

# DOA Estimation in Impulsive Noise via Low-Rank Matrix Approximation and Weakly Convex Optimization

QI LIU 

City University of Hong Kong, Hong Kong

YUANTAO GU , Senior Member, IEEE

Tsinghua University, Beijing, China

HING CHEUNG SO , Fellow, IEEE

City University of Hong Kong, Hong Kong

**Conventional direction-of-arrival (DOA) estimators are vulnerable to impulsive noise. In this paper, to tackle this issue, a class of weakly convex-inducing penalties is introduced for robust DOA estimation via low-rank matrix approximation, where  $\ell_{2,1}$ -norm is adopted as the metric for suppressing the outliers. Two iterative algorithms are developed to construct the noise-free data matrix. To avoid determining the number of sources, the DOAs are estimated by exploiting the special joint diagonalization structure of the constructed signal covariance matrix. Compared with several existing algorithms, the proposed methods enjoy faster computation, similar DOA estimation performance against impulsive noise and requiring no *a priori* information of the source number. Numerical experiments are included to demonstrate the outlier-resistance of our solutions.**

Manuscript received June 28, 2018; revised December 24, 2018; released for publication March 24, 2019. Date of publication April 9, 2019; date of current version December 5, 2019.

DOI No. 10.1109/TAES.2019.2909728

Refereeing of this contribution was handled by H. Mir.

This work was supported in part by the NSFC/RGC Joint Research Scheme under a Grant sponsored by the National Natural Science Foundation of China and in part by the Research Grants Council of Hong Kong (61531166005 and N\_CityU 104/15).

Authors' addresses: Q. Liu and H. C. So are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, E-mail: (lq\_skyven@126.com; hcso@ee.cityu.edu.hk); Y. Gu is with the Department of Electronic Engineering and Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China, E-mail: (gyt@tsinghua.edu.cn). (*Corresponding author: Hing Cheung So.*)

0018-9251 © 2019 IEEE

## I. INTRODUCTION

Narrowband far-field source localization using sensor arrays has been playing a fundamental role in many areas including radar, sonar, and wireless communications [1]–[3]. In array signal processing, this problem is also known as direction-of-arrival (DOA) estimation. The challenge of direction finding arises from the fact that the observed snapshots are nonlinear functions of the DOAs. The well-known subspace-based methods, e.g., MUSIC [4], PUMA [5], and Capon's method [6], require that the number of sources is known *a priori*. Unfortunately, the source number is usually not available in practice, and source enumeration is not an easy task [7]. Furthermore, these estimators cannot perform well in the presence of outlier-contaminated observations as their derivations are based on additive Gaussian noise assumption.

In practical applications, the measurement noise may be of different kinds and exhibits non-Gaussian properties, where impulsive noise is a typical case [8], e.g., Gaussian mixture model (GMM) [9], compound Gaussian model (CGM) [10], and symmetric  $\alpha$ -stable ( $S\alpha S$ ) distribution [11]. Since the impulsive noise can model outliers well, it has also been widely studied in robust statistics [12], [13]. Compared with the Gaussian noise, the probability density function (PDF) of impulsive noise has heavier tails, which exceeds a few standard deviations than the Gaussian distribution. Therefore, the performance of the conventional second-order statistics-based DOA estimators [14], [15] may greatly degrade when encountering the impulsive noise. That is to say, they are no longer appropriate in a non-Gaussian environment, largely due to their nonrobustness against even a small number of outliers. Based on the fractional lower-order statistics, various DOA estimators have been proposed to deal with the outliers, including ROC-MUSIC [16], FLOM-MUSIC [17], and SCM/TCM-MUSIC [18]. However, these algorithms are suboptimal and offer a satisfactory DOA estimation performance at the expense of requiring a large number of sample sizes. Other robust DOA estimation algorithms such as MM-MUSIC [19], MT-MUSIC [20], EM-MUSIC [9], and  $\ell_p$ -MUSIC [21], are proposed to compute the sample covariance matrix and then apply conventional subspace-type technique for direction finding, which achieve higher resolution than the fractional lower-order statistics-based algorithms. Note that the  $\ell_p$ -MUSIC exhibits a number of advantages over the conventional subspace-type algorithms and several outlier-resistant approaches. However, the  $\ell_p$ -MUSIC cannot work without the prior knowledge of the number of sources, and so is the robust G-MUSIC [22]. A DOA estimator [23] via algebraic structure of the noise subspace has been developed for source enumeration, but it is not robust against the impulsive noise.

Low-rank matrix approximation (LRMA) has attracted considerable research interests in many important areas, such as machine learning, signal and image processing, and computer vision, especially with high-dimensional datasets, which is also widely used in principal component

analysis (PCA) [24]. The purpose of LRMA is to extract the low-dimensional subspaces or principal components of the matrices constructed from these signals. Although PCA, which can be easily realized by the truncated singular value decomposition (SVD), is a standard tool to tackle this task, it fails to work in the presence of non-Gaussian disturbances and/or outliers. It is because the design of PCA basically utilizes the  $\ell_2$ -norm minimization or least squares, indicating that it is only perfect for additive Gaussian noise, which motivates the development of the robust PCA (RPCA) [25], [26]. However, the RPCA is computationally demanding because full SVD is required. Since the covariance matrix of uniform noise is an arbitrary unknown diagonal matrix instead of a scaled identity matrix, to circumvent this problem, IRMD-MUSIC [27] is proposed to determine the noise-free covariance matrix by nuclear norm minimization. Furthermore, Pal and Vaidyanathan [28] have suggested using a low-rank matrix denoising approach followed by a MUSIC-like subspace method to estimate the DOAs. In [29], by exploiting the Toeplitz structure of the covariance matrix of the array output, a low-rank matrix reconstruction method for DOA estimation is devised via applying alternating direction method of multipliers (ADMM) [30] as the solver. However, in the presence of impulsive noise, their performance of DOA estimation will degrade substantially.

In this paper, by adopting the LRMA framework, our developed methods are able to achieve faster computation, but enjoy comparable DOA estimation performance in impulsive noise. Our contributions lie in combining the concept of weakly convexity with LRMA framework to derive two iterative methods for noise-free data matrix construction in the presence of impulsive noise. Note that there is no need for the source number information as we exploit the special joint diagonalization structure of the constructed signal covariance matrix. Also, no assumptions on the models of noises and signals, and sparsity of the sensor array, are required. Furthermore, our schemes are computationally attractive and able to compute the spatial spectrum without rank information.

The rest of this paper is organized as follows. In Section II, the signal model for DOA estimation is established and the corresponding preliminaries are introduced, including weakly convex sparsity-inducing function and LRMA. The proposed methods for DOA estimation are derived in Section III. Numerical examples are provided in Section IV, followed by conclusions in Section V.

## II. PROBLEM FORMULATION AND PRELIMINARIES

### A. Signal Model

Consider  $K$  narrow-band far-field uncorrelated sources  $s_k$ ,  $k \in \{1, \dots, K\}$ , impinging on a uniform linear array (ULA) of omnidirectional  $(2M + 1)$  sensors from directions  $\theta_k \in [-90^\circ, 90^\circ]$ , where the distance between adjacent sensors is  $\bar{d}$ . To avoid phase ambiguity, the intersensor spacing is set to be half the wavelength  $\bar{\lambda}$ , i.e.,  $\bar{d} = \bar{\lambda}/2$ . Let  $N$  be the number of snapshots, at the  $t$ th time instant, the

$(2M + 1) \times 1$  array observation output is written as

$$\mathbf{x}(t) = \sum_{k=1}^K \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t) + \mathbf{q}(t), \quad t = 1, \dots, N \quad (1)$$

where  $\mathbf{x}(t) = [x_{-M}(t), \dots, x_M(t)]^T$  with superscript  $T$  being the transpose operator,  $\mathbf{a}(\theta) = [a_1(\theta), \dots, a_M(\theta)]^T$  is the array response to a unit-amplitude source in direction  $\theta$ ,  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^T$  is the source direction vector,  $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$  is the  $(2M + 1) \times K$  array steering matrix with  $\mathbf{a}(\theta_k) = [e^{j2\pi M\tau(k)}, \dots, 1, \dots, e^{-j2\pi M\tau(k)}]^T$ ,  $\tau(k) = \bar{d}/\bar{\lambda} \sin(\theta_k)$ ,  $j = \sqrt{-1}$ ,  $\mathbf{s}(t) = [s_1(t), \dots, s_K(t)]^T$  contains the source complex signal amplitudes at time  $t$ , and  $\mathbf{q}(t) = [q_1(t), \dots, q_M(t)]^T$  is the additive noise vector at time  $t$ .

Collecting the  $N$  snapshots, the matrix form of (1) can be written as

$$\mathbf{X} = \mathbf{A}(\boldsymbol{\theta}) \mathbf{S} + \mathbf{Q} \quad (2)$$

where  $\mathbf{X} = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)] \in \mathbb{C}^{(2M+1) \times N}$ ,  $\mathbf{S} = [\mathbf{s}(t_1), \dots, \mathbf{s}(t_N)] \in \mathbb{C}^{K \times N}$ , and  $\mathbf{Q} = [\mathbf{q}(t_1), \dots, \mathbf{q}(t_N)] \in \mathbb{C}^{(2M+1) \times N}$ . In the presence of impulsive noise, the PDF of  $\mathbf{q}(t)$  has heavier tails than the Gaussian distribution, which has a few large values. These large values are considered as outliers. In this study, the noise data matrix  $\mathbf{Q}$  is assumed a column-wise sparse matrix. It is because when the jammer signals arrive at some sensors at time  $t$ , these sensor outputs corrupted by the jammers at the  $t$ th snapshot will turn to large values, so that the element modulus of the  $t$ th column of  $\mathbf{Q}$  will be large with a high probability and  $\mathbf{Q}$  is finally a column-wise sparsity matrix when collecting all snapshots. Since the array steering matrix  $\mathbf{A}(\boldsymbol{\theta})$  is of full column rank, it can be easily verified that  $\text{rank}(\mathbf{A}(\boldsymbol{\theta})\mathbf{S}) = r \leq K$  for  $K < (2M + 1) < N$ . Consequently, the noise-free data matrix  $\mathbf{P} \triangleq \mathbf{A}(\boldsymbol{\theta})\mathbf{S}$  is of low-rank. We are interested in recovering the noise-free data matrix  $\mathbf{P}$  from the noisy  $\mathbf{X}$  corrupted by  $\mathbf{Q}$ . The noise-free data covariance matrix  $\mathbf{R}$  is computed as

$$\mathbf{R} = \frac{1}{N} E\{\mathbf{P}\mathbf{P}^H\} = \mathbf{A}(\boldsymbol{\theta}) \mathbf{R}_s \mathbf{A}^H(\boldsymbol{\theta}) \quad (3)$$

where  $\mathbf{R}_s = E\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \text{diag}(\sigma_s^2)$  and  $\boldsymbol{\sigma}_s^2 = [\sigma_1^2, \dots, \sigma_K^2]^T$ . Here,  $E\{\cdot\}$ ,  $^H$ , and  $\text{diag}(\cdot)$  denote the expectation, Hermitian transpose, and diagonal matrix with elements being its diagonal, respectively.

