IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 February 2024, accepted 26 February 2024, date of publication 8 March 2024, date of current version 21 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374118

RESEARCH ARTICLE

Joint Channel Estimation, Data Decoding, and Group-Sparse Impulsive Noise Estimation for Slowly Time-Varying Single Carrier Underwater Acoustic Communications

MINHAO ZHANG[1,2,3], JIANG ZHU[2], (Member, IEEE), YIFAN WANG[1,2,3], YANBING FU[1,2,3], YAN WEI[1,2,3], (Member, IEEE), XINGBIN TU[1,2,3], (Member, IEEE), AND FENGZHONG QU[1,2,3,4], (Senior Member, IEEE)

[1]Key Laboratory of Ocean Observation-Imaging Testbed of Zhejiang Province, Zhejiang University, Zhoushan 316021, China
[2]Ocean College, Zhejiang University, Zhoushan 316021, China
[3]Engineering Research Center of Oceanic Sensing Technology and Equipment, Ministry of Education, Zhejiang University, Zhoushan 316021, China
[4]Hainan Institute, Zhejiang University, Sanya 572025, China

Corresponding author: Yan Wei (redwine447@zju.edu.cn)

This work was supported in part by the Key Research and Development Project of Hainan Province under Grant ZDYF2023GXJS010, and in part by the Science and Technology Projects of Zhoushan Municipal Science and Technology Bureau for Zhejiang University under Grant 2023C81002.

ABSTRACT The Gaussian noise assumption is usually made for slowly time-varying underwater acoustic communication (UWAC) systems, which is entirely adequate for most cases. However, certain human activities such as geophysical surveys, multi-beam echosounders and iceberg breakup generate impulsive noise. Consequently, the impulsive noise degrades the designed receiver based on the Gaussian noise model significantly. An optimization formulation which takes the slowly time-varying channel impulse response (CIR), the group-sparse impulsive noise and the encoder structure of the data symbol into account is proposed. To tackle the optimization problem, we propose a joint channel estimation, data decoding and impulsive noise estimation (JCDI) algorithm consisting of two modules named module A and module B, where module A performs the joint channel estimation, data estimation and impulsive noise estimation, while module B uses the low density parity check (LDPC) decoder to decode the encoded data symbols. In particular, the block coordinate descent (BCD) algorithm is proposed, and a smoothed fast iterative threshold soft algorithm (FISTA) is proposed to estimate the group-sparse impulsive noise. Substantial numerical simulations and real experiments are conducted to show the effectiveness of JCDI under impulsive noise, i.e., JCDI achieves lower coded BER than conventional receivers.

INDEX TERMS Block coordinate descent, group sparse, underwater acoustic communications, impulsive noise.

I. INTRODUCTION

Underwater acoustic (UWA) communications is vital to various commercial and scientific research activities in oceans and has many applications such as military, oceanographic, commercial shipping, offshore oil and gas industry and underwater vehicles [1]. However, due to the propagation characteristics, underwater acoustic channels are considered to be one of the most challenging communication channels. In particular, the low propagation speed and the reflection of acoustic wave from surface and bottom of the ocean, along with the refraction caused by nonuniformity of sound speed in the water, result in significant multipath delay spread, which leads to severe inter-symbol-interference (ISI)

The associate editor coordinating the review of this manuscript and approving it for publication was Yougan Chen.

especially in high symbol rate single-carrier system. Besides, due to the low speed of sound, coupled with the motion of transceiver platforms as well as the fluctuation of propagation medium, Doppler effect [2] is almost inevitable in UWA communication, which results in time-varying channel. There are different modulation techniques for underwater acoustic communications including direct sequence spread spectrum (DSSS), frquency shift keying (FSK), phase shift keying (PSK) and so on. They all have pros and cons. For example, DSSS and FSK are suitable for low date rate and high reliable communications and coherent modulation like PSK and orthogonal frequency division multiplexing (OFDM) [3] is suitable for high date rate communications. In this paper, we focus on the single carrier PSK modulation which faces the challenge of ISI.

To address the challenge of ISI, there are mainly two categories of approaches, i.e., channel estimation (CE)-based approaches and direct adaptive (DA) approaches. The CE-based approaches such as the minimum mean square error (MMSE) decision feedback equalizer (DFE) [4], the iterative block DFE in the frequency domain (IB-DFE) [5] and the frequency domain generalized approximate message passing (FD-GAMP) [6] need to acquire channel information first. To get precise channel information, complicated algorithms are usually involved. The DA approaches are more computationally efficient. The mostly widely used DA approaches include the least mean square (LMS) equalizer [7] and the recursive least squares (RLS) [10] equalizer. There are various variants of the LMS equalizer such as the normalized LMS (NLMS) equalizer [8] and the sparsity-aware improved proportionate NLMS (IPNLMS) equalizer [9]. They only require linear time complexity. Compared with the LMS equalizer, the RLS equalizer has faster convergence but requires more complexity. To address the Doppler, multiple methods such as direct resampling, phase locked loop (PLL) [7] and joint carrier frequency offset, gridless channel estimation and data detection (JCCED-VALSE) [11] are proposed. In [12] and [13], Gaussian Markov model is used to model the time-varying characteristic of channel in two adjacent sub-blocks. In [14], the RLS predictor is adopted to track the delay-Doppler spreading function block by block.

