cramér–Rao bound (CRB) with known sensor errors [63] 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 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 Figs. 9 and 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

Fig. 7.   RMSE versus SNR.

Fig. 9.   RMSE versus number of snapshots.

Fig. 8.   Resolution probability versus SNR.

Fig. 10.   Resolution probability versus number of snapshots.

of this subsection. The RMSE and resolution probability versus angular separation are displayed in Figs. 11 and 12, respectively. These again indicate that the IRLS algorithm outperforms the SVT, APG, and ADMM algorithms in terms of RMSE and resolution probability.

E. DISTORTED SENSOR DETECTION PERFORMANCE

Parallel to the three examples in Section V.D, we now examine the performance of the detection of distorted sensors of the SVT, APG, ADMM, and IRLS algorithms. The threshold in Algorithm 3 is set as $h = 10d$. We utilize the success detection rate as a metric, which is defined as

$$\text{DetecRate} = N_{\text{detec}}/Q.$$

N_{detec} denotes 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 Figs. 13, 14, and 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.

VI. CONCLUSION

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 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.

Fig. 11.   RMSE versus source separation angle.

Fig. 13.   Success detection rate versus SNR.

Fig. 12.   Resolution probability versus source separation angle.

Fig. 14.   Success detection rate versus number of snapshots.

APPENDIX

A. Proof of Theorem 1

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

$$
\begin{aligned}
& f(\mathbf{Z}_k) - f(\mathbf{Z}_{k+1}) \\
= \; & \|[\mathbf{Z}_k, \mu\mathbf{I}]\|_* - \|[\mathbf{Z}_{k+1}, \mu\mathbf{I}]\|_* \\
& + \lambda \left(\|[\mathbf{Y} - \mathbf{Z}_k, \mu\mathbf{1}]\|_{2,1} - \|[\mathbf{Y} - \mathbf{Z}_{k+1}, \mu\mathbf{1}]\|_{2,1} \right) \\
= \; & \operatorname{trace}\!\left((\mathbf{Z}_k\mathbf{Z}_k^{\mathrm{H}} + \mu^2\mathbf{I})^{\frac{1}{2}} \right) - \operatorname{trace}\!\left((\mathbf{Z}_{k+1}\mathbf{Z}_{k+1}^{\mathrm{H}} + \mu^2\mathbf{I})^{\frac{1}{2}} \right) \\
& + \lambda \left(\|[\mathbf{Y} - \mathbf{Z}_k, \mu\mathbf{1}]\|_{2,1} - \|[\mathbf{Y} - \mathbf{Z}_{k+1}, \mu\mathbf{1}]\|_{2,1} \right) \\
\geq \; & \operatorname{trace}\!\left(\frac{1}{2}\left(\mathbf{Z}_k\mathbf{Z}_k^{\mathrm{H}} - \mathbf{Z}_{k+1}\mathbf{Z}_{k+1}^{\mathrm{H}} \right)\mathbf{P}_k \right) \\
& + \frac{\lambda}{2}\operatorname{trace}\!\left(\mathbf{Q}_k\left[(\mathbf{Y} - \mathbf{Z}_k)(\mathbf{Y} - \mathbf{Z}_k)^{\mathrm{H}} \right.\right. \\
& \qquad\qquad \left.\left. - (\mathbf{Y} - \mathbf{Z}_{k+1})(\mathbf{Y} - \mathbf{Z}_{k+1})^{\mathrm{H}} \right] \right) \qquad (16) \\
= \; & \operatorname{trace}\!\left(\frac{1}{2}(\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}}\mathbf{P}_k \right) \\
& + \operatorname{trace}\!\left((\mathbf{Z}_k - \mathbf{Z}_{k+1})\mathbf{Z}_{k+1}^{\mathrm{H}}\mathbf{P}_k \right)
\end{aligned}
$$

