Correspondence

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

Distorted sensors could occur randomly and may lead to the breakdown of a sensor array system. In this article, 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. 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.

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, and wireless communications, to name just a few [1], [2], [3]. Many classical approaches have been proposed, including multiple signal classification (MUSIC) [4], estimation of signal parameters via rotational invariance techniques [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

Manuscript received 9 September 2022; revised 9 December 2022; accepted 30 January 2023. Date of publication 3 February 2023; date of current version 9 August 2023.

DOI. No. 10.1109/TAES.2023.3241886

Refereeing of this contribution was handled by D. A. Garren.

The work was supported in part by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt, in part by the National Natural Science Foundation of China under Grant 62202174, and in part by the Guangdong Provincial Key Laboratory of Human Digital Twin under Grant 2022B1212010004.

Authors' addresses: Huiping Huang was with the Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, 64283 Darmstadt, Germany, and is now with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden, E-mail: (huiping@chalmers.se); Qi Liu is with School of Future Technology, South China University of Technology, China, and also with Pazhou Lab, Guangzhou 510330, China, E-mail: (drliuqi@scut.edu.cn); Hing Cheung So is with Department of Electrical Engineering, City University of Hong Kong, 999077 Hong Kong SAR, China, E-mail: (hcso@ee.cityu.edu.hk); Abdelhak M. Zoubir is with Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, 64283 Darmstadt, Germany, E-mail: (zoubir@spg.tudarmstadt.de). *(Corresponding author: Qi Liu.)*

array encounters distortions [8], [9], [10], [11], 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., [12], [13], [14], and [15]. 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 [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]. In [16], the genetic algorithm was applied for array failure correction. A minimal resource allocation network was used for DOA estimation under array sensor failure [28], which requires a training procedure with no failed sensors. A Bayesian compressive sensing approach was proposed in [18], which needs a noise-free array as a reference. Methods using difference coarray were developed in [19], [20], and [21]. The idea of [19] 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 coarray. On the other hand, [20] and [21] restricted the array to some special sparse structures, such as coprime and nested arrays. Approaches based on precalibrated sensors have been well-documented in the past decades [23], [24], [25]. 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) [29], [30], [31]. 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 [32], [33], [34], [35]. Another special structure of LRSD is low-rank and column-sparse decomposition (LRCSD) [36], [37], [38], [39], [40], also known as RPCA-outlier pursuit [41], [42], [43], [44], 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, e.g., when a fraction of the data vectors are grossly corrupted by outliers [38], [40].

Several algorithms have been contributed to solve the LRSD and LRCSD problems, such as singular value thresholding (SVT) [45], accelerated proximal gradient (APG) [46], alternating direction method of multipliers (ADMM) [34], [40], and iteratively reweighted least squares (IRLS) [39], [40], [43], [44]. 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) at 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 [47], [48], [49]. In this sense, the IRLS is more efficient in solving the corresponding problems. Another related research direction that uses alternating optimization was introduced in [50] and [51], where the received data are approximated in a reduced-dimension space, and an auxiliary parameter vector is introduced to construct the output power spectrum. That is, alternating updates between the decomposition matrix and auxiliary parameter vector are performed. Differently, the IRLS algorithm alternates the low-rank matrix with the row-sparse matrix; see Section IV for more details.

Therefore, in the present work, we develop an IRLS algorithm for joint DOA estimation and distorted sensor detection. The main contributions are as follows.

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

The rest of this article 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. Finally, Section VI concludes this article.

Notation: In this article, 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, $\{\|\mathbf{V}_{i,:}\|_2\} = \{\|\mathbf{V}_{1,:}\|_2, \|\mathbf{V}_{2,:}\|_2, \ldots, \|\mathbf{V}_{M,:}\|_2\}$, and $\mathbf{V}_{i,:}$ 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

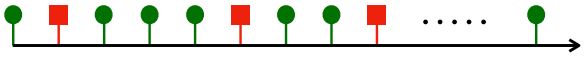

Fig. 1. Illustration of array structure of interest.

same dimensions, we define their Frobenius inner product as $\langle \mathbf{X}, \mathbf{Y} \rangle \triangleq \text{trace}(\mathbf{X}^{\text{H}}\mathbf{Y})$, where $\text{trace}(\cdot)$ 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 $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^{\text{T}}$. 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 green circles stand for *perfect* sensors and the red boxes refer to distorted ones. The red 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 $t = 1, 2, \ldots, T$ denotes the time index, T is the total number of available snapshots, and $\mathbf{s}(t) \in \mathbb{C}^K$ and $\mathbf{n}(t) \in \mathbb{C}^M$ are signal and noise vectors, respectively. The steering matrix $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$ has steering vectors as columns, where the steering vector $\mathbf{a}(\theta_k)$ is a function of θ_k, for $k = 1, 2, \ldots, K$. 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, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_M]^{\text{T}}$, being a sparse vector. Specifically, for $m = 1, 2, \ldots, M$

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

The nonzero γ_m denotes sensor gain and phase error, namely, $\gamma_m = \rho_m e^{j\phi_m}$, 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 $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(T)] \in \mathbb{C}^{M\times T}$ contains the measurements, $\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), \ldots, \mathbf{s}(T)] \in \mathbb{C}^{K\times T}$ is the signal matrix, and $\mathbf{N} = [\mathbf{n}(1), \mathbf{n}(2), \ldots, \mathbf{n}(T)] \in \mathbb{C}^{M\times T}$ 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 nonzero) 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 e.g., [45]. By adapting the SVT algorithm to our problem, we need to solve

$$\min_{\mathbf{Z},\mathbf{V},\mathbf{W}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1} + \frac{1}{2\tau}\|\mathbf{Z}\|_{\text{F}}^2 + \frac{1}{2\tau}\|\mathbf{V}\|_{\text{F}}^2 + \frac{1}{\tau}\langle \mathbf{W}, \mathbf{Y}-\mathbf{Z}-\mathbf{V}\rangle \tag{3}$$

where λ is a tuning parameter, and 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 $\mathbf{V}$ [45]:

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

$$\mathcal{S}_\kappa(\mathbf{C}) = \arg\min_{\mathbf{X}} \kappa\|\mathbf{X}\|_{2,1} + \frac{1}{2}\|\mathbf{X} - \mathbf{C}\|_{\text{F}}^2$$

where $\mathbf{L}\mathbf{S}\mathbf{R}^{\text{H}}$ 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 [34].

