Robust Harmonic Retrieval via Block Successive Upper-Bound Minimization

Cheng Qian, *Member, IEEE*, Yunmei Shi, Lei Huang, *Senior Member, IEEE*, and Hing Cheung So, *Fellow, IEEE*

***Abstract*—Harmonic retrieval (HR) is a problem of significance with numerous applications. Many existing algorithms are explicitly or implicitly developed under Gaussian noise assumption, which, however, are not robust against non-Gaussian noise such as impulsive noise or outliers. In this paper, by employing the ℓ_p-fitting criterion and block successive upper-bound minimization (BSUM) technique, a variant of the classical RELAX algorithm named as BSUM-RELAX is devised for robust HR. It is revealed that the BSUM-RELAX successively performs alternating optimization along coordinate directions, i.e., it updates one harmonic by fixing the other $(K-1)$ components, such that the whole problem is split into K single-tone HR problems, which are then solved by creating a surrogate function that majorizes the objective function of each subproblem. To further refine the frequency component, the Newton's method that takes linear complexity $\mathcal{O}(N)$ is derived for updating the frequency estimates. We prove that under the single-tone case, BSUM-RELAX converges to a Karush-Kuhn-Tucker point. Furthermore, the BSUM-RELAX is extended to the multidimensional HR case. Numerical results show that the proposed algorithm outperforms the state-of-the-art methods in heavy-tailed noise scenarios.**

***Index Terms*—Harmonic retrieval, RELAX, robust estimation, impulsive noise/outliers, majorization minimization.**

I. Introduction

HARMONIC retrieval (HR) (also known as spectrum estimation) is a classical signal processing problem with diverse applications such as synthetic aperture radar imaging [2], wireless communications [3] and source localization [4], to name a few. In HR, fast and accurate estimation of sinusoidal parameters from noisy data is crucial for system reliability. For example, in the frequency division duplex massive multiple-input multiple-output (MIMO) system, the mobile station must estimate the channel state information (CSI) and then feed it back to the base station for achieving reliable communication with high data rates. However, due to the large dimension of the channel matrix, it is impossible to feed back the whole channel matrix through a low rate uplink channel. To overcome this issue, one seeks to estimate the multipath parameters first and then sends them back instead of the whole channel matrix to reduce the feedback overhead. These parameters include direction-of-arrival (DOA), direction-of-departure (DOD), propagation delays and path-loss, where each DOA or DOD may be characterized by a pair of azimuth and elevation angles if two-dimensional (2-D) arrays are deployed at either transmitter or receiver [5]. In such cases, multidimensional HR comes into play.

Numerous HR methods have been developed in the past few decades. Many of them are devised by assuming white Gaussian noise, e.g., [6]–[12]. Under this assumption, the maximum likelihood (ML) method [6] is asymptotically statistically efficient when the number of samples tends to infinity. Nevertheless, due to the multidimensional optimization, ML is usually computationally intensive, which prohibits its practical implementation. To circumvent this problem, suboptimal algorithms, e.g., RELAX [7]–[9], iterative quadratic ML [10], iterative adaptive approach (IAA) [11] and tensor principal-singular-vector utilization for modal analysis (TPUMA) [12], are proposed to balance the complexity and estimation accuracy. However, in real-world applications, the acquired data may frequently contain a portion of unusual measurements, which are corrupted much more significantly than the others. Such corruption is caused by the impulsive noise/outliers with heavy-tailed distribution, which cannot be handled using the conventional Gaussian assumption based techniques. In the presence of outliers, the aforementioned algorithms will suffer severe performance loss and fail to provide satisfactory performance.

The ℓ_p-norm has been successfully employed for sparse signal recovery, e.g., the iterative reweighted least squares algorithm [13]. Since outliers are sparse in nature, the ℓ_p-fitting criterion is of effectiveness in dealing with them [14]–[20]. For 1-D HR, Zeng *et al.* [14] proposed to use the ℓ_p-norm criterion for DOA estimation in impulsive noise. Li studied the Laplace periodogram [19] and ℓ_p-norm periodogram with $p \geq 1$ [20], where the performance for the latter is asymptotically analyzed. These two approaches have shown better

Manuscript received February 17, 2018; revised August 11, 2018; accepted September 19, 2018. Date of publication October 11, 2018; date of current version November 1, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Remy Boyer. This work was supported by the National Science Foundation of China under Grant U1713217 and Grant U1501253. This paper was presented in part at the 10th IEEE Sensor Array and Multichannel Signal Processing Workshop, Sheffield U.K., July 8–11, 2018. *(Corresponding author: Lei Huang.)*

C. Qian is with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA (e-mail: alextoqc@gmail.com).

Y. Shi is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: shiyunmeidamei@gmail.com).

L. Huang is with the Guangdong Key Laboratory of Intelligent Information Processing, College of Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: dr.lei.huang@ieee.org).

H. C. So is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: hcso@ee.cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2875394

performance than the standard ℓ_2-norm based periodogram when there is heavy-tailed noise. In the multidimensional HR case, the iteratively reweighted higher-order singular value decomposition (IR-HOSVD) technique is treated as effective tools for handling outliers [16]. However, this approach is a subspace method which does not exploit the Vandermonde structure in the optimization procedures; hence, it is suboptimal in updating frequencies. Moreover, its identifiability, i.e., the maximum number of identifiable harmonics, is limited by the smallest dimension of the data, which means that once the number of harmonics exceeds the minimum dimension of the data, IR-HOSVD will fail to work. This is of course not preferred in practice. For example, in the downlink channel estimation problem, the receiver may have only two antennas while the number of multipath can easily exceed two. In such scenarios, IR-HOSVD can only resolve one multipath, which is obvious not enough. Note that an ℓ_p-fitting based iterative adaptive approach (ℓ_p-IAA) has relatively higher identifiability than the method in [16]. But its drawbacks mainly arise in the low-resolution ability and huge complexity. This is because ℓ_p-IAA is inherently a dictionary based method. It must employ a finely designed dictionary to reduce the bias of off-grid frequency estimates, and there is no closed-form solution for updating the frequencies. In other words, the frequencies are estimated via constructing a pseudo-spectrum through searching the dictionary grids. Therefore, it is of great interest to develop novel HR algorithm that is not only robust against outliers but also has strong identifiability.

In this work, the robust HR issue is revisited. By using the ℓ_p-norm fitting criterion and block successive upper-bound minimization (BSUM), the RELAX method is reformulated for robust HR, ending up with a novel BSUBM-RELAX approach. The robustness of BSUM-RELAX is achieved based on the ℓ_p-norm fitting criterion, which is well known for handling outliers [18]. BSUM-RELAX alternately updates each harmonic by fixing the remaining $(K-1)$ of them, where K is the number of harmonics, such that the original optimization problem is split into K subproblems, and each subproblem inherits a nonlinear ℓ_p-fitting criterion. The majorization minimization technique is then employed to build a quadratic approximation of the ℓ_p-norm. Unlike the standard RELAX which does not specify an optimization algorithm for estimating frequencies, BSUM-RELAX exploits Newton's method to refine frequencies. It is worth noting that the Newton's method takes linear time complexity and thus is very efficient. The convergence of the BSUM-RELAX for single-tone case is also proved. Furthermore, the BSUM-RELAX is generalized to the multidimensional HR case in Section III. Extensive simulation results in Section IV show that BSUM-RELAX approaches the Cramér-Rao bound (CRB) in Gaussian mixture noise case and outperforms the state-of-the-art algorithms.

Part of the work has been presented in [1], where the derivation of the proposed method for robust 1-D harmonic retrieval is reported. The multidimensional extension and proof of convergence are provided only in this journal version, which naturally includes more comprehensive simulation results.

Notation: Throughout the paper, we use boldface lowercase letters for vectors and boldface uppercase letters for matrices. Superscripts $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^{-1}$ represent transpose, complex conjugate, conjugate transpose, and inverse, respectively. The ℓ_p-norm is defined as $\|\mathbf{x}\|_p^p = \sum_{i=1}^N |x_i|^p$, where $|\cdot|$ represents the absolute value. The $\circ$, $\otimes$, $\circledast$ and $\odot$ are outer, Kronecker, Khatri-Rao and element-wise products, respectively. The $\hat{a}$ denotes an estimate of a, $\mathbf{1}_m$ is a one vector with length m, Re$(\cdot)$ takes the real part of its argument, diag$(\cdot)$ is a diagonal matrix and vec$(\cdot)$ is the vectorization operator.

II. BSUM-RELAX FOR 1-D HR

In this section, we first give a brief review of the 1-D HR with impulsive noise, and then bring forth our motivation. We subsequently tailor the ℓ_p-norm minimization for 1-D HR and solve it via BSUM-RELAX.

A. Signal Model

In the 1-D HR problem, the signal is modeled as a sum of K sinusoids:

$$x(t) = \sum_{k=1}^{K} \gamma_k e^{j\omega_k t} + n(t),\ t = 0, \ldots, N-1 \tag{1}$$

where ω_k and γ_k are the frequency and complex amplitude of the kth sinusoid, respectively, N is the number of uniformly-sampled measurements and $n(t)$ is the additive noise, which is assumed to be outliers with unknown heavy-tailed distribution.

The problem of interest is to estimate the mixed spectrum, i.e., the K unknown amplitudes and frequencies of sinusoids from N samples $\mathbf{x} = [x(t), \ldots, x(N)]^T$ that are possibly corrupted by outliers. When $n(t)$ is Gaussian distributed, γ_k and ω_k can be obtained by solving the following nonlinear least squares (LS) problem:

$$\min_{\gamma_k, \omega_k} \left\| \mathbf{x} - \sum_{k=1}^{K} \gamma_k \mathbf{a}(\omega_k) \right\|_2^2 \tag{2}$$

where

$$\mathbf{a}(\omega_k) = \left[1\ e^{j\omega_k}\ \cdots\ e^{j(N-1)\omega_k}\right]^T. \tag{3}$$

When the noise is white, the problem in (2) is indeed the ML formulation for HR, and the standard RELAX method is actually an alternative scheme for solving (2). In the past few decades, there are a plenty of algorithms that have been devised to solve (2), e.g., [6], [7]. However, (2) is only optimal for Gaussian noise. When outliers appear, the HR estimators based on this cost function may fail to work properly [14]–[16]. Outliers/impulsive noise exists in a variety of real-life environments, crossing various frequency bands [21]. There are many sources producing such a type of noise, for example thunderstorms and man-made noises like electricity substation or power system switching. These noises can significantly perturb the data and then degrade the performance of conventional HR algorithms. Thus, our goal here is to design a robust, efficient and easily implemented optimization framework for HR in the presence of outliers.

B. Algorithm Design

1) Single-Tone Case: Subject to the fact that outliers are sparse, we may employ the ℓ_0-norm for robustness, i.e.,

$\|\mathbf{x} - \gamma\mathbf{a}(\omega)\|_0$. However, the straightforward minimization of ℓ_0-norm is recognized undoable. Motivated by [14]–[18], where good results in practice are achieved when the quasi ℓ_p-norm is selected to approximate the ℓ_0-norm, we propose to use the ℓ_p-fitting criterion instead of the ℓ_2-norm for robust HR.

To simplify the analysis, let us start from the single-tone case. Under this setting, the optimization problem can be described as

$$\min_{\gamma,\omega} \sum_{i=1}^{N} \left(|x_i - \gamma a_i(\omega)|^2 + \epsilon\right)^{p/2} \tag{4}$$

where $0 < \epsilon < 1$ is a small regularization parameter that keeps the cost function differentiable [13], [17]. When $1 < p \leq 2$, we can simply set $\epsilon = 0$ since the ℓ_p-norm is differentiable. However, when $p \leq 1$ and $\epsilon = 0$, the cost becomes non-smooth and non-differentiable. In this case, choosing a small ϵ is helpful. Typically, $\epsilon \in [10^{-8}, 10^{-6}]$ suffices to serve the purpose. We note that the above optimization problem is difficult to be optimized, since the objective function is non-convex. Moreover, our target of interest is embedded in the exponential term, meaning that the signal model is inherently nonlinear in $\{\omega, \gamma\}$. This further increases the difficulties.

To deal with (4), we first consider the following lemma–Young's inequality [22] which holds for a general function $g(\cdot)$: $\mathbb{R} \to \mathbb{R}$

Lemma 2.1: Assuming $0 < p \leq 2$, $x > 0$, $y > 0$, we have

$$x^p \leq \frac{p}{2}\frac{x^2}{y^{2-p}} + \frac{2-p}{2} y^{\frac{2}{2-p}} \tag{5}$$

where the equality holds true if and only if $x = y$.