The underwater ambient noise also has great effects on communication systems. The approaches mentioned above all assume that the noise obeys Gaussian distribution. However, such an assumption is not always hold in UWA channels. Many UWA environments have impulsive noise sources [15], including human industrial activities, marine life noise, heavy rain, strong winds [16]. In such a scenario, many assumptions held in the above approaches are no longer true, thus leading to performance degradation. The impulse noise is typically modeled by the following model: (1) student t distribution [17], (2) symmetric alpha-Stable distribution [16], [18], (3) Gaussian mixture model (GMM) [19], (4) sparse vector [20], [21], [22], (5) Bernoulli-Gaussian hidden Markov model (BGHMM) [23]. With respect to these noise models, various approaches have been proposed. One is to use a simple threshold test, usually used in OFDM system, since the time-domain OFDM samples can be modeled as i.i.d Gaussian. This approach pre-processes the received signal via clipping or blanking techniques [24]. In [25], recursive least sign algorithm (RLSA) using the l_1-norm of the estimated error is applied to construct a robust estimator in impulsive environment. However, it only considers the channel estimation. Thus, it is expected that their method is effective for data decoding only when the impulsive noise is absent. Besides, there are approaches that can perform joint symbol detection and impulsive-noise estimation, usually based on Bayesian algorithms. In [26], sparse Bayesian learning is used to estimate and subtract the noise impulses in powerline communications. In [23], a near-optimal yet computationally tractable approach combing generalized approximate message passing (GAMP) with soft-input soft-output decoding is proposed for OFDM transmissions in impulsive noise environments. For deep-sea vertical underwater acoustic communications, an enhanced iterative receiver based on vector approximate message passing is proposed to suppress the loud impulsive noise [19].

This paper utilizes the sparse and group sparse property of the impulsive noise and proposes a receiver which aims to maximize a posteriori probability of symbols, channel and impulsive noise. The contributions of this paper are listed as follows: Firstly, a joint channel estimation, data decoding and impulsive noise estimation receiver (JCDI) is proposed. The joint optimization problem is approximately solved by alternatively optimizing symbols, channel and impulsive noise. Secondly, for the impulsive noise, an efficient smoothing fast iterative soft threshold algorithm (S-FISTA) is proposed exploiting the sparse and group sparse property of the impulsive noise. In addition, we adopt the efficient majorization-minimization (MM) algorithm to estimate the symbol with constraint. Finally, numerical simulations and experiments are conducted to demonstrate the effectiveness of the proposed receiver. Compared with conventional receivers, the proposed receiver achieves lower coded bit error rate (BER) under impulsive noise.

The following notations are adopted. The superscript T and H indicate transpose and conjugate transpose, respectively. $\|\boldsymbol{x}\|$ and $\|\boldsymbol{x}\|_1$ indicate l_2-norm and l_1-norm of the vector $\boldsymbol{x}$, respectively. $\|\boldsymbol{X}\|$ indicates spectral norm of the matrix $\boldsymbol{X}$. $\mathbb{C}^a$ and $\mathbb{C}^{a\times b}$ denote a $a \times 1$ dimensional complex vector space and a $a \times b$ dimensional complex vector space, respectively. $\mathbb{R}^{a\times b}$ denote a $a \times b$ dimensional real vector space. $\boldsymbol{x}(a)$ indicates the a-th element of the vector $\boldsymbol{x}$. $\boldsymbol{x}(a:b)$ indicates a new vector whose elements take values from the a-th element to the b-th element of the vector $\boldsymbol{x}$. $\boldsymbol{X}(:, a:b)$ indicates a new matrix which takes column vector from the a-th column vector to the b-th column vector of the matrix $\boldsymbol{X}$. $\angle x$ denotes the phase angle of the complex variable x. $\text{sign}\{\cdot\}$ is a sign function. $\min\{a, b\}$ returns the minimum value of a and b. $\Re\{x\}$ and $\Im\{x\}$ output the real and imaginary part of x,

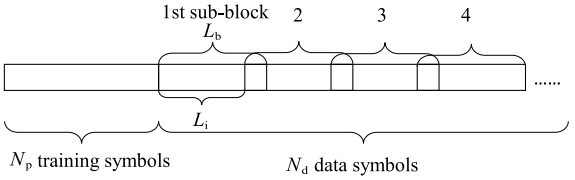

FIGURE 1. The received data block divided into several sub-blocks.

respectively. Italics denotes variable and bold denotes vectors or matrices. j denotes $\sqrt{-1}$. ∇ denotes the gradient operator. $\{a_1, a_2, \cdots, a_M\}^N$ denotes a N dimensional vector space, the element of which takes the value from $a_1, a_2, \cdots, a_M$. $a \triangleq b$ denotes that a is defined as b. $\mathcal{CN}(\mu, 6)$ denotes complex Gaussian distribution with expectation of μ and covariance matrix of 6.

II. SYSTEM MODEL

Let the vector $\boldsymbol{b} \in \{0,1\}^{N_b}$ be the transmitted (equi-probable) information bits. These bits are coded by a rate-R encoder to get the length-N_b/R vector $\boldsymbol{c} = \mathcal{C}(\boldsymbol{b})$, where $\mathcal{C} : \{0,1\}^{N_b} \to \{0,1\}^{N_b/R}$ is a coding function which is a low density parity check (LDPC) code in this paper. Let $\mathcal{M} : \{0,1\}^Q \to \mathbb{A}$ denote the 2^Q-ary mapping, where $\mathbb{A}$ is the quadrature phase shift keying (QPSK) alphabet in this paper. The data block contains N_p training symbols denoted by $\boldsymbol{x}_p$ and N_d data symbols denoted by $\boldsymbol{x}$.