The second method is APG, whose updating equation can be given as [46]

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

where subscript $\cdot_k$ denotes the variable at the kth iteration; $h(\mathbf{Z}, \mathbf{V}) \triangleq p(\mathbf{Z}_k, \mathbf{V}_k) + \langle \nabla_{\mathbf{Z}_k} p(\mathbf{Z}, \mathbf{V}_k); \mathbf{Z} - \mathbf{Z}_k\rangle + \langle \nabla_{\mathbf{V}_k} p(\mathbf{Z}_k, \mathbf{V}), \mathbf{V} - \mathbf{V}_k\rangle + \mu M\|\mathbf{Z} + \mathbf{V} - \mathbf{Z}_k - \mathbf{V}_k\|_{\text{F}}^2 + q(\mathbf{Z}, \mathbf{V})$, with $p(\mathbf{Z}, \mathbf{V}) \triangleq \frac{1}{\mu}\|\mathbf{Y} - \mathbf{Z} - \mathbf{V}\|_{\text{F}}^2$, $q(\mathbf{Z}, \mathbf{V}) \triangleq \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1}$, and μ being a small positive scalar. The detailed algorithm can be found in [46] and also [34].

As for ADMM, we consider the following problem:

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

and its augmented Lagrangian function is $\mathcal{L}_\mu(\mathbf{Z}, \mathbf{V}, \mathbf{W}) = \|\mathbf{Z}\|_* + \lambda\|\mathbf{V}\|_{2,1} + \langle \mathbf{W}, \mathbf{Y}-\mathbf{Z}-\mathbf{V}\rangle + \frac{\mu}{2}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_{\text{F}}^2$, 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 $\mathcal{L}_\mu(\mathbf{Z}, \mathbf{V}, \mathbf{W})$ w.r.t $\mathbf{Z}$ (resp. $\mathbf{V}$) while keeping $\mathbf{V}$ (resp. $\mathbf{Z}$) and $\mathbf{W}$ unchanged; $\mathbf{W}$ is updated as $\mathbf{W} = \mathbf{W} + \mu(\mathbf{Y} - \mathbf{Z} - \mathbf{V})$ [52].

All the aforementioned 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.

IV. PROPOSED METHOD

In this section, we develop an IRLS algorithm for the task of jointly estimating DOAs of sources and detecting distorted sensors. Following the data model in (2), the problem to be solved can be given as

$$\min_{\mathbf{Z},\mathbf{V}} \frac{1}{2}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_{\mathrm{F}}^2+\lambda_1\|\mathbf{Z}\|_*+\lambda_2\|\mathbf{V}\|_{2,1} \tag{6}$$

where λ_1 and λ_2 are two tuning parameters. To handle the nonsmoothness of the nuclear norm and the $\ell_{2,1}$ mixed-norm, we introduce a smoothing parameter μ. That is, Problem (6) is transferred to

$$\min_{\mathbf{Z},\mathbf{V}} f(\mathbf{Z},\mathbf{V}) \tag{7}$$

with $f(\mathbf{Z},\mathbf{V}) \triangleq \frac{1}{2}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_{\mathrm{F}}^2+\lambda_1\|[\mathbf{Z},\mu\mathbf{I}]\|_*+\lambda_2\|[\mathbf{V},\mu\mathbf{1}]\|_{2,1}$, where $\mathbf{1}$ is an all-ones vector of appropriate length. 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 $\mathbf{P} \triangleq \left(\mathbf{Z}\mathbf{Z}^{\mathrm{H}}+\mu^2\mathbf{I}\right)^{-\frac{1}{2}}$ and

$$\mathbf{Q} \triangleq \begin{bmatrix} \frac{1}{\sqrt{\|\mathbf{V}_{1,:}\|_2^2+\mu^2}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\|\mathbf{V}_{M,:}\|_2^2+\mu^2}} \end{bmatrix}. \tag{8}$$

According to the Karush–Kuhn–Tucker (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

$$\begin{cases} \mathbf{Z}_{k+1}=(\mathbf{I}+\lambda_1\mathbf{P}_k)^{-1}(\mathbf{Y}-\mathbf{V}_k) \\ \mathbf{V}_{k+1}=(\mathbf{I}+\lambda_2\mathbf{Q}_k)^{-1}(\mathbf{Y}-\mathbf{Z}_{k+1}) \end{cases} \tag{9}$$

where $\mathbf{P}_k$ and $\mathbf{Q}_k$ depend on $\mathbf{Z}_k$ and $\mathbf{V}_k$, respectively. The IRLS algorithm is summarized in Algorithm 1. The algorithm is based on alternating optimization, and it is terminated when a stopping condition is reached.

Algorithm 1: IRLS Algorithm for Solving Problem (7).

Input : $\mathbf{Y} \in \mathbb{C}^{M\times T}$, $\lambda_1, \lambda_2, \mu, \epsilon, k_{\max}$
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**
  $k \leftarrow k+1$
  calculate $\mathbf{P}_k$ and $\mathbf{Q}_k$
  update $\mathbf{Z}_k$ using $\mathbf{Z}=(\mathbf{I}+\lambda_1\mathbf{P})^{-1}(\mathbf{Y}-\mathbf{V})$
  update $\mathbf{V}_k$ using $\mathbf{V}=(\mathbf{I}+\lambda_2\mathbf{Q})^{-1}(\mathbf{Y}-\mathbf{Z})$
  converged $\leftarrow \begin{cases} k \geq k_{\max} \text{ or} \\ \frac{|f(\mathbf{Z}_k,\mathbf{V}_k)-f(\mathbf{Z}_{k-1},\mathbf{V}_{k-1})|}{|f(\mathbf{Z}_k,\mathbf{V}_k)|} \leq \epsilon \end{cases}$
end while
$\widehat{\mathbf{Z}} \leftarrow \mathbf{Z}_k$, $\widehat{\mathbf{V}} \leftarrow \mathbf{V}_k$

A. Convergence Analysis for IRLS

In this part, the monotonicity and boundedness of the objective function $f(\mathbf{Z},\mathbf{V})$ in (7) are proved in Theorems 1 and 2, respectively.