### B. Low-Rank Matrix Approximation

In LRMA, it is assumed that the data lie near some low-dimensional subspace, then the matrix should have (approximately) low-rank. In order to recover the low-rank matrix  $\mathbf{P}$  from the given observation matrix  $\mathbf{X}$  corrupted by the noise matrix  $\mathbf{Q}$ , it is straightforward to consider the regularized rank minimization problem, which seeks the best rank- $r$   $\mathbf{P}$  by solving

$$\min_{\mathbf{P}} \text{rank}(\mathbf{P}), \quad \text{s.t. } \mathbf{X} = \mathbf{P} + \mathbf{Q}. \quad (4)$$

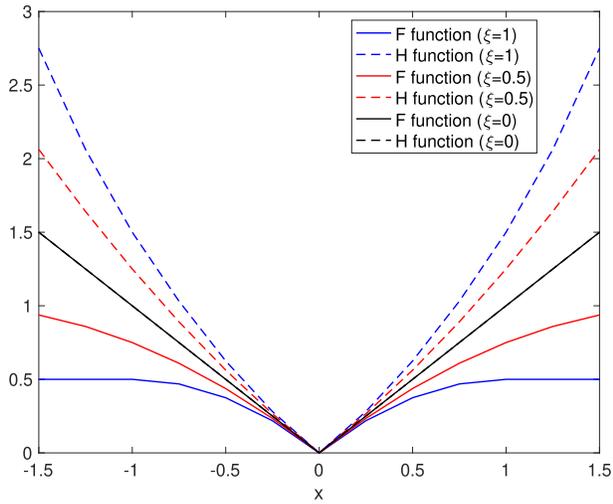


Fig. 1. Relationship between  $F(\cdot)$  and  $H(\cdot)$ .

Unfortunately, the problem of rank minimization in (4) is NP-hard in general because the rank is discrete and nonconvex. Analogous to the strategy of employing the  $\ell_1$ -norm instead of the  $\ell_0$ -norm for sparse signal recovery [31], convex relaxation of (4) leads to the nuclear norm minimization

$$\min_{\mathbf{P}} \|\mathbf{P}\|_*, \quad \text{s.t. } \mathbf{X} = \mathbf{P} + \mathbf{Q} \quad (5)$$

where the nuclear norm  $\|\mathbf{P}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{P})$  denotes the sum of singular values of  $\mathbf{P}$ . In the presence of the outliers, RPCA [26] is applied to modify (5) as

$$\min_{\mathbf{P}, \mathbf{Q}} \|\mathbf{P}\|_* + \lambda \|\mathbf{Q}\|_1, \quad \text{s.t. } \mathbf{X} = \mathbf{P} + \mathbf{Q} \quad (6)$$

where  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm of a matrix and  $\lambda$  is a regularization parameter. Although (6) is a convex optimization and global minimization is guaranteed, it has a high computational cost even fast algorithms are employed because full SVD is required [24], [32]–[34].

Motivated by the fact that a weakly convex function as an approximation of the  $\ell_0$ -norm is able to better induce sparsity than the  $\ell_1$ -norm [13], [35], we propose to extend a weakly convex function in LRMA. Another possible way to relax (4) with function  $J$  is

$$\min_{\mathbf{P}} J(\mathbf{P}), \quad \text{s.t. } \mathbf{X} = \mathbf{P} + \mathbf{Q} \quad (7)$$

where  $J(\cdot)$  belongs to a class of weakly convex sparsity inducing functions, and is defined as

$$J(\mathbf{P}) \triangleq \sum_{i=1}^M F(\sigma_i(\mathbf{P})) \quad (8)$$

with  $F(\cdot)$  being a weakly convex sparseness measure. That is,  $F(\cdot)$  is weakly convex if and only if there exists a convex function  $H(p) = F(p) + \xi p^2$  when  $\xi > 0$ .  $F(\cdot)$  at  $\xi = 0$  is the special case of  $\ell_1$ -norm. Taking one of weakly convex penalties, namely, minimax concave penalty (MCP) [35] as an example, we can see that both  $F(\cdot)$  and  $H(\cdot)$  are convex at  $\xi = 1$ , while they are weakly convex at  $\xi = 0.5$ , in Fig. 1.

A class of weakly convex (nonconvex) functions has been proposed to induce sparsity [35]. The definition is:

DEFINITION 1

- 1)  $F(0) = 0$ ,  $F(\cdot)$  is even and not identically zero.
- 2)  $F(\cdot)$  is nondecreasing on  $[0, +\infty)$ .
- 3)  $F(\cdot)$  is concave on  $[0, +\infty)$ .
- 4)  $F(\cdot)$  is weakly convex on  $[0, +\infty)$ .

From [35, Lemma 1.1],  $F(p)/p \rightarrow \bar{\alpha}$  as  $p \rightarrow 0^+$  for  $\bar{\alpha} > 0$ . Hence, according to Definition 1,  $\rho \triangleq \xi/\bar{\alpha}$  characterizes the nonconvexity of  $F(\cdot)$  and  $J(\cdot)$ , where  $\xi$  divided by  $\bar{\alpha}$  is to remove the scaling effect on the penalty. The  $J(\cdot)$  belongs to a class of weakly convex sparsity inducing functions in Definition 1, and possesses some “favorable” properties so that (7) can be solved by tractable algorithms and results in low-rank solutions. Compared to (4), the main advantage of (7) lies in the fact that it does not require the rank information, i.e.,  $K$ . Thus, works [35]–[38] adopting other low-rank inducing penalties in (7) empirically demonstrate better recovery performance when compared with the nuclear norm.

In this paper, motivated by the fact that  $\ell_{2,1}$ -norm can detect outliers with column-wise sparsity in the presence of impulsive noise, we apply it as the metric for the residual error to propose the following robust formulation:

$$\min_{\mathbf{P}, \mathbf{Q}} J(\mathbf{P}) + \lambda \|\mathbf{Q}\|_{2,1}, \quad \text{s.t. } \|\mathbf{X} - \mathbf{P} - \mathbf{Q}\|_F^2 \leq \epsilon \quad (9)$$

where the  $\ell_{2,1}$ -norm is defined as  $\|\mathbf{Q}\|_{2,1} \triangleq \sum_{j=1}^N \sqrt{\sum_{i=1}^{2M+1} ([\mathbf{Q}]_{i,j})^2}$  and  $\epsilon$  is error tolerance parameter. In this formulation, large errors due to the outliers for each data point are  $\|[\mathbf{Q}]_{\cdot,j}\|$ , which are not squared. Herein, the large errors do not dominate the objective function. Note that it is easily proved that  $\ell_{2,1}$ -norm is a norm because it satisfies three conditions, namely, positive scalability, triangle inequality, and existence of a zero vector [39].

### III. PROPOSED DOA ESTIMATORS

#### A. LRMA-ADMM

In this section, to stably recover  $\mathbf{P}$  and  $\mathbf{Q}$ , instead of directly solving (9), we solve the following problem:

$$\min_{\mathbf{P}, \mathbf{Q}} J(\mathbf{P}) + \lambda \|\mathbf{Q}\|_{2,1} + \frac{\mu}{2} \|\mathbf{X} - \mathbf{P} - \mathbf{Q}\|_F^2. \quad (10)$$

It is well demonstrated in [40] and [41] that (10) is equivalent to (9) for some pairs of  $\lambda$  and  $\mu$ , where  $\mu$  is the penalty parameter. As a variant of the augmented Lagrangian method, ADMM is well-suited for distributed optimization, which utilizes a variable-splitting scheme to decouple components in the cost function and solves the optimization problem effectively in an alternating minimization manner. We apply ADMM to solve (10). Toward this goal, the augmented Lagrangian of (10) is written as

$$\begin{aligned} \mathcal{L}(\mathbf{P}, \mathbf{Q}, \Gamma) &= J(\mathbf{P}) + \lambda \|\mathbf{Q}\|_{2,1} + \langle \Gamma, \mathbf{X} - \mathbf{P} - \mathbf{Q} \\ &> + \frac{\mu}{2} \|\mathbf{X} - \mathbf{P} - \mathbf{Q}\|_F^2 \end{aligned} \quad (11)$$

where  $\Gamma$  is the dual variable, and  $\langle \cdot, \cdot \rangle$  denotes the standard trace inner product, i.e.,  $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}^H \mathbf{B})$ . Note that the last term of (10) is combined into  $\frac{\mu}{2} \|\mathbf{X} - \mathbf{P} - \mathbf{Q}\|_F^2$ . Thus, the separable function defined in (8) can be expressed as the Moreau envelope [42] of the weakly convex function, i.e.,

$$J^\#(\mathbf{P}) = \min_{\mathbf{V}} \left\{ J(\mathbf{V}) + \frac{\lambda}{2} \|\mathbf{P} - \mathbf{V}\|_F^2 \right\}. \quad (12)$$

If the nonconvexity parameter  $\rho$  is small enough such that  $\rho < \frac{1}{2\lambda}$  [35], then the objective function in (12) is strongly convex, and the minimizer is unique.

Now, we proceed to unveil each step of the ADMM for problem (10). Specifically, we optimize problem (11) with respect to one variable while fixing the others, which results in the following subproblems.

In the  $(t+1)$ th iteration, the update of  $\mathbf{P}$  is

$$\mathbf{P}^{t+1} = \arg \min_{\mathbf{P}} J(\mathbf{P}) + \frac{\mu^t}{2} \|\mathbf{P} - (\mathbf{X} - \mathbf{Q}^t + \Gamma^t / \mu^t)\|_F^2. \quad (13)$$

Since  $F(\cdot)$  is concave on  $[0, +\infty)$  in Definition 1, by the definition of the supergradient [43] and Lemma 1, we have

$$F(\sigma_i(\mathbf{P})) \leq F(\sigma_i(\mathbf{P}^t)) + f(\sigma_i(\mathbf{P}^t))(\sigma_i(\mathbf{P}) - \sigma_i(\mathbf{P}^t)) \quad (14)$$

where  $f(\sigma_i(\mathbf{P}^t)) \in \partial F(\sigma_i(\mathbf{P}^t))$ , and Lemma 1 is given by [44, Proposition 4.8].

LEMMA 1 Let  $F(\cdot)$  be a weakly convex function with parameter  $\rho$  on  $S \in \{1, 2, \dots, N\}$ . Then, the inequality

$$F(p_2) \geq F(p_1) + \langle f(p_1), p_2 - p_1 \rangle + \rho \|p_1 - p_2\|^2 \quad (15)$$

holds for all  $p_1 \in \text{int}S$ ,  $p_2 \in S$ , and  $f(p_1) \in \partial F(p_1)$ , where  $\|\cdot\|$  denotes the induced norm of the inner product in  $S$ , and the interior of  $S$  is denoted as  $\text{int}S$ .