It follows that given the estimates $\gamma^{(\ell)}$ and $\omega^{(\ell)}$ at the ℓth iteration, we have

$$\left(|x_i - \gamma a_i(\omega)|^2 + \epsilon\right)^{p/2} \leq \frac{p}{2}\frac{|x_i - \gamma a_i(\omega)|^2}{\left(|x_i - \gamma^{(\ell)} a_i(\omega^{(\ell)})|^2 + \epsilon\right)^{1-p/2}} + \text{const} \tag{6}$$

where x_i denotes the ith element in $\mathbf{x}$, and likewise for a_i. Thus, an upper bound of the objective function in (4) can be found

$$\|\mathbf{x} - \gamma\mathbf{a}(\omega)\|_p^p \leq \sum_{i=1}^{N} w_i^{(\ell)} |x_i - \gamma a_i(\omega)|^2 + \text{const.} \tag{7}$$

where $w_i^{(\ell)} = \frac{p}{2}\left(|x_i - \gamma^{(\ell)} a_i(\omega^{(\ell)})|^2 + \epsilon\right)^{p/2-1}$.

Instead of solving (4) directly, we majorize its objective function and minimize the resulting upper bound at the right-hand-side of (6) iteratively. Note that this step corresponds to the *majorization minimization* technique [23]–[25]. Then the problem for the $(\ell+1)$th iteration is expressed as

$$\min_{\gamma,\omega} \left\| \sqrt{\mathbf{w}^{(\ell)}} \odot (\mathbf{x} - \gamma\mathbf{a}(\omega)) \right\|_2^2 \tag{8}$$

where the weighting vector is

$$\begin{aligned} \mathbf{w}^{(\ell)} &= \left[w_1^{(\ell)} \cdots w_N^{(\ell)}\right]^T \\ &= \frac{p}{2}\left(\left|\mathbf{x} - \gamma^{(\ell)}\mathbf{a}(\omega^{(\ell)})\right|^2 + \epsilon\right)^{p/2-1}. \end{aligned} \tag{9}$$

We note here that (8) is a weighted nonlinear LS problem, and its optimal solution is related to the frequency corresponding to the peak of the periodogram of $\mathbf{w}^{(\ell)} \odot \mathbf{x}$. Moreover, the optimal γ can be analytically derived as a function of ω. In other words, for a given ω, the cost in (8) can be concentrated w.r.t. γ [26]. This results in a LS estimate of γ, that is,

$$\gamma = \frac{\mathbf{a}^H(\omega)(\mathbf{w}^{(\ell)} \odot \mathbf{x})}{\sum_{i=1}^{N} w_i^{(\ell)}}. \tag{10}$$

Substituting it back in (8), the problem for ω can be equivalently expressed as

$$\omega^{(\ell+1)} = \arg\max_{\omega} \left|\mathbf{a}^H(\omega)(\mathbf{w}^{(\ell)} \odot \mathbf{x})\right|^2 \tag{11}$$

i.e., finding ω that maximizes the spectrum of $(\mathbf{w}^{(\ell)} \odot \mathbf{x})$.

Problem (11) only contains one unknown variable, so the associated gradient and Hessian are both scalars. As we will see later, the calculations of the gradient and Hessian are very simple, which only take about $\mathcal{O}(N)$ flops,[1] respectively. On the other hand, for the single-tone HR, when the signal-to-noise ratio (SNR) is relatively high, it is almost sure that the highest peak of the spectrum reflects the actual location of the unknown frequency [6]. This enables to find a good initialization of ω through calculating the spectrum of $\mathbf{w}^{(\ell)} \odot \mathbf{x}$. These advantages motivate us to employ the Newton's method to cope with (11). Specifically, solving (11) is equivalent to minimizing its negative objective and the Newton's update is written as

$$\omega^{(r+1)} = \omega^{(r)} - \mu^{(r)} H^{-1}(\omega^{(r)}) g(\omega^{(r)}). \tag{12}$$

where $\mu^{(r)}$ is the step-size, (r) denotes the rth inner iteration for updating ω, the gradient $g(\omega)$ and Hessian $H(\omega)$ are given by

$$g(\omega) = -2\text{Re}\left((\mathbf{t} \odot \mathbf{a}(\omega))^H (\mathbf{w}^{(\ell)} \odot \mathbf{x})(\mathbf{w}^{(\ell)} \odot \mathbf{x})^H \mathbf{a}(\omega)\right) \tag{13}$$

$$\begin{aligned} H(\omega) = &-2\text{Re}\Big((\mathbf{t} \odot \mathbf{t} \odot \mathbf{a}(\omega))^H (\mathbf{w}^{(\ell)} \odot \mathbf{x})(\mathbf{w}^{(\ell)} \odot \mathbf{x})^H \mathbf{a}(\omega) \\ &+ \left|(\mathbf{t} \odot \mathbf{a}(\omega))^H (\mathbf{w}^{(\ell)} \odot \mathbf{x})\right|^2\Big) \end{aligned} \tag{14}$$

with $\mathbf{t} = j\,[0, 1, \ldots, (N-1)]^T$. One can see clearly in (13) that the complexity for calculating the gradient mainly lies in the computations of two inner products, which is about $\mathcal{O}(N)$ flops. The computational cost of the Hessian in (14) is slightly higher than the gradient, where there are triple inner products with each of which requiring $\mathcal{O}(N)$ flops. Hence, as we mentioned before, the complexity for Newton's update is around $\mathcal{O}(N)$ flops, which is marginal.

When the Newton's update is terminated, we return its output as an estimate of $\omega^{(\ell+1)}$. Finally, by substituting $\omega^{(\ell+1)}$ into (10), the amplitude γ is refined as

$$\gamma^{(\ell+1)} = \frac{\mathbf{a}^H(\omega^{(\ell+1)})(\mathbf{w}^{(\ell)} \odot \mathbf{x})}{\sum_{i=1}^{N} w_i^{(\ell)}}. \tag{15}$$

[1] Here, $\mathcal{O}(\cdot)$ is defined as the order of number of complex multiplications.

Algorithm 1: BSUM-RELAX For 1-D HR With Single-Tone.

1: **function** BSUM-RELAX $\mathbf{x}, p$
2: Set $\ell = 1$
3: Initialize $\omega^{(0)}, \gamma^{(0)}$ using oversampled FFT
4: **while** stopping criterion has not been reached **do**
5: $\omega^{(\ell)} \leftarrow$ solve (11)
6: $\gamma^{(\ell)} \leftarrow$ (15)
7: $\ell = \ell + 1$
8: **end while**
9: Return $\hat{\omega}$ and $\hat{\gamma}$
10: **end function**

We iteratively refine the weights $\mathbf{w}$, ω and γ until a stopping criterion is reached. The steps for implementing BSUM-RELAX are summarized in Algorithm 1.

Algorithm 1 deals with a non-convex optimization problem, where the subproblems with respect to (w.r.t.) the frequency is non-linear and non-convex too. One may be interested in the convergence of Algorithm 1: whether this algorithm converges to a stationary or Karush Kuhn Tucker (KKT) point or not. The following proposition shows the convergence of Algorithm 1.

Proposition 2.1: Assume that $0 < p < 2$ and $\epsilon > 0$. Then, the solution sequence produced by Algorithm 1 converges to a set that contains all the KKT points of Problem (4). ■

Proof: See Appendix A.

Remark 2.1

- We can employ the backtracking line search technique to determine the step-size for Newton's method. In the high SNR scenarios, the periodogram often has only one dominant peak located around the true frequency. This means that we may employ the over-sampled FFT to initialize the Newton's method. According to our experience, with $\mu = 1$, one iteration is usually sufficient to obtain satisfactory performance. Also note that it is not necessary to run FFT in every outer iteration, since the current frequency estimate can be reused for the next iteration.
- The standard RELAX algorithm does not specify a certain method to compensate the off-grid frequencies caused by FFT. Although we may employ zoom FFT (Z-FFT) to reduce the off-grid gap, the resulting complexity is very high. This is mainly caused by the fact that Z-FFT requires a few iterations to obtain satisfactory accuracy, where each iteration involves a FFT operation. Fortunately, Newton's method can be directly applied to accelerate the standard RELAX algorithm by simply setting $p = 2$ and replacing the weighting vector $\mathbf{w}^{(\ell)}$ with $\mathbf{1}_N$.

2) Multiple-Tone Case: Next we consider a scenario where the spectrum is mixed with $K > 1$ sinusoids. Similar to the single-tone case, we propose to solve

$$\min_{\boldsymbol{\gamma}, \boldsymbol{\omega}} \sum_{i=1}^{N} \left(\left| x_i - \sum_{k=1}^{K} \gamma_k a_i(\omega_k) \right|^2 + \epsilon \right)^{p/2} \tag{16}$$

where $\boldsymbol{\gamma} = [\gamma_1 \ \cdots \ \gamma_K]^T$ and $\boldsymbol{\omega} = [\omega_1 \ \cdots \ \omega_K]^T$.

Suppose that after ℓ iterations, there are K sinusoidal estimates available. In the next iteration, we set up K inner iterations to alternately refine the sinusoids. In particular, the kth inner iteration updates the kth sinusoid component–$\{\omega_k, \gamma_k\}$ and the resulting estimates are denoted by $\{\omega_k^{(\ell+1)}, \gamma_k^{(\ell+1)}\}$. This way, the kth sinusoid can be refined with Algorithm 1 by subtracting the other $(K-1)$ sinusoids from the data. Mathematically, this is expressed as

$$\min_{\omega_k, \gamma_k} \sum_{i=1}^{N} \left(\left| y_{i,k}^{(\ell+1)} - \gamma_k a_i(\omega_k) \right|^2 + \epsilon \right)^{p/2} \tag{17}$$

where

$$\mathbf{y}_k^{(\ell+1)} = \mathbf{x} - \sum_{i=1}^{k-1} \gamma_i^{(\ell+1)} \mathbf{a}(\omega_i^{(\ell+1)}) - \sum_{i=k+1}^{K} \gamma_i^{(\ell)} \mathbf{a}(\omega_i^{(\ell)}) \tag{18}$$

and $y_{i,k}^{(\ell+1)}$ is the kth element of $\mathbf{y}_k^{(\ell+1)}$. It is obvious that Problem (17) is a single-tone HR problem, such that it can be solved using Algorithm 1. We refine all K harmonic estimates in the same way as described above.

It is interesting to determine the source number K; however, the source number detection problem is out of the scope of this paper. Here, we recommend the readers to use the following Bayesian information criterion (BIC) [7] to estimate K, i.e.,

$$\mathrm{BIC}(\eta) = 2N \ln \left(\left\| \mathbf{y} - \sum_{k=1}^{\eta} \hat{\gamma}_k \mathbf{a}(\hat{\omega}_k) \right\|_2^2 \right) + 3\eta \ln(2N). \tag{19}$$

The detailed steps for implementing BSUM-RELAX with BIC are summarized in Algorithm 2. Although (19) is based on Gaussian assumption, it may not be proper for our setting. But our experience shows that BIC works well even there is outliers, especially in high SNRs; as we will see in Section IV-B.

We note that BSUM-RELAX for multiple tones belongs to a class of BSUM algorithm [27], [28]. However, due to the non-convexity of each subproblem w.r.t. $\{\omega_k, \gamma_k\}$, the convergence of BSUM-RELAX for multiple tones case is still unknown. Currently, we can only guarantee the decrease of the objective function in (16). The related convergence proof for the multiple tones case will be our future work.