THEOREM 1 The IRLS procedure (9) produces a nonincreasing objective function defined in (7), i.e., $f(\mathbf{Z}_k,\mathbf{V}_k) \geq f(\mathbf{Z}_{k+1},\mathbf{V}_{k+1})$ for $k=0,1,2,\cdots$. Besides, 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 A1. ■

THEOREM 2 The objective function $f(\mathbf{Z},\mathbf{V})$ defined in (7) is bounded below by $|\mu|(\lambda_1\sqrt{M}+\lambda_2 M)$.

PROOF See Appendix A2. ■

THEOREM 3 Any limit point of the sequence $\{(\mathbf{Z}_k,\mathbf{V}_k)\}$ generated by (9) is a stationary point of Problem (7), and moreover, the stationary point is globally optimal.

PROOF See Appendix A3. ■

REMARK The differences between our work and [39] are stated as follows.

1) The problem formulation in [39] is column-sparse, while we have row-sparsity of $\mathbf{V}$. This leads to differences in matrix multiplication and matrix derivative.
2) The authors in[39] consider the noiseless case. We consider the noisy case, which is more practical.
3) To update $\mathbf{Z}$ using matrices $\mathbf{P}$ and $\mathbf{Q}$, the approach in [39] involves a Sylvester equation and utilizes the MATLAB command `lyap`. However, our method admits a closed-form formula; see (9).
4) The proofs of convergence are not exactly the same. The authors in [39] prove the monotonicity of the objective and the boundedness of the sequence $\{\mathbf{Z}_k\}$. We prove the monotonicity and the boundedness of the objective, and we also show the boundedness of $\{(\mathbf{Z}_k,\mathbf{V}_k)\}$.

B. 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 $\mathbf{Z}=\mathbf{A}\mathbf{S}$ can be viewed as a noise-free data model. DOAs can be found via subspace-based methods, such as MUSIC, whose spatial spectrum is calculated as

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

The SVD of $\widehat{\mathbf{Z}}$ is $\widehat{\mathbf{Z}}=\mathbf{L}\mathbf{\Sigma}\mathbf{R}^{\mathrm{H}}$, 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

Algorithm 2: Detection of Distorted Sensors.

Input : $\widehat{\mathbf{V}} \in \mathbb{C}^{M\times T}$, h
Output: M_{fail}
calculate $\mathbf{v} = \left[\|\widehat{\mathbf{V}}_{1,:}\|_2, \|\widehat{\mathbf{V}}_{2,:}\|_2, \ldots, \|\widehat{\mathbf{V}}_{M,:}\|_2\right]^{\mathrm{T}}$
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** $i = 3, 4, \ldots, M$ **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: $M_{\text{fail}} \leftarrow M - i_{\text{fail}} + 1$

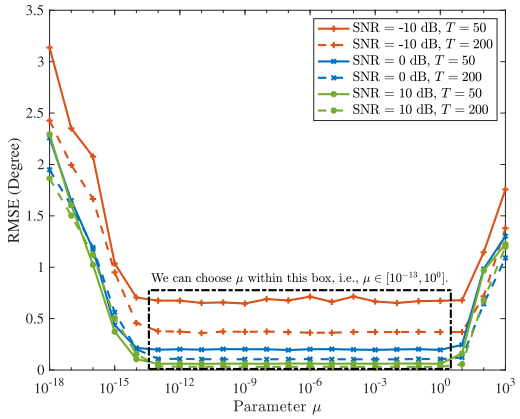

Fig. 2. RMSE versus μ, with $M = 10$ sensors (four of which fail), $K = 2$ sources, $\lambda_1 = 2$, and $\lambda_2 = 0.2$.

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 $\|\widehat{\mathbf{V}}_{i,:}\|_2$, $i = 1, 2, \ldots, M$. Algorithm 2 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 $i = 3, 4, \ldots, M$, 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 $M_{\text{fail}} = M - i_{\text{fail}} + 1$.

Analysis of the maximal number of distorted sensors to be detected by the IRLS algorithm is left as an open question. Related work can be found in [36].

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 (7) used in the IRLS algorithm. We set $\epsilon = 10^{-16}$, $k_{\max} = 1000$, and $\mathbf{Z}_{\text{init}} = \mathbf{V}_{\text{init}} = \mathbf{O}$, where $\mathbf{O}$ denotes the $M \times T$ all-zeros

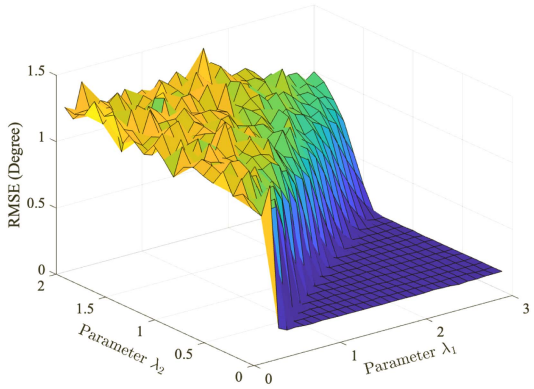

Fig. 3. RMSE versus λ_1 and λ_2, with $M = 10$ sensors (four of which fail), $K = 2$ sources, $T = 100$ snapshots, SNR $= 0$ dB, and $\mu = 0.01$.