Without loss of generality, for  $\mathbf{P} \in \mathbb{R}^{(2M+1) \times N}$  with rank  $K$ , we adopt the convention that its singular values are arranged in nonincreasing order

$$\sigma_1(\mathbf{P}) \geq \dots \geq \sigma_K(\mathbf{P}) > 0 = \sigma_{K+1}(\mathbf{P}) = \dots = \sigma_{2M+1}(\mathbf{P}).$$

Due to the antimonotone property of supergradient, we have

$$0 \leq f(\sigma_1(\mathbf{P}^t)) \leq \dots \leq f(\sigma_{2M+1}(\mathbf{P}^t)). \quad (16)$$

Hence, it motivates us to minimize the right-hand side (RHS) of (13) instead of  $F(\sigma_i(\mathbf{P}))$ . Moreover

$$\begin{aligned} \mathbf{P}^{t+1} &= \arg \min_{\mathbf{P}} \sum_{i=1}^{2M+1} F(\sigma_i(\mathbf{P}^t)) + f(\sigma_i(\mathbf{P}^t))(\sigma_i(\mathbf{P}) - \sigma_i(\mathbf{P}^t)) \\ &\quad + \frac{\mu^t}{2} \|\mathbf{P}^t - (\mathbf{X} - \mathbf{Q}^t + \Gamma^t / \mu^t)\|_F^2 \\ &= \arg \min_{\mathbf{P}} \sum_{i=1}^{2M+1} f(\sigma_i(\mathbf{P}^t)) \sigma_i(\mathbf{P}) \\ &\quad + \frac{\mu^t}{2} \|\mathbf{P}^t - (\mathbf{X} - \mathbf{Q}^t + \Gamma^t / \mu^t)\|_F^2 \\ &= \mathbf{U}_r T_{\lambda F}(\Sigma_r) \mathbf{V}_r^H \end{aligned} \quad (17)$$

where the global optimal solution is given by the singular value thresholding  $\mathbf{U}_r T_{\lambda F}(\Sigma_r) \mathbf{V}_r^H$  in [45, Lemma 3] and  $T_{\lambda F}(\cdot)$  is the proximal operator of the weakly convex function. The SVD of  $(\mathbf{X} - \mathbf{Q}^t + \Gamma^t / \mu^t)$  is defined as  $(\mathbf{X} - \mathbf{Q}^t + \Gamma^t / \mu^t) \triangleq \mathbf{U}[\Sigma_r, \mathbf{0}]\mathbf{V}^H$ ,  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices and  $\Sigma_r = \text{diag}(\sigma_1^2, \dots, \sigma_r^2)$  with  $\{\sigma_i^2\}_{i=1}^r$  being the singular values. The corresponding truncated SVD is computed as  $\mathbf{U}_r \Sigma_r \mathbf{V}_r^H$ , where  $\mathbf{U}_r$  and  $\mathbf{V}_r$  are composed of singular vectors corresponding to the  $r$  largest singular values in  $\mathbf{U}$  and  $\mathbf{V}$ , respectively.

REMARK 1 Note that the rank information is needed in SVD computation in (13). However, the rank information is usually not available in practice. Here, to use prediction rule to determine the rank information [46], we only need those singular values that are larger than a given threshold and their corresponding singular vectors. The prediction rule is defined as

$$\hat{r}^{t+1} = \begin{cases} r^t + 1, & \text{if } r^t < \hat{r}^t \\ \min(r^t + \text{round}(0.1\bar{d}), \bar{d}), & \text{if } r^t = \hat{r}^t \end{cases} \quad (18)$$

where  $\bar{d} = \min(2M+1, N)$ ,  $\hat{r}^t$  is the predicted rank, and  $r^t$  is the number of singular values in the  $\hat{r}^t$  singular values that are larger than the inverse of  $\mu^t$ . We initialize  $\hat{r}^t$  as  $\hat{r}^0 = 9$ , where  $\text{sgn}(\cdot)$  denotes the sign of a quantity with  $\text{sgn}(0) = 0$ .

The update of  $\mathbf{Q}$  is

$$\mathbf{Q}^{t+1} = \arg \min_{\mathbf{Q}} \lambda \|\mathbf{Q}\|_{2,1} + \frac{\mu^t}{2} \|\mathbf{Q} - (\mathbf{X} - \mathbf{P}^{t+1} + \Gamma^t / \mu^t)\|_F^2. \quad (19)$$

LEMMA 2 Let  $\mathbf{W}^t = \mathbf{X} - \mathbf{P}^{t+1} + \Gamma^t / \mu^t$ . Since the problem (19) is convex with  $\mathbf{Q}$ , it has a global solution  $\mathbf{Q}^*$ , where its  $j$ th column is [47]

$$\mathbf{Q}^*(:, j) = \begin{cases} \frac{\|\mathbf{W}^t(:, j)\|_2 - \frac{\lambda}{\mu^t}}{\|\mathbf{W}^t(:, j)\|_2} \mathbf{W}^t(:, j), & \text{if } \frac{\lambda}{\mu^t} < \|\mathbf{W}^t(:, j)\|_2 \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

It is well known that  $\ell_{2,1}$ -norm is a sparsity-inducing norm defined as the  $\ell_1$ -norm of the columns of  $\mathbf{Q}$ .

The update of the dual variable  $\Gamma$  is

$$\Gamma^{t+1} = \Gamma^t + \mu^t (\mathbf{X} - \mathbf{P}^{t+1} - \mathbf{Q}^{t+1}) \quad (21)$$

where  $\mu^{t+1}$  is updated by  $\min(\mu_{\max}^t, \rho \mu^t)$  and  $\rho > 1$ . It is worth mentioning that a choice of stepsize to be increasing (but upper bounded) is effective in improving the convergence of ADMM for nonconvex optimization [48]. The steps of LRMA-ADMM are summarized in Algorithm 1.

We now analyze the convergence of LRMA-ADMM.

From Definition 1, we observe that the nonconvex regularizer is continuous, concave, smooth, differentiable, and monotonically increasing on  $[0, +\infty)$ , and its gradient is nonnegative and monotonically decreasing.

---

**Algorithm 1:** LRMA-ADMM.

---

**Input:**  $\mathbf{X}, M, N$   
**Initialize:**  $\lambda, \hat{r}^0, \mu^0, \rho$   
**for**  $t = 0, 1, \dots$  **do**  
    **repeat**  
        Update  $\mathbf{P}^{t+1}$  according to (17);  
        Update  $\mathbf{Q}^{t+1}$  according to (20);  
        Update  $\Gamma^{t+1}$  according to (21);  
         $\mu^{t+1} = \min(\mu_{\max}^t, \rho\mu^t)$ ;  
    **until** termination condition satisfied.  
**end for**  
**Output:**  $\hat{\mathbf{P}}$

---

**THEOREM 1** Suppose  $\sum_{i=1}^t \frac{\mu^i + \mu^{i-1}}{2(\mu^{i-1})^2} < \infty$  and arbitrary starting point  $\{\mathbf{P}^0, \Gamma^0\}$ , the sequence  $\{\mathbf{P}^t, \mathbf{Q}^t, \Gamma^t\}$  generated by LRMA-ADMM via (17), (20), and (21) is bounded and converges to a stationary point of the problem (11) if  $\lim_{t \rightarrow \infty} (\mathbf{P}^t - \mathbf{P}^{t+1}) = \mathbf{0}$ .

**PROOF** See Appendix. ■

### B. LRMA-IRLS

In this section, we consider a special case of weakly convex function, i.e., Schatten  $p$ -quasi norm of a matrix. Using the Schatten  $p$ -quasi norm  $\|\mathbf{P}\|_{S_p}^p = \sum_i \sigma_i^p(\mathbf{P})$  with  $0 < p < 1$ , i.e., the  $l_p$ -norm of the singular values, instead of the nuclear norm could further improve the performance. When  $p = 1$ , the Schatten  $p$ -norm  $\|\mathbf{P}\|_{S_p}^p$  is the trace norm or nuclear norm. When  $p \rightarrow 0$ , the Schatten  $p$ -norm  $\|\mathbf{P}\|_{S_p}^p$  will approximate the rank of  $\mathbf{P}$ .

**REMARK 2** It needs to be emphasized that the widely used  $p$ -quasi norm ( $0 \leq p < 1$ ) in the literature for sparse recovery does not belong to the class of sparsity-inducing penalties considered in this paper [35]. This is due to the fact that the function  $|x|^p$ ,  $p \in [0, 1)$  goes against Definition 1 4) and Lemma 1. However, with approximation to  $|x|^p$  to avoid infinite derivative around zero point, i.e.,  $F(x) = |x|(|x| + \epsilon)^{p-1}$ ,  $p \in [0, 1)$ ,  $\epsilon > 0$ ,  $F(x)$  satisfies Definition 1. Similarly, extending the  $p$ -quasi norm of a vector to the Schatten  $p$ -quasi norm of a matrix, Schatten  $p$ -quasi norm still goes against the Definition 1 4) but satisfies 1)–3), which has been proved in [49].

We exploit the fact

$$\begin{aligned}
 \|\mathbf{P}\|_{S_p}^p &= \sum_{i=1}^{2M+1} (\sigma_i(\mathbf{P}^T \mathbf{P}))^{\frac{p}{2}} \leq \sum_{i=1}^{2M+1} (\sigma_i(\mathbf{P}^T \mathbf{P}) + \epsilon^2)^{\frac{p}{2}} \\
 &= \sum_{i=1}^{2M+1} (\sigma_i(\mathbf{P}^T \mathbf{P} + \epsilon^2 \mathbf{I}))^{\frac{p}{2}} \\
 &= \left\| \begin{bmatrix} \mathbf{P} \\ \epsilon \mathbf{I} \end{bmatrix} \right\|_{S_p}^p \\
 &\triangleq J(\mathbf{P}). \tag{22}
 \end{aligned}$$

For  $p \geq 1$ ,  $J(\mathbf{P})$  is convex with respect to  $\mathbf{P}$  for a given  $\epsilon$ , which can be easily proved by using the convexity of Schatten  $p$ -quasi norm when  $p \geq 1$ . Note that in sparse recovery [50], [51] and matrix completion [36], [52], (22) is popular to use for approximating the nonsmooth functions, e.g., the  $p$ -quasi norm when  $0 < p \leq 1$ .

**REMARK 3** Since  $J(\mathbf{P})$  in (22) is concave on  $[0, +\infty)$  and smooth, the inequality in Lemma 1 holds when  $\rho = 0$ , which is the result of

$$(\epsilon^2 + |x|^2)^{\frac{p}{2}} - (\epsilon^2 + |y|^2)^{\frac{p}{2}} - \frac{py(x-y)}{(\epsilon^2 + |x|)^{1-p/2}} \geq 0$$

extending to matrices.

**LEMMA 3** (SEE TH. 4.4 IN [18]) Let  $F(\cdot)$  be a function satisfying Definition 1 1)–3). For any  $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{(2M+1) \times N}$

$$\sum_{i=1}^K F(\sigma_i(\mathbf{P}_1 - \mathbf{P}_2)) \geq \sum_{i=1}^K (F(\sigma_i(\mathbf{P}_1)) - F(\sigma_i(\mathbf{P}_2))) \tag{23}$$

holds for all  $K = 1, \dots, 2M + 1$ . Lemma 3 shows the sub-additivity of singular values, and this is a generalization of the result  $F(p_1 + p_2) \leq F(p_1) + F(p_2)$  for  $p_1, p_2 \geq 0$  in [35]. Based on Lemma 2,  $J(\mathbf{P})$  defined in (22) satisfies the weakly convex function of Definition 1 4) [35].