C. Complexity Analysis

Since Algorithm 2 is based on Algorithm 1, let us first analyze the complexity of Algorithm 1 for single-source case. In each iteration, the complexities for the computations of $\mathbf{w}$ and γ are similar, i.e., $\mathcal{O}(N)$ flops. We employ the Newton's method to estimate ω. It follows from (13) and (14) that the calculations of gradient and Hessian are related to three terms, i.e., $(\mathbf{t} \odot \mathbf{a}(\omega))^H (\mathbf{w} \odot \mathbf{x})$, $(\mathbf{w} \odot \mathbf{x})^H \mathbf{a}(\omega)$ and $(\mathbf{t} \odot \mathbf{t} \odot \mathbf{a}(\omega))^H (\mathbf{w} \odot \mathbf{x})$, where each term costs $\mathcal{O}(N)$ flops. Moreover, the first two terms are reusable, such that the complexities for gradient and Hessian can be further reduced. We mention that according to our convergence analysis, an inexact update for the frequency estimate in each iteration can guarantee convergence. This implies that one iteration is enough for Newton's method. Thus, the complexity for Newton's update is about $\mathcal{O}(N)$. The bottleneck

Algorithm 2: BSUM-RELAX For Multiple Tones.

function BSUM-RELAX $\mathbf{x}, p, K_{\max}$
  if the source number K is known **then**
    Set $K_{\max} = K$
  end if
  for $i = 1, \ldots, K_{\max}$ **do**
    Set $\ell = 0$
    while stopping criterion has not been reached **do**
      $\ell = \ell + 1$
      for $k = i, 1, 2, \ldots, i-1$ **do**
        $\mathbf{y}_k^{(\ell)} \leftarrow$ (18), where K in (18) is set to be i.
        $(\omega_k^{(\ell)}, \gamma_k^{(\ell)}) \leftarrow$ Run Algorithm 1 on $\mathbf{y}_k^{(\ell)}$ with a fixed number of iterations. (A default setting can be one iteration.)
      end for
    end while
    Set $\hat{\boldsymbol{\omega}} = [\omega_1^{(L)}, \ldots, \omega_i^{(L)}]$ and $\hat{\boldsymbol{\gamma}} = [\gamma_1^{(L)}, \ldots,]$, where L is the maximum number of iterations in the while loop
    Calculate $\mathrm{BIC}(i)$ via (19)
  end for
  if the source number is known **then**
    Set $\hat{K} = K$
  else
    Estimate the source number as the index of the minimum BIC value, which is expressed as $[\mathrm{BIC}_{\min}, \hat{K}] = \min([\mathrm{BIC}(1), \ldots, \mathrm{BIC}(K_{\max})])$ in MATLAB.
  end if
  Return $\hat{\boldsymbol{\omega}}(1:\hat{K})$ and $\hat{\boldsymbol{\gamma}}(1:\hat{K})$
end function

here is the initialization of ω. Suppose that we choose the initial point from oversampled FFT. The complexity may reach $\mathcal{O}(LN + N\log N)$ flops, where L is the number of iterations for convergence.

In the multiple-tone case, there are K sources estimated via Algorithm 1. Then the complexity of our method in each iteration is $\mathcal{O}(K(LN + N\log N))$ flops. Obviously, the complexity is proportional to both K and N, indicating that the proposed method is efficient for small K but it could be a bit more complicated when K is large.

We here carry out a simulation to showcase the complexity of BSUM-RELAX as a function of K, where the results are plotted in Fig. 1. The RELAX method is compared with the BSUM-RELAX. Note that for RELAX, the MATLAB function fminbnd is employed to update frequency in (11) with a default setting. Moreover, we replace fminbnd in RELAX by the presented Newton's method to verify its efficiency. A two-component Gaussian mixture distribution is used to generate noise, where the first component has a weight 0.9 and variance 1 while the second one has a weight 0.1 and variance 100. SNR is 5 dB, N is 64, and K is increased from 1 to 6. The frequency of the kth harmonic is generated as $\omega_k = 0.2\pi k, \forall k = 1, \ldots, K$, and γ_k has unit magnitude and random phase. The $p = 1$ is

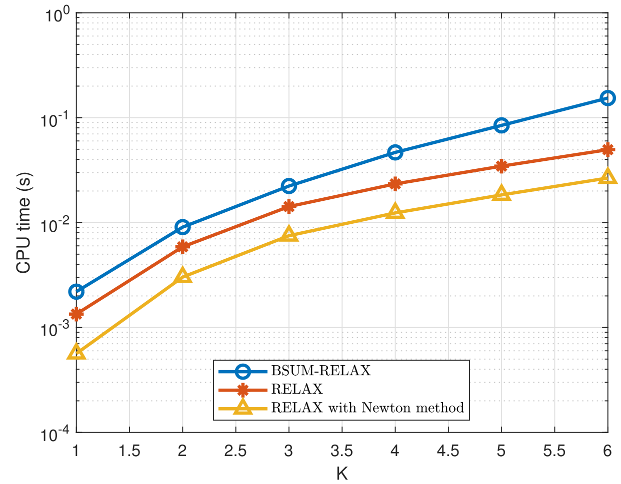


Fig. 1. Complexity comparison as a function of K.

chosen for BSUM-RELAX. It is seen that the complexity of BSUM-RELAX is the largest since it requires more iterations to achieve robustness. RELAX with Newton's method has lower complexity than the classical RELAX, which confirms the efficiency of our Newton's method.

Remark 2.2: Although our method is derived based on the undamped sinusoidal model, it can be generalized to estimate damped sinusoidal parameters. We refer the interested readers to [9] for further details.

III. BSUM-RELAX FOR MULTIDIMENSIONAL HR

Although 1-D HR seems to be most common, many applications of multidimensional HR can be found in practice. For example, in wireless communications, especially in the frequency division duplex based massive MIMO system, one should report the CSI oftentimes. This requires to jointly estimate the multipath parameters like path-loss, arrival and departure azimuth/elevation angles, all of which can be viewed as parameters of multidimensional harmonics. Moreover, in radar or wireless communication systems, impulsive noise shown as short duration interference can occur from many ways, e.g., lightning in the atmosphere or radio frequency interference [21]. This motivates us to develop a robust algorithm for the multidimensional HR problem.

A. Signal Model

We proceed to extend the BSUM-RELAX algorithm to multidimensional HR case. Let us consider a single snapshot multidimensional HR model that is formatted as[2]

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}} \in \mathbb{C}^{N_1 \times \cdots \times N_M} \tag{20}$$

[2]Note that when $M = 2$, $\boldsymbol{\mathcal{X}}$ in (20) reduces to a matrix and the problem becomes 2-D HR. Our analysis in this section holds for the 2-D HR problem as well.

where N_m is the sample size of the mth dimension, $\boldsymbol{\mathcal{N}}$ stands for the noise and

$$\boldsymbol{\mathcal{S}} = \sum_{k=1}^{K} \alpha_k \mathbf{a}_{1,k} \circ \cdots \circ \mathbf{a}_{M,k} \tag{21}$$

represents the signal component with its $(i_1, \ldots, i_M)$ entry as

$$s_{i_1,\ldots,i_M} = \sum_{k=1}^{K} \alpha_k \prod_{m=1}^{M} e^{j\nu_{m,k}(i_m - 1)}. \tag{22}$$

In $\boldsymbol{\mathcal{S}}$, there are K multidimensional harmonics, each of which is parameterized by a complex amplitude α_k and M frequencies $\boldsymbol{\nu}_k = [\nu_{1,k} \ \ldots \ \nu_{M,k}]^T \in \mathbb{R}^M$, $\mathbf{a}_{m,k}$ is Vandermonde containing frequency $\nu_{m,k}$, i.e., $\mathbf{a}_{m,k} = [1 \ e^{j\nu_{m,k}} \ \cdots \ e^{j(N_m-1)\nu_{m,k}}]^T$. The problem here is to estimate $\{\alpha_k, \boldsymbol{\nu}_k\}_{k=1}^K$ from $\boldsymbol{\mathcal{X}}$.

A plethora of techniques have been developed to estimate multiple multidimensional frequencies from $\boldsymbol{\mathcal{X}}$, including Fourier-based non-parametric methods and modern parametric methods, such as multidimensional MUSIC, multidimensional (unitary) ESPRIT [29], TPUMA [12], multidimensional folding (MDF) [30] and improved MDF (IMDF) [31]. Among these methods, the multidimensional (unitary) ESPRIT, MDF and IMDF are all rotational invariance based techniques, which do not have statistical efficiency; while the multidimensional MUSIC and TPUMA may achieve statistical efficiency when all the harmonics are statistically independent. Generally speaking, TPUMA has higher estimation accuracy than any of the mentioned methods in many challenging scenarios including the sample starving and coherent sources cases. However, due to the Gaussian assumption behind these methods, none of them are robust to outliers.

For robust HR, a family of ℓ_p-based estimators have been proposed, e.g., ℓ_p-MUSIC [14] and HOSVD based robust HR methods including ℓ_p-HOSVD [16]. Nevertheless, their identifiability, i.e., the maximum number of identifiable harmonics, is limited by the smallest dimension of the multidimensional HR. In other words, to apply ℓ_p-HOSVD or ℓ_p-MUSIC, one has to assume $K \le \min(N_1, \ldots, N_M)$. In certain applications such an assumption may not hold true, e.g., downlink channel estimation in massive MIMO. More recently, a non-parametric method named as ℓ_p-iterative adaptive approach (ℓ_p-IAA) is developed for robust HR. Unlike ℓ_p-HOSVD/MUSIC, the ℓ_p-IAA has higher identifiability [15]. However, this method is inherently a dictionary based technique, where the estimation accuracy is limited by the minimum resolution of frequency grids. Furthermore, ℓ_p-IAA is a searching-based estimator, i.e., it searches over the frequency grids to obtain the spectrum. Thus, its computational complexity is inevitably high. To overcome the aforementioned flaws, we propose a novel efficient algorithm by generalizing the BSUM-RELAX method to the multidimensional HR case.

B. Algorithm Design

1) Single-Tone Case: For the purpose of analysis, we start from the single-tone case, where unlike the 1-D HR problem, we now have a multidimensional sinusoid which is parameterized by M frequencies and one amplitude. After applying the ℓ_p-fitting criterion, we have

$$\min_{\alpha,\boldsymbol{\nu}} \sum_{i=1}^{\breve{N}} \left(|\breve{x}_i - \alpha \breve{a}_i(\boldsymbol{\nu})|^2 + \epsilon\right)^{p/2} \tag{23}$$

where $\breve{N} = \prod_{i=1}^{M} N_i$, $\breve{\mathbf{x}} = \mathrm{vec}(\boldsymbol{\mathcal{X}})$ and $\breve{\mathbf{a}}(\boldsymbol{\nu}) = \mathbf{a}_M \otimes \cdots \otimes \mathbf{a}_1$.

It follows from Lemma 2.1 that the cost function in (23) is upper bounded by

$$\sum_{i=1}^{\breve{N}} \left(|\breve{x}_i - \alpha \breve{a}_i(\boldsymbol{\nu})|^2 + \epsilon\right)^{p/2} \le \sum_{i=1}^{\breve{N}} \breve{w}_i^{(\ell)} |\breve{x}_i - \alpha \breve{a}_i(\boldsymbol{\nu})|^2 + \text{const.} \tag{24}$$

where $\breve{w}_i^{(\ell)} = \frac{p}{2}\left(|\breve{x}_i - \breve{\gamma}^{(\ell)} \breve{a}_i(\nu^{(\ell)})|^2 + \epsilon\right)^{p/2-1}$. Thus, by minimizing the majorized cost at the right-hand-side of (24), we have

$$\min_{\alpha,\boldsymbol{\nu}} \left\| \sqrt{\breve{\mathbf{w}}^{(\ell)}} \odot \left(\breve{\mathbf{x}} - \alpha \breve{\mathbf{a}}(\boldsymbol{\nu})\right) \right\|_2^2 \tag{25}$$

where $\breve{\mathbf{w}}^{(\ell)}$ is the weighting vector characterized by $(\alpha^{(\ell)}, \boldsymbol{\nu}^{(\ell)})$.

Following the analysis of (10), at the $(\ell+1)$th iteration, the ML formulation w.r.t. $\boldsymbol{\nu}$ is cast as

$$\boldsymbol{\nu}^{(\ell+1)} = \min_{\boldsymbol{\nu}} \left| \breve{\mathbf{a}}(\boldsymbol{\nu})^H \left(\breve{\mathbf{w}}^{(\ell)} \odot \breve{\mathbf{x}}\right) \right|^2. \tag{26}$$

The above problem is merely a conventional M-D single-tone HR problem, where the measurement vector is $\breve{\mathbf{w}}^{(\ell)} \odot \breve{\mathbf{x}}$. This subproblem can be optimally solved using a M-D ML method generalized from the ML for 1-D HR [6]. However, implementing the ML algorithm in each subproblem is very computationally intensive. An alternative idea is to derive a Newton's method for updating the frequencies.

