A Robust Iteratively Reweighted ℓ_2 Approach for Spectral Compressed Sensing in Impulsive Noise

Zhen-Qing He, Hongbin Li, Zhi-Ping Shi, Jun Fang, and Lei Huang

***Abstract*—This letter concentrates on the problem of spectral compressed sensing in impulsive noise, which aims to recover a spectrally sparse signal from its contaminated and undersampled measurements. We propose a robust formulation for joint sparse signal and frequency recovery, which includes the generalized ℓ_p-norm $(0 < p < 2)$ data-fidelity fitting term added to a log-sum sparsity-promoting regularizer. To handle this intractable issue, we develop an iteratively reweighted ℓ_2 approach via majorizing the original objective function by a quadratic surrogate function. Simulation results illustrate that the proposed approach attains a significant performance improvement over the existing methods under impulsive noise.**

***Index Terms*—Compressed sensing (CS), grid mismatch, impulsive noise, line spectra estimation.**

I. Introduction

INFERRING the frequency and phase/amplitude information of complex sinusoids from noise-contaminated observations is a problem of paramount importance and finds a wide variety of applications in, e.g., source localization [1], inverse scattering imaging [2], and channel estimation [3], to name a few. Over recent years, the compressed sensing (CS) paradigm has demonstrated that under certain conditions, it is able to accurately recover a spectrally sparse signal from far fewer measurements than traditional methods use. This has stirred up a renewed research thrust in spectrum analysis, named the spectral CS [4]. To address the basis mismatch issue which is usually incurred by discretizing the frequency space, several approaches are introduced, mainly including atomic-norm-based methods that involve continuously infinite dictionaries [5]–[7] and joint sparse signal recovery methods with frequency parameter learning [8]–[12].

Despite attractive features, e.g., fewer sample request and superresolution, of the aforementioned methods, most of them explicitly or implicitly rely on the Gaussian noise assumption and may fall short or even break down when measurement noises are non-Gaussian. In many practical scenarios, measurements may contain, besides thermal (Gaussian) noise, a small portion of large outliers owning to some abnormal behaviors in data acquisition, such as sensor failures, amplifier saturation, and malicious attacks. This type of non-Gaussian disturbance is usually referred to as the impulsive noise, which has been frequently encountered in signal processing and wireless communication systems (see, e.g., [13]–[15]).

In this letter, we aim to recover a spectrally sparse signal from partially observed samples under impulsive noise from a robust statistical fitting perspective. Unlike most of the prior works utilizing the ℓ_2-norm error-fitting criterion that is statistical optimal for Gaussian noise, we propose a robust formulation for joint sparse signal and frequency recovery, which includes the ℓ_p $(0 < p < 2)$ data-fidelity term added to a log-sum sparsity-promoting regularizer. In order to solve the resulting intractable problem, we develop an iteratively reweighted approach via iteratively minimizing an ℓ_2 surrogate function that majorizes the original objective function. This leads to an interweaved iterative fashion for robust joint sparse signal recovery and dictionary parameter refinement.

Related work: The proposed approach is an extension of the recent iteratively reweighted ℓ_2 method in [12] to the impulsive noise environment. The difference is that we introduce an additional ℓ_p-based metric for a residual error to seek resistance against the impulsive noise. This extension, however, leads to a technically more intractable nonconvex problem, which involves, besides the unknown sparse signal, a highly nonlinear frequency parameter. Note that our work is also related to the conventional ℓ_2 iteratively reweighted methods [16]–[19] that concern only the sparse recovery, but besides this aim, we also consider the ℓ_2 iteratively reweighted scheme to the residual error. To the best of our knowledge, the literature for superresolution spectral CS under impulsive noise is less abundant. Although recent works [20], [21] tackle this problem by leveraging the atomic-norm denoising to demix the spectral lines and sparse outliers via semidefinite programming, two regularization parameters are needed to be tuned for robust recovery of the missing samples, and the model order is also required for final frequency estimation.