Next, based on the smoothed version of Schatten  $p$ -quasi norm in (22), the iteratively reweighted least squares (IRLS) method is employed as a fast solver to deal with the robust formulation (9). The main challenge for solving (9) is that the residual error term is nonsmooth. By introducing a regularization term, (9) is relaxed as

$$\min_{\mathbf{P}} J(\mathbf{P}) + \lambda \left\| \begin{bmatrix} \mathbf{X} - \mathbf{P} \\ \epsilon \mathbf{1}^T \end{bmatrix} \right\|_{2,1} \tag{24}$$

where  $\mathbf{1} \in \mathbb{R}^M$  is the vector of all ones. The term  $\epsilon \mathbf{1}^T$  makes the objective function smooth.

Equivalently, (24) is rewritten as

$$\min_{\mathbf{P}} \text{trace}((\mathbf{P}^H \mathbf{P} + \epsilon^2 \mathbf{I})^{\frac{p}{2}}) + \lambda \sum_{i=1}^{2M+1} (\|\mathbf{X} - \mathbf{P}\|_2^2 + \epsilon^2)^{\frac{1}{2}}. \tag{25}$$

Since the objective function  $\mathcal{J}(\mathbf{P}, \epsilon) \triangleq \text{trace}((\mathbf{P}^H \mathbf{P} + \epsilon^2 \mathbf{I})^{\frac{p}{2}}) + \lambda \sum_{i=1}^{2M+1} (\|\mathbf{X} - \mathbf{P}\|_2^2 + \epsilon^2)^{\frac{1}{2}}$  in (25) is smooth, we take its derivative

$$\frac{\partial \mathcal{J}(\mathbf{P}, \epsilon)}{\partial \mathbf{P}} = p \mathbf{P} (\mathbf{P}^H \mathbf{P} + \epsilon^2 \mathbf{I})^{\frac{p-2}{2}} + \lambda (\mathbf{X} - \mathbf{P}) \mathbf{D} \tag{26}$$

$$\mathbf{D}_{ii} = (\|\mathbf{X} - \mathbf{P}\|_2^2 + \epsilon^2)^{\frac{1}{2}} \tag{27}$$

where  $\mathbf{D}$  is a diagonal matrix with  $i$ th diagonal entry being  $\mathbf{D}_{ii}$ . Therefore, setting (26) to zero results in

$$p \mathbf{P} (\mathbf{P}^H \mathbf{P} + \epsilon^2 \mathbf{I})^{\frac{p-2}{2}} + \lambda (\mathbf{X} - \mathbf{P}) \mathbf{D} = \mathbf{0}. \tag{28}$$

As a consequence, we have

$$\mathbf{B} \triangleq (\mathbf{P}^H \mathbf{P} + \epsilon^2 \mathbf{I})^{\frac{p-2}{2}} \tag{29}$$

---

**Algorithm 2:** LRMA-IRLS.

---

**Input:**  $\mathbf{X}, M, N$ **Initialize:**  $\lambda, \epsilon, \rho, p$ **for**  $t = 0, 1, \dots, \text{do}$ **repeat**Compute the derivative of  $\mathcal{J}(\mathbf{P}, \epsilon)$  in (25);

1)  $\mathbf{B}^{t+1} \leftarrow (\mathbf{P}^H \mathbf{P} + \epsilon^2 \mathbf{I})^{\frac{p-2}{2}}$ ;

2)  $\mathbf{P}^{t+1} \leftarrow (1/\lambda)\mathbf{I} - \mathbf{B}(\lambda\mathbf{D} + \mathbf{B})^{-1}\mathbf{X}$ ;

where  $\mathbf{D}_{ii} = (\|\mathbf{X} - \mathbf{P}\|_2^2 + \epsilon^2)^{\frac{1}{2}}$ ;

**until** termination condition satisfied.**end for****Output:**  $\hat{\mathbf{P}}$ 

---

$$\begin{aligned} \mathbf{P} &= (\lambda\mathbf{I} + \mathbf{B}\mathbf{D}^{-1})^{-1}\lambda\mathbf{X} \\ &= (1/\lambda)\mathbf{I} - \mathbf{B}(\lambda\mathbf{D} + \mathbf{B})^{-1}\mathbf{X} \end{aligned} \quad (30)$$

where  $^{-1}$  denotes the inverse operator. However, it should be pointed out that (30) is not the solution to the problem (24) for fixed  $\mathbf{B}$  and  $\mathbf{D}$ . In fact, (30) denotes the relationship under which the critical point exists. Therefore, it is natural to estimate  $\mathbf{D}$ ,  $\mathbf{B}$ , and  $\mathbf{P}$  iteratively according to (27), (29), and (30). An iterative procedure is adopted, which is shown in Algorithm 2.

**REMARK 4** For any given  $\epsilon > 0$ , any limit point of the sequence  $\{\mathbf{P}^t\}$  generated by LRMA-IRLS is a stationary point of problem (24). The proof of IRLS has been completed in low-rank representation [36], [37], [52] with the constraint function  $\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$ . The proof of LRMA-IRLS is similar to [37, Th. 2] by setting  $\mathbf{X}\mathbf{Z} = \mathbf{P}$  and  $\mathbf{E} = \mathbf{Q}$ , and thus omitted here.

### C. DOA Estimation

Once the estimated noise-free data matrix  $\hat{\mathbf{P}}$  is obtained, the DOAs can be estimated using the subspace-type DOA estimation methods, such as MUSIC. According to the principle of MUSIC, the spatial spectrum is

$$P_{\text{MUSIC}} \triangleq \frac{1}{\mathbf{a}^H(\theta)(\mathbf{I} - \mathbf{U}_s \mathbf{U}_s^H)\mathbf{a}(\theta)} \quad (31)$$

where  $\hat{\mathbf{P}} = \mathbf{U}_s \Sigma_s \mathbf{V}_s^H$ , the columns of  $\mathbf{U}_s$  and  $\mathbf{V}_s$  contain the left and right orthonormal base vectors of  $\hat{\mathbf{P}}$ , respectively, and  $\Sigma_s$  is a diagonal matrix whose diagonal elements are the singular values arranged in descending order. Under the assumption that the number of sources is known *a priori*, the DOAs are determined by searching for the maxima of spatial spectrum.

**REMARK 5** Using the subspace-type DOA estimation methods to obtain DOAs, the number of sources must be known for separating the noise and signal subspaces. When the source number is unknown, it can be determined using conventional methods such as Akaike information criterion [53] and minimum description length [54], but they are effective only when the noise is spatially white. Nevertheless,

a mismatch in the source number information will lead to performance degradation of DOA estimation [55].

Now, we estimate DOAs even when there is no *a priori* information of the source number. Since it is assumed that the source signals are uncorrelated,  $\mathbf{R}_s$  is diagonal. When  $\hat{\mathbf{P}}$  is obtained, the  $(m, n)$  entry of the signal covariance matrix  $\mathbf{R}$  can be expressed as [56]

$$\mathbf{R}(m, n) = \sum_{k=1}^K d_{m,k} e^{j2\pi\tau_k k}, \quad m, n = -M, \dots, 0, \dots, M \quad (32)$$

where  $d_{m,k} = \sigma_k^2 e^{j2\pi\tau_k k}$ . Based on the  $m$ th row of  $\mathbf{R}$ , we form the Toeplitz matrix as follows:

$$\begin{aligned} \mathbf{R}_m &= \begin{bmatrix} \mathbf{R}(m, 0) & \mathbf{R}(m, 1) & \cdots & \mathbf{R}(m, M) \\ \mathbf{R}(m, -1) & \mathbf{R}(m, 0) & \cdots & \mathbf{R}(m, M-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}(m, -M) & \mathbf{R}(m, -M+1) & \cdots & \mathbf{R}(m, 0) \end{bmatrix} \\ &= \bar{\mathbf{A}}(\theta) \{\mathbf{R}_s\}_m \bar{\mathbf{A}}^H(\theta) \\ &= \sum_{k=1}^K d_{m,k} \bar{\mathbf{a}}(\theta_k) \bar{\mathbf{a}}^H(\theta_k) \end{aligned} \quad (33)$$

where  $\bar{\mathbf{A}}(\theta)$  denotes a steering matrix with the  $k$ th steering vector being  $\bar{\mathbf{a}}(\theta) = [1, e^{-j2\pi\tau_k}, \dots, e^{-j2\pi M\tau_k}]^T$ , and  $\{\mathbf{R}_s\}_m = \text{diag}\{d_{m,1}, \dots, d_{m,K}\}$ . It means that  $\mathbf{R}_m$  have the joint diagonalization structure and span the same range space of  $\bar{\mathbf{A}}(\theta)$ , i.e.,

$$\text{range}(\mathbf{R}_m) = \text{range}(\bar{\mathbf{A}}(\theta)). \quad (34)$$

Therefore, we can utilize these  $(M+1)$  matrices  $\mathbf{R}_m$  to identify the range space of the array manifold matrix  $\bar{\mathbf{A}}(\theta)$  and estimate the DOAs. For the  $k$ th source, there always exists a vector  $\mathbf{b}_k$  that is orthogonal to the range space spanned by the steering vectors except for  $\bar{\mathbf{a}}(\theta_k)$ , i.e.,

$$\mathbf{b}_k \perp \text{range}\{\bar{\mathbf{a}}(\theta_1), \dots, \bar{\mathbf{a}}(\theta_{k-1}), \bar{\mathbf{a}}(\theta_{k+1}), \dots, \bar{\mathbf{a}}(\theta_K)\}. \quad (35)$$

Equivalently, we have

$$\bar{\mathbf{a}}^H(\theta_i) \mathbf{b}_k = \begin{cases} \bar{\mathbf{a}}^H(\theta_k) \mathbf{b}_k, & i = k \\ 0, & i \neq k \end{cases}. \quad (36)$$