$$
\begin{aligned}
& + \frac{\lambda}{2}\operatorname{trace}\!\left(2\mathbf{Q}_k\mathbf{Y}(\mathbf{Z}_{k+1} - \mathbf{Z}_k)^{\mathrm{H}} \right) \\
& + \frac{\lambda}{2}\operatorname{trace}\!\left(\mathbf{Q}_k(\mathbf{Z}_k\mathbf{Z}_k^{\mathrm{H}} - \mathbf{Z}_{k+1}\mathbf{Z}_{k+1}^{\mathrm{H}}) \right) \\
= \; & \operatorname{trace}\!\left(\frac{1}{2}(\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}}\mathbf{P}_k \right) \\
& + \operatorname{trace}\!\left((\mathbf{Z}_k - \mathbf{Z}_{k+1})\mathbf{Z}_{k+1}^{\mathrm{H}}\mathbf{P}_k \right) \\
& + \frac{\lambda}{2}\operatorname{trace}\!\left(2\mathbf{Q}_k\mathbf{Y}(\mathbf{Z}_{k+1} - \mathbf{Z}_k)^{\mathrm{H}} \right) \\
& + \frac{\lambda}{2}\operatorname{trace}\!\left(\mathbf{Q}_k(\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}} \right) \\
& + \lambda\operatorname{trace}\!\left(\mathbf{Q}_k(\mathbf{Z}_k - \mathbf{Z}_{k+1})\mathbf{Z}_{k+1}^{\mathrm{H}} \right) \\
\geq \; & \operatorname{trace}\!\left((\mathbf{Z}_k - \mathbf{Z}_{k+1})\mathbf{Z}_{k+1}^{\mathrm{H}}\mathbf{P}_k \right) \\
& + \operatorname{trace}\!\left(\lambda\mathbf{Q}_k\mathbf{Y}(\mathbf{Z}_{k+1} - \mathbf{Z}_k)^{\mathrm{H}} \right) \\
& + \operatorname{trace}\!\left(\lambda\mathbf{Q}_k(\mathbf{Z}_k - \mathbf{Z}_{k+1})\mathbf{Z}_{k+1}^{\mathrm{H}} \right) \qquad (17) \\
= \; & \operatorname{trace}\!\left((\mathbf{Z}_k - \mathbf{Z}_{k+1})\mathbf{Z}_{k+1}^{\mathrm{H}}(\mathbf{P}_k + \lambda\mathbf{Q}_k) \right) \\
& + \operatorname{trace}\!\left(\lambda\mathbf{Q}_k\mathbf{Y}(\mathbf{Z}_{k+1} - \mathbf{Z}_k)^{\mathrm{H}} \right) \\
= \; & \operatorname{trace}\!\left((\mathbf{P}_k + \lambda\mathbf{Q}_k)\mathbf{Z}_{k+1}(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}} \right) \\
& + \operatorname{trace}\!\left(\lambda\mathbf{Q}_k\mathbf{Y}(\mathbf{Z}_{k+1} - \mathbf{Z}_k)^{\mathrm{H}} \right)
\end{aligned}
$$

Fig. 15. Success detection rate versus source separation angle.

$$= \text{trace}\big(\lambda \mathbf{Q}_k \mathbf{Y}(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}}\big)$$
$$+ \text{trace}\big(\lambda \mathbf{Q}_k \mathbf{Y}(\mathbf{Z}_{k+1} - \mathbf{Z}_k)^{\text{H}}\big) \qquad (18)$$
$$= 0,$$

which indicates that $f(\mathbf{Z})$ is a non-increasing function with sequence $\{\mathbf{Z}_k\}$ generated by the IRLS procedure (10). The first equality is based on the definition of the objective function in (9), and the second equality uses $\|\mathbf{Z}\|_* = \text{trace}\left(\big(\mathbf{Z}\mathbf{Z}^{\text{H}}\big)^{-\frac{1}{2}}\right)$ when $M < T$. Inequality (16) holds thanks to Lemmata 1 and 2. Inequality (17) holds because of $\text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}}\mathbf{P}\big) \geq 0$ and $\text{trace}\big(\mathbf{Q}(\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}}\big) \geq 0$, which result from the fact that $\mathbf{P}$ and $\mathbf{Q}$ are symmetric matrices and $\text{trace}\big(\mathbf{X}\mathbf{X}^{\text{H}}\big) = \|\mathbf{X}\|_{\text{F}}^2 \geq 0$. Equality (18) 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