II. Problem Formulation

Consider the following spectral line model:

$$\tilde{y}_t = \sum_{c=1}^{C} \tilde{\alpha}_c e^{-j2\pi \tilde{f}_c t}, \quad t = 1, \ldots, \widetilde{T} \tag{1}$$

Manuscript received February 20, 2017; revised April 22, 2017; accepted April 23, 2017. Date of publication April 27, 2017; date of current version May 9, 2017. This work was supported by the National Natural Science Foundation of China under Grant 61671128, the National Science Foundation under Grant ECCS-1408182 and Grant ECCS-1609393, and the Chinese Scholarship Council. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael A. Lexa. (*Corresponding author: Zhen-Qing He.*)

Z.-Q. He, Z.-P. Shi, and J. Fang are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhenqinghe@outlook.com; szp@uestc.edu.cn; JunFang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu).

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

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

Digital Object Identifier 10.1109/LSP.2017.2698721

where $j = \sqrt{-1}$, $\{\tilde{\alpha}_c \in \mathbb{C}\}$ and $\{\tilde{f}_c \in [0,1]\}$ are the complex amplitudes and normalized frequencies of the C sinusoidal components, respectively, and $\mathbb{C}$ denotes the complex space. In this letter, we assume the same measurement model as that considered in [5], [7], and [8], i.e., randomly extracting M ($M \leq \widetilde{T}$) samples from $\{\tilde{y}_t\}$. Using this sampling strategy, however, would make no difference for our algorithm development to other data acquisition schemes. Let $\{t_1, \ldots, t_M\} \subset \{1, \ldots, \widetilde{T}\}$ be the sampling index and $\boldsymbol{y} \triangleq [\tilde{y}_{t_1}, \ldots, \tilde{y}_{t_M}]^T$ the undersampled measurement, where $(\cdot)^T$ denotes the transpose. As a result, the goal of the conventional spectral CS is to recover $\{\tilde{\alpha}_c, \tilde{f}_c\}$ or $\{\tilde{y}_t\}$ from $\boldsymbol{y}$ by solving the linear inverse sparse system

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2}$$

Here, $\boldsymbol{\beta} \triangleq [\beta_1, \ldots, \beta_N]^T$ is an unknown sparse vector, which encompasses the amplitude information, $\boldsymbol{\epsilon} \in \mathbb{C}^M$ is the measurement noise, and $\boldsymbol{A}(\boldsymbol{\theta}) \triangleq [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_N)] \in \mathbb{C}^{M\times N}$ ($M \ll N$) is the overcomplete dictionary with

$$\boldsymbol{a}(\theta_n) \triangleq \left[e^{-j2\pi\theta_n t_1}, \ldots, e^{-j2\pi\theta_n t_M}\right]^T \in \mathbb{C}^M \tag{3}$$

where $\boldsymbol{\theta} \triangleq \{\theta_i, i = 1, \ldots, N\}$ is a finite set of discretized grid points covering the frequency parameter space. In practice, the true frequency parameters $\{\tilde{f}_c\}$ do not necessarily lie on the discretized grid $\boldsymbol{\theta}$, thereby leading to the so-called basis mismatch problem. To deal with this, a superresolution approach in [12] considers the following sparsity-regularization formulation with an unknown parametric dictionary:

$$\min_{\boldsymbol{\beta},\, \mathbf{0}\preceq\boldsymbol{\theta}\preceq\mathbf{1}} \sum_{i=1}^{N} \rho(\beta_i) + \lambda^{-1}\|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}\|_2^2 \tag{4}$$

where $\mathbf{0}$ and $\mathbf{1}$ are M-dimensional vectors consisting of 0 and 1, respectively, $\preceq$ denotes $\leq$ with an elementwise operation, $\rho(\beta_i) \triangleq \log(|\beta_i|^2 + \varepsilon)$ with $\varepsilon > 0$ is a sparsity-encouraging function, and $\lambda > 0$ is a regularization parameter, which balances the residual error and sparsity.