Unlike the 1-D HR, there are M frequencies in each dimension. Hence, the gradient and Hessian w.r.t. $\boldsymbol{\nu}$ are no longer scalars but

$$\breve{\mathbf{g}}(\boldsymbol{\nu}) = [\breve{g}(\nu_1) \ \cdots \ \breve{g}(\nu_M)]^T \in \mathbb{R}^M \tag{27}$$

$$\breve{\mathbf{H}}(\boldsymbol{\nu}) = \frac{\partial \breve{\mathbf{g}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^T} \in \mathbb{R}^{M \times M} \tag{28}$$

where the mth element of $\breve{\mathbf{g}}$ and the (m,n) entry of $\breve{\mathbf{H}}$ are, respectively, given as

$$\breve{g}_m = -2\mathrm{Re}\left((\breve{\mathbf{d}}(\nu_m))^H \breve{\mathbf{p}}^{(\ell)} (\breve{\mathbf{p}}^{(\ell)})^H \breve{\mathbf{a}}(\nu_m)\right) \tag{29}$$

$$\breve{H}_{m,n} = -2\mathrm{Re}\Big((\breve{\mathbf{t}}_n \odot \breve{\mathbf{d}}(\nu_m))^H \breve{\mathbf{p}}^{(\ell)} (\breve{\mathbf{p}}^{(\ell)})^H \breve{\mathbf{a}}(\nu_m) + (\breve{\mathbf{d}}(\nu_m))^H \breve{\mathbf{p}}^{(\ell)} (\breve{\mathbf{p}}^{(\ell)})^H \breve{\mathbf{a}}(\nu_n)\Big) \tag{30}$$

in which $\breve{\mathbf{d}}(\nu_m) = \breve{\mathbf{t}}_m \odot \breve{\mathbf{a}}(\nu_m)$, $\breve{\mathbf{p}}^{(\ell)} = \breve{\mathbf{w}}^{(\ell)} \odot \breve{\mathbf{x}}$ and $\breve{\mathbf{t}}_m = \mathbf{1}_{N_M} \otimes \cdots \otimes (j\,[0, 1, \ldots, N_m - 1]^T) \otimes \cdots \otimes \mathbf{1}_{N_1}$.

Algorithm 3: BSUM-RELAX For M-D HR With Single-Tone.

function BSUM-RELAX $\mathbf{x}, p, \alpha^{(0)}, \boldsymbol{\nu}^{(0)}$
  Set $\ell = 1$.
  while stopping criterion has not been reached **do**
    $\boldsymbol{\nu}^{(\ell)} \leftarrow$ solve(26)
    $\alpha^{(\ell)} \leftarrow (32)$
    $\ell = \ell + 1$
  end while
  Return $\hat{\alpha}$ and $\hat{\boldsymbol{\nu}}$
end function

With the gradient and Hessian, the Newton's update for $\boldsymbol{\nu}^{(\ell+1)}$ boils down to

$$\boldsymbol{\nu}^{(r+1)} = \boldsymbol{\nu}^{(r)} - \mu^{(r)} \breve{\mathbf{H}}^{-1}(\boldsymbol{\nu}^{(r)}) \breve{\mathbf{g}}(\boldsymbol{\nu}^{(r)}). \tag{31}$$

Based on $\boldsymbol{\nu}^{(\ell+1)}$, we update

$$\alpha^{(\ell+1)} = \frac{\breve{\mathbf{a}}^H(\boldsymbol{\nu}^{(\ell+1)})(\breve{\mathbf{w}}^{(\ell)} \odot \breve{\mathbf{x}}^{(\ell)})}{\sum_{i=1}^{\breve{N}} \breve{w}_i^{(\ell)}}. \tag{32}$$

The details of BSUM-RELAX are summarized in Algorithm 3.

2) Multiple-Tone Case: Similar to the 1-D case, if the number sources is known, we propose to solve

$$\min_{\{\alpha_k, \boldsymbol{\nu}_k\}_{k=1}^K} \left\| \breve{\mathbf{x}} - \sum_{k=1}^K \alpha_k \breve{\mathbf{a}}(\boldsymbol{\nu}_k) \right\|_p^p. \tag{33}$$

By denoting

$$\breve{\mathbf{y}}_k^{(\ell+1)} = \breve{\mathbf{x}} - \sum_{i=1}^{k-1} \alpha_i^{(\ell+1)} \breve{\mathbf{a}}(\boldsymbol{\nu}_i^{(\ell+1)}) - \sum_{i=k+1}^{K} \alpha_i^{(\ell)} \breve{\mathbf{a}}(\boldsymbol{\nu}_i^{(\ell)}) \tag{34}$$

the subproblem w.r.t. the kth sinusoid at the $(\ell+1)$th iteration is expressed as

$$\min_{\alpha_k, \boldsymbol{\nu}_k} \left\| \breve{\mathbf{y}}_k^{(\ell+1)} - \alpha_k \breve{\mathbf{a}}(\boldsymbol{\nu}_k) \right\|_p^p \tag{35}$$

which is a M-D single-tone HR problem, and thus can be solved using Algorithm 3.

If the number of sources is unknown, we may use the BIC criterion for assistance, i.e.,

$$\mathrm{BIC}(\eta) = 2\breve{N} \ln \left(\left\| \breve{\mathbf{x}} - \sum_{k=1}^K \alpha_k \breve{\mathbf{a}}(\boldsymbol{\nu}_k) \right\|_2^2 \right) + 3\eta \ln(2\breve{N}). \tag{36}$$

The explicit steps are summarized in Algorithm 4.

Algorithm 4: BSUM-RELAX For M-D HR With Multiple Tones.

function BSUM-RELAX $\boldsymbol{\mathcal{X}}, p, \boldsymbol{\nu}^{(0)}, \{\alpha_k^{(0)}\}, K$
  if the source number K is known **then**
    Set $K_{\max} = K$
  end if
  for $i = 1, \ldots, K_{\max}$ **do**
    Set $\ell = 0$
    while stopping criterion has not been reached **do**
      $\ell = \ell + 1$
      for $k = i, 1, 2, \ldots, i-1$ **do**
        $\breve{\mathbf{y}}_k^{(\ell)} \leftarrow$ (34), where K in (34) is set to be i.
        $(\boldsymbol{\nu}_k^{(\ell)}, \alpha_k^{(\ell)}) \leftarrow$ Run Algorithm 3 on $\breve{\mathbf{y}}_k^{(\ell)}$ with a fixed number of iterations. (A default setting can be one iteration.)
      end for
    end while
    Set $\hat{\boldsymbol{\nu}}_k = [\nu_{1,k}^{(L)}, \ldots, \nu_{M,k}^{(L)}], \forall k = 1, \ldots, i$ and $\hat{\boldsymbol{\alpha}} = [\alpha_1^{(L)}, \ldots, \alpha_i^{(L)}]$, where L is the number of iterations in the while loop
    Calculate $\mathrm{BIC}(i)$ via (36) by setting $N = \prod_{i=1}^M N_i$
  end for
  if the source number is known **then**
    Set $\hat{K} = K$
  else
    Estimate the source number as the index of the minimum BIC value, which is expressed as $[\mathrm{BIC}_{\min}, \hat{K}] = \min([\mathrm{BIC}(1), \ldots, \mathrm{BIC}(K_{\max})])$ in MATLAB.
  end if
  Return $\{\hat{\boldsymbol{\nu}}_1, \ldots, \hat{\boldsymbol{\nu}}_{\hat{K}}\}$ and $\{\hat{\alpha}_1, \ldots, \hat{\alpha}_{\hat{K}}\}$
end function

IV. Simulation Results

In this section, we evaluate the performance of BSUM-RELAX by comparing it with the state-of-the-art algorithms in terms of root mean square error (RMSE) of frequency estimates, probability of resolution (PR) and CPU time, where 200 Monte-Carlo tests are used to produce the RMSE and PR, which are defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{200} \sum_{i=1}^{200} \|\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}\|_2^2} \tag{37}$$

while PR is calculated based on a binary hypothesis. In particular, $\Theta := [\omega_1 - \delta, \omega_1 + \delta] \cup \cdots \cup [\omega_K - \delta, \omega_K + \delta]$ is the hypothesis, where $\omega_1 < \cdots < \omega_K$, δ is a small threshold, namely, $\delta = 0.25 \min(|\omega_2 - \omega_1|, \ldots, |\omega_K - \omega_{K-1}|)$, and $[\omega_k - \delta, \omega_k + \delta]$ is the kth frequency sector. If all the frequency estimates are successfully localized in their own sectors, we say the frequencies are successfully resolved. For illustration, assuming that after 200 independent tests, if there are T times that all the frequency estimates are in Θ, the PR is calculated as $T/200$. Moreover, for BSUM-RELAX, the step-size used in Newton's method is $\mu = 1$. When $p \geq 1$, we initialize BSUM-RELAX using $32N$-FFT. When $p < 1$, we first choose $p = 1$ and run BSUM-RELAX with 10 iterations, and then use the output to initialize the BSUM-RELAX.

A. Impulsive Noise Model

Unlike Gaussian noise, the probability density functions (PDFs) of outliers (also known as impulsive noise) are usually heavy-tailed. In this section, we consider two widely used impulsive noise models to examine the performance of the proposed method, i.e., Gaussian mixture model (GMM) and α-stable distribution.

1) GMM: We consider a two-component GMM to generate impulsive noise, whose PDF is expressed as [32]

$$p(\mathcal{X}) = \sum_{i=1}^{2} \frac{\varrho_i}{\pi\sigma_i^2} \exp\left(-\frac{\mathcal{X}^2}{\sigma_i^2}\right)$$

where σ_i^2 is the variance of the ith term, and $0 \leq \varrho_i \leq 1$ is the probability of occurrence of the ith component. Here, $\varrho_1 + \varrho_2 = 1$, and we use the second component to represent outliers, such that $\varrho_2 < \varrho_1$ and $\sigma_2^2 > \sigma_1^2$. In the following, we choose $\sigma_1^2 = 1$ and $\sigma_2^2 = 100$, which corresponds to a situation where strong outliers are present with probability ϱ_2. The SNR is defined as $\mathrm{SNR} = \frac{\|\mathbf{A}\boldsymbol{\gamma}\|^2}{N\sum_{i=1}^{2}\varrho_i\sigma_i^2}$. The CRB for GMM noise is provided in Appendix B.

2) α-Stable: The PDF of the α-stable distribution is generally not available, but its characteristic function can be written in closed-form, i.e.,

$$\phi(t;\alpha,\beta,c,\mu) = \exp\left(jt\mu - c^{\alpha}|t|^{\alpha}\left(1 - j\beta\mathrm{sgn}(t)\Phi(\alpha)\right)\right)$$

where $\Phi(\alpha) = \tan(\alpha\pi/2)$, $0 < \alpha \leq 2$ is the stability parameter, $-1 \leq \beta \leq 1$ is a measure of asymmetry, $c > 0$ is a scale factor which measures the width of the distribution and μ is a shift parameter. There are two special cases that admit closed-form PDFs, that is, $\alpha = 1$ and $\alpha = 2$ which yield the Cauchy and Gaussian distributions, respectively. When $\alpha < 2$, the corresponding stable distribution exhibits heavy tails, and thus is considered suitable for modeling impulsive noise. The parameter α controls the density of impulses. In the following, we set $\alpha = 0.8$, and the other parameters are $\beta = 0$, $c = 2$ and $\mu = 0$, resulting in a symmetric α-stable (SαS) distribution with zero-shift. In the presence of α-stable noise, the standard SNR is no longer appropriate to quantify the relative strength between signal and noise. Instead, a generalized SNR (GSNR) defined as [14] $\mathrm{GSNR} = \frac{\|\mathbf{A}\boldsymbol{\gamma}\|^2}{Nc^{\alpha}}$, is employed.