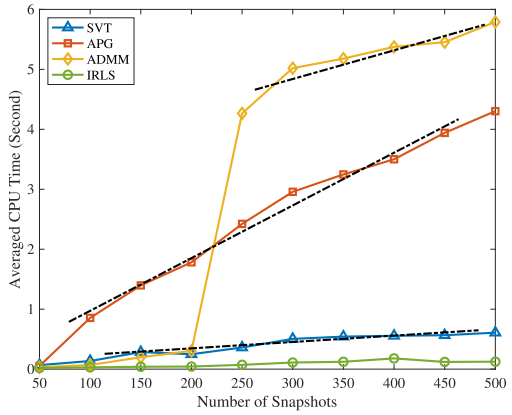

Fig. 4. Computational complexity versus number of snapshots.

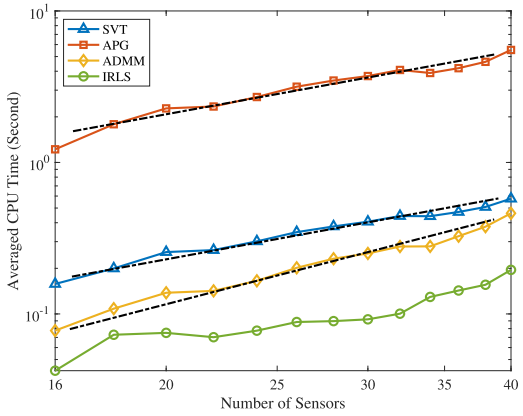

Fig. 5. Computational complexity versus number of sensors.

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.

Consider a uniform linear array (ULA) of $M = 10$ sensors, four of which at random positions are distorted by gain and phase errors, receiving $K = 2$ signals with DOAs $\boldsymbol{\theta} = [-10^\circ, 10^\circ]^{\mathrm{T}}$. The sensor gain and phase errors are

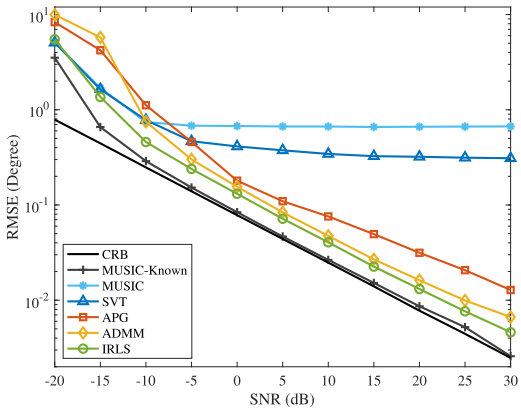

Fig. 6. rmse versus SNR.

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 [45]	493.8653	0.1205	**5**
APG [46]	123.3227	0.8403	22
ADMM [34]	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 [45]	1965.5994	0.6166	25
APG [46]	282.5776	4.3472	32
ADMM [34]	282.5776	5.8497	78
IRLS	**282.2776**	**0.1063**	**10**

The bold values denote the smallest values in each column.

randomly generated by drawing from uniform distributions on $[0, 10]$ and $[-15^\circ, 15^\circ]$, respectively. In the first example, we test six 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 six 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 μ. Note that in Fig. 3, our goal is to find a pair of (λ_1, λ_2) such that it 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 the IRLS 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, four of which at random positions are distorted, receives $K = 2$ signals from -10° and 10°. The objective function value, CPU time, and number of iterations are tabulated in Table I (upper) for SNR $= 0$ 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.

TABLE II
Computational Complexity

Algorithm	Complexity
SVT [45]	$K_{\text{svt}}\mathcal{O}(TM^2)$
APG [46]	$K_{\text{apg}}\mathcal{O}(TM^2)$
ADMM [34]	$K_{\text{admm}}\mathcal{O}(TM^2)$
IRLS	$K_{\text{irls}}\mathcal{O}(M^3)$

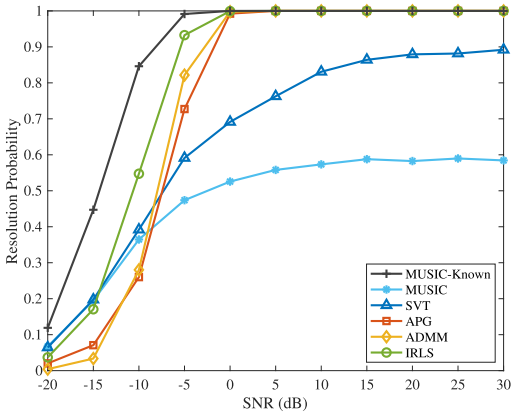

Fig. 7. Resolution probability versus SNR.

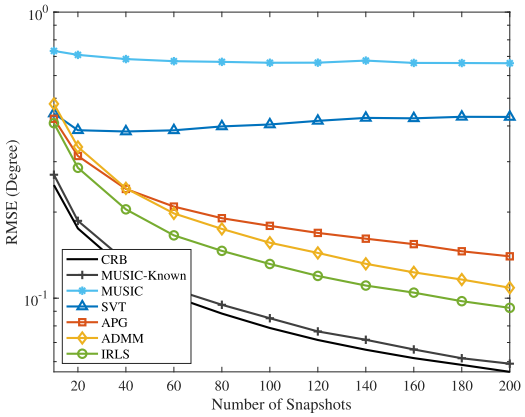

Fig. 8. rmse versus number of snapshots.

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.

Fig. 4 plots the averaged CPU time against the number of snapshots at $M = 10$ sensors (four 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

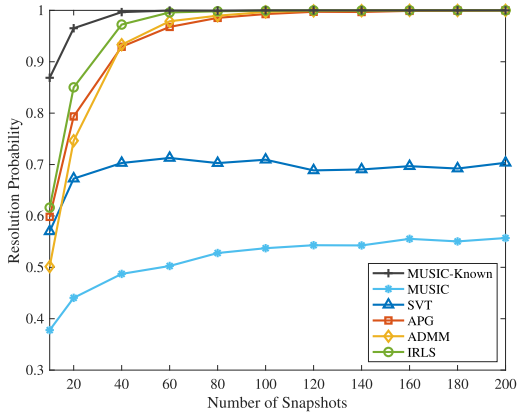

Fig. 9. Resolution probability versus number of snapshots.