In the following, we establish the system model and present the problem formulation. We divide N_d data symbols into I sub-blocks. There is an overlap between adjacent sub-blocks. The length of the overlap is $L_b - L_i$, where L_b is the length of each sub-block and L_i is the interval between adjacent sub-blocks, see Fig. 1 for an illustration.

The channel is assumed to be slowly time-varying. Define $\boldsymbol{h}_i = [h_i(0), h_i(1), \cdots, h_i(L_c-1)]^T$ where $h_i(l)$ denotes the l-th tap of the channel impulsive response (CIR) of the i-th sub-block and L_c is the length of the channel.[1] We assume that the channels between the adjacent sub-blocks follow the Gaussian-Markov chain assumption, i.e.,

$$\boldsymbol{h}_i = \alpha_i \boldsymbol{h}_{i-1} + \boldsymbol{w}_i, i = 1, 2, \cdots, N_{sb}, \tag{1}$$

where N_{sb} is the number of sub-blocks and $\boldsymbol{h}_i \in \mathbb{C}^{L_c}$ is the CIR within the i-th sub-block, which is assumed time-invariant. α_i is the first-order autoregressive (AR) coefficient, $\boldsymbol{w}_i$ is the white Gaussian noise and $\boldsymbol{h}_0$ is the channel estimate using the training symbols. The received signal of the i-th sub-block under the time-invariant channel is modeled as

$$\begin{aligned} y_i(n) = &\sum_{l=0}^{L_c-1} h_i(l)\, x_i(n-l) + v((i-1)L_i + n) \\ &+ o((i-1)L_i + n), n = 1, 2, \cdots, L_b \end{aligned} \tag{2}$$

[1] A more detailed modeling of the channel can be done using physical underwater acoustics theory. However, the channel modeling based on physical underwater acoustics theory is out of the scope of the paper. Since this paper focuses on the impulsive noise, further research which includes underwater environment and underwater propagation models will be investigated in the future.

where the subscript i denotes the i-th sub-block, n denotes instant, $x_i(n)$ denotes the n-th transmitted symbol in the i-th sub-block, v denotes the additive white Gaussian noise (AWGN) and o denotes the impulsive noise. The impulsive noise often has the sparsity and group sparsity [25], i.e., most elements of the impulsive noise and the elements of the first order difference of the impulsive noise are nearly zero. The heavy tails of the Laplace distribution make it well-suited for modeling sparse distributions, as it allows for a higher probability of small values and occasional large values. Here we exploit the sparsity of signals by assuming a Laplacian prior on the signal coefficients, similar to the least absolute shrinkage and selection operator (LASSO) or basis pursuit does. By imposing an additional Laplacian distribution on the first order difference of the impulsive noise, the sparsity both at the individual level and group level is promoted, and finer control over the sparsity pattern is allowed. In this way, the distribution of the impulsive noise is supposed to follow

$$p(\boldsymbol{o}) \propto \mathrm{e}^{-\lambda_1 \|\boldsymbol{o}\|_1 - \lambda_2 \|\boldsymbol{D}\boldsymbol{o}\|_1}, \tag{3}$$

where λ_1 and λ_2 are the known parameters of the distribution and $\boldsymbol{D} \in \mathbb{R}^{(N_d+L_c-2)\times(N_d+L_c-1)}$ is defined as

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}. \tag{4}$$

It is worth noting that the distribution (3) also generalize the sparse impulsive noise case by setting $\lambda_2 = 0$.

Define

$$\boldsymbol{x}_{i-1,\text{tail}} = \begin{bmatrix} x_{i-1}(L_i - L_c + 2) \\ x_{i-1}(L_i - L_c + 3) \\ \vdots \\ x_{i-1}(L_i) \end{bmatrix} \in \mathbb{C}^{L_c-1}. \tag{5}$$

Note that in order to cancel the ISI between adjacent sub-blocks, $\boldsymbol{x}_{i-1,\text{tail}}$ is assumed known when we detect the i-th sub-block.

Define

$$\boldsymbol{y}_i = [y_i(1), y_i(2), \cdots, y_i(L_b)]^T \in \mathbb{C}^{L_b}, \tag{6}$$

$$\boldsymbol{x}_i = [x_i(1), x_i(2), \cdots, x_i(L_b)]^T \in \mathbb{C}^{L_b}, \tag{7}$$

$$\boldsymbol{o}_i = \begin{bmatrix} o((i-1)L_i + 1) \\ o((i-1)L_i + 2) \\ \vdots \\ o((i-1)L_i + L_b) \end{bmatrix} \in \mathbb{C}^{L_b}, \tag{8}$$

$$\boldsymbol{x} = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_{N_{sb}}^T\right]^T \in \mathbb{C}^{N_d}, \tag{9}$$

$$\alpha = \left[\alpha_1, \alpha_2, \cdots, \alpha_{N_{sb}}\right]^T \in \mathbb{C}^{N_{sb}}, \tag{10}$$

$$\boldsymbol{x}_{i,e} = \left[\boldsymbol{x}_{i-1,\text{tail}}^T, \boldsymbol{x}_i^T\right]^T \in \mathbb{C}^{L_b+L_c-1}, \tag{11}$$