B. Parameter Selection for BSUM-RELAX

1) How to initialize: The optimization problem of the proposed method is non-convex, and thus the initial point is critical for our method to produce good frequency estimates. There are many ways to initialize our method, e.g., using an existing robust HR algorithm. However, we should take the complexity of initialization into consideration. To balance the complexity and accuracy, an efficient initialization method has been stated in Algorithm 2. Specifically, in the first outer iteration, we implement Algorithm 1 to estimate the first harmonic component from $\mathbf{x}$, denoted by $\hat{\mathbf{x}}_1$. Then we subtract the first estimated harmonic from the data and estimate the second component via Algorithm 1. We refine the two components alternately for a few iterations by fixing one and solving for the other. After that we

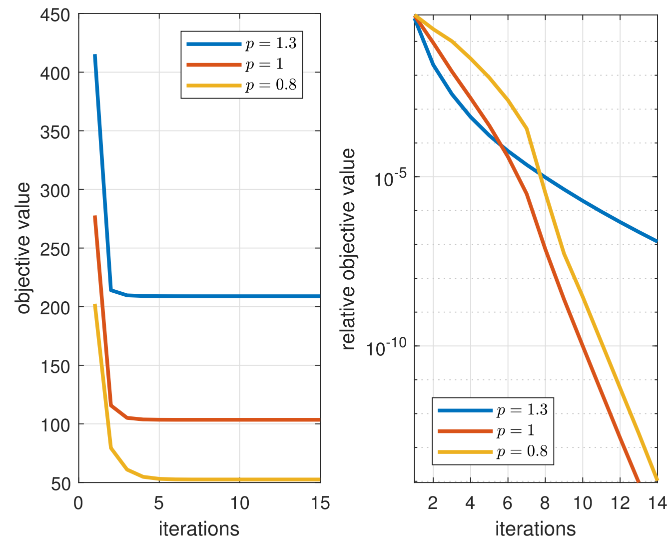


Fig. 2. Objective function value in (4) versus number of iterations.

subtract the two estimated harmonics from the data to estimate the third harmonic, where a few iterations are implemented to refine the three harmonics. We do the same for the remaining harmonics until the number of harmonics reaching K. The previous $(K-1)$ outer iterations are for the final estimates. This idea was suggested by Li *et al.* [7], and it works well in our case, as we will see in the following examples, where unless otherwise specified, we initialize BSUM-RELAX like this.

2) When to stop: Since BSUM-RELAX is an iterative method, it is instructive to know its convergence property before comparing it with the other techniques. Let us consider a case where there are two sinusoids with $N = 128$ samples. Among these samples, $\varrho = 0.1$ of them are corrupted by outliers. The parameters for sinusoids are $\omega_1 = 0.1\pi$, $\omega_2 = 0.2\pi$, $\gamma_1 = 3e^{2j}$ and $\gamma_2 = 2e^{3j}$. We first plot the real and imaginary parts of signal $\sum_{k=1}^{K}\gamma_k\mathbf{a}(\omega_k)$ and noise $\mathbf{n}$ in Fig. 3 to illustrate the relative strength between signal and noise. We examine the convergence of BSUM-RELAX under $p = 0.8, 1$ and 1.3. Fig. 2 plots the objective function value in (16) and its relative value defined as $|f^{(\ell)} - f^{(\ell-1)}|/f^{(\ell)}$ as a function of iterations, where $f^{(\ell)}$ denotes the objective value at iteration ℓ. It is shown that BSUM-RELAX with different p's converge after 8 iterations. We note here that in many cases, BSUM-RELAX can provide comparable performance within 8 iterations. In the following, we terminate for BSUM-RELAX when the relative objective function value is smaller than 10^{-6} or 10 iterations are reached.

3) How to choose ϵ and p: We study how p and ϵ affect the performance of BSUM-RELAX. The parameter setting for generating signal model is the same as those in Fig. 2. We vary p from 0.1 to 2 and choose four ϵ's for BSUM-RELAX, i.e., $\epsilon \in \{10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}\}$. It is worth noting that when $p < 1$, the optimization problem for our method is more difficult to be solved than the case with $p \geq 1$. If we still use the initial point for $p \geq 1$ to initialize BSUM-RELAX with $p < 1$, our method can easily stuck at local minimum and fail to estimate the parameters. To make our method perform well when $p < 1$, we first run BSUM-RELAX with $p = 1$ for 10 iterations, and

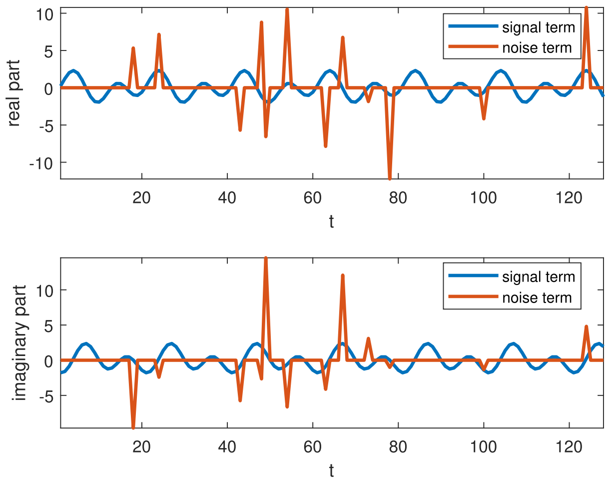
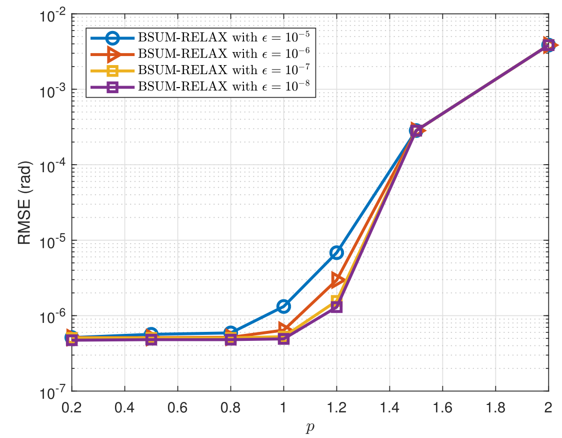
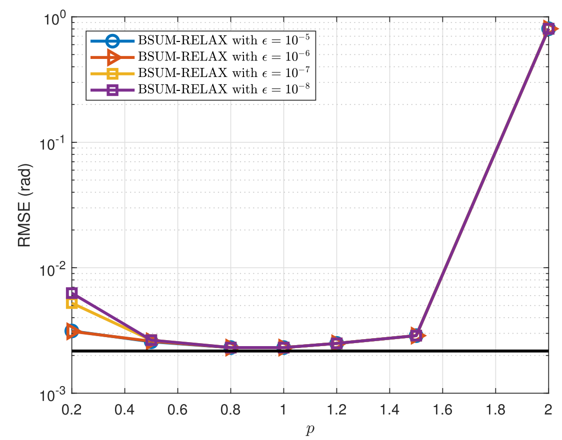


Fig. 3. Real and imaginary parts of $\sum_{k=1}^{K} \gamma_k \mathbf{a}(\omega_k)$ and additive GMM noise under SNR $= -5$ dB.



(a) outliers only



(b) GMM noise

Fig. 4. RMSE of BSUM-RELAX versus p under different ϵ's.

then use its estimate to initialize BSUM-RELAX with $p < 1$. Fig. 4(a) plots the RMSE comparison as a function of p, where SNR is -5 dB and 10% of the samples are corrupted by Gaussian noise with variance 100 while the remaining samples are noiseless. It is seen in Fig. 4(a) that under different ϵ's, BSUM-RELAX produces very similar RMSE results. Moreover, BSUM-RELAX with a smaller p has lower RMSE than that with a relatively larger p, which means that in this case, small p helps improving the robustness of the algorithm.

In Fig. 4(b), we study the impact of p in GMM noise with parameters $\varrho_1 = 0.9, \varrho_2 = 0.1, \sigma_1 = 1$ and $\sigma_2 = 100$. In this scenario, ϵ plays an important role when p is extremely small. BSUM-RELAX with $\epsilon = 10^{-7}$ and $\epsilon = 10^{-8}$ shows slightly large variance when $p = 0.2$. Fortunately, apart from such a small p, our method works well with a wide range of p and ϵ. Generally speaking, our method can provide comparable performance when p is close to 1, for which a small ϵ may offer better performance, e.g., Fig. 4(a). Therefore, unless stated otherwise, we choose $p = 1$ and $\epsilon = 10^{-8}$ for BSUM-RELAX in the following examples.

4) How to Determine K: This question is indeed related to the source number detection problem which is also known as model order selection. We here examine how well BIC works in some tough scenarios. We consider a 1-D HR model where there are $K = 2$ equal power sinusoids with length $N = 32$ in the GMM noise. Both closely- and widely-spaced frequencies are simulated, where the former includes a pair of frequencies $\omega_1 = 0.1\pi$ and $\omega_2 = 0.15\pi$ while the latter are $\omega_1 = 0.1\pi$ and $\omega_2 = 0.3\pi$. We set the maximum number of signals used in Algorithm 2, i.e., $K_{\max}$, as four. The probability of detection (PD) is computed as PD $= \frac{1}{500}\sum_{i=1}^{500} 1_{\hat{K}^{(i)}=K}$, where $\hat{K}^{(i)}$ is the source number estimate from the ith Monte-Carlo test and $1_{\text{condition}} = 1$ if and only if the condition is true. We choose $p = 1.3$, 1 and 0.8 for BSUM-RELAX. Fig. 5 shows the results, where the PDs of BSUM-RELAX under different p's are close to one when SNR ≥ 4 dB in both closely- and widely-spaced cases. In addition, BSUM-RELAX has higher PD when

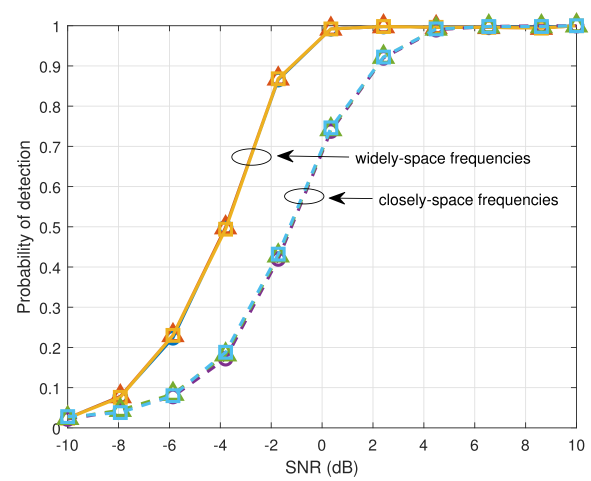


Fig. 5. Probability of detection of BSUM-RELAX using BIC, where □, △ and O stand for BSUM-RELAX with $p = 1.3, p = 1$ and $p = 0.8$, respectively.

frequencies are widely-spaced. We must clarify that among the 500 independent tests, BIC does not guarantee to offer an exact source number estimate, even in high SNRs. Actually, the PDs in this example are very close to one (e.g., 0.996 in SNR = 8 dB) but not one. However, these are optimistic and promising results, which implies that BIC works remarkably well. In the following examples, we would like to avoid using BIC to determine the source number since there is a possibility that BIC outputs an incorrect source number estimate which brings challenge when we calculate the RMSEs. To avoid such scenarios, we therefore assume that the actual source number is known to all the competitors.

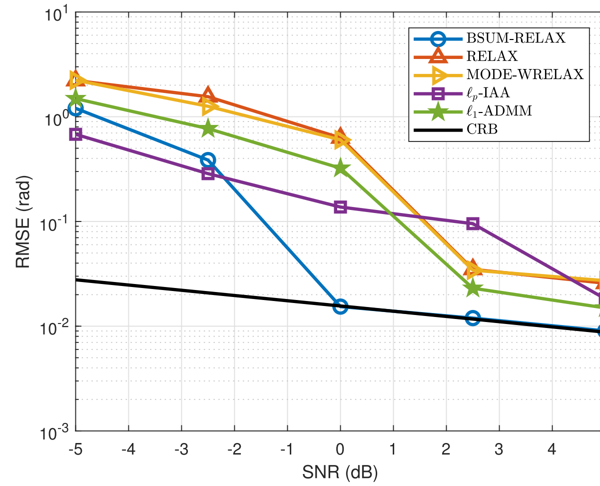
C. Performance Comparison: 1-D HR

1) RMSE and PR versus SNR: In the 1-D case, we compare our method with RELAX [7], MODE-WRELAX [8], ℓ_p-IAA [15], ℓ_1-ADMM [33] in terms of RMSE and PR. The dictionary for ℓ_1-ADMM is constructed by uniformly dividing $[0, 2\pi]$ with 1024 grids, such that the minimum resolution is $2\pi/1024 \approx 0.006$. We choose $p = 1$ for BSUM-RELAX and $p = 1.3$ for ℓ_p-IAA.[3] The stopping criterion for RELAX and MODE-WRELAX is the same as BSUM-RELAX, while the one for ℓ_p-IAA and ℓ_1-ADMM is set as either relative objective function value smaller than 10^{-6} or reaching 100 iterations. Both the GMM and α-stable noises are tested. We first consider the GMM noise scenario where ϱ_2 is fixed at 0.1, i.e., 10% of the data are corrupted by outliers. The other parameters for generating signal model are $N = 32, K = 2, \omega_1 = 0.20\pi, \omega_2 = 0.28\pi, \gamma_1 = 3e^{2j}$ and $\gamma_2 = 2e^{3j}$.