Substituting (36) into (33) yields

$$\mathbf{R}_m \mathbf{b}_k = \sum_{k=1}^K d_{m,k} \bar{\mathbf{a}}(\theta_k) \bar{\mathbf{a}}^H(\theta_k) \mathbf{b}_k = g_m \bar{\mathbf{a}}(\theta_k) \quad (37)$$

where the geometric interpretation of (37) is that when  $\theta$  is one of the true DOAs, there exists a scalar  $g_m = d_{m,k} \bar{\mathbf{a}}^H(\theta_k) \mathbf{b}_k$  making  $\mathbf{R}_m \mathbf{b}_k$  and  $\bar{\mathbf{a}}(\theta_k)$  parallel, i.e.,

$$\mathbf{R}_m \mathbf{b} = g_m \bar{\mathbf{a}}(\theta), \quad -M \leq m \leq 0 \quad (38)$$

which leads to the following optimization problem:

$$\min_{\theta} \sum_{m=-M}^0 \|\mathbf{R}_m \mathbf{b} - g_m \bar{\mathbf{a}}(\theta)\|^2, \quad \text{s.t. } \|\mathbf{g}\| = 1 \quad (39)$$

where the constraint  $\|\mathbf{g}\| = 1$  is to avoid the trivial solution of (39), i.e.,  $\{\mathbf{b} = \mathbf{0}, \mathbf{g} = \mathbf{0}\}$ . However, (39) is difficult to solve since only  $\theta$  is of interest while  $\mathbf{b}$  and  $\mathbf{g}$  are unknown nuisance parameters. To cope with this issue, we attempt to simplify (39) via expanding the objective function

$$\begin{aligned} & \min_{\theta} \mathbf{b}^H \left( \sum_{m=-M}^0 \mathbf{R}_m^H \mathbf{R}_m \right) \mathbf{b} - \mathbf{b}^H \left( \sum_{m=-M}^0 g_m \mathbf{R}_m^H \bar{\mathbf{a}}(\theta) \right) \\ & - \left( \sum_{m=-M}^0 g_m^* \bar{\mathbf{a}}^H(\theta) \mathbf{R}_m \right) \mathbf{b} + \bar{\mathbf{a}}^H(\theta) \bar{\mathbf{a}}(\theta) \sum_{m=-M}^0 |g_m|^2 \\ & \text{s.t. } \|\mathbf{g}\| = 1. \end{aligned} \quad (40)$$

where  $\bar{\mathbf{a}}^H(\theta) \bar{\mathbf{a}}(\theta) = M + 1$  and  $\sum_{m=-M}^0 |g_m|^2 = \|\mathbf{g}\|^2 = 1$ . Denoting  $\mathbf{F} \triangleq \sum_{m=-M}^0 \mathbf{R}_m^H \mathbf{R}_m$  and  $\mathbf{G}(\theta) \triangleq [\mathbf{R}_{-M}^H \bar{\mathbf{a}}(\theta), \dots, \mathbf{R}_0^H \bar{\mathbf{a}}(\theta)]$ , (40) is rewritten as

$$\begin{aligned} & \min_{\theta} \mathcal{J}(\theta, \mathbf{b}, \mathbf{g}) \triangleq \mathbf{b}^H \mathbf{F} \mathbf{b} - \mathbf{b}^H \mathbf{G}(\theta) \mathbf{g} - \mathbf{g}^H \mathbf{G}^H(\theta) \mathbf{b} + M + 1 \\ & \text{s.t. } \|\mathbf{g}\| = 1. \end{aligned} \quad (41)$$

Fixing  $\theta$  and  $\mathbf{g}$ , we set the partial derivative of  $\mathcal{J}(\theta, \mathbf{b}, \mathbf{g})$  with respect to  $\mathbf{b}^*$  to zero to obtain

$$\frac{\partial \mathcal{J}(\theta, \mathbf{b}, \mathbf{g})}{\partial \mathbf{b}^*} = \mathbf{F} \mathbf{b} - \mathbf{G}(\theta) \mathbf{g} = \mathbf{0} \quad (42)$$

and the solution is

$$\mathbf{b}^* = \mathbf{F}^\dagger \mathbf{G}(\theta) \mathbf{g} \quad (43)$$

where  $\dagger$  denotes the Moore–Penrose pseudoinverse. Substituting (43) into (41), the optimization problem is transformed into

$$\min_{\theta} M + 1 - \mathbf{g}^H \mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta) \mathbf{g}, \quad \text{s.t. } \|\mathbf{g}\| = 1. \quad (44)$$

which is also equal to

$$\max_{\theta} \mathbf{g}^H \mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta) \mathbf{g}, \quad \text{s.t. } \|\mathbf{g}\| = 1. \quad (45)$$

Thus

$$\begin{aligned} \max_{\theta} \mathbf{g}^H \mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta) \mathbf{g} &= \max_{\theta} \mathbf{g}^H \sum_{i=1}^{M+1} \lambda_i \mathbf{u}_i \mathbf{u}_i^H \mathbf{g} \\ &= \max_{\theta} \sum_{i=1}^{M+1} \lambda_i |\mathbf{g}^H \mathbf{u}_i|^2 \\ &= \lambda_{\max} \{ \mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta) \} \end{aligned} \quad (46)$$

where  $\lambda_i$  are the eigenvalues of  $\mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta)$  with corresponding eigenvectors  $\mathbf{u}_i$ . Therefore, for (44), DOAs are estimated by finding the maxima of power spectrum  $P(\theta)$

$$P(\theta) \triangleq \frac{1}{M + 1 - \lambda_{\max} \{ \mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta) \}} \quad (47)$$

where the source number information is not needed. The procedure is described in Algorithm 3.

#### IV. SIMULATION RESULTS

In this section, the performance of the proposed methods is compared with MM-MUSIC [19],  $\ell_p$ -MUSIC [21], and

---

#### Algorithm 3.

---

**Input:**  $\hat{\mathbf{P}}, M, N$

1) Calculate the noise-free data covariance matrix  $\mathbf{R}$ :

$$\mathbf{R} = \frac{1}{N} E \{ \hat{\mathbf{P}} \hat{\mathbf{P}}^H \};$$

2) Choose the first  $M + 1$  rows of  $\mathbf{R}$  and each row is utilized to form the Toeplitz matrix as (33);

3) Construct  $\mathbf{F}$  and  $\mathbf{G}(\theta)$ :

$$\mathbf{F} = \sum_{m=-M}^0 \mathbf{R}_m^H \mathbf{R}_m,$$

$$\mathbf{G}(\theta) = [\mathbf{R}_{-M}^H \mathbf{a}(\theta), \dots, \mathbf{R}_0^H \mathbf{a}(\theta)];$$

4) Find the maxima of power spectrum  $P(\theta)$ :

$$P(\theta) \triangleq \frac{1}{M+1-\lambda_{\max} \{ \mathbf{G}^H(\theta) \mathbf{F}^\dagger \mathbf{G}(\theta) \}}.$$

**Output:** DOAs are obtained from the peaks of  $P(\theta)$ .

---

average conditional Cramer-Rao bound [9], [57], [58] for DOA estimation. The impulsive noise is taken to model the additive noise. Three types of impulsive noise, namely,  $\text{S}\alpha\text{S}$ , GMM, and CGM distributions, are adopted in  $\mathbf{q}(t)$ .

##### A. $\text{S}\alpha\text{S}$

The  $\text{S}\alpha\text{S}$  distribution with zero-location, whose characteristic function is expressed as

$$\varphi(\omega) = \exp(-\gamma^\alpha |\omega|^\alpha) \quad (48)$$

where  $0 < \alpha \leq 2$  is called the characteristic exponent that describes the tail of the distribution, and  $\gamma > 0$  is the scale. When  $\alpha = 2$ , the  $\alpha$ -stable distribution reduces to the Gaussian distribution and  $\gamma^2$  is similar to the variance of the Gaussian distribution. When  $\alpha = 1$ , the  $\alpha$ -stable distribution becomes the Cauchy distribution. When  $\alpha < 2$ ,  $\alpha$ -stable noise shows heavy tails and hence is impulsive. The smaller the value of  $\alpha$ , the more impulsive the noise is. Since the second-order and higher-order moments of the  $\alpha$ -stable distribution are infinite for  $\alpha < 2$ , the commonly used signal-to-noise (SNR) is meaningless in this case. Instead, we employ the generalized SNR (GSNR) [21], [58], [59]:  $\text{GSNR} = E\{|s(n)|^2\} / \gamma^\alpha$ . In our simulations,  $\alpha = 1.8$  and  $\gamma = 0.1$ .

##### B. GMM

The PDF of the two-term Gaussian mixture noise is

$$p_n(n) = \sum_{i=1}^2 \frac{c_i}{\pi \sigma_i^2} \exp\left(-\frac{|n|^2}{\sigma_i^2}\right) \quad (49)$$

where  $0 \leq c_i \leq 1$  and  $\sigma_i^2$  are the probability and variance of  $i$ th term, respectively, with  $c_1 + c_2 = 1$ . If  $\sigma_2^2 \gg \sigma_1^2$  and  $c_2 < c_1$  are selected, large noise samples of variance  $\sigma_2^2$  occurring with a smaller probability  $c_2$  can be viewed as outliers embedded in Gaussian background noise of variance  $\sigma_1^2$ . Therefore, GMM can well model the phenomenon in the presence of both Gaussian noise and outliers. In our simulations,  $\sigma_2^2 = 100\sigma_1^2$ ,  $c_2 = 0.1$ , and  $\text{SNR} = 30$  dB. Hence, there are around 10% noise samples that are considered as outliers.

### C. CGM

It is modeled by a product for two random variables:  $z = \sqrt{x}y$ , whose envelope characteristic function is given by

$$\Psi(\omega) = \left(1 + \frac{\omega^2}{\beta^2}\right)^{-\alpha}. \quad (50)$$

This corresponds to the  $K$ -distributed envelope whose PDF is [10]

$$f_R(z) = \frac{2\beta}{\Gamma(\alpha)} \left(\frac{\beta z}{2}\right)^\alpha K_{\alpha-1}(\beta z) u(z) \quad (51)$$

where  $\alpha$  is the shape parameter,  $\beta$  denotes the rate parameter, and  $u(z)$  is the unit step function. The PDF of  $x$  is a Gamma distribution with a shape parameter  $\alpha$  and a scale parameter  $\beta$ , and  $y$  is a Gaussian random variable with zero mean and variance  $\sigma_y^2$ .  $z$  has  $K$ -distribution envelope. The variances of Gaussian noise and  $K$ -distributed noise are, respectively, denoted as  $\sigma_1^2$  and  $\sigma_2^2$ , where  $\sigma_2^2 = \kappa\sigma_1^2$  with  $\kappa$  being a constant. In our simulations,  $\kappa = 4$  and  $\alpha = \beta = 0.1$ .

To evaluate the robustness of the proposed methods, unless stated otherwise, all simulations are carried out with the settings: the numbers of sensors in ULA and snapshots are  $(2M + 1) = 9$  and  $N = 300$ , respectively. The signal waveforms are independent quadrature phase-shift keying signals with equal power, where the DOAs of two incoming signals are  $\theta_1 = 20.2^\circ$  and  $\theta_2 = 30.8^\circ$ .  $\mu$  is initiated as  $\mu^0 = 0.5/\|\text{sgn}(\mathbf{X})\|_2$ . In each experiment, 500 Monte Carlo trials are performed. The function  $F(\cdot)$  forming the low-rank inducing penalty in LRMA-ADMM is chosen as

$$F(t) = (|t| - \rho|t|^2)\mathcal{X}_{0 \leq |t| \leq \frac{1}{2\rho}} + \left(\frac{1}{4\rho}\right)\mathcal{X}_{|t| > \frac{1}{2\rho}} \quad (52)$$

where  $\mathcal{X}_p$  denotes the indicator function and  $\rho$  is set as 2 so that  $\bar{\alpha} = 1$  [35]. Hence, we calculate that when  $\lambda < \frac{1}{2\rho}$ , the proximal operator of (52) is

$$\text{prox}_F(v, \lambda) = \frac{v - \lambda \text{sign}(v)}{1 - 2\tau\rho} \mathcal{X}_{\lambda \leq |v| \leq \frac{1}{2\rho}} + v_i \mathcal{X}_{|v| > \frac{1}{2\rho}}. \quad (53)$$

The intuition behind the choice of the MCP in (52) is twofold. First, MCP provides the convexity of the penalized loss in sparse regions to the greatest extent given certain thresholds for variable selection and unbiasedness, compared with other weakly convex penalties. Second, since the proximal operator of MCP in (53) has closed-form solution, it is easy to deal with the nonconvex minimization problem.