Given the above system model, the posterior probability density function (PDF) of the data symbol, the channel and

the impulsive noise is

$$\begin{aligned} & p\left(\boldsymbol{x}, \boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_{\mathrm{sb}}}, \boldsymbol{o} \mid \boldsymbol{y} ; \alpha\right) \\ & \quad \propto p\left(\boldsymbol{x}, \boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_{\mathrm{sb}}}, \boldsymbol{o}, \boldsymbol{y} ; \alpha\right) \\ & \quad \stackrel{a}{=} p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_{\mathrm{sb}}}, \boldsymbol{o}\right) p(\boldsymbol{x}) \\ & \qquad \times p\left(\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_{\mathrm{sb}}} ; \alpha\right) p(\boldsymbol{o}) \\ & \quad \stackrel{b}{=} p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_{\mathrm{sb}}}, \boldsymbol{o}\right) p(\boldsymbol{x}) \\ & \qquad \times \prod_{i=1}^{N_{\mathrm{sb}}} p\left(\boldsymbol{h}_i \mid \boldsymbol{h}_{i-1} ; \alpha_i\right) p(\boldsymbol{o}) \\ & \quad \stackrel{c}{=} \prod_{i=1}^{N_{\mathrm{sb}}} p\left(\boldsymbol{y}_i \mid \boldsymbol{x}_{i,\mathrm{e}}, \boldsymbol{h}_i, \boldsymbol{o}_i\right) p(\boldsymbol{x}) \prod_{i=1}^{N_{\mathrm{sb}}} p\left(\boldsymbol{h}_i \mid \boldsymbol{h}_{i-1} ; \alpha_i\right) p(\boldsymbol{o}), \end{aligned} \tag{12}$$

where $\stackrel{a}{=}$ is due to the independence of the symbol $\boldsymbol{x}$, the channel $\boldsymbol{h}$ and the impulsive noise $\mathbf{o}$, $\stackrel{b}{=}$ is due to the Gaussian-Markov chain assumption of the channel in all the sub-blocks shown in (1) and $\stackrel{c}{=}$ approximately holds as the overlap length of each sub-block $L_{\mathrm{b}} - L_{\mathrm{i}}$ (in our setting, $L_{\mathrm{b}} - L_{\mathrm{i}} = 9$) is small.

Define $\boldsymbol{h} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N_{\mathrm{sb}}}]$. Consequently, taking the logarithm of (12), the maximum a posteriori (MAP) estimation of the data symbol, the channel and the impulsive noise can be formulated as

$$\begin{aligned} \max_{\boldsymbol{x},\boldsymbol{h},\alpha,\boldsymbol{o}} & \sum_{i=1}^{N_{\mathrm{sb}}} \ln p\left(\boldsymbol{y}_i \mid \boldsymbol{x}_{i,\mathrm{e}}, \boldsymbol{h}_i, \boldsymbol{o}\right) + \ln p(\boldsymbol{x}) \\ & + \sum_{i=1}^{N_{\mathrm{sb}}} \ln p\left(\boldsymbol{h}_i \mid \boldsymbol{h}_{i-1} ; \alpha_i\right) + \ln p(\boldsymbol{o}). \end{aligned} \tag{13}$$

The above optimization problem (13) is difficult to solve due to the non-convex structure. In fact, the explicit expression of the priori distribution $p(\boldsymbol{x})$ is hard to obtain due to the encoding structure of the data symbols. This also increases the difficulty of solving the optimization problem. In the ensuing section, we propose a tractable scheme to jointly perform the channel estimation, symbol detection and impulsive noise estimation.

III. THE JOINT CHANNEL ESTIMATION, DATA DECODING, AND IMPULSIVE NOISE ESTIMATION ALGORITHM

In this section, we propose an algorithm named JCDI which solves (13) approximately. The algorithm consists of two modules named module A and module B as shown in Fig.2, where module A performs the joint channel estimation, symbol estimation and impulsive noise estimation, and outputs the soft information, i.e., the log-likelihood ratio (LLR) of the data symbols to feed into the module B. The module B is a LDPC decoder which performs symbol decoding and the decoded symbol is fed into the module A. The two modules iterate until the decoded symbol is error-free or the number of the maximum iterations is reached.

For module A, we drop the prior distribution $\ln p(\boldsymbol{x})$ in (13) but incorporating the constraint of the symbol $\boldsymbol{x}$,

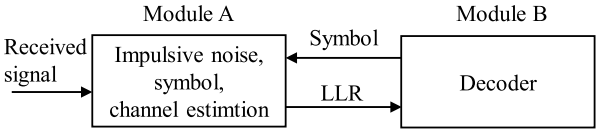

FIGURE 2. The block diagram of the proposed iterative JCDI receiver.

i.e., $x_i(n) \in \mathcal{A}$, $1 \le n \le L_b$, where $\mathcal{A}$ is a convex set. Consequently, a relaxed optimization problem can be formulated as

$$\begin{aligned} \max_{\boldsymbol{x},\boldsymbol{h},\alpha,\boldsymbol{o}} & \sum_{i=1}^{N_{\mathrm{sb}}} \ln p\left(\boldsymbol{y}_i \mid \boldsymbol{x}_{i,\mathrm{e}}, \boldsymbol{h}_i, \boldsymbol{o}\right) \\ & + \sum_{i=1}^{N_{\mathrm{sb}}} \ln p\left(\boldsymbol{h}_i \mid \boldsymbol{h}_{i-1} ; \alpha_i\right) + \ln p(\boldsymbol{o}) \\ \text{s.t.} \quad & x_i(n) \in \mathcal{A},\ 1 \le n \le L_b, 1 \le i \le N_{sb}. \end{aligned} \tag{14}$$

The specific form of $\mathcal{A}$ can be designed according to the constellation. For example, $\mathcal{A}$ can be designed as $|x_i(n)| \le A$ for PSK constellation, $|\Re\{x_i(n)\}| \le A$, $|\Im\{x_i(n)\}| \le A$ for square-shaped QPSK constellation, or

$$|\Re\{x_i(n)\} + \Im\{x_i(n)\}| \le A,\ |\Re\{x_i(n)\} - \Im\{x_i(n)\}| \le A \tag{15}$$

for diamond-shaped QPSK constellation where A is the amplitude constraint.