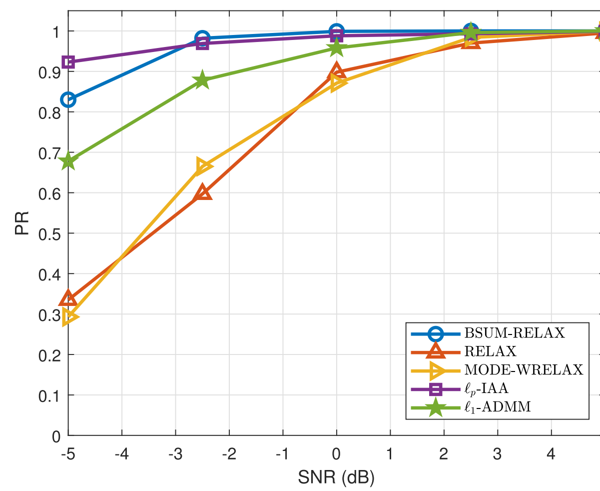
Fig. 6 depicts the RMSE and PR results for the multiple-tone case. We see that BSUM-RELAX performs the best when SNR > -2 dB. ℓ_p-IAA has the smallest RMSE in SNR $= -5$ dB, but it is outperformed by BSUM-RELAX and ℓ_1-ADMM as SNR increases. The ℓ_1-ADMM has high thresholding performance compared to the other competitors. This is mainly caused by some "outages", which correspond to the local minima that are not close to the ground truth frequencies. Fig. 6(b) helps to explain our analysis, where the PR of ℓ_1-ADMM does not reach 1 when SNR < 2 dB, implying that in most Monte-Carlo tests, the frequency estimates obtained from ℓ_1-ADMM have small variances, while in the remaining small number of tests, it fails to resolve the two harmonics and produces outages that cause the RMSE very large. Moreover, we find that the ℓ_1-ADMM algorithm is sensitive to the hyper-parameter, which might be the main reason for its bad performance. Although ℓ_p-IAA is non-parametric, it is intrinsically a dictionary based technique, which is a biased estimator and cannot handle off-grid frequencies. Since ℓ_p-IAA requires matrix inverse, its complexity is very high. We emphasize that the proposed method is hyper-parameter free and it can easily handle off-grid frequencies. All of these are the praiseworthy advantages of BSUM-RELAX over the existing robust compressive sensing approaches such as ℓ_1-ADMM.

Now we test the robustness in the presence of α-stable noise. The samples of α-stable noise are generated by modulating a complex correlated Gaussian sequence with a non-negative sequence with skewed stable distribution [34]. The parameter setting for the α-stable distribution is $\alpha = 1.3$, $\beta = 0$, $c = 0.8$ and $\mu = 0$. The parameters for harmonics are kept the same as Fig. 6. It is seen from Fig. 7 that our method has the best

[3]Since ℓ_p-IAA performs unstably for small p, we choose $p = 1.3$ for comparison.



(a) RMSE versus SNR.

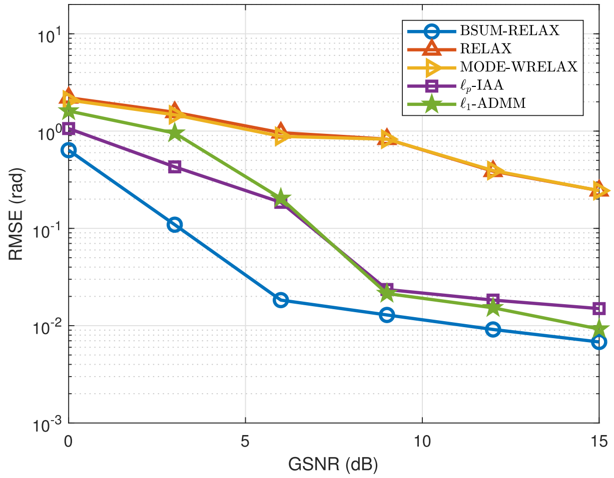


(b) PR versus SNR.

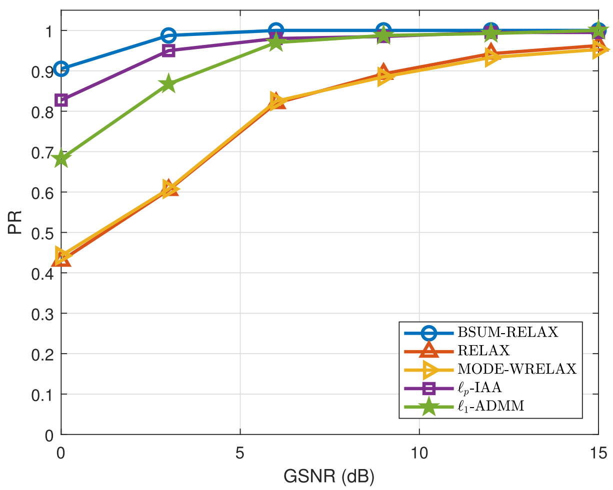
Fig. 6. RMSE and PR performance versus SNR for 1-D HR with GMM noise.

performance. When GSNR < 3 dB, the RMSE of ℓ_1-ADMM is relatively higher than BSUM-RELAX, which indicates that BSUM-RELAX is able to skip some local minima and finally arrives the neighborhood of the ground truth point. Fig. 7(b) shows such an evidence, where the minimum GSNR for BSUM-RELAX reaching 100% PR is about 6 dB, while ℓ_p-IAA and ℓ_1-ADMM require more than 10 dB. Since RELAX and MODE-WRELAX are not robust to outliers, they exhibit the worst performance throughout of the GSNR region.

2) RMSE and Complexity versus N: In this example, we compare the complexities of BSUM-RELAX, RELAX MODE-WRELAX, ℓ_p-IAA and ℓ_1-ADMM. In addition, similar to Fig. 1, we also include "RELAX with Newton's method" for comparison. Assume that there are two closely-spaced harmonics with parameters $\omega_1 = 0.2$, $\omega_2 = 0.23$, $\gamma_1 = 3e^{j2}$ and $\gamma_2 = 2e^{3j}$. GMM noise is considered with $\varrho_2 = 0.1$. SNR is -5 dB. The sample size N is $\{32, 64, 96, 128, 192, 256\}$. For ℓ_p-IAA and ℓ_1-ADMM, we choose a small number of grids when N is small, and vise versa. Specifically, the number of grids corresponding to N is chosen as $\{2^7, 2^9, 2^{11}, 2^{13}, 2^{14}, 2^{14}\}$, such that the resolution ability matches the CRB.

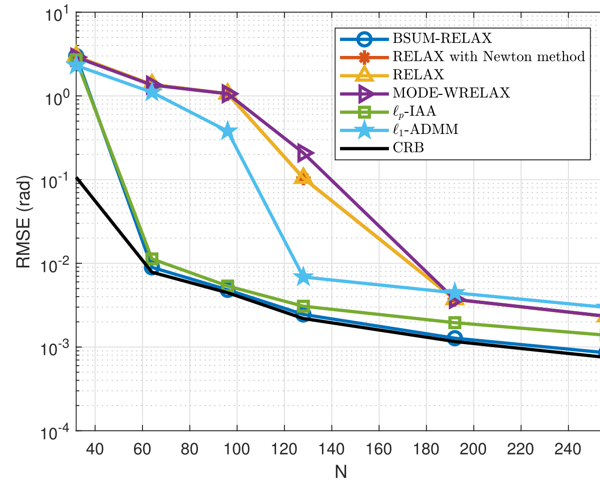


(a) RMSE versus SNR.

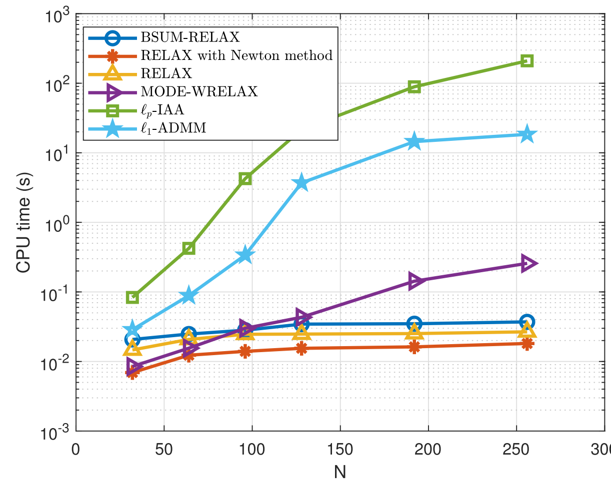


(b) PR versus SNR.

Fig. 7. RMSE and PR performance versus SNR for 1-D HR with α-stable noise.



(a) RMSE versus N.



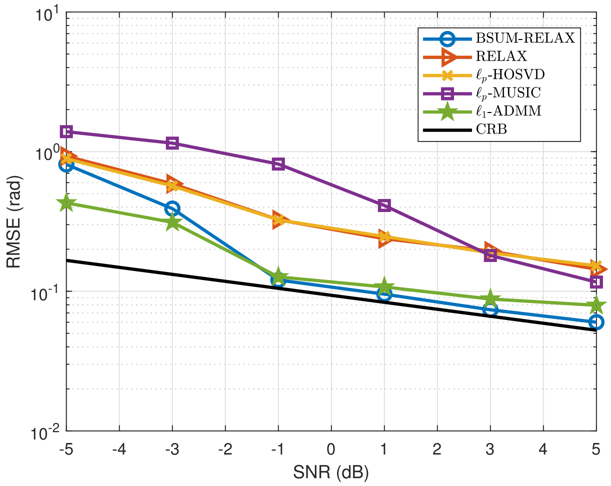
(b) CPU time versus N.

Fig. 8. RMSE and CPU time versus N for 1-D HR.

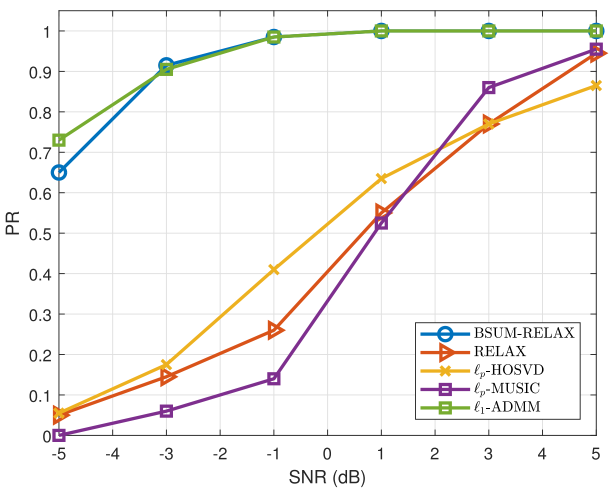
The results are shown in Fig. 8, where the RMSEs of BSUM-RELAX are the smallest when $N \geq 64$, which almost attain the CRB. ℓ_p-IAA has the second best performance and is followed by ℓ_1-ADMM. Regarding of complexity, the proposed method is substantially faster than ℓ_p-IAA and ℓ_1-ADMM, especially for large N. The main reason for the high complexity of the latter two methods is that their dictionaries are too large, for which a simple matrix multiplication may take huge complexities. And our experience is that in many cases, due to the problematic solver, the ℓ_p-IAA does not converge in many cases, while ℓ_1-ADMM converges slowly since it is a gradient descent based method. To make ℓ_1-ADMM work, we need to tune its hyper-parameter carefully, which costs a lot of time. MODE-WRELAX is faster than RELAX when $N < 100$ but it becomes slower as N increases. This is because when N is small, MODE is able to provide good initial point with relatively low complexity, and such an initial point reduces the number of iterations for convergence. However, when N is towards 256, the singular value decomposition in MODE takes $\mathcal{O}(N^3)$ flops. This dominates the complexity of MODE-WRELAX and makes the method slower than RELAX. It is worth highlighting that the proposed method only has two user determined parameters, i.e., p and ϵ. To be honest, these two parameters are not hyper-parameters. We do not need to tune them. Thus, our method is indeed a non-parametric algorithm, which is a great advantage over the existing robust HR methods. Although RELAX performs slightly faster than our method, its estimation accuracy is much worse. We find that RELAX can be accelerated by using the presented Newton's method. Therefore, instead of using RELAX, we recommend the readers to try the proposed one, which offers better results.