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

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

Besides, combining (10) and (16) yields

$$f(\mathbf{Z}_k) - f(\mathbf{Z}_{k+1})$$
$$\geq \text{trace}\left(\frac{1}{2}\big(\mathbf{Z}_k\mathbf{Z}_k^{\text{H}} - \mathbf{Z}_{k+1}\mathbf{Z}_{k+1}^{\text{H}}\big)\mathbf{P}_k\right)$$
$$+ \frac{\lambda}{2}\text{trace}\Big(\mathbf{Q}_k\big[(\mathbf{Y} - \mathbf{Z}_k)(\mathbf{Y} - \mathbf{Z}_k)^{\text{H}}$$
$$- (\mathbf{Y} - \mathbf{Z}_{k+1})(\mathbf{Y} - \mathbf{Z}_{k+1})^{\text{H}}\big]\Big)$$
$$= \frac{1}{2}\text{trace}\big((\mathbf{P}_k + \lambda\mathbf{Q}_k)(\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}}\big)$$
$$\geq \frac{1}{2}\sum_{i=1}^{M}\zeta_i(\mathbf{P}_k + \lambda\mathbf{Q}_k)\zeta_{M-i+1}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}}\big) \quad (19)$$
$$\geq \frac{1}{2}\zeta_M(\mathbf{P}_k + \lambda\mathbf{Q}_k)\|\mathbf{Z}_k - \mathbf{Z}_{k+1}\|_{\text{F}}^2 \geq \frac{1}{2}\zeta_{\min} \times \|\mathbf{Z}_k - \mathbf{Z}_{k+1}\|_{\text{F}}^2,$$

where $\zeta_i(\cdot)$ denotes the ith largest eigenvalue of its input Hermitian matrix, inequality (19) follows from the fact that $\text{trace}(\mathbf{X}\mathbf{Y}) \geq \sum_{i=1}^{M} \zeta_i(\mathbf{X})\zeta_{M-i+1}(\mathbf{Y})$ holds for any two positive semi-definite matrices $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{C}^{M \times M}$ [49], 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

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

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

B. Proof of Theorem 2

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

$$\|[\mathbf{Y} - \mathbf{Z}, \mu\mathbf{1}]\|_{2,1} = \sum_{i=1}^{M}\sqrt{\|(\mathbf{Y} - \mathbf{Z})_{i,:}\|_2^2 + \mu^2}$$
$$\geq \sum_{i=1}^{M}|\mu| = |\mu|M$$
$$\|[\mathbf{Z}, \mu\mathbf{I}]\|_* = \text{trace}\left(\big(\mathbf{Z}\mathbf{Z}^{\text{H}} + \mu^2\mathbf{I}\big)^{\frac{1}{2}}\right)$$
$$\geq \big(\text{trace}(\mathbf{Z}\mathbf{Z}^{\text{H}} + \mu^2\mathbf{I})\big)^{\frac{1}{2}} \qquad (20)$$
$$= \big(\text{trace}(\mathbf{Z}\mathbf{Z}^{\text{H}}) + M\mu^2\big)^{\frac{1}{2}}$$
$$\geq \big(M\mu^2\big)^{\frac{1}{2}} = |\mu|\sqrt{M}.$$

Inequality (20) holds because $\text{trace}\left(\mathbf{X}^{\frac{1}{2}}\right) = \sum_i\sqrt{\zeta_i} \geq \sqrt{\sum_i\zeta_i} = (\text{trace}(\mathbf{X}))^{\frac{1}{2}}$ for any symmetric matrix $\mathbf{X}$, with ζ_i being the eigenvalue of $\mathbf{X}$. Therefore, the objective function in (9) is bounded below as $f(\mathbf{Z}) = \|[\mathbf{Z}, \mu\mathbf{I}]\|_* + \lambda\|[\mathbf{Y} - \mathbf{Z}, \mu\mathbf{1}]\|_{2,1} \geq |\mu|(\sqrt{M} + \lambda M)$.

C. Proof of Theorem 3

Denote the limit point of sequence $\{\mathbf{Z}_k\}$ as $\mathbf{Z}_{k+1}$. Then, according to $\lim_{k\to\infty}\|\mathbf{Z}_k - \mathbf{Z}_{k+1}\|_{\text{F}} = 0$ in Theorem 1 and (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 (9) is convex w.r.t. $\mathbf{Z}$, the stationary point is globally optimal.

D. Proof of Theorem 4

Similar to the proof of Theorem 1, we calculate the difference between the objective function values in two successive iterations as