Define $\boldsymbol{H}_{i,\mathrm{e}} \in \mathbb{C}^{L_{\mathrm{b}} \times (L_{\mathrm{b}} + L_{\mathrm{c}} - 1)}$ as

$$\boldsymbol{H}_{i,\mathrm{e}} = \begin{bmatrix} h_i(L_{\mathrm{c}}-1) & \cdots & h_i(0) & 0 & \cdots & 0 & 0 \\ 0 & \cdots & h_i(1) & h_i(0) & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & h_i(1) & h_i(0) \end{bmatrix}. \tag{16}$$

Consequently, based on the received signal model (2), the Gaussian-Markov chain channel model (1) and the impulsive noise distribution (3), the optimization problem (14) can be equivalently simplified as

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{h},\alpha,\boldsymbol{o}} & \underbrace{\sum_{i=1}^{N_{\mathrm{sb}}} \sigma_v^{-2} \|\boldsymbol{y}_i - \boldsymbol{H}_{i,\mathrm{e}} \boldsymbol{x}_{i,\mathrm{e}} - \boldsymbol{o}_i\|^2}_{\text{the first term}} \\ & + \underbrace{\sum_{i=1}^{N_{\mathrm{sb}}} \sigma_w^{-2} \|\boldsymbol{h}_i - \alpha_i \boldsymbol{h}_{i-1}\|^2}_{\text{the second term}} \\ & + \underbrace{\lambda_1 \|\boldsymbol{o}\|_1}_{\text{the third term}} + \underbrace{\lambda_2 \|\boldsymbol{D}\boldsymbol{o}\|_1}_{\text{the fourth term}} \\ \text{s.t.} \quad & x_i(n) \in \mathcal{A},\ 1 \le n \le L_b, 1 \le i \le N_{sb}. \end{aligned} \tag{17}$$

The optimization problem (17) is separable with respect to some variables which can be utilized to design efficient algorithms. The first term is separable w.r.t. the

symbol, the channel and the impulsive noise at different sub-blocks, i.e., $\{\boldsymbol{x}_1, \boldsymbol{h}_1, \boldsymbol{o}_1\}, \{\boldsymbol{x}_2, \boldsymbol{h}_2, \boldsymbol{o}_2\}, \cdots, \{\boldsymbol{x}_{N_{\text{sb}}}, \boldsymbol{h}_{N_{\text{sb}}}, \boldsymbol{o}_{N_{\text{sb}}}\}$. The third term of the objective function is separable w.r.t. the impulsive noise $o(1), o(2), \cdots, o(N_{\text{d}} + L_{\text{c}} - 1)$, while the fourth term is non-separable. For every sub-block, with the channel $\boldsymbol{h}_i$ and the impulsive noise $\boldsymbol{o}_i$ being fixed, the objective function is non-separable w.r.t. the symbol $x_i(1), x_i(2), \cdots, x_i(L_{\text{b}})$. Fixing all the variables except $\boldsymbol{h}_i$, the first term and the second term is a quadratic function of $\boldsymbol{h}_i$. Besides, the second term is a quadratic function of α_i. Thus both $\boldsymbol{h}_i$ and α_i can be optimized via the least squares method. We adopt block coordinate descent (BCD) algorithm to optimize the impulsive noise, the symbol, the channel. In addition, the idea of the successive interference cancellation (SIC) scheme is also utilized and we perform the symbol estimation sub-block-by-sub-block. In the following, we present these details.

A. ESTIMATION OVER THE IMPULSIVE NOISE

In this subsection, we perform the estimation over the impulsive noise and propose a S-FISTA to solve the subproblem, i.e., estimate the impulsive noise.

As we fix all the variables except the impulsive noise $\boldsymbol{o}$, the second term in (17) is irrelevant and can be dropped. The impulsive noise $\boldsymbol{o}$ is optimized over the whole block. Then, we have a simplified optimization problem

$$\min_{\boldsymbol{o}} \underbrace{\sigma_v^{-2}\|\boldsymbol{y}_{\text{in}} - \boldsymbol{o}\|^2}_{\text{the first term}} + \underbrace{\lambda_2\|\boldsymbol{D}\boldsymbol{o}\|_1}_{\text{the second term}} + \underbrace{\lambda_1\|\boldsymbol{o}\|_1}_{\text{the third term}} \triangleq q(\boldsymbol{o}) \tag{18}$$

where σ_v^2 is the variance of the additive noise and $\boldsymbol{y}_{\text{in}} \in \mathbb{C}^{N_{\text{d}}+L_{\text{c}}-1}$ defined as

$$\boldsymbol{y}_{\text{in}} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} \tag{19}$$

is the residual received signal. The channel convolution matrix $\boldsymbol{H}$ is construted based on the channel estimate of the last iteration between module A and B shown in Fig. 2. The symbol $\boldsymbol{x}$ adopts the hard decision of the symbol estimate from the decoder.