D. Performance Comparison for 3-D HR

We note that the generalization of ℓ_p-IAA and MODE-WRELAX to the multidimensional HR model is not straightforward, and there is no related work available in the literature. So we do not include them for comparison in this subsection. Instead, we will compare several existing robust algorithms which are designed specifically or can be easily generalized for multidimensional HR. These estimators include ℓ_p-HOSVD [16], ℓ_p-MUSIC [14] and ℓ_1-ADMM [33]. Specifically, for the



(a) RMSE versus SNR.



(b) PR versus SNR.

Fig. 9. RMSE and PR versus SNR for 3-D HR in GMM noise.

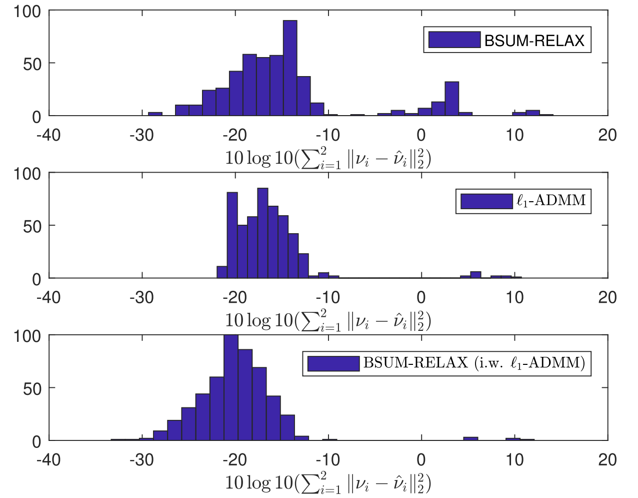


Fig. 10. Histograms of BSUM-RELAX and ℓ_1-ADMM in α-stable noise with GSNR $= -1$ dB.

ℓ_1-ADMM, its dictionary has a form of Khatri-Rao product of three Vandermonde "fat" matrices, where the ith Vandermonde dictionary corresponds to the ith dimension and consists of 2^8 grids that uniformly divide $[0, 2\pi]$. Therefore, the size for the dictionary of ℓ_1-ADMM in this case is 125×2^{24}. Similar to the previous simulations, we consider both GMM and α-stable noise in the following examples.

We first employ the GMM to model outliers, where 10% of the data are corrupted by outliers with mean zeros and variance 100 while the remaining are with complex normal distribution. The signal has dimension $5\times 5\times 5$ consisting of two equal power 3-D sinusoids, where their parameters are $\boldsymbol{\nu}_1 = [0.2\pi, 0.4\pi, 1\pi]^T$, $\boldsymbol{\nu}_2 = [0.8\pi, 1.6\pi, 1.4\pi]^T$, $\alpha_1 = e^{j0.8\pi}$ and $\alpha_2 = e^{j0.9\pi}$. Additionally, $p = 1$ is used for BSUM-RELAX, while $p = 1.3$ is selected for ℓ_p-HOSVD and ℓ_p-MUSIC for the purpose of convergence. The simulation results are plotted in Fig. 9. We see that the ℓ_1-ADMM works slightly better than our method in low SNRs. However, this method cannot handle off-grid frequencies and its performance is inferior to BSUM-RELAX when SNR > -1 dB. The other three algorithms do not show very good performance since their RMSEs are a bit far from the CRB. In Fig. 9(b), we observe similar results, where the PRs of BSUM-RELAX and ℓ_1-ADMM are much higher than the other estimators.

Unlike the GMM case, we consider a more difficult scenario for all the competitors. Two sinusoids with closely-spaced frequencies, i.e., $\boldsymbol{\nu}_1 = [0.2\pi, 0.8\pi, 1\pi]^T$ and $\boldsymbol{\nu}_2 = [0.26\pi, 1.3\pi, 1.54\pi]^T$ are considered. Their amplitudes keep the same as the GMM case. The α-stable noise is generated in a same way as Fig. 7(a). We note that due to the corruption of α-stable noise and the small dimension of the signal which is only $5\times 5\times 5$, it is not easy for all the algorithms to resolve two closely-spaced frequencies. We find that during the Monte-Carlo tests, the ℓ_1-ADMM and BSUM-RELAX sometimes failed to work. Such an observation is shown in Fig. 10, i.e., the histograms of BSUM-RELAX and ℓ_1-ADMM based on 500 Monte-Carlo tests at GSNR $= -1$ dB. The x-axis denotes the variance in dB which is calculated through $10\log 10(\sum_{i=1}^{2}\|\boldsymbol{\nu}_i - \hat{\boldsymbol{\nu}}_i\|_2^2)$. Note that except for the BSUM-RELAX stated in Algorithm 4, we also employ the ℓ_1-ADMM to initialize our method. We observed that most of the frequency estimates of BSUM-RELAX has small variances concentrated around -16 dB, while some are extremely large. Compared to BSUM-RELAX, the ℓ_1-ADMM performs much better since it has a smaller number of outages.[4] Moreover, BSUM-RELAX initialized with (i.w.) ℓ_1-ADMM works well. It has the smallest number of outages and it has a median variance around -20 dB. From this example, we know that initialization is very important for BSUM-RELAX.

Since outages take large values, after averaging, the resulting RMSEs will be very large, which cannot be used to distinguish the performance of algorithms. To overcome this issue, we use the median to calculate RMSEs, which is defined as

$$\text{median} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd} \\ (x_{n/2} + x_{n/2+1})/2, & \text{if } n \text{ is even} \end{cases} \tag{38}$$

[4]Here, outage stands for the frequency estimates with their $10\log 10$ $(\sum_{i=1}^{2}\|\boldsymbol{\nu}_i - \hat{\boldsymbol{\nu}}_i\|_2^2)$ greater than -10.

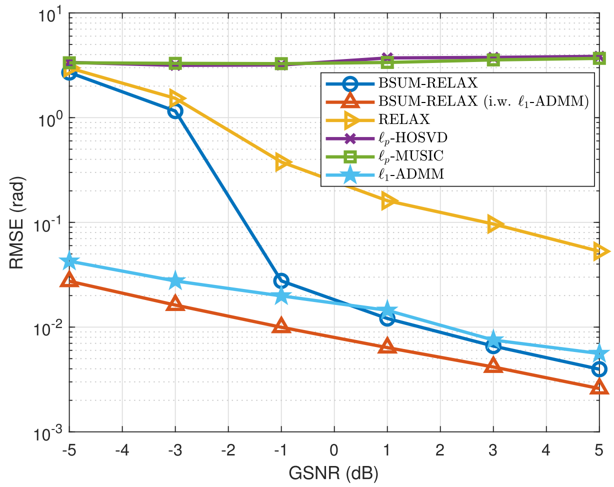


Fig. 11. RMSE versus GSNR for 3-D HR in α-stable noise.

where $x_1 \leq x_2 \leq \cdots \leq x_n$. It is seen in Fig. 11 that BSUM-RELAX has a larger variance than the ℓ_1-ADMM approach when GSNR is smaller than 0 dB. As GSNR increases, BSUM-RELAX outperforms ℓ_1-ADMM since it can handle off-grid frequencies. It is unsurprised that with a good initialization from ℓ_1-ADMM, BSUM-RELAX provides the best performance, but it is at the expense of increased complexity. The ℓ_p-HOSVD and ℓ_p-MUSIC fail to work properly since their resolution ability is limited under such a small dimensional signal model.

V. Conclusion

In this paper, the problem of 1-D and M-D HR in the presence of grossly corrupted data, i.e., outliers, has been addressed. We started from the 1-D case and formulated this problem as an ℓ_p-norm minimization problem, where $0 < p < 2$. An algorithmic framework that is based on BSUM was proposed to minimize ℓ_p-fitting criterion, resulting in a *BSUM-RELAX* algorithm. We then generalized the BSUM-RELAX to the robust M-D HR problem. Numerical results showed that the proposed BSUM-RELAX method has asymptotically statistical efficiency and outperforms the state-of-the-art algorithms.

Appendix A
Proof of Proposition 2.1

Define the objective function in (4) as

$$f(\omega,\gamma) = \sum_{i=1}^{N} \left(|x_i - \gamma a_i(\omega)|^2 + \epsilon\right)^{p/2} \tag{39}$$

and let the upper bound of $f(\omega,\gamma)$ at the point $(\omega^{(\ell)},\gamma^{(\ell)})$ be

$$u(\omega,\gamma) = \left\| \sqrt{\mathbf{w}^{(\ell)}} \odot (\mathbf{x} - \gamma \mathbf{a}(\omega)) \right\|_2^2 + \text{const}. \tag{40}$$

Thus, we have

$$u(\omega,\gamma) \geq f(\omega,\gamma) \tag{41}$$

where the equality holds true at $(\omega^{(\ell)},\gamma^{(\ell)})$, i.e.,

$$u(\omega^{(\ell)},\gamma^{(\ell)}) = f(\omega^{(\ell)},\gamma^{(\ell)}). \tag{42}$$

Let $\mathbf{r} = \mathbf{x} - \gamma\mathbf{a}(\omega)$. Then the derivative of f w.r.t. $\mathbf{r}^*$ is

$$\frac{\partial f}{\partial \mathbf{r}^*} = \mathbf{w} \odot \mathbf{r} \tag{43}$$

where $\mathbf{w} = \frac{p}{2}[|r_1|^{p-2} \ \cdots \ |r_N|^{p-2}]^T$. It is easy to compute

$$\frac{\partial \mathbf{r}}{\partial \omega} = -\gamma \mathbf{t} \odot \mathbf{a}(\omega). \tag{44}$$

Thus, the derivative of f w.r.t. ω is

$$\begin{aligned} \nabla_\omega f(\omega,\gamma) &= \left(\frac{\partial \mathbf{r}}{\partial \omega}\right)^T \frac{\partial f}{\partial \mathbf{r}^*} \\ &= -\gamma \left(\mathbf{t} \odot \mathbf{a}(\omega)\right)^T \left(\mathbf{w} \odot (\mathbf{x} - \gamma \mathbf{a}(\omega))\right). \end{aligned} \tag{45}$$

On the other hand, the derivative of u w.r.t. ω is

$$\nabla_\omega u(\omega,\gamma) = -\gamma \left(\mathbf{t} \odot \mathbf{a}(\omega)\right)^T \left(\mathbf{w}^{(\ell)} \odot (\mathbf{x} - \gamma \mathbf{a}(\omega))\right). \tag{46}$$

It is obvious that at the point $(\omega^{(\ell)},\gamma^{(\ell)})$, we have

$$\nabla_\omega f(\omega^{(\ell)},\gamma^{(\ell)}) = \nabla_\omega u(\omega^{(\ell)},\gamma^{(\ell)}). \tag{47}$$

Minimizing the upper-bound $u(\omega,\gamma)$ is related to Problem (8). As we have claimed in the Section II-B that in each iteration, the optimal ω in (8) and (11) are equivalent, i.e.,

$$\min_{\omega} u(\omega,\gamma) \Leftrightarrow \max_{\omega} \left|\mathbf{a}^H(\omega)(\mathbf{w}^{(\ell)} \odot \mathbf{x})\right|^2 \tag{48}$$

where the problem on the right-hand-side can be solved using Newton's method with backtracking linear search when given $\mathbf{w}^{(\ell)} \odot \mathbf{x}$.[5] Once ω is found, γ is updated subsequently using (10). The resulting estimate, i.e., $(\omega^{(\ell+1)},\gamma^{(\ell+1)})$, becomes the minimizer of the problem on the left-hand-side of (48). Therefore, we have

$$u(\omega^{(\ell)},\gamma^{(\ell)}) \geq u(\omega^{(\ell+1)},\gamma^{(\ell)}) \geq u(\omega^{(\ell+1)},\gamma^{(\ell+1)}) \tag{49}$$

and then

$$\begin{aligned} f(\omega^{(\ell)},\gamma^{(\ell)}) &= u(\omega^{(\ell)},\gamma^{(\ell)}) && \text{(50a)} \\ &\geq u(\omega^{(\ell+1)},\gamma^{(\ell+1)}) && \text{(50b)} \\ &= f(\omega^{(\ell+1)},\gamma^{(\ell+1)}) && \text{(50c)} \end{aligned}$$

where (50a) and (50c) follow (42).