It is known from the theory of robust statistics [22] that the ℓ_2-norm-based residual fitting criterion in (4) performs well only in Gaussian noise and is optimum in terms of maximum likelihood. However, this rapidly increasing squared penalty cannot characterize some large errors, and it is, therefore, prone to impulsive noise. To this end, we propose the following modified formulation for spectrally sparse signal recovery:

$$\min_{\boldsymbol{\beta},\, \mathbf{0}\preceq\boldsymbol{\theta}\preceq\mathbf{1}} G(\boldsymbol{\beta},\boldsymbol{\theta}) \triangleq \sum_{i=1}^{N} \rho(\beta_i) + \lambda^{-1} \sum_{m=1}^{M} \phi_\nu^p([\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}]_m) \tag{5}$$

where $[\cdot]_m$ is its mth element and $\phi_\nu^p(x)$ is defined as [23]

$$\phi_\nu^p(x) = \begin{cases} \frac{p}{2}\nu^{p-2}|x|^2, & |x| \leq \nu \\ |x|^p - \frac{2-p}{2}\nu^p, & |x| > \nu. \end{cases} \tag{6}$$

Here, the generalized ℓ_p penalty $\phi_\nu^p(x)$ is a smoothed approximation of $|x|^p$ with a local quadratic function around the nonsmoothed region $|x| = 0$ and has been employed to promote sparsity when $0 < p \leq 1$ [23]. Obviously, formulation (5) generalizes problem (4) and includes the latter as a special case when $p = 2$. The reason for considering $\phi_\nu^p(x)$ as a fitting metric to suppress the impulsive noise is essentially twofold. First, it is computational tractable due to its continuity and differentiability. Second, it is versatile to accommodate various kinds of impulsive noise because of a lot of flexibility provided by the robustness-controlling constants ν and p.

III. Robust Iteratively Reweighted ℓ_2 Approach

To solve optimization (5), we resort to the majorization–minimization (MM) framework [24] and develop an iteratively reweighted ℓ_2 approach. The main idea is to iteratively minimize a weighted ℓ_2 surrogate function, which majorizes the objective function evaluated at the previous iteration. Before deriving the proposed method, we need to introduce two useful lemmas for constructing the ℓ_2 surrogate function.

Lemma 1 (see [18]): For $x \geq 0$ and $\varepsilon > 0$, the function $\log(x+\varepsilon)$ can be upperbounded as

$$\log(x+\varepsilon) \leq \log(y+\varepsilon) + \frac{x-y}{y+\varepsilon}, \quad \forall\, y \geq 0 \tag{7}$$

with equality achieved at $x = y$.

Lemma 2: For $0 < p < 2$ and $\nu > 0$, the function $\phi_\nu^p(x)$ can be upperbounded as[1]

$$\phi_\nu^p(x) \leq \tilde{\phi}_\nu^p(x;\bar{x}) \triangleq \frac{p}{2}|\omega|^{p-2}|x|^2 + \frac{2-p}{2}(|\omega|^p - \nu^p) \tag{8}$$

where

$$\omega = \begin{cases} \nu, \; |\bar{x}| \leq \nu \\ |\bar{x}|, |\bar{x}| > \nu \end{cases}, \quad \forall\, x, \bar{x} \in \mathbb{R} \tag{9}$$

with equality achieved at $x = \bar{x}$.

We now discuss how to majorize the objective function $G(\boldsymbol{\beta},\boldsymbol{\theta})$ defined in (5). With Lemmas 1 and 2, it can be shown that the upper-bound functions of $\rho(\beta_i)$ and $\phi_\nu^p([\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}]_m)$ can, respectively, be selected as

$$\rho(\beta_i) \leq \tilde{\rho}\big(\beta_i; \beta_i^{(k)}\big) \triangleq \frac{|\beta_i|^2}{|\beta_i^{(k)}|^2 + \epsilon} + \text{const.} \tag{10}$$

and

$$\begin{aligned}\phi_\nu^p([\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}]_m) &\leq \tilde{\phi}_\nu^p\big([\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}]_m; [\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}^{(k)}]_m\big) \\ &\triangleq w_m(\boldsymbol{\theta})\big|[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}]_m\big|^2 + \text{const.}\end{aligned} \tag{11}$$

where $\boldsymbol{\beta}^{(k)} \triangleq [\beta_1^{(k)}, \ldots, \beta_N^{(k)}]^T$ represents an estimate of $\boldsymbol{\beta}$ at iteration k and

$$w_m(\boldsymbol{\theta}) \triangleq \begin{cases} \frac{p}{2}\nu^{p-2}, & \big|[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}^{(k)}]_m\big| \leq \nu \\ \frac{p}{2}\big|[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}^{(k)}]_m\big|^{p-2}, & \big|[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}^{(k)}]_m\big| > \nu. \end{cases}$$