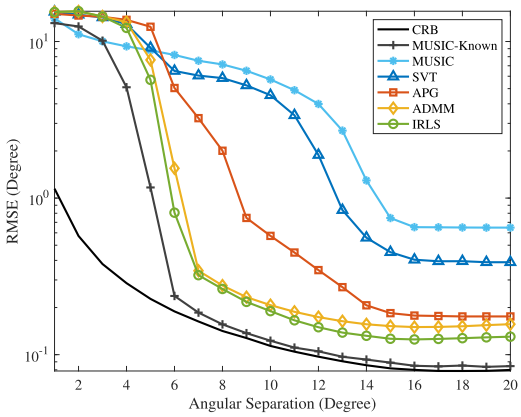

Fig. 10. RMSE versus source separation angle.

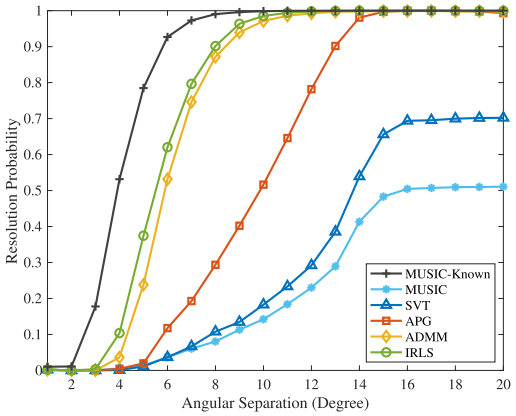

Fig. 11. Resolution probability versus source separation angle.

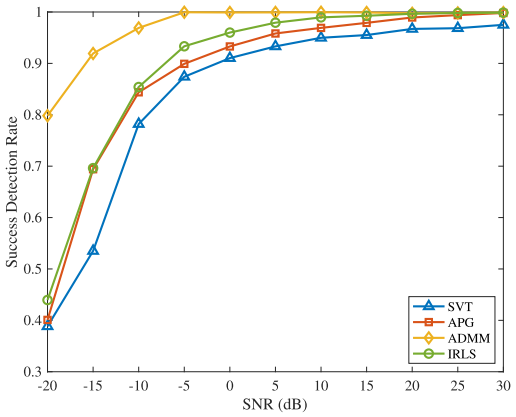

Fig. 12. Success detection rate versus SNR.

ADMM[1] algorithms are nearly linearly increasing with T. This is consistent with the theoretical analysis in Table II. Fig. 5 displays the CPU time versus the number of sensors with $T = 100$ snapshots and the other parameters are the same as those in Fig. 4. We see that the curves of the CPU time of the SVT, APG, and ADMM algorithms are approximately linearly correlated to M in a log scale, which again matches the theoretical calculations.

D. 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: $\max_k\{|\hat{\theta}_k - \theta_k|\} \leq 0.5^\circ$.

In the first example, we consider a ULA of $M = 10$ sensors, three 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. 6 and 7, respectively. The traditional Cramér–Rao bound with known sensor errors [53] 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. 8 and 9, 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

[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}.

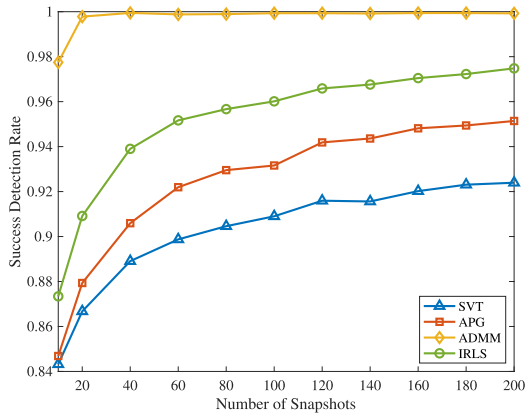

Fig. 13. Success detection rate versus number of snapshots.

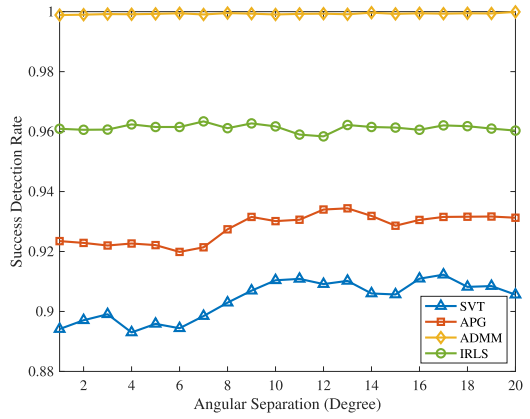

Fig. 14. Success detection rate versus source separation angle.

separation are displayed in Figs. 10 and 11, 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 2 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 Figs. 12, 13, and 14, 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

In this article, we studied the problem of simultaneously estimating 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 (LR^{2}SD), and solved by IRLS. 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.

APPENDIX A

A. Proof of Theorem 1

We first have the following two lemmata:

LEMMA 1 (LEMMA 1 IN [39]) For any matrices $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{C}^{M\times T}$, we have $\|\mathbf{Y}\|_{2,1} - \|\mathbf{X}\|_{2,1} \geq \frac{1}{2}\text{trace}\big(\mathbf{H}(\mathbf{Y}\mathbf{Y}^{\text{H}} - \mathbf{X}\mathbf{X}^{\text{H}})\big)$, where $\mathbf{H}$ is a diagonal matrix with main diagonal $[1/\|\mathbf{Y}_{1,:}\|_2, \ldots, 1/\|\mathbf{Y}_{M,:}\|_2]$.

LEMMA 2 (LEMMA 2 IN [39]) For any two symmetric positive definite matrices $\mathbf{X}$ and $\mathbf{Y}$, it holds that $\text{trace}\Big(\mathbf{Y}^{\frac{1}{2}}\Big) - \text{trace}\Big(\mathbf{X}^{\frac{1}{2}}\Big) \geq \text{trace}\Big(\frac{1}{2}(\mathbf{Y}-\mathbf{X})^{\text{H}}\mathbf{Y}^{-\frac{1}{2}}\Big)$.