Notice that the second and the third term of (18) are non-differentiable. Although subgradient methods can be adopted to solve such a problem, their convergence speed is not satisfactory. We seek to use fast first-order methods to efficiently solve the optimization problem (18). However, the fast methods cannot be directly applied to the optimization problem because of the nondifferentiable terms, i.e., the second and the third term. Therefore, we first smooth the objective function.

Also, we notice that the first term and the third term of (18) are separable w.r.t. to the impulsive noise $o(1), o(2), \cdots, o(N_{\text{d}} + L_{\text{c}} - 1)$, while the second term is not separable. The above properties of the objective function $q(\boldsymbol{o})$ can be utilized when solving the optimization problem (18).

Before solving the optimization problem (18), some concepts are introduced in Appendix VI to make the paper self-contained. The definition of smoothness and smoothable functions are introduced in Definition 1 and 2, respectively. Then, two theorems of calculating the smooth approximation according to the structure of the function are presented in Theorem 1 and Theorem 2.

To solve optimization problem (18), we draw on the principles of smoothing first-order gradient descent methods. We first smooth the second term in (18) with the Moreau envelope. Then, the proximal operator is adopted to solve the smoothed optimization problem. Besides, the Nesterov momentum is adopted to accelarate the convergence. It should be noted that the variable of the optimization problem (18) is complex-valued. Therefore, before optimizing the impulsive noise $\boldsymbol{o}$, we transform the optimization problem with variables being complex-valued into that with variables being real-valued. Next, we present the details of the steps of the algorithm.

1) TRANSFORMING THE OPTIMIZATION PROBLEM WITH VARIABLES BEING COMPLEX-VALUED INTO THAT WITH VARIABLES BEING REAL-VALUED

For ease of representation, define

$$L_y = N_{\text{d}} + L_{\text{c}} - 1. \tag{20}$$

Define $\check{\boldsymbol{y}}_{\text{in}} \in \mathbb{C}^{2L_y}$ and $\check{\boldsymbol{o}} \in \mathbb{C}^{2L_y}$ where

$$\check{\boldsymbol{y}}_{\text{in}}(2n-1:2n) = \left[\Re\{\boldsymbol{y}_{\text{in}}(n)\}, \Im\{\boldsymbol{y}_{\text{in}}(n)\}\right]^{\text{T}} \in \mathbb{C}^2, \tag{21}$$

$$\check{\boldsymbol{o}}(2n-1:2n) = [\Re\{o(n)\}, \Im\{o(n)\}]^{\text{T}} \in \mathbb{C}^2. \tag{22}$$

Define

$$\boldsymbol{B} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 4}, \tag{23}$$

$$\boldsymbol{C}_n = \begin{bmatrix} \boldsymbol{0}_{4\times 2(n-2)} & \mathbf{I}_{4\times 4} & \boldsymbol{0}_{4\times(2L_y-2n)} \end{bmatrix} \in \mathbb{R}^{4\times 2L_y}, \tag{24}$$

and $\boldsymbol{A}_n = \boldsymbol{B}\boldsymbol{C}_n$. By using the structure of $\boldsymbol{D}$ defined in (4), it is not hard to know that the l_1-norm in the second term of (18) can be formulated as

$$\|\boldsymbol{D}\boldsymbol{o}\|_1 = \sum_{n=1}^{L_y-1} |\boldsymbol{D}(n,:)\,\boldsymbol{o}| = \sum_{n=2}^{L_y} \|\boldsymbol{A}_n\check{\boldsymbol{o}}\|. \tag{25}$$

Substituting (21), (22), (23), (24) and (25) into the complex-valued optimization problem (18) leads to a new optimization problem with real-valued variables

$$\min_{\check{\boldsymbol{o}}} \underbrace{\sigma_v^{-2}\left\|\check{\boldsymbol{y}}_{\text{in}} - \check{\boldsymbol{o}}\right\|^2}_{\text{the first term}} + \underbrace{\lambda_2 \sum_{n=2}^{L_y} \left\|\boldsymbol{A}_n\check{\boldsymbol{o}}\right\|}_{\text{the second term}} + \underbrace{\lambda_1 \sum_{n=1}^{L_y} \left\|\check{\boldsymbol{o}}(2n-1:2n)\right\|}_{\text{the third term}}. \tag{26}$$

2) SMOOTHING THE SECOND TERM OF THE OBJECTIVE FUNCTION IN (26)

For ease of representation, define the second term of (26) as $h\left(\check{\boldsymbol{o}}\right)$. We adopt the Moreau envelop to obtain the smoothed function of $h\left(\check{\boldsymbol{o}}\right)$. The Moreau envelope of the l_2-norm $\|\check{\boldsymbol{o}}\|$ is the Huber function $H_\mu(\check{\boldsymbol{o}})$ which is a $\frac{1}{\mu}$-smooth approximation of $\|\check{\boldsymbol{o}}\|$ with parameters $\left(1, \frac{1}{2}\right)$ [27, Chapter 10, p. 309]. $H_\mu(\check{\boldsymbol{o}})$ is defined as

$$H_\mu\left(\check{\boldsymbol{o}}\right) = \begin{cases} \frac{1}{2\mu}\left\|\check{\boldsymbol{o}}\right\|^2, \left\|\check{\boldsymbol{o}}\right\| \le \mu \\ \left\|\check{\boldsymbol{o}}\right\| - \frac{\mu}{2}, \left\|\check{\boldsymbol{o}}\right\| > \mu \end{cases}. \tag{27}$$