Note that equalities in (10) and (11) hold at $\beta_i = \beta_i^{(k)}$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^{(k)}$, respectively. Now, it can be readily obtained from (10) and (11) that the objective function $G(\boldsymbol{\beta},\boldsymbol{\theta})$ can be majorized by the following weighted surrogate function:

$$\begin{aligned}\tilde{G}(\boldsymbol{\beta},\boldsymbol{\theta};\boldsymbol{\beta}^{(k)}) \triangleq\; &\boldsymbol{\beta}^H \boldsymbol{D}^{(k)}\boldsymbol{\beta} + \lambda^{-1}\big[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}\big]^H \\ &\times \boldsymbol{W}^{(k)}(\boldsymbol{\theta})\big[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta}\big] + \text{const.}\end{aligned} \tag{12}$$

[1] When $0 < p \leq 1$, the result can be proved by the first-order concavity [23]. A unified proof for $0 < p < 2$ is omitted here due to lack of space.

where $(\cdot)^H$ denotes the conjugate transpose, and

$$\boldsymbol{D}^{(k)} \triangleq \operatorname{diag}\left\{\frac{1}{|\beta_1^{(k)}|^2+\varepsilon}, \ldots, \frac{1}{|\beta_N^{(k)}|^2+\varepsilon}\right\} \tag{13}$$

$$\boldsymbol{W}^{(k)}(\boldsymbol{\theta}) \triangleq \operatorname{diag}\left\{w_1(\boldsymbol{\theta}), \ldots, w_M(\boldsymbol{\theta})\right\}. \tag{14}$$

For brevity, exact expressions on the constant terms in (10)–(12) are omitted here as they have no effect on the final solution.

The surrogate function $\tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\beta}^{(k)})$, reminiscent of the computational advantage of quadratic optimization, is much easier to minimize than the intractable objective function $G(\boldsymbol{\beta}, \boldsymbol{\theta})$. Motivated from the fact that

$$\min_{\boldsymbol{\beta},\, \mathbf{0} \preceq \boldsymbol{\theta} \preceq \mathbf{1}} \tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\beta}^{(k)}) = \min_{\mathbf{0} \preceq \boldsymbol{\theta} \preceq \mathbf{1}} \left\{ \min_{\boldsymbol{\beta}} \tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\beta}^{(k)}) \right\}, \tag{15}$$

one can first minimize $\tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\beta}^{(k)})$ with respect to (w.r.t.) $\boldsymbol{\beta}$ and then w.r.t. the remaining $\boldsymbol{\theta}$. The optimal solution of minimizing $\tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\beta}^{(k)})$ w.r.t. $\boldsymbol{\beta}$, as a preparation for the next iteration, is given by

$$\begin{aligned} \boldsymbol{\beta}^{(k+1)} = & \left(\lambda \boldsymbol{D}^{(k)} + \boldsymbol{A}^H(\boldsymbol{\theta}) \boldsymbol{W}^{(k)}(\boldsymbol{\theta}) \boldsymbol{A}(\boldsymbol{\theta})\right)^{-1} \\ & \times \boldsymbol{A}^H(\boldsymbol{\theta}) \boldsymbol{W}^{(k)}(\boldsymbol{\theta}) \boldsymbol{y}. \end{aligned} \tag{16}$$

Now, by replacing $\boldsymbol{\beta}$ with $\boldsymbol{\beta}^{(k+1)}$ in (15) and ignoring terms independent of $\boldsymbol{\theta}$, we arrive at the following problem related to the update of $\boldsymbol{\theta}$:

$$\begin{aligned} \min_{\mathbf{0} \preceq \boldsymbol{\theta} \preceq \mathbf{1}} F^{(k)}(\boldsymbol{\theta}) \triangleq & -\boldsymbol{y}^H \boldsymbol{W}^{(k)}(\boldsymbol{\theta}) \boldsymbol{A}(\boldsymbol{\theta}) \left(\lambda \boldsymbol{D}^{(k)} + \boldsymbol{A}^H(\boldsymbol{\theta}) \right. \\ & \left. \times \boldsymbol{W}^{(k)}(\boldsymbol{\theta}) \boldsymbol{A}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{A}^H(\boldsymbol{\theta}) \boldsymbol{W}^{(k)}(\boldsymbol{\theta}) \boldsymbol{y}. \end{aligned} \tag{17}$$