For LRMA-IRLS,  $p$  is set to a value ( $p = 0.1$ ) smaller than 1, then the resultant problem will better approximate the original problem. All our methods require the selection of the regularization parameter  $\lambda$ , which balances the fidelity and sparsity of the solution. Although there are many ways for its determination based on the statistical information of the noise, such as cross-validation [60], we simply set  $\lambda = 0.2$  for IRMA-ADMM, which aligns with the choice of  $\lambda \leq (2M + 1)^{-1/2}$  suggested in [61], and we

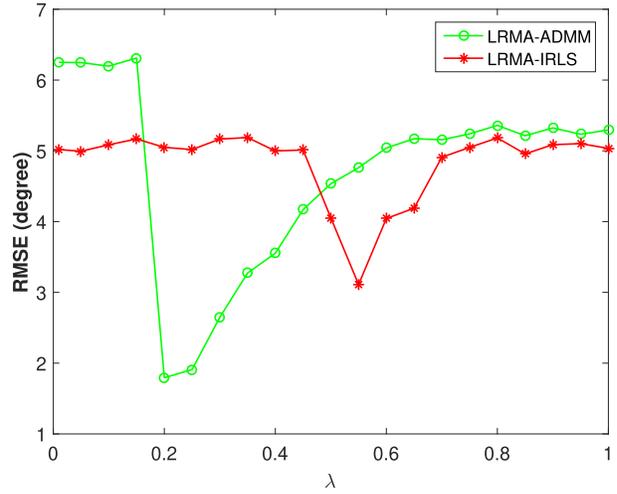


Fig. 2. RMSE versus  $\lambda$ .

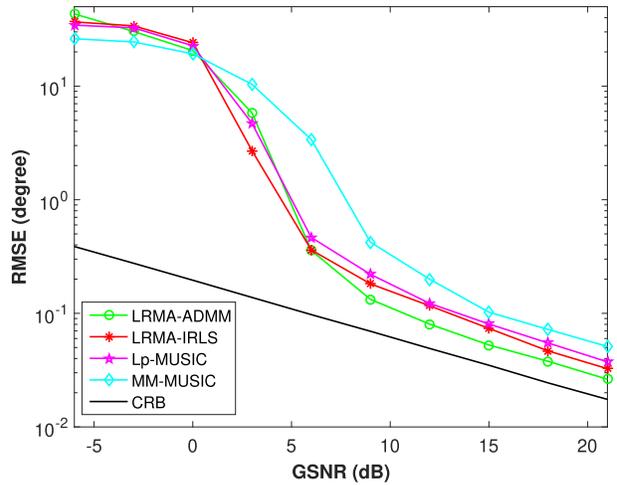


Fig. 3. RMSE versus GSNR in  $S\alpha S$  noise.

observe that IRMA-IRLS with  $\lambda = 0.55$  works well in Fig. 2. The effect of the tuning parameter  $\lambda$  on the performance of the proposed methods has been studied at SNR = 0 dB. All simulations are performed using MATLAB R2015b on a personal computer with 3.40 GHz Intel core i7 CPU and 4 GB RAM, under a 64-bit Microsoft Windows 7 operating system.

In the first experiment, we test the statistical performance of the proposed methods in DOA estimation. All three types of impulsive noise are considered. For  $S\alpha S$  noise, the root mean squared error (RMSE) versus SNR is plotted in Fig. 3. It is observed that the LRMA-ADMM achieves the best DOA estimation performance in  $S\alpha S$  noise, especially when GSNR > 5 dB. Figs. 4 and 5 show the RMSE of the DOA estimators in GMM and CGM noises versus SNR, respectively. We see that the proposed methods and  $\ell_p$ -MUSIC outperform MM-MUSIC, while our solutions enjoy comparable performance with  $\ell_p$ -MUSIC in GMM noise.

In the second experiment, the resolution probability of resolving two closely-spaced sources is examined. The resolution probability is computed as the ratio between the

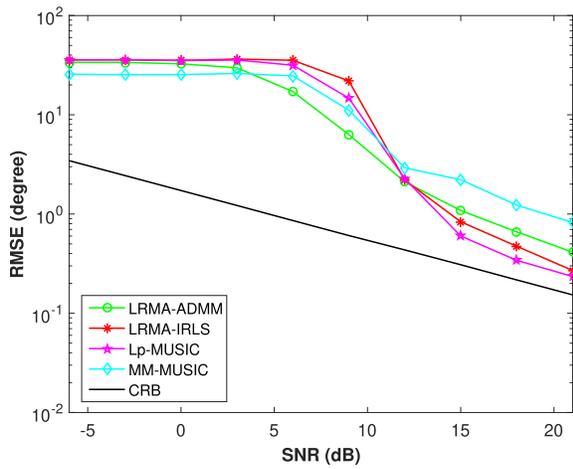


Fig. 4. RMSE versus SNR in GMM noise.

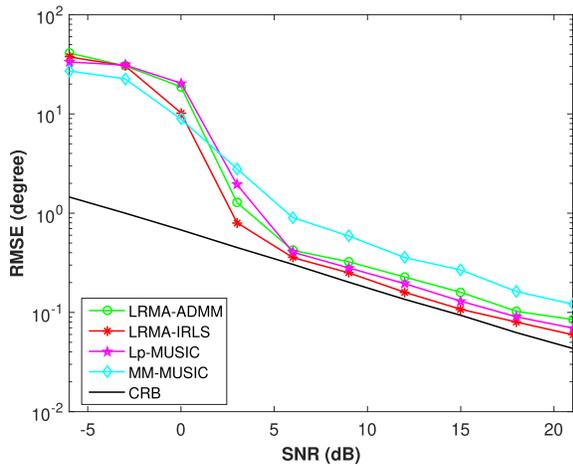


Fig. 5. RMSE versus SNR in CGM noise.

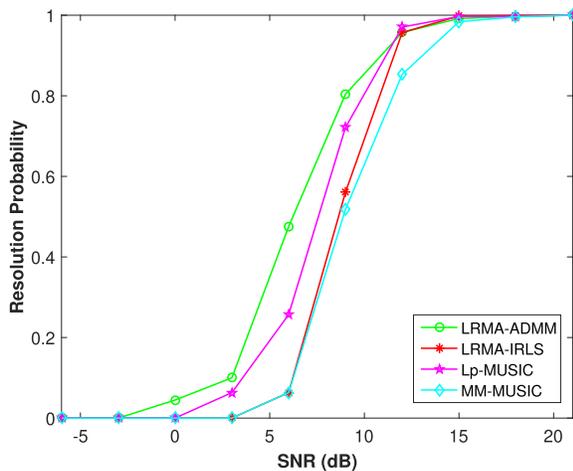


Fig. 6. Resolution probability versus SNR in GMM noise.

number of successful runs and total number of the independent runs. A trial is regarded as success when the maximum value of the absolute deviation between the estimated and true DOAs is less than the half deviation, i.e.,  $\max\{\theta_k - \hat{\theta}_k\} \leq |\theta_2 - \theta_1|/2, k = 1, 2$  where  $\theta_1 = 20.2^\circ$  and  $\theta_2 = 30.8^\circ$ . The results are shown in Figs. 6–8. The resolution

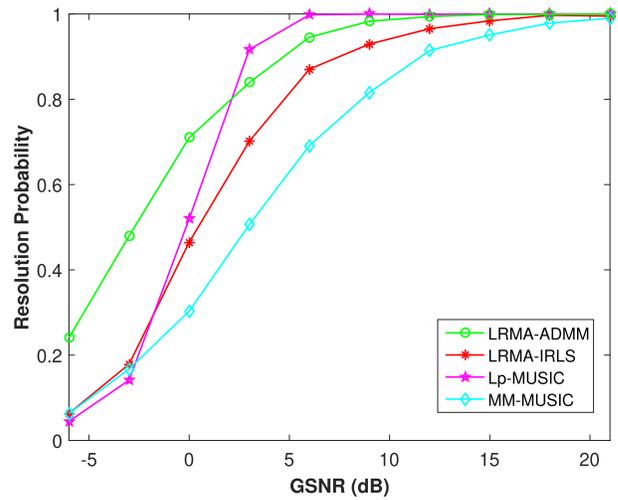


Fig. 7. Resolution probability versus GSNR in  $S\alpha S$  noise.

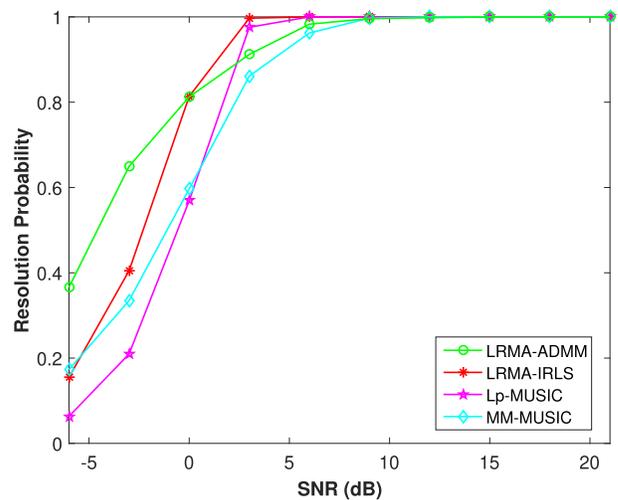


Fig. 8. Resolution probability versus SNR in CGM noise.

performance of the proposed schemes is superior to the MM-MUSIC. LRMA-ADMM outperforms other methods in the presence of non-Gaussian noise, except  $\ell_p$ -MUSIC in the case of  $S\alpha S$  and CGM noises.

The third experiment tests the resolution probability of DOA estimators in terms of the angle separation between two targets, where  $\text{SNR}/\text{GSNR} = 10$  dB. The two uncorrelated targets are considered with DOAs  $\theta_1 = 0^\circ$  and  $\theta_2 = 0^\circ + \Delta\theta$ , where  $\Delta\theta$  varies from  $2^\circ$  to  $20^\circ$ . As indicated in Figs. 9–11, the proposed methods tend to become unbiased when the angular separation is more than  $10^\circ$  apart. From Figs. 10 and 11, in the case of  $S\alpha S$  and CGM noises, the proposed methods have higher resolution probability in terms of the angle separation compared with  $\ell_p$ -MUSIC.

The last experiment compares computational time, and the results are tabulated in Table I. The average CPU runtime is used as the performance metric, although the runtime gives only a rough estimation of complexity. It is demonstrated that both solutions are much faster than the other investigated methods. LRMA-ADMM achieves the best

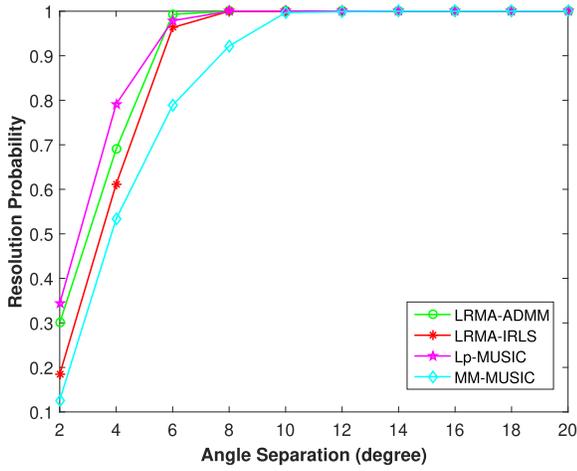


Fig. 9. Resolution probability versus angle separation in GMM noise.