Then, we calculate the difference between the objective function values at 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\|_{\text{F}}^2 - \frac{1}{2}\|\mathbf{Y}-\mathbf{Z}_{k+1}-\mathbf{V}_{k+1}\|_{\text{F}}^2 \\
&\quad + \lambda_1\|[\mathbf{Z}_k, \mu\mathbf{I}]\|_* - \lambda_1\|[\mathbf{Z}_{k+1}, \mu\mathbf{I}]\|_* \\
&\quad + \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 \\
&\quad + \lambda_1\text{trace}\left(\frac{1}{2}\left(\mathbf{Z}_k\mathbf{Z}_k^{\text{H}} - \mathbf{Z}_{k+1}\mathbf{Z}_{k+1}^{\text{H}}\right)\mathbf{P}_k\right) \\
&\quad + \frac{\lambda_2}{2}\text{trace}\big(\mathbf{Q}_k\left(\mathbf{V}_k\mathbf{V}_k^{\text{H}} - \mathbf{V}_{k+1}\mathbf{V}_{k+1}^{\text{H}}\right)\big) \\
&= \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 \\
&\quad + \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)^{\text{H}}\mathbf{P}_k\right) \\
&\quad + \lambda_1\text{trace}\big(\left(\mathbf{Z}_k - \mathbf{Z}_{k+1}\right)\mathbf{Z}_{k+1}^{\text{H}}\mathbf{P}_k\big) \\
&\quad + \frac{\lambda_2}{2}\text{trace}\Big(\mathbf{Q}_k\left(\mathbf{V}_k - \mathbf{V}_{k+1}\right)\left(\mathbf{V}_k - \mathbf{V}_{k+1}\right)^{\text{H}}\Big) \\
&\quad + \lambda_2\text{trace}\big(\mathbf{Q}_k(\mathbf{V}_k - \mathbf{V}_{k+1})\mathbf{V}_{k+1}^{\text{H}}\big) \\
&\geq \frac{1}{2}\text{trace}\big((\mathbf{Z}_k - \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}} \\
&\quad + (\mathbf{V}_k - \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\text{H}}\big) \\
&\quad + \text{trace}\big((-\mathbf{Y} + \mathbf{Z}_{k+1})(\mathbf{Z}_k - \mathbf{Z}_{k+1})^{\text{H}} \\
&\quad + (-\mathbf{Y} + \mathbf{V}_{k+1})(\mathbf{V}_k - \mathbf{V}_{k+1})^{\text{H}} \\
&\quad + \mathbf{Z}_k\mathbf{V}_k^{\text{H}} - \mathbf{Z}_{k+1}\mathbf{V}_{k+1}^{\text{H}}\big) \\
&\quad + \lambda_1\text{trace}\big(\left(\mathbf{Z}_k - \mathbf{Z}_{k+1}\right)\mathbf{Z}_{k+1}^{\text{H}}\mathbf{P}_k\big) \\
&\quad + \lambda_2\text{trace}\big(\mathbf{Q}_k(\mathbf{V}_k - \mathbf{V}_{k+1})\mathbf{V}_{k+1}^{\text{H}}\big)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2}\text{trace}\big((\mathbf{Z}_k-\mathbf{Z}_{k+1})(\mathbf{Z}_k-\mathbf{Z}_{k+1})^{\text{H}}\\
&\quad+(\mathbf{V}_k-\mathbf{V}_{k+1})(\mathbf{V}_k-\mathbf{V}_{k+1})^{\text{H}}\big)\\
&\quad+\text{trace}\big(\mathbf{Z}_k\mathbf{V}_k^{\text{H}}-\mathbf{Z}_{k+1}\mathbf{V}_{k+1}^{\text{H}}\big)\\
&\quad+\text{trace}\big(-\mathbf{V}_k(\mathbf{Z}_k-\mathbf{Z}_{k+1})^{\text{H}}\big)\\
&\quad+\text{trace}\big(-\mathbf{Z}_k(\mathbf{V}_k-\mathbf{V}_{k+1})^{\text{H}}\big)\\
&= \frac{1}{2}\text{trace}\big((\mathbf{Z}_k-\mathbf{Z}_{k+1})(\mathbf{Z}_k-\mathbf{Z}_{k+1})^{\text{H}}\\
&\quad+(\mathbf{V}_k-\mathbf{V}_{k+1})(\mathbf{V}_k-\mathbf{V}_{k+1})^{\text{H}}\big)\\
&\quad+\text{trace}\big((\mathbf{Z}_k-\mathbf{Z}_{k+1})(\mathbf{V}_{k+1}-\mathbf{V}_k)^{\text{H}}\big)\\
&= \frac{1}{2}\text{trace}\big((\mathbf{Z}_k-\mathbf{Z}_{k+1}-\mathbf{V}_k+\mathbf{V}_{k+1})\cdot\\
&\quad(\mathbf{Z}_k-\mathbf{Z}_{k+1}-\mathbf{V}_k+\mathbf{V}_{k+1})^{\text{H}}\big)\\
&= \frac{1}{2}\|\mathbf{Z}_k-\mathbf{Z}_{k+1}-\mathbf{V}_k+\mathbf{V}_{k+1}\|_{\text{F}}^2 \geq 0 \qquad (10)
\end{aligned}
$$