It is not easy to obtain an analytical solution to (17) as $F^{(k)}(\boldsymbol{\theta})$ is highly nonlinear in $\boldsymbol{\theta}$. Nevertheless, given $\boldsymbol{\theta}^{(k)}$, we can find the next iteration $\boldsymbol{\theta}^{(k+1)}$ by the gradient descent scheme, viz.,

$$\boldsymbol{\theta}^{(k+1)} = \operatorname{mod}\left(\boldsymbol{\theta}^{(k)} - \gamma^k \nabla F^{(k)}(\boldsymbol{\theta}^{(k)}), 1\right) \tag{18}$$

where γ^k is an appropriate step size and mod is a remainder operator to restrict the solution on to $\mathbf{0} \preceq \boldsymbol{\theta} \preceq \mathbf{1}$, such that

$$F^{(k)}(\boldsymbol{\theta}^{k+1}) \leq F^{(k)}(\boldsymbol{\theta}^{(k)}). \tag{19}$$

Here, the gradient $\nabla F^{(k)}(\boldsymbol{\theta}^{(k)})$ can be readily calculated by the chain rule. It follows from (12), (15), and (17) that through iteratively minimizing the surrogate function $\tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\beta}^{(k)})$, the functional sequence $\left\{G(\boldsymbol{\theta}^{(k)}, \boldsymbol{\beta}^{(k)})\right\}$ can be guaranteed to be nonincreasing and eventually convergent to a finite value since $G(\boldsymbol{\beta}, \boldsymbol{\theta})$ is bounded from below.

Discussion: Although the sequence $\{(\boldsymbol{\theta}^{(k)}, \boldsymbol{\beta}^{(k)})\}$ might converge to a stationary point of $G(\boldsymbol{\beta}, \boldsymbol{\theta})$, it is only valid for fixed λ and ε. Inappropriate choices of these parameters would likely cause the sequence to get stuck in an undesirable local minima, quite possibly, without any guarantee of sparsity on the estimation of $\boldsymbol{\beta}$. As such, similar to [25], we therefore recommend a monotonically decreasing sequence $\{\varepsilon^{(k)}\}$, i.e.,

$$\varepsilon^{(k)} = \varepsilon^{(k-1)}/\varrho, \quad \varrho > 1 \tag{20}$$

to ensure a stable sparse recovery. Note that the regularization parameter λ plays an important role in indicating a tradeoff between the sparsity and residual error. Although some of the existing approaches, e.g., the cross validation and the discrepancy principle, can be available to determine λ, they need a large number of experiments or reoptimization processes to select an appropriate λ in a fixed (variance or power) noise environment, and it is, therefore, impractical for ever-changing noise scenario. To deal with this, we suggest the heuristic Bayesian approach proposed in [12] for Gaussian noise, which can automatically update this parameter based on the previous iteration, i.e.,

$$\lambda^{(k+1)} = \|\Delta_k\|_2^2 / (dM) \tag{21}$$

where $\Delta_k \triangleq \left(\boldsymbol{W}^{(k)}(\boldsymbol{\theta}^{(k)})\right)^{\frac{1}{2}} \left[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta}^{(k)})\boldsymbol{\beta}^{(k)}\right]$ and $d > 0$ is a constant scaling factor. The rationale behind this update rule is that the weighted matrix $\left(\boldsymbol{W}^{(k)}(\boldsymbol{\theta}^{(k)})\right)^{\frac{1}{2}}$ can contribute to downweight large elements of $\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta}^{(k)})\boldsymbol{\beta}^{(k)}$, that is, large residuals will be penalized less than small ones. Therefore, the weighted residual Δ_k, reminiscent of the connection between the Gaussian likelihood and quadratic error-fitting criterion, can be approximately modeled as Gaussian noise. Numerical results show that this update rule, although supoptimal, is quite effective and provides superior performance.

For clarification, the proposed approach is summarized in the flow chart below, which is referred to as the robust superresolution iteratively reweighted (RSURE-IR) algorithm.

Algorithm 1: RSURE-IR Algorithm.