$$f(\mathbf{Z}_k, \mathbf{V}_k) - f(\mathbf{Z}_{k+1}, \mathbf{V}_{k+1})$$
$$= \frac{1}{2}\|\mathbf{Y} - \mathbf{Z}_k - \mathbf{V}_k\|_{\text{F}}^2 - \frac{1}{2}\|\mathbf{Y} - \mathbf{Z}_{k+1} - \mathbf{V}_{k+1}\|_{\text{F}}^2$$
$$+ \lambda_1\|[\mathbf{Z}_k, \mu\mathbf{I}]\|_* - \lambda_1\|[\mathbf{Z}_{k+1}, \mu\mathbf{I}]\|_*$$
$$+ \lambda_2\|[\mathbf{V}_k, \mu\mathbf{1}]\|_{2,1} - \lambda_2\|[\mathbf{V}_{k+1}, \mu\mathbf{1}]\|_{2,1}$$
$$\geq \frac{1}{2}\|\mathbf{Y} - \mathbf{Z}_k - \mathbf{V}_k\|_{\text{F}}^2 - \frac{1}{2}\|\mathbf{Y} - \mathbf{Z}_{k+1} - \mathbf{V}_{k+1}\|_{\text{F}}^2$$

$$+ \lambda_1 \, \text{trace}\!\left(\frac{1}{2} \left(\mathbf{Z}_k \mathbf{Z}_k^{\mathrm{H}} - \mathbf{Z}_{k+1} \mathbf{Z}_{k+1}^{\mathrm{H}} \right) \mathbf{P}_k \right)$$

$$+ \frac{\lambda_2}{2} \text{trace}\!\left(\mathbf{Q}_k \left(\mathbf{V}_k \mathbf{V}_k^{\mathrm{H}} - \mathbf{V}_{k+1} \mathbf{V}_{k+1}^{\mathrm{H}} \right) \right)$$

$$= \frac{1}{2} \| \mathbf{Y} - \mathbf{Z}_k - \mathbf{V}_k \|_{\mathrm{F}}^2 - \frac{1}{2} \| \mathbf{Y} - \mathbf{Z}_{k+1} - \mathbf{V}_{k+1} \|_{\mathrm{F}}^2$$

$$+ \lambda_1 \, \text{trace}\!\left(\frac{1}{2} \left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right) \left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right)^{\mathrm{H}} \mathbf{P}_k \right)$$

$$+ \lambda_1 \, \text{trace}\!\left(\left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right) \mathbf{Z}_{k+1}^{\mathrm{H}} \mathbf{P}_k \right)$$

$$+ \frac{\lambda_2}{2} \text{trace}\!\left(\mathbf{Q}_k \left(\mathbf{V}_k - \mathbf{V}_{k+1} \right) \left(\mathbf{V}_k - \mathbf{V}_{k+1} \right)^{\mathrm{H}} \right)$$

$$+ \lambda_2 \, \text{trace}\!\left(\mathbf{Q}_k (\mathbf{V}_k - \mathbf{V}_{k+1}) \mathbf{V}_{k+1}^{\mathrm{H}} \right) \qquad (21)$$

$$\geq \frac{1}{2} \text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}}$$

$$+ (\mathbf{V}_k - \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\mathrm{H}} \big)$$

$$+ \text{trace}\big((-\mathbf{Y} + \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}}$$

$$+ (-\mathbf{Y} + \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\mathrm{H}}$$

$$+ \mathbf{Z}_k \mathbf{V}_k^{\mathrm{H}} - \mathbf{Z}_{k+1} \mathbf{V}_{k+1}^{\mathrm{H}} \big)$$

$$+ \lambda_1 \, \text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1}) \mathbf{Z}_{k+1}^{\mathrm{H}} \mathbf{P}_k \big)$$

$$+ \lambda_2 \, \text{trace}\big(\mathbf{Q}_k (\mathbf{V}_k - \mathbf{V}_{k+1}) \mathbf{V}_{k+1}^{\mathrm{H}} \big)$$

$$= \frac{1}{2} \text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}}$$

$$+ (\mathbf{V}_k - \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\mathrm{H}} \big)$$

$$+ \text{trace}\big(\mathbf{Z}_k \mathbf{V}_k^{\mathrm{H}} - \mathbf{Z}_{k+1} \mathbf{V}_{k+1}^{\mathrm{H}} \big)$$

$$+ \text{trace}\big(-\mathbf{V}_k (\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}} \big)$$

$$+ \text{trace}\big(-\mathbf{Z}_k (\mathbf{V}_k - \mathbf{V}_{k+1})^{\mathrm{H}} \big)$$

$$= \frac{1}{2} \text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}}$$

$$+ (\mathbf{V}_k - \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\mathrm{H}} \big)$$

$$+ \text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{V}_{k+1} - \mathbf{V}_k)^{\mathrm{H}} \big)$$

$$= \frac{1}{2} \text{trace}((\mathbf{Z}_k - \mathbf{Z}_{k+1} - \mathbf{V}_k + \mathbf{V}_{k+1}) \cdot$$

$$(\mathbf{Z}_k - \mathbf{Z}_{k+1} - \mathbf{V}_k + \mathbf{V}_{k+1})^{\mathrm{H}})$$

$$= \frac{1}{2} \| \mathbf{Z}_k - \mathbf{Z}_{k+1} - \mathbf{V}_k + \mathbf{V}_{k+1} \|_{\mathrm{F}}^2 \geq 0,$$