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

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

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 (9) and (10) yields

$$
\begin{aligned}
&f(\mathbf{Z}_k,\mathbf{V}_k)-f(\mathbf{Z}_{k+1},\mathbf{V}_{k+1})\\
&\geq \frac{\lambda_1}{2}\text{trace}\Big(\left(\mathbf{Z}_k-\mathbf{Z}_{k+1}\right)\left(\mathbf{Z}_k-\mathbf{Z}_{k+1}\right)^{\text{H}}\mathbf{P}_k\Big)\\
&\quad+\frac{\lambda_2}{2}\text{trace}\Big(\mathbf{Q}_k\left(\mathbf{V}_k-\mathbf{V}_{k+1}\right)\left(\mathbf{V}_k-\mathbf{V}_{k+1}\right)^{\text{H}}\Big)\\
&\quad+\frac{1}{2}\text{trace}\big((\mathbf{Z}_k-\mathbf{Z}_{k+1})(\mathbf{Z}_k-\mathbf{Z}_{k+1})^{\text{H}}\big)\\
&\geq \frac{\lambda_1}{2}\text{trace}\Big(\left(\mathbf{Z}_k-\mathbf{Z}_{k+1}\right)\left(\mathbf{Z}_k-\mathbf{Z}_{k+1}\right)^{\text{H}}\mathbf{P}_k\Big)\\
&\quad+\frac{\lambda_2}{2}\text{trace}\Big(\mathbf{Q}_k\left(\mathbf{V}_k-\mathbf{V}_{k+1}\right)\left(\mathbf{V}_k-\mathbf{V}_{k+1}\right)^{\text{H}}\Big)\\
&\geq \frac{\lambda_1}{2}\sum_i^M\zeta_i\left(\mathbf{P}_k\right)\zeta_{M-i+1}\left(\left(\mathbf{Z}_k-\mathbf{Z}_{k+1}\right)\left(\mathbf{Z}_k-\mathbf{Z}_{k+1}\right)^{\text{H}}\right)\\
&\quad+\frac{\lambda_2}{2}\sum_i^M\zeta_i\left(\mathbf{Q}_k\right)\zeta_{M-i+1}\left(\left(\mathbf{V}_k-\mathbf{V}_{k+1}\right)\left(\mathbf{V}_k-\mathbf{V}_{k+1}\right)^{\text{H}}\right)\\
&\geq \frac{\lambda_1}{2}\zeta_{\min}^{(\mathbf{P})}\times\|\mathbf{Z}_k-\mathbf{Z}_{k+1}\|_{\text{F}}^2+\frac{\lambda_1}{2}\zeta_{\min}^{(\mathbf{Q})}\times\|\mathbf{V}_k-\mathbf{V}_{k+1}\|_{\text{F}}^2
\end{aligned}
$$

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

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

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

B. Proof of Theorem 2

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

$$
\begin{aligned}
\|[\mathbf{V},\mu\mathbf{1}]\|_{2,1} &= \sum_{i=1}^M\sqrt{\|\mathbf{V}_{i,:}\|_2^2+\mu^2} \geq \sum_{i=1}^M|\mu| = |\mu|M\\
\|[\mathbf{Z},\mu\mathbf{I}]\|_* &= \text{trace}\Big(\left(\mathbf{Z}\mathbf{Z}^{\text{H}}+\mu^2\mathbf{I}\right)^{\frac{1}{2}}\Big)\\
&\geq \left(\text{trace}\left(\mathbf{Z}\mathbf{Z}^{\text{H}}+\mu^2\mathbf{I}\right)\right)^{\frac{1}{2}}\\
&= \left(\text{trace}\left(\mathbf{Z}\mathbf{Z}^{\text{H}}\right)+M\mu^2\right)^{\frac{1}{2}}\\
&\geq \left(M\mu^2\right)^{\frac{1}{2}} = |\mu|\sqrt{M}. \qquad (11)
\end{aligned}
$$

Inequality (11) 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 $\mathbf{X}$. Considering the above two inequalities and $\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_{\text{F}}^2\geq 0$, we can prove that the objective function in (7) is bounded below as $f(\mathbf{Z},\mathbf{V})=\frac{1}{2}\|\mathbf{Y}-\mathbf{Z}-\mathbf{V}\|_{\text{F}}^2+\lambda_1\|[\mathbf{Z},\mu\mathbf{I}]\|_*+\lambda_2\|[\mathbf{V},\mu\mathbf{1}]\|_{2,1}\geq|\mu|(\lambda_1\sqrt{M}+\lambda_2 M)$.

C. Proof of Theorem 3

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 $\lim_{k\to\infty}\|\mathbf{Z}_k-\mathbf{Z}_{k+1}\|_{\text{F}}=0$ and $\lim_{k\to\infty}\|\mathbf{V}_k-\mathbf{V}_{k+1}\|_{\text{F}}=0$ in Theorem 1 and (9), 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 (7). Since Problem (7) is convex w.r.t. $\mathbf{Z}$ and $\mathbf{V}$, the stationary point is globally optimal.

HUIPING HUANG
Chalmers University of Technology, Gothenburg, Sweden

QI LIU, *Member, IEEE*
South China University of Technology, Guangzhou, Guangdong, China
Pazhou Lab, Guangzhou, China

HING CHEUNG SO, *Fellow, IEEE*
City University of Hong Kong, Hong Kong, China

ABDELHAK M. ZOUBIR, *Fellow, IEEE*
Darmstadt University of Technology, Darmstadt, Germany

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[2] H. L. Van Trees, "Chapter 1 - Introduction," in Hoboken, NJ, USA:*Optimum Array Processing*. Wiley, 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. Amsterdam, The Netherlands: Elsevier, 2014, vol. 3, pp. 463–502.

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

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

[7] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[8] B. Wang, Y. D. Zhang, and W. Wang, "Robust DOA estimation in the presence of miscalibrated sensors," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 1073–1077, Jul. 2017.

[9] Q. Wang, T. Dou, H. Chen, W. Yan, and W. Liu, "Effective block sparse representation algorithm for DOA estimation with unknown mutual coupling," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2622–2625, Dec. 2017.

[10] Z. Yang, R. C. de Lamare, and W. Liu, "Sparsity-based STAP using alternating direction method with gain/phase errors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 6, pp. 2756–2768, Dec. 2017.

[11] H. Huang, M. Fauß, and A. M. Zoubir, "Block sparsity-based DOA estimation with sensor gain and phase uncertainties," in *Proc. Eur. Signal Process. Conf.*, A. Coruna, Spain, 2019, pp. 1–5.

[12] B. Lin, J. Liu, M. Xie, and J. Zhu, "Direction-of-arrival tracking via low-rank plus sparse matrix decomposition," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 1302–1305, 2015.

[13] A. Das, "A Bayesian sparse-plus-low-rank matrix decomposition method for direction-of-arrival tracking," *IEEE Sensors J.*, vol. 17, no. 15, pp. 4894–4902, Aug. 2017.