Require: $\boldsymbol{y} \in \mathbb{C}^M$, $\nu > 0$, $p \in (0, 2)$, $d > 0$, $\varrho > 0$.
1: Set $k = 0$, given initializations: $\boldsymbol{\theta}^{(0)} \in [0, 1]^N$, $\boldsymbol{\beta}^{(0)} \in \mathbb{C}^N$, $\varepsilon > 0$ and $\lambda > 0$.
2: **Repeat**
3: Compute $\boldsymbol{D}^{(k)}$ via (13) and $\boldsymbol{W}^{(k)}(\boldsymbol{\theta}^{(k)})$ via (14).
4: Compute the regularization parameter λ via (21).
5: Update $\boldsymbol{\theta}^{(k+1)}$ via (18) when satisfying (19).
6: Update $\boldsymbol{\beta}^{(k+1)}$ via (16) based on $\boldsymbol{\theta}^{(k+1)}$.
7: $\varepsilon \leftarrow \varepsilon/\varrho$.
8: $k \leftarrow k + 1$.
9: **until** satisfy certain stopping criterion.
Return: $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\beta}^{(k)}$

Remark 1: The selections of p and ν are key for robustness and convergence guarantee. In general, the closer the upper-bound function turns out to be with its original function, the faster it tends to be in terms of convergence. It is obvious that when p is close to 2 or for large ν, the proposed method has a relatively faster convergence since in this case $\phi_\nu^p(x)$ is much closer to the surrogate function $\tilde{\phi}_\nu^p(x, \bar{x})$. Meanwhile, the cutoff parameter ν in $\phi_\nu^p(x)$ decides how much we retain the quadratic penalty that is less sensitive to small thermal noise, and a small p can achieve a better robustness against gross outliers. Therefore, pursuing fast convergence by increasing p or ν may lose its robustness to outliers. Our empirical experiments suggest that choosing p and ν within $[0.5, 1.1]$ and $[0.1, 1]$, respectively, can lead to a better performance.

IV. Simulation Results

We test the performance of the proposed RSURE-IR method by conducting numerical experiments under impulsive noise

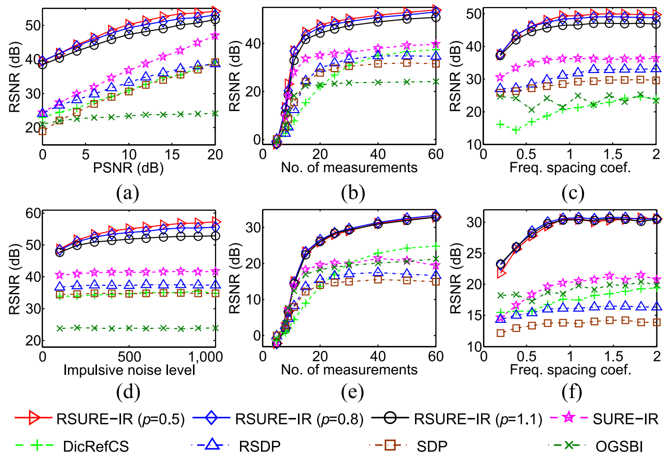

Fig. 1. RSNRs versus various parameters in two impulsive noise environments: (a)–(d) with Gaussian mixture noise; (e)–(f) with α-stable noise.

environments. The measurement $\boldsymbol{y}$ is obtained by randomly choosing M ($M \leq \widetilde{T}$) entries from $\tilde{\boldsymbol{y}} \triangleq [\tilde{y}_1, \ldots, \tilde{y}_{\widetilde{T}}]^T$ via standard uniform sampling. For RSURE-IR, we set $\nu = 0.1$, $d = 1$, $\varepsilon^{(0)} = 1$, $\lambda^{(0)} = 0.001$, and $\varrho = 10$. For comparison, the benchmark methods include the Bayesian learning with dictionary refinement algorithm (DicRefCS) [12], the atomic-norm minimization via the semidefinite programming (SDP) approach [5] as well as its robust SDP (RSDP) variant [20], the off-grid sparse Bayesian inference (OGSBI) algorithm [9], and the superresolution iteratively reweighted approach (SURE-IR) [12]. Except for the last experiment, all simulations are obtained by averaging the results of 10^3 trials with $\widetilde{T} = 64$ and $|\tilde{\alpha}_1| = \cdots = |\tilde{\alpha}_C| = 5$, and the frequencies $\{\tilde{f}_c\}$ are uniformly generated within $[0, 1)$. The reconstruction signal-to-noise ratio (RSNR) is employed to measure the recovery accuracy, which is defined as $\mathrm{RSNR} \triangleq 20\log_{10}\left(\|\tilde{\boldsymbol{y}}\|_2 / \|\tilde{\boldsymbol{y}} - \hat{\tilde{\boldsymbol{y}}}\|_2\right)$ where $\hat{\tilde{\boldsymbol{y}}}$ denotes the reconstructed signal of $\tilde{\boldsymbol{y}}$.