According to Theorem 2, the Moreau envelope of $\|\boldsymbol{A}_n\check{\boldsymbol{o}}\|$ is $H_\mu\left(\boldsymbol{A}_n\check{\boldsymbol{o}}\right)$ with parameters $(\|\boldsymbol{A}_n\|^2, \frac{1}{2})$. According to Theorem 1, the Moreau envelope of $\lambda_2\sum_{n=2}^{L_y}\left\|\boldsymbol{A}_n\check{\boldsymbol{o}}\right\|$ is

$$h_\mu(\check{\boldsymbol{o}}) = \lambda_2\sum_{n=2}^{L_y} H_\mu\left(\boldsymbol{A}_n\check{\boldsymbol{o}}\right) \tag{28}$$

with parameters $(\lambda_2\sum_{n=2}^{L_y}\|\boldsymbol{A}_n\|^2, \lambda_2\frac{L_y-1}{2})$. It can be easily proved that the maximum singular value of $\boldsymbol{A}_n$ is $\sqrt{2}$ for arbitrary n. Therefore, the smooth parameters are simplified to $\left(\sqrt{2}\lambda_2\left(L_y-1\right), \lambda_2\left(L_y-1\right)/2\right)$.

3) PROXIMAL MINIMIZATION

Define the first and third terms in (26) as $f(\check{\boldsymbol{o}})$ and $g(\check{\boldsymbol{o}})$, respectively. Define the smoothing function

$$F_\mu(\check{\boldsymbol{o}}) = f(\check{\boldsymbol{o}}) + h_\mu(\check{\boldsymbol{o}}). \tag{29}$$

After smoothing, the optimization problem (26) is approximated by

$$\min_{\check{\boldsymbol{o}}} F_\mu(\check{\boldsymbol{o}}) + g(\check{\boldsymbol{o}}). \tag{30}$$

This optimization problem can be solved by the proximal gradient method. The solution at the k-th iteration is

$$\check{\boldsymbol{o}}^{k+1} = \text{prox}_{\frac{1}{\tilde{L}}g}\left(\check{\boldsymbol{o}}^k - \frac{1}{\tilde{L}}\nabla F_\mu\left(\check{\boldsymbol{o}}^k\right)\right) \tag{31}$$

with $\tilde{L} = L_f + \frac{\sqrt{2}\lambda_2(L_y-1)}{\mu}$ where $L_f = 2\sigma_v^{-2}$ and L_y is defined in (20). The proximal operator $\text{prox}_{\frac{1}{\tilde{L}}g}\left(\check{\boldsymbol{o}}^k - \frac{1}{\tilde{L}}\nabla F_\mu\left(\check{\boldsymbol{o}}^k\right)\right)$ is defined as

$$\underset{\check{\boldsymbol{o}}^{k+1}}{\text{argmin}}\left\{\frac{1}{\tilde{L}}g(\check{\boldsymbol{o}}^{k+1}) + \frac{1}{2}\|\check{\boldsymbol{o}}^k - \frac{1}{\tilde{L}}\nabla F_\mu\left(\check{\boldsymbol{o}}^k\right)\|^2\right\}. \tag{32}$$

The superscript k denotes the k-th iteration. Based on (28), (29) and (27), the gradient of $F_\mu(\check{\boldsymbol{o}})$ is

$$\nabla F_\mu\left(\check{\boldsymbol{o}}\right) = 2\sigma_v^{-2}\left(\check{\boldsymbol{o}} - \check{\boldsymbol{y}}_{\text{in}}\right) + \lambda_2\sum_{n=2}^{L_y}\nabla H_\mu\left(\boldsymbol{A}_n\check{\boldsymbol{o}}\right), \tag{33}$$

where

$$\nabla H_\mu\left(\boldsymbol{A}_n\check{\boldsymbol{o}}\right) = \begin{cases} \frac{1}{\mu}\boldsymbol{A}_n^{\text{T}}\boldsymbol{A}_n\check{\boldsymbol{o}}, \left\|\boldsymbol{A}_n\check{\boldsymbol{o}}\right\| \le \mu \\ \frac{\boldsymbol{A}_n^{\text{T}}\boldsymbol{A}_n\check{\boldsymbol{o}}}{\left\|\boldsymbol{A}_n\check{\boldsymbol{o}}\right\|}, \left\|\boldsymbol{A}_n\check{\boldsymbol{o}}\right\| > \mu \end{cases}. \tag{34}$$

Furthermore, according to the definition of the proximal operator, substituting the gradient of the smoothing function $\nabla F_\mu(\check{\boldsymbol{o}})$ into (31) yields

$$o(n) = \begin{cases} \left(|a(n)| - \frac{1}{\tilde{L}}\lambda_1\right)\exp\left\{\angle a\left(n\right)\right\}, |a(n)| > \frac{1}{\tilde{L}}\lambda_1 \\ 0, |a(n)| \le \frac{1}{\tilde{L}}\lambda_1 \end{cases}, \tag{35}$$

where $a\left(n\right) = \check{\boldsymbol{a}}(2n-1) + j\check{\boldsymbol{a}}\left(2n\right)$ and

$$\check{\boldsymbol{a}} = \check{\boldsymbol{o}}^k - \frac{1}{\tilde{L}}\nabla F_\mu\left(\check{\boldsymbol{o}}^k\right) \in \mathbb{R}^{2L_y}. \tag{36}$$

The steps of obtaining (35) is shown in Appendix VI.