[14] 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., May 2014, vol. 9109, pp. 130–140, International Society for Optics and Photonics.

[15] P. P. Markopoulos, N. Tsagkarakis, D. A. Pados, and G. N. Karystinos, "Realified L1-PCA for direction-of-arrival estimation: Theory and algorithms," *EURASIP J. Adv. Signal Process.*, vol. 30, pp. 1–16, 2019.

[16] B.-K. Yeo and Y. Lu, "Array failure correction with a genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 47, no. 5, pp. 823–828, May 1999.

[17] 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 *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, 2012, pp. 3709–3712.

[18] G. Oliveri, P. Rocca, and A. Massa, "Reliable diagnosis of large linear arrays-a bayesian compressive sensing approach," *IEEE Trans. Antennas Propag.*, vol. 60, no. 10, pp. 4627–4636, Oct. 2012.

[19] C. Zhu, W.-Q. Wang, H. Chen, and H. C. So, "Impaired sensor diagnosis, beamforming, and DOA estimation with difference coarray processing," *IEEE Sensors J.*, vol. 15, no. 7, pp. 3773–3780, Jul. 2015.

[20] M. Wang, Z. Zhang, and A. Nehorai, "Direction finding using sparse linear arrays with missing data," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, New Orleans, LA, USA, 2017, pp. 3066–3070.

[21] 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 Trans. Signal Process.*, vol. 67, no. 12, pp. 3213–3226, Jun. 2019.

[22] B. Ng, J. P. Lie, M. Er, and A. Feng, "A practical simple geometry and gain/phase calibration technique for antenna array processing," *IEEE Trans. Antennas Propag.*, vol. 57, no. 7, pp. 1963–1972, Jul. 2009.

[23] M. Pesavento, A. B. Gershman, and K. M. Wong, "Direction finding in partly calibrated sensor arrays composed of multiple subarrays," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2103–2115, Sep. 2002.

[24] B. Liao and S. C. Chan, "Direction finding with partly calibrated uniform linear arrays," *IEEE Trans. Antennas Propag.*, vol. 60, no. 2, pp. 922–929, Feb. 2012.

[25] W. Suleiman, P. Parvazi, M. Pesavento, and A. M. Zoubir, "Non-coherent direction-of-arrival estimation using partly calibrated arrays," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5776–5788, Nov. 2018.

[26] F. Afkhaminia and M. Azghani, "Sparsity-based direction of arrival estimation in the presence of gain/phase uncertainty," in *Proc. Eur. Signal Process. Conf.*, Kos, Greece, 2017, pp. 2616–2619.

[27] H. Huang and A. M. Zoubir, "Low-rank and sparse decomposition for joint DOA estimation and contaminated sensors detection with sparsely contaminated arrays," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, Toronto, Canada, 2021, pp. 4615–4619.

[28] S. Vigneshwaran, N. Sundararajan, and P. Saratchandran, "Direction of arrival (DoA) estimation under array sensor failures using a minimal resource allocation neural network," *IEEE Trans. Antennas Propag.*, vol. 55, no. 2, pp. 334–343, Feb. 2007.

[29] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.

[30] N. Vaswani, Y. Chi, and T. Bouwmans, "Rethinking PCA for modern data sets: Theory, algorithms, and applications," *Proc. IEEE*, vol. 106, no. 8, pp. 1274–1276, Aug. 2018.

[31] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust PCA in image and video processing," *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, Aug. 2018.

[32] 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 *Proc. Int. Conf. Neural Inf. Process. Syst.*, Red Hook, USA, 2009, pp. 2080–2088.

[33] 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 *Proc. Conf. Comput. Vis. Pattern Recognit.*, Colorado Spring, USA, 2011, pp. 1673–1680.

[34] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, Granada, Spain, 2011, pp. 1–9.

[35] Y. Bando et al., "Speech enhancement based on bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 2, pp. 215–230, Feb. 2018.

[36] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3047–3064, May 2012.

[37] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, Madison, USA, 2010, pp. 663–670.

[38] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[39] C. Lu, Z. Lin, and S. Yan, "Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 646–654, Feb. 2015.

[40] Q. Liu, Y. Gu, and H. C. So, "DOA estimation in impulsive noise via low-rank matrix approximation and weakly convex optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3603–3616, Dec. 2019.

[41] H. Zhang, Z. Lin, C. Zhang, and E. Chang, "Exact recoverability of robust PCA via outlier pursuit with tight recovery bounds," in *Proc. AAAI Conf. Artif. Intell.*, Austin Texas, USA, 2015, vol. 29, no. 1, pp. 3143–3149.

[42] X. Li, J. Ren, S. Rambhatla, Y. Xu, and J. Haupt, "Robust PCA via dictionary based outlier pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, Calgary, Canada, 2018, pp. 4699–4703.

[43] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression," in *Proc. Int. Conf. Pattern Recognit.*, Tsukuba, Japan, 2012, pp. 2805–2808.

[44] P. Rodrêguez and B. Wohlberg, "Performance comparison of iterative reweighting methods for total variation regularization," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, 2014, pp. 1758–1762.

[45] J.-F. Cai, E. J. Candés, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[46] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[47] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2009.

[48] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 183–195, Jan. 2014.

[49] D. Straszak and N. K. Vishnoi, "Iteratively reweighted least squares and slime mold dynamics: Connection and convergence," *Math. Program.*, vol. 194, pp. 685–717, 2022.

[50] L. Wang, R. C. de Lamare, and M. Haardt, "Direction finding algorithms based on joint iterative subspace optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2541–2553, Oct. 2014.

[51] L. Qiu, Y. Cai, R. C. de Lamare, and M. Zhao, "Reduced-rank DOA estimation algorithms based on alternating low-rank decomposition," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 565–569, May 2016.

[52] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends in Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[53] J. P. Delmas, "Chapter 16 - Performance Bounds and Statistical Analysis of DOA Estimation," in *Academic Press Library in Signal Processing: 3*, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds. New York, NY, USA: Elsevier, 2014, vol. 3, pp. 719–764.