First, we adopt the Gaussian mixture model (GMM) to synthesize the impulsive noise, that is, each element of $\boldsymbol{\epsilon}$ is generated according to the mixed circularly symmetric complex normal ($\mathcal{CN}$) distribution with probability density as $\tau_1\, \mathcal{CN}(0, \sigma_1^2) + \tau_2\, \mathcal{CN}(0, \sigma_2^2)$, where $0 \leq \tau_i \leq 1$ $(i = 1, 2)$, $\tau_1 + \tau_2 = 1$, and σ_i^2 is the variance of the ith component. If $\sigma_2^2 \gg \sigma_1^2$ and $\tau_2 < \tau_1$, noise samples with variance σ_1^2 and occurrence probability τ_1 are more frequently observed, while the remaining samples (with variance σ_2^2 and probability τ_2) can be viewed as the impulsive outliers. The observation quality is measured by peak signal-to-noise ratio (PSNR): $\mathrm{PSNR} \triangleq 10\log_{10}(\frac{\max\{|\tilde{\alpha}_c|\}_{c=1}^C}{\tau_1\sigma_1^2 + \tau_2\sigma_2^2})$. In Fig. 1(a)–(c), we depict the RSNR versus the PSNR ($C = 3, M = 30$), the number of measurements M ($C = 3, \mathrm{PSNR} = 10\,\mathrm{dB}$), and the frequency spacing coefficient ($C = 2, \mathrm{PSNR} = 10\,\mathrm{dB}, M = 25$), respectively. The common conditions are $\sigma_2^2 = 200\sigma_1^2$ and $\tau_2 = 0.2$. In Fig. 1(c), the frequency spacing coefficient is defined as $\mu \triangleq 64|\tilde{f}_1 - \tilde{f}_2|$, where μ ranges from 0.2 to 2. In Fig. 1(d), we plot the RSNR versus the impulsive noise level $\kappa \triangleq \sigma_2^2/\sigma_1^2$ with $\tau_2 = 0.1$, $C = 3$, and $\mathrm{PSNR} = 10\,\mathrm{dB}$. It is observed from Fig. 1(a)–(d) that the proposed approach yields a uniformly better performance than the other compared approaches, e.g., with significant 10 dB gain over the SURE-IR algorithm when $M > 20$ or $\mathrm{PSNR} < 15\,\mathrm{dB}$; in particular, $p = 0.5$ gives the best performance. Second, to further test the behavior of our approach, we now consider the α-stable noise process [26], and the noise samples are generated from the complex symmetric α-stable distribution whose characteristic function is $\psi(\cdot) = \exp\left(-\tilde{\gamma}^{\alpha} |\cdot|^{\alpha}\right)$. Here, we define the PSNR as $10\log_{10}\left(\frac{\max\{|\tilde{\alpha}_c|\}_{c=1}^C}{\tilde{\gamma}^{\alpha}}\right)$ and set $\alpha = 1.2$. To compare with the GMM noise, we depict the RNSR versus the number of measurements and the frequency spacing coefficient in Fig. 1(e) and (f), respectively, where $\mathrm{PSNR} = 15\,\mathrm{dB}$ and the other conditions are kept unchanged as those of Fig. 1(b) and (c). Again, it is seen that the proposed RSURE-IR approach outperforms other algorithms by a big margin when $M > 20$. Unlike the GMM noise case, $p = 0.8$ presents a relatively good performance among all the selected values of p, whereas $p = 0.5$ and $p = 1.1$ exhibit a very similar performance.