Assume that there exists a convergent subsequence of $\{\omega^{(\ell)}, \gamma^{(\ell)}\}$, which has a limited point $(\omega^{(*)},\gamma^{(*)})$. The subsequence is indexed as $\{r_j\}_{j=1,\ldots,\infty}$. Then

$$\begin{aligned} u(\omega,\gamma^{(r_j)}) &\geq u(\omega^{(r_j+1)},\gamma^{(r_j)}) && \text{(51a)} \\ &\geq f(\omega^{(r_j+1)},\gamma^{(r_j+1)}) && \text{(51b)} \\ &= u(\omega^{(r_j+1)},\gamma^{(r_j+1)}) && \text{(51c)} \\ &\geq u(\omega^{(r_{j+1})},\gamma^{(r_j+1)}) && \text{(51d)} \\ &\geq u(\omega^{(r_{j+1})},\gamma^{(r_{j+1})}) && \text{(51e)} \\ &= f(\omega^{(r_{j+1})},\gamma^{(r_{j+1})}). && \text{(51f)} \end{aligned}$$

[5]The maximization can also be done via one dimensional search with small enough searching grid, or equivalently using oversampled FFT and refined with Newton's method.

Let $j \to \infty$, we obtain

$$u(\omega, \gamma^*) \geq u(\omega^{(*)}, \gamma^{(*)}) \tag{52}$$

which indicates that $\omega^{(*)}$ is a block-wise minimizer of $u(\omega, \gamma^{(*)})$. Hence, it meets the partial KKT condition w.r.t. ω, i.e., $\nabla_\omega u(\omega^{(*)}, \gamma^{(*)}) = 0$. With (47), we have

$$\nabla_\omega f(\omega^{(*)}, \gamma^{(*)}) = 0. \tag{53}$$

In a similar manner, we can also show that

$$u(\omega^{(r_j)}, \gamma) \geq u(\omega^{(r_j)}, \gamma^{(r_j)}) \tag{54}$$

which by taking $j \to \infty$ produces

$$u(\omega^{(*)}, \gamma) \geq u(\omega^{(*)}, \gamma^{(*)}). \tag{55}$$

Together with (53), we obtain that $\gamma^{(*)}$ satisfies the partial conditional KKT conditions w.r.t. γ. Therefore, we conclude that $(\omega^{(*)}, \gamma^{(*)})$ is a KKT point of the original problem.

This completes the proof.

Appendix B
A Brief Derivation of CRB in GMM Noise

To make the paper self-contained, we provide a brief derivation for the CRB of M-D HR in GMM noise. First, we vectorize $\boldsymbol{\mathcal{S}}$ as

$$\mathbf{s} = \mathrm{vec}(\boldsymbol{\mathcal{S}}) = \breve{\mathbf{A}}\boldsymbol{\alpha} \tag{56}$$

where $\breve{\mathbf{A}} = \mathbf{A}_M \circledast \cdots \circledast \mathbf{A}_1 = [\breve{\mathbf{a}}(\boldsymbol{\nu}_1) \ \cdots \ \breve{\mathbf{a}}(\boldsymbol{\nu}_K)]$ and $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_K]^T$. Then it follows from [35] that the CRB matrix for frequencies $\boldsymbol{\nu} = [\nu_{1,1} \ \cdots \ \nu_{M,1} \ \cdots \ \nu_{1,K} \ \cdots \ \nu_{M,K}]^T$ is

$$\mathrm{CRB}(\boldsymbol{\nu}) = \frac{1}{I_c}\left(\mathrm{Re}\left(\left(\mathbf{D}^H \mathbf{P}_{\breve{\mathbf{A}}}^{\perp} \mathbf{D}\right) \odot \left(\mathbf{I}_M \otimes (\boldsymbol{\alpha}\boldsymbol{\alpha}^H)\right)\right)\right)^{-1} \tag{57}$$

where $\mathrm{CRB}(\boldsymbol{\nu}) \in \mathbb{R}^{MK \times MK}$ and

$$\mathbf{P}_{\breve{\mathbf{A}}}^{\perp} = \mathbf{I}_{\prod_{m=1}^{M} N_m} - \breve{\mathbf{A}}\breve{\mathbf{A}}^{\dagger} \tag{58}$$

$$\mathbf{D} = \left[\partial \breve{\mathbf{a}}_1 / \partial \breve{\boldsymbol{\nu}}_1^T \ \cdots \ \partial \breve{\mathbf{a}}_K / \partial \breve{\boldsymbol{\nu}}_K^T\right] \in \mathbb{C}^{(\prod_{m=1}^{M} N_m) \times MK}. \tag{59}$$

Additionally, I_c can be determined as follows. Suppose that the PDF of Gaussian mixture noise is [32]

$$p(\mathcal{X}) = \sum_{l=1}^{L} \frac{\varrho_l}{\pi \sigma_l^2} \exp\left(-\frac{\mathcal{X}^2}{\sigma_l^2}\right) \tag{60}$$

where $\sum_{l=1}^{L} \varrho_l = 1$. Then I_c is calculated as [35]

$$\begin{aligned} I_c &= \pi \int_0^{\infty} \frac{(\partial p(\mathcal{X}) / \partial \mathcal{X})^2}{p(\mathcal{X})} \mathcal{X} d\mathcal{X} \\ &= \int_0^{\infty} \frac{\sum_{l=1}^{L} \sum_{q=1}^{L} \frac{4 \varrho_l \varrho_q}{\sigma_l^4 \sigma_q^4} \exp\left(-\mathcal{X}^2 \left(\frac{1}{\sigma_l^2} + \frac{1}{\sigma_q^2}\right)\right)}{\sum_{l=1}^{L} \frac{\varrho_l}{\sigma_l^2} \exp\left(-\frac{\mathcal{X}^2}{\sigma_l^2}\right)} \mathcal{X}^3 d\mathcal{X}. \end{aligned} \tag{61}$$

Acknowledgment

The authors would like to thank reviewers for their valuable comments, Prof. J. Li for her discussions on the realization of RELAX algorithm, Dr. Y. Chen for the MATLAB code of ℓ_p-IAA.

References

[1] C. Qian, Y. Shi, L. Huang, and H.-C. So, "Robust spectrum estimation via majorization minimization," in *Proc. IEEE 10th Sensor Array Multichannel Signal Process. Workshop*, Sheffield, U.K., 2018, pp. 529–532.

[2] J. Saeedi and K. Faez, "Synthetic aperture radar imaging using nonlinear frequency modulation signal," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 99–110, Feb. 2016.

[3] J. A. Mahal, A. Khawar, A. Abdelhadi, and T. C. Clancy, "Spectral coexistence of MIMO radar and MIMO cellular system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 655–668, Apr. 2017.

[4] S. Miron, Y. Song, D. Brie, and K. T. Wong, "Multilinear direction finding for sensor-array with multiple scales of invariance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 2057–2070, Jul. 2015.

[5] A. Kammoun, H. Khanfir, Z. Altman, M. Debbah, and M. Kamoun, "Preliminary results on 3D channel modeling: From theory to standardization," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1219–1229, Jun. 2014.

[6] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 5, pp. 591–598, Sep. 1974.

[7] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.

[8] R. Wu, J. Li, and Z.-S. Liu, "Super resolution time delay estimation via MODE-WRELAX," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 294–307, Jan. 1999.

[9] Z.-S. Liu, J. Li, and P. Stoica, "Relax-based estimation of damped sinusoidal signal parameters," *Signal Process.*, vol. 62, no. 3, pp. 311–321, 1997.

[10] Y. Hua, "The most efficient implementation of the IQML algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2203–2204, Aug. 1994.

[11] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, Jan. 2010.

[12] W. Sun and H.-C. So, "Accurate and computationally efficient tensor-based subspace approach for multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5077–5088, Oct. 2012.

[13] R. Chartrand and W. Yin, "Iterative reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 3869–3872.

[14] W.-J. Zeng, H.-C. So, and L. Huang, "ℓ_p-music: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.

[15] Y. Chen, H.-C. So, and W. Sun, "ℓ_p-norm based iterative adaptive approach for robust spectral analysis," *Signal Process.*, vol. 94, pp. 144–148, 2014.

[16] F. Wen and H. C. So, "Robust multi-dimensional harmonic retrieval using iteratively reweighted HOSVD," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2464–2468, Dec. 2015.

[17] X. Fu, K. Huang, W.-K. Ma, N. D. Sidiropoulos, and R. Bro, "Joint tensor factorization and outlying slab suppression with applications," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6315–6328, Dec. 2015.

[18] C. Qian, X. Fu, N. D. Sidiropoulos, L. Huang, and J. Xie, "Inexact alternating optimization for phase retrieval in the presence of outliers," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 6069–6082, Nov. 2017.

[19] T.-H. Li, "Laplace periodogram for time series analysis," *J. Amer. Statistical Assoc.*, vol. 103, no. 482, pp. 757–768, 2008.

[20] T. Li, "A nonlinear method for robust spectral analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2466–2474, May 2010.

[21] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sep. 1993.

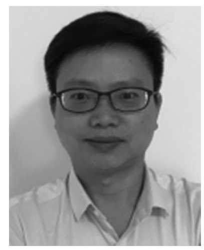
[22] W. H. Young *et al.*, "On classes of summable functions and their Fourier series," *Proc. Roy Soc. London A*, vol. 87, no. 594, pp. 225–229, 1912.

[23] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.
[24] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2411–2422, Oct. 2007.
[25] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for $\ell_2 - \ell_0$ image regularization," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 563–591, 2013.
[26] N. D. Sidiropoulos and E. Tsakonas, "Signal processing and optimization tools for conference review and session assignment," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 141–155, May 2015.
[27] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
[28] M. Hong, M. Razaviyayn, Z. Luo, and J. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, Jan. 2016.
[29] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.
[30] X. Liu and N. D. Sidiropoulos, "Almost sure identifiability of constant modulus multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2366–2368, Sep. 2002.
[31] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4543–4556, Dec. 2006.
[32] A. van den Bos, "The multivariate complex normal distribution-a generalization," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 537–539, Mar. 1995.
[33] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.
[34] W. J. Szajnowski and J. B. Wynne, "Simulation of dependent samples of symmetric alpha-stable clutter," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 151–152, May 2001.
[35] R. J. Kozick 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.

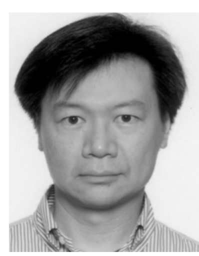
Cheng Qian (S'11–M'17) was born in Deqing, Zhejiang, China. He received the B.E. degree in communication engineering from Hangzhou Dianzi University, Hangzhou, China, in 2011, and the M.S. and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2013 and 2017, respectively. He was a visiting Ph.D. student with the Department of Electrical and Computer Engineering (ECE), University of Minnesota Twin Cities (UMN), Minneapolis, MN, USA, from 2014 to 2016. After his graduation, he returned back to the group in UMN as a Postdoc for half a year. Then, he moved to Charlottesville and joined the University of Virginia, Charlottesville, VA, USA, as a Postdoctoral Associate with the Department of ECE. His research interests include signal processing, tensor decomposition, and bioinformatics.

Yunmei Shi was born in Shandong, China. She received the B.Sc. degree in electronic and information engineering from the Harbin Institute of Technology, Weihai, China, in 2012, the M.Sc. degree in electronic and communication engineering from the Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China, and is currently working toward the Ph.D. degree in information and communication engineering with the Harbin Institute of Technology, Harbin, China.

From 2016 to 2017, she was a Research Scholar with the Department of Electrical and Computer Engineering, Digital Technology Center, University of Minnesota. Her research interests include array signal processing and optimization.



Lei Huang (M'07–SM'14) was born in Guangdong, China. He received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Xidian University, Xian, China, in 2000, 2003, and 2005, respectively. From 2005 to 2006, he was a Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. From 2009 to 2010, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong, and a Research Associate with the Department of Electronic Engineering, The Chinese University of Hong Kong. From 2011 to 2014, he was a Professor with the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School. Since 2014, he has been with the College of Information Engineering, Shenzhen University, Shenzhen, China, where he is currently a Distinguished Professor. His research interests include spectral estimation, array signal processing, statistical signal processing, and their applications in radar, and navigation and wireless communications. He has been on the Editorial Boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2015-present), Elsevier-*Digital Signal Processing* (2012-present), and *IET Signal Processing* (2017-present). He has been an elected member of the Sensor Array and Multichannel Technical Committee of the IEEE Signal Processing Society (2016-present). He was elected an IET Fellow in 2018.



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, Kowloon Tong, 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. During 1995–1996, 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.

He has been on the Editorial Boards of the IEEE SIGNAL PROCESSING MAGAZINE (2014–2017), IEEE TRANSACTIONS ON SIGNAL PROCESSING (2010–2014), *Signal Processing* (2010–), and *Digital Signal Processing* (2011–). He was also the 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 of the 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).