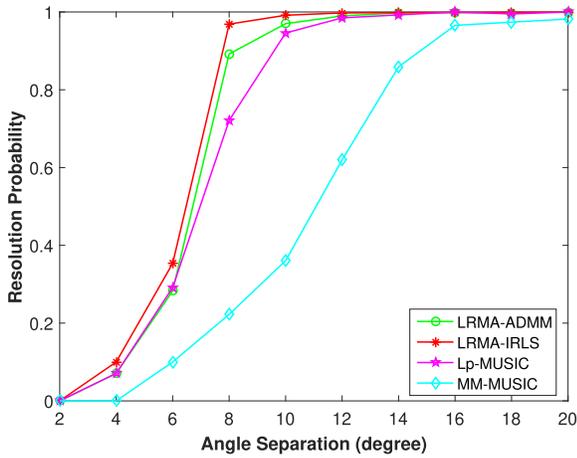


Fig. 10. Resolution probability versus angle separation in SαS noise.

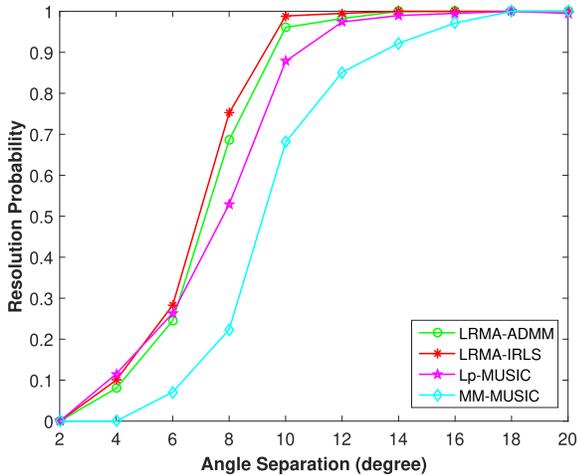


Fig. 11. Resolution probability versus angle separation in CGM noise.

runtime performance among all methods, which is efficient since the ADMM updates each variable separately only once in each iteration.

TABLE I  
Runtime Comparison

Algorithm	Time(s)	SNR/GSNR = 5 dB, $2M + 1 = 9$ , $N = 300$		
		SαS	GMM	CGM
LRMA-ADMM		0.0261	0.0244	0.0256
LRMA-IRLS		0.1567	0.1635	0.1567
$\ell_p$ -MUSIC		14.5390	13.9789	14.7632
MM-MUSIC		10.3758	8.3901	8.3759

## V. CONCLUSION

In this paper, we have proposed two iterative methods via combination of LRMA technique and a class of weakly convex inducing penalties to tackle the problem of DOA estimation in impulsive noise. Although  $\ell_p$ -MUSIC outperforms the proposed methods, especially in GMM noise, our methods can achieve a comparable DOA estimation performance with faster computation and without requiring *a priori* information of the source number. Simulation results evaluate the direction finding performance of the proposed methods, indicating a good balance between accuracy and complexity.

## APPENDIX

From (21) and the definition of  $\mathcal{L}$ , we have

$$\begin{aligned}
 & \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^t, \mu^t) \\
 &= \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^{t-1}, \mu^{t-1}) + \langle \Gamma^t - \Gamma^{t-1}, \mathbf{X} - \mathbf{P}^t - \mathbf{Q}^t \rangle \\
 &> + \frac{\mu^t - \mu^{t-1}}{2} \|\mathbf{X} - \mathbf{P}^t - \mathbf{Q}^t\|_F^2 \\
 &= \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^{t-1}, \mu^{t-1}) + \frac{\mu^t + \mu^{t-1}}{2(\mu^{t-1})^2} \|\Gamma^t - \Gamma^{t-1}\|_F^2.
 \end{aligned} \tag{54}$$

Moreover, it follows that  $\mathbf{Q}^{t+1}$  is a global solution to (19) and  $\mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^t)$  is monotonically decreasing

$$\begin{aligned}
 & \mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^{t+1}, \Gamma^t, \mu^t) \\
 &\leq \mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^t, \Gamma^t, \mu^t) \leq \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^t, \mu^t) \\
 &= \mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^{t-1}, \mu^{t-1}) + \frac{\mu^t + \mu^{t-1}}{2(\mu^{t-1})^2} \|\Gamma^t - \Gamma^{t-1}\|_F^2.
 \end{aligned} \tag{55}$$

Applying the chain rule for  $t$  times, then

$$\begin{aligned}
 & \mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^{t+1}, \Gamma^t, \mu^t) \\
 &\leq \mathcal{L}(\mathbf{P}^1, \mathbf{Q}^1, \Gamma^0, \mu^0) + \sum_{i=1}^t \frac{\mu^i + \mu^{i-1}}{2(\mu^{i-1})^2} \|\Gamma^i - \Gamma^{i-1}\|_F^2.
 \end{aligned} \tag{56}$$

With  $t \rightarrow \infty$ , if  $\sum_{i=1}^t \frac{\mu^i + \mu^{i-1}}{2(\mu^{i-1})^2} < \infty$ , the left-hand side (LHS) of (56) is bounded. Indeed,  $\Gamma^t$  is bounded. Recall

$$\partial_{\mathbf{Q}} \|\mathbf{Q}\|_{2,1} = \begin{cases} 0, & \text{if } \|\mathbf{Q}^t(:, j)\|_2 = 0 \\ \frac{\mathbf{W}^t(:, j)}{\|\mathbf{W}^t(:, j)\|_2}, & \text{otherwise} \end{cases}. \tag{57}$$

It is easily proved that  $\partial_{\mathbf{Q}}\|\mathbf{Q}\|_{2,1}$  is bounded. The optimal  $\mathbf{Q}^{t+1}$  needs to satisfy the first-order optimality condition, i.e.,

$$\begin{aligned} 0 &\in \partial_{\mathbf{Q}}\mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}, \Gamma^t, \mu^t) \\ &= \partial_{\mathbf{Q}}\|\mathbf{Q}\|_{2,1} - \Gamma^t - \mu^t(\mathbf{X} - \mathbf{P}^{t+1} - \mathbf{Q}^{t+1}) \\ &= \partial_{\mathbf{Q}}\|\mathbf{Q}\|_{2,1} - \Gamma^{t+1} \end{aligned} \quad (58)$$

Thus,  $\Gamma^t$  is bounded. Adding  $\frac{1}{2\mu^t}\|\Gamma^t\|_F^2$  to the augmented Lagrangian gives

$$\begin{aligned} \mathcal{L}(\mathbf{P}^{t+1}, \mathbf{Q}^{t+1}, \Gamma^t, \mu^t) + \frac{1}{2\mu^t}\|\Gamma^t\|_F^2 = \\ J(\mathbf{P}^{t+1}) + \lambda\|\mathbf{Q}^{t+1}\|_{2,1} + \frac{\mu^t}{2}\|\mathbf{X} - \mathbf{P}^{t+1} - \mathbf{Q}^{t+1}\|_F^2 + \frac{\Gamma^t}{\mu^t}\|_F^2. \end{aligned} \quad (59)$$

Since  $\{\Gamma^t\}$  is bounded, the LHS is bounded. Thus, each term on the RHS is bounded; then,  $\{\mathbf{P}^t\}$  and  $\{\mathbf{Q}^t\}$  are bounded.

By Bolzano–Weierstrass theorem, let  $\{\mathbf{P}^*, \mathbf{Q}^*\}$  be any accumulation point of problem (11). Thus, due to the assumption of  $\lim_{t \rightarrow \infty}(\mathbf{P}^t - \mathbf{P}^{t+1}) = \mathbf{0}$  and  $\sum_{i=1}^t \frac{\mu^i + \mu^{i-1}}{2(\mu^{i-1})^2} < \infty$ , accordingly, there exists  $\Gamma^*$  such that the following optimality conditions are satisfied:

$$\partial_{\mathbf{Q}}\|\mathbf{Q}\|_{2,1} = \Gamma^*, \quad -\Gamma^* \in \partial(-F(\sigma_i(\mathbf{P}))), \quad \mathbf{X} = \mathbf{P}^* + \mathbf{Q}^*. \quad (60)$$

We conclude that  $\{\mathbf{P}^*, \mathbf{Q}^*, \Gamma^*\}$  satisfies the Karush–Kuhn–Tucker conditions of  $\mathcal{L}(\mathbf{P}^t, \mathbf{Q}^t, \Gamma^t)$ . Thus,  $\{\mathbf{P}^*, \mathbf{Q}^*\}$  is a stationary point of problem (11) [62].

#### ACKNOWLEDGMENT

The authors would like to thank the three anonymous reviewers and the associate editor for their helpful comments.

#### 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. C. So  
Source localization: Algorithms and analysis  
In *Handbook of Position Location: Theory, Practice and Advances*, S. A. Zekavat and M. Buehrer, Eds. New York, NY, USA: Wiley, 2011, ch. 2.
- [3] Q. Liu, W. Wang, D. Liang, and X. P. Wang  
Real-valued reweighted  $\ell_1$  norm minimization method based on data reconstruction in MIMO radar  
*IEICE Trans. Commun.*, vol. E98-B, no. 11, pp. 2307–2313, Nov. 2015.
- [4] R. O. Schmidt  
Multiple emitter location and signal parameter estimation  
*IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
- [5] C. Qian, L. Huang, M. Cao, H. C. So, and J. Xie  
PUMA: An improved realization of MODE for DOA estimation  
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 5, pp. 2128–2139, Mar. 2017.
- [6] J. Capon  
High-resolution frequency-wavenumber spectrum analysis  
*Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [7] L. Huang, C. Qian, H. C. So, and J. Fang  
Source enumeration for large array using shrinkage-based detectors with small samples  
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 334–357, Apr. 2015.
- [8] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma  
Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts  
*IEEE Trans. Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.
- [9] R. J. Kozyck and B. M. Sadler  
Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures  
*IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.
- [10] M. Rangaswamy, D. Weiner, and A. Ozturk  
Computer generation of correlated non-Gaussian radar clutter  
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 106–116, Jan. 1995.
- [11] C. L. Nikias and M. Shao  
*Signal Processing With Alpha-stable Distributions and Applications*. New York, NY, USA: Wiley, 1995.
- [12] P. J. Huber  
*Robust Statistics*, 2nd ed. Berlin, Germany: Wiley, 2009.
- [13] Q. Liu, C. Yang, Y. Gu, and H. C. So  
Robust sparse recovery via weakly convex optimization in impulsive noise  
*Signal Process.*, vol. 152, pp. 84–89, Nov. 2018.
- [14] B. Liao and S. C. Chan  
Direction-of-arrival estimation in subarrays-based linear sparse arrays with gain/phase uncertainties  
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 4, pp. 2268–2280, Oct. 2013.
- [15] Q. Liu, H. C. So, and Y. Gu  
Off-grid DOA estimation with nonconvex regularization via joint sparse representation  
*Signal Process.*, vol. 140, pp. 171–176, Nov. 2017.
- [16] P. Tsakalides and C. L. Nikias  
The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments  
*IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1623–1633, Jul. 1996.
- [17] T.-H. Liu and J. M. Mendel  
A subspace-based direction finding algorithm using fractional lower order statistics  
*IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1605–1613, Aug. 2001.
- [18] S. Visuri, H. Oja, and V. Koivunen  
Subspace-based direction-of-arrival estimation using nonparametric statistics  
*IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2060–2073, Sep. 2001.
- [19] 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. Int. Conf. Acoust., Speech, Signal Process.*, Kyoto, Japan, 2012, pp. 3709–3712.
- [20] K. Todros and A. O. Hero  
Robust multiple signal classification via probability measure transformation  
*IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1156–1170, Mar. 2015.
- [21] W.-J. Zeng, H. C. So, and L. Huang  
 $\ell_p$ -MUSIC: Robust direction-of-arrival estimator for impulsive noise environments  
*IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, May 2013.