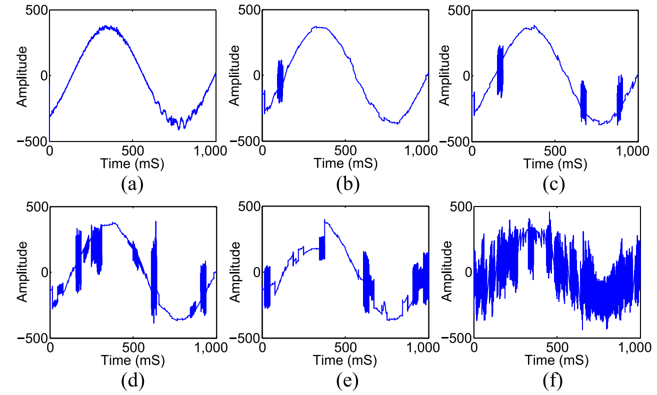

Fig. 2. (a)–(f) Original message of an AM signal and its reconstructions by respective algorithms, $M = 30$.

The last experiment tests the performance of various algorithms using a real-world amplitude modulated (AM) signal [4], [27], which encodes the original message appearing in Fig. 2(a). The signal was transmitted from a communication device using carrier frequency 8.2 kHz, and the received signal was sampled by an analog–digital converter at a rate of 32 kHz. The sampled signal has a total number of 32768 samples and is divided into a number of short-time half overlapped 64 segments, each with 1024 samples. For each segment, we randomly select $M = 30$ samples and 20% of them are added to outliers from a zero mean uniform distribution within $[-400, 400]$. Then, we use respective algorithms to recover the AM signal in each segment (the SDP and RSDP methods were not included here due to prohibitive computational complexity). After all segments are reconstructed, we perform AM demodulation to recover the original message, and the results are displayed in Fig. 2(b)–(f). It can be seen that our proposed approach provides a better visual recovery quality (the best one with $p = 0.8$) than other three methods.

V. Conclusion

We introduced a robust formulation for spectral CS in impulsive noise by exploiting a generalized ℓ_p-norm $(0 < p < 2)$ error-fitting criterion. Based on the MM scheme, we developed an iteratively reweighted ℓ_2 approach for robust joint sparse signal recovery and frequency parameter learning. Numerical results showed that the proposed approach presents a substantial performance advantage over other existing methods under impulsive noise.

REFERENCES

[1] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[2] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," *Inverse Probl.*, vol. 18, no. 5, pp. 1247–1279, 2002.

[3] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[4] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, Jul. 2013.

[5] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[6] Z. Lu, R. Ying, S. Jiang, P. Liu, and W. Yu, "Distributed compressed sensing off the grid," *IEEE Signal Process. Lett.*, vol. 22, no. 1, pp. 105–109, Jan. 2015.

[7] M. Cho, K. V. Mishra, J.-F. Cai, and W. Xu, "Block iterative reweighted algorithms for super-resolution of spectrally sparse signals," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2319–2313, Dec. 2015.

[8] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3809–3822, Jul. 2012.

[9] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[10] J. Fang, J. Li, Y. Shen, H. Li, and S. Li, "Super-resolution compressed sensing: An iterative reweighted algorithm for joint parameter learning and sparse signal recovery," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 761–765, Jun. 2014.

[11] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4997–5008, Oct. 2014.

[12] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.

[13] 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.

[14] 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.

[15] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.

[16] 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, Dec. 2008.

[17] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, Jun. 2010.

[18] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

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

[20] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *arXiv preprint arXiv:1609.02247*, 2016.

[21] G. Tang, P. Shah, B. N. Bhaskar, and B. Recht, "Robust line spectral estimation," in *Proc. 48th AsilomarConf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 301–305.

[22] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.

[23] J. Song, P. Babu, and D. P. Palomar, "Sparse generalized eigenvalue problem via smooth optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1627–1642, Apr. 2015.

[24] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[25] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, USA, Apr. 2008, pp. 3869–3872.

[26] E. Ollila, J. Eriksson, and V. Koivunen, "Complex elliptically symmetric random variables—Generation, characterization, and circularity tests," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 58–69, Jan. 2011.

[27] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.