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

$$\min\{\lambda_1, \lambda_2\} \left(\| [\mathbf{Z}_k, \mu\mathbf{I}] \|_* + \| [\mathbf{V}_k, \mu\mathbf{1}] \|_{2,1} \right)$$

$$\leq \ \lambda_1 \| [\mathbf{Z}_k, \mu\mathbf{I}] \|_* + \lambda_2 \| [\mathbf{V}_k, \mu\mathbf{1}] \|_{2,1}$$

$$\leq \ f(\mathbf{Z}_k, \mathbf{V}_k) \ \leq \ f(\mathbf{Z}_0, \mathbf{V}_0).$$

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 indicates that the sequence $\{(\mathbf{Z}_k, \mathbf{V}_k)\}$ is bounded.

Besides, combining (15) and (21) yields

$$f(\mathbf{Z}_k, \mathbf{V}_k) - f(\mathbf{Z}_{k+1}, \mathbf{V}_{k+1})$$

$$\geq \frac{\lambda_1}{2} \, \text{trace}\!\left(\left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right) \left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right)^{\mathrm{H}} \mathbf{P}_k \right)$$

$$+ \frac{\lambda_2}{2} \text{trace}\!\left(\mathbf{Q}_k \left(\mathbf{V}_k - \mathbf{V}_{k+1} \right) \left(\mathbf{V}_k - \mathbf{V}_{k+1} \right)^{\mathrm{H}} \right)$$

$$+ \frac{1}{2} \text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}} \big)$$

$$\geq \ \frac{\lambda_1}{2} \, \text{trace}\!\left(\left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right) \left(\mathbf{Z}_k - \mathbf{Z}_{k+1} \right)^{\mathrm{H}} \mathbf{P}_k \right)$$

$$+ \frac{\lambda_2}{2} \text{trace}\!\left(\mathbf{Q}_k \left(\mathbf{V}_k - \mathbf{V}_{k+1} \right) \left(\mathbf{V}_k - \mathbf{V}_{k+1} \right)^{\mathrm{H}} \right)$$

$$\geq \ \frac{\lambda_1}{2} \sum_i^M \zeta_i(\mathbf{P}_k) \zeta_{M-i+1}((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\mathrm{H}})$$

$$+ \frac{\lambda_2}{2} \sum_i^M \zeta_i(\mathbf{Q}_k) \zeta_{M-i+1}((\mathbf{V}_k - \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\mathrm{H}})$$

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

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 $\lim_{k \to \infty} \| \mathbf{V}_k - \mathbf{V}_{k+1} \|_{\mathrm{F}} = 0$.

E. Proof of Theorem 5

Considering $\| \mathbf{Y} - \mathbf{Z} - \mathbf{V} \|_{\mathrm{F}}^2 \geq 0$ and the inequalities in Appendix VI.B, we can prove that the objective function in (13) is bounded below as $f(\mathbf{Z}, \mathbf{V}) = \frac{1}{2} \| \mathbf{Y} - \mathbf{Z} - \mathbf{V} \|_{\mathrm{F}}^2 + \lambda_1 \| [\mathbf{Z}, \mu\mathbf{I}] \|_* + \lambda_2 \| [\mathbf{V}, \mu\mathbf{1}] \|_{2,1} \geq |\mu|(\lambda_1 \sqrt{M} + \lambda_2 M)$.

F. Proof of Theorem 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 and (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 (13). Since Problem (13) is convex w.r.t. $\mathbf{Z}$ and $\mathbf{V}$, the stationary point is globally optimal.

REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

[6] J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, vol. 11, no. 4, pp. 329–337, December 1986.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] B. Liao and S. C. Chan, "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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] I. T. Jolliffe, "Outlier detection, influential observations and robust estimation of principal components," in *Principal Component Analysis*. Springer New York, 1986, pp. 173–198.

[42] 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.

[43] 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.

[44] 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

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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, USA, February 2015.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

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

[61] 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.

[62] 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.

[63] 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.