- [22] R. Couillet  
Robust spiked random matrices and a robust G-MUSIC estimator  
*J. Multivariate Anal.*, vol. 140, pp. 139–161, 2015.
- [23] M. W. Morency, S. A. Vorobyov, and G. Leus  
Joint detection and localization of an unknown number of sources using the algebraic structure of the noise subspace  
*IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4685–4700, Sep. 2018.
- [24] I. Jolliffe  
*Principal Component Analysis*. New York, NY, USA: Springer, 1986.
- [25] C. Croux, P. Filzmoser, and M. R. Oliveira  
Algorithms for projection–pursuit robust principal component analysis  
*Chemometrics Intell. Lab. Syst.*, vol. 87, no. 2, pp. 218–225, 2007.
- [26] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma  
Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization  
*Adv. Neural Inf. Process. Syst.*, vol. 22, pp. 2080–2088, 2009.
- [27] B. Liao, C. Guo, and H. C. So  
Direction-of-arrival estimation in nonuniform noise via low-rank matrix decomposition  
*In Proc. 22nd Int. Conf. Digit. Signal Process.*, London, U.K., 2017, pp. 1–4.
- [28] P. Pal and P. P. Vaidyanathan  
A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising  
*IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 737–741, Mar. 2014.
- [29] X. Wu, W. P. Zhu, and J. Yan  
A Toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation  
*IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8223–8237, Apr. 2017.
- [30] 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 Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [31] E. J. Candès, J. K. Romberg, and T. Tao  
Stable signal recovery from incomplete and inaccurate measurements  
*Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Mar. 2006.
- [32] E. J. Candès and B. Recht  
Exact matrix completion via convex optimization  
*Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [33] E. J. Candès, J. Romberg, and T. Tao  
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information  
*IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Jan. 2006.
- [34] X. Jiang, Z. Zhong, X. Liu, and H. C. So  
Robust matrix completion via alternating projection  
*IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 579–583, Mar. 2017.
- [35] L. Chen and Y. Gu  
The convergence guarantees of a non-convex approach for sparse recovery  
*IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3754–3767, Jun. 2014.
- [36] K. Mohan and M. Fazel  
Iterative reweighted algorithms for matrix rank minimization  
*J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 3441–3473, Jan. 2012.
- [37] 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, Dec. 2015.
- [38] G. Marjanovic and V. Solo  
On  $l_q$  optimization and matrix completion  
*IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5714–5724, Aug. 2012.
- [39] C. Ding, D. Zhou, X. He, and H. Zha  
R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization  
*In Proc. Int. Conf. Mach. Learn.*, vol. 148, 2006, pp. 281–288.
- [40] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma  
Stable principal component pursuit  
*In Proc. Int. Symp. Inf. Theory*, Austin, TX, USA, Jul. 2010, pp. 1518–1522.
- [41] R. Paffenroth, P. du Toit, R. Nong, L. Scharf, A. P. Jayasumana, and V. Bandara  
Space-time signal processing for distributed pattern detection in sensor networks  
*IEEE J. Sel. Topics Signal Process.* vol. 7, no. 1, pp. 38–49, Jan. 2013.
- [42] N. Parikh and S. Boyd  
Proximal algorithms  
*Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
- [43] K. Border  
The supergradient of a concave function  
2001. [Online]. Available: <http://www.hss.caltech.edu/kcb/Notes/Supergrad.pdf>
- [44] J.-P. Vial  
Strong and weak convexity of sets and functions  
*Math. Operations Res.*, vol. 8, no. 2, pp. 231–259, 1983.
- [45] C. Lu, J. Tang, S. Yan, and Z. Lin  
Generalized nonconvex nonsmooth low-rank minimization  
*In Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2014, pp. 4130–4137.
- [46] Z. Lin, M. Chen, and Y. Ma  
The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices  
Coordinated Science Laboratory, UIUC, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, Sep. 2010.
- [47] J. Yang, W. Yin, Y. Zhang, and Y. Wang  
A fast algorithm for edge-preserving variational multichannel image restoration  
*SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 569–592, May 2009.
- [48] C. Lu, J. Feng, Z. Lin, and S. Yan  
Nonconvex sparse spectral clustering by alternating direction method of multipliers and its convergence analysis  
*In Proc. AAAI Conf. Artif. Intell.*, New Orleans, Louisiana, USA, Feb. 2018.
- [49] C. Lu, J. Tang, S. Yan, and Z. Lin  
Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm  
*IEEE Trans. Image Process.* vol. 25, no. 2, pp. 829–839, Feb. 2016.
- [50] E. J. Candès, M. B. Wakin, and S. P. Boyd  
Enhancing sparsity by reweighted  $l_1$  minimization  
*J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.
- [51] F. Wen, P. Liu, Y. Liu, R. C. Qiu, and W. Yu  
Robust sparse recovery in impulsive noise via  $\ell_p$ - $\ell_1$  optimization  
*IEEE Trans. Signal Process.* vol. 65, no. 1, pp. 105–118, Jan. 2017.
- [52] M. Fornasier, H. Rauhut, and R. Ward  
Low-rank matrix recovery via iteratively reweighted least squares minimization  
*SIAM J. Optim.*, vol. 21, no. 4, pp. 1614–1640, 2011.

- [53] M. Wax and T. Kailath  
Detection of signals by information theoretic criteria  
*IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
- [54] A. Barron, J. Rissanen, and B. Yu  
The minimum description length principle in coding and modeling  
*IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, Oct. 1998.
- [55] W. Wang, R. Wu, J. Liang, and H. C. So  
Phase retrieval approach for DOA estimation with array errors  
*IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 5, pp. 2610–2620, May 2017.
- [56] W.-J. Zeng, X.-L. Li, and H. C. So  
Direction-of-arrival estimation based on spatial-temporal statistics without knowing the source number  
*Signal Process.*, vol. 93, no. 12, pp. 3479–3486, Dec. 2013.
- [57] P. Stoica and A. Nehorai  
Performance study of conditional and unconditional direction-of-arrival estimation  
*IEEE Trans. Acoust. Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.
- [58] B. M. Sadler, R. J. Kozyck, and T. Moore  
Performance analysis for direction finding in non-Gaussian noise  
In *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, USA, 1999, pp. 2857–2860.
- [59] X. Zhong, A. B. Premkumar, and A. S. Madhukumar  
Particle filtering for acoustic source tracking in impulsive noise with alpha-stable process  
*IEEE Sensors J.*, vol. 13, no. 2, pp. 589–600, Feb. 2013.
- [60] K. Nordhausen  
*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Stanford, CA, USA: Springer, 2009.
- [61] Z. Lin, M. Chen, and Y. Ma  
The augmented Lagrange multiple method for exact recovery of a corrupted low-rank matrices  
Dept. Elect. Comput. Eng., UIUC, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, 2009.
- [62] J. Yang and Y. Zhang  
Alternating direction algorithms for  $\ell_1$ -problems in compressed sensing  
*SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.



**Qi Liu** received the B.E. degree in measuring and control technology and instrumentations and the M. Sc. degree in control science and engineering from the College of Automation, Harbin Engineering University, Harbin, China, in 2013 and 2016, respectively. He is currently working toward the Ph.D. degree at the Department of Electronic Engineering, City University of Hong Kong, Hong Kong.

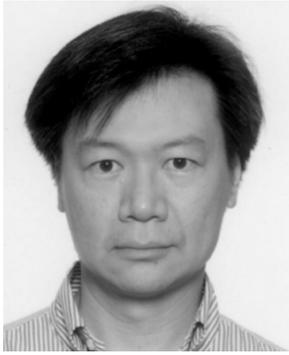
From Oct. 2018 to Apr. 2019, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California, Davis. His research interests include robust compressed sensing, sparse recovery, nonconvex optimization, array signal processing, parameter estimation, and image denoising/inpainting.



**Yuantao Gu** (S'02–M'03–SM'17) received the B.E. degree from Xi'an Jiaotong University, Xi'an, China, in 1998, and the Ph.D. degree with honors from Tsinghua University, Beijing, China, in 2003, both in electronic engineering.

In 2003, he joined the faculty of Tsinghua University where he is currently a Full Professor with the Department of Electronic Engineering. From 2005 to 2006, he was a Visiting Scientist with Microsoft Research Asia, from 2012 to 2013, with the Research Laboratory of Electronics, Massachusetts Institute of Technology, and in 2015, with the Department of Electrical Engineering and Computer Science, University of Michigan in Ann Arbor. His research interests include high-dimensional statistics, sparse signal recovery, temporal-space and graph signal processing, related topics in wireless communications, and information networks.

Dr. Gu has been a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2019 and an elected member of the IEEE Signal Processing Theory and Methods Technical Committee since 2017. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2015 to 2019 and a Handling Editor for Elsevier *Digital Signal Processing* from 2015 to 2017. He was the recipient of the Best Paper Award of IEEE Global Conference on Signal and Information Processing (GlobalSIP) in 2015, the Award for Best Presentation of Journal Paper of IEEE International Conference on Signal and Information Processing (ChinaSIP) in 2015, and Zhang Si-Ying (CCDC) Outstanding Youth Paper Award (with his student) in 2017.



**Hing Cheung So** (S'90–M'95–SM'07–F'15) was born in Hong Kong. He received the B.Eng. degree from the City University of Hong Kong, Hong Kong, and the Ph.D. degree from The Chinese University of Hong Kong, Hong Kong, both in electronic engineering, in 1990 and 1995, respectively.

From 1990 to 1991, he was an Electronic Engineer with the Research and Development Division, Everex Systems Engineering Ltd., Hong Kong. From 1995 to 1996, he was a Postdoctoral Fellow with The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor. His research interests include detection and estimation, fast and adaptive algorithms, multidimensional harmonic retrieval, robust signal processing, source localization, and sparse approximation.

Dr. So has been on the editorial boards of the IEEE SIGNAL PROCESSING MAGAZINE (2014–2017), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2010–2014), *Signal Processing* (2010–), and *Digital Signal Processing* (2011–). He was also Lead Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, special issue on “Advances in Time/Frequency Modulated Array Signal Processing” in 2017. In addition, he was an elected member in Signal Processing Theory and Methods Technical Committee (2011–2016) of the IEEE Signal Processing Society where he was the Chair in the awards subcommittee (2015–2016).