4) THE NESTEROV MOMENTUM ACCELERATION

The Nesterov momentum is adopted to accelerate the convergence of the proximal gradient method. At the k-th iteration, the Nesterov momentum replaces the initial point $\check{\boldsymbol{o}}$ in (31) by $\boldsymbol{z}^k$, i.e.,

$$\check{\boldsymbol{o}}^{k+1} = \text{prox}_{\frac{1}{\tilde{L}}g}\left(\boldsymbol{z}^k - \frac{1}{\tilde{L}}\nabla F_\mu\left(\boldsymbol{z}^k\right)\right), \tag{37}$$

where the superscript k denotes the k-th iteration. $\boldsymbol{z}^{k+1}$ is updated by

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \tag{38}$$

$$\boldsymbol{z}^{k+1} = \check{\boldsymbol{o}}^{k+1} + \frac{t_k-1}{t_{k+1}}\left(\check{\boldsymbol{o}}^{k+1} - \check{\boldsymbol{o}}^k\right). \tag{39}$$

Up to this point, we have presented a comprehensive description of the proposed S-FISTA for estimating the impulsive noise. The details of the S-FISTA are also summarized in Table 1 where μ is selected so that an ε-optimal solution will be achieved [27, p.314].

It should be noted that σ_v^2 is estimated by maximum likelihood (ML) estimator based on training symbols and is assumed time-variant within the whole block. At the 0-th iteration when the symbol decision and channel is not available, the impulsive noise estimation module outputs zero impulsive noise to the next module.

B. ESTIMATION OVER THE SYMBOL

In this subsection, we fix the channel and the impulsive noise and optimize the symbol. The symbol is estimated by the MM algorithm.

When optimizing the symbol, all terms except the first term in the joint optimization problem (17) are irrelevant and are dropped. Notice that the first term is separable w.r.t. the symbol in different sub-blocks. We estimate the symbol

TABLE 1. The proposed complex S-FISTA algorithm for impulsive noise estimation.

Input: $\boldsymbol{y}_o$, λ_2, L_y, σ_v^2, ε.
Initialize: $\boldsymbol{z}^0 = \boldsymbol{0} \in \mathbb{R}^{2L_y}$.
(1) Calculate the smoothing parameter μ:
$\alpha = \sqrt{2}\lambda_2 (L_y - 1)$
$\beta = \lambda_2 (L_y - 1)/2$
$L_f = 2\sigma_v^{-2}$
$\mu = \sqrt{\frac{\alpha}{\beta}} \frac{\varepsilon}{\sqrt{\alpha\beta} + \sqrt{\alpha\beta + L_f \varepsilon}}$
(2) Calculate the Lipschitz constant of smoothed function $F_\mu(\breve{\boldsymbol{o}}^k)$:
$\tilde{L} = L_f + \frac{\lambda_2 (L_y - 1)\sqrt{2}}{\mu}$
For $k = 0 : K$
(3) Proximal minimization: $\breve{\boldsymbol{o}}^{k+1} = \text{prox}_{\frac{1}{\tilde{L}} g}\left(\boldsymbol{z}^k - \frac{1}{\tilde{L}} \nabla F_\mu (\boldsymbol{z}^k)\right)$
(4) Update initial point:
$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$
$\boldsymbol{z}^{k+1} = \breve{\boldsymbol{o}}^{k+1} + \frac{t_k - 1}{t_{k+1}} (\breve{\boldsymbol{o}}^{k+1} - \breve{\boldsymbol{o}}^k)$
End

and the channel sub-block by sub-block sequentially. Define $\boldsymbol{H}_{i-1,\text{tail}} = \boldsymbol{H}_{i,\text{e}}(:, 1 : L_\text{c} - 1)$ and $\boldsymbol{H}_i = \boldsymbol{H}_{i,\text{e}}(:, L_\text{c} : L_\text{b} + L_\text{c} - 1)$ where $\boldsymbol{H}_{i,\text{e}}$ is defined in (16). For the i-th sub-block, the simplified optimization subproblem is

$$\min_{\boldsymbol{x}_i} \sigma_v^{-2} \|\boldsymbol{y}_i - \boldsymbol{H}_{i-1,\text{tail}} \boldsymbol{x}_{i-1,\text{tail}} - \boldsymbol{H}_{i-1} \boldsymbol{x}_i - \boldsymbol{o}_i\|^2 \triangleq q(\boldsymbol{x}_i)$$
$$\text{s.t.} \quad x_i(n) \in \mathcal{A}, \ 1 \le n \le L_b. \tag{40}$$

It should be noted that as the channel $\boldsymbol{h}_i$ of the i-th sub-block is not available, we replace the $\boldsymbol{h}_i$ with $\boldsymbol{h}_{i-1}$ based on the assumption of slowly time-varying channel.

The MM algorithm is adopted to solve (40). Since (40) is a convex optimization problem, global minimization can be guaranteed. The idea of the MM algorithm is described as follows. First, an initial point $\boldsymbol{x}_i^k$ is introduced where the superscript denotes the k-th MM iteration. Construct a surrogate upper bound of the objective function $q(\boldsymbol{x}_i|\boldsymbol{x}_i^k)$ which satisfies $q(\boldsymbol{x}_i|\boldsymbol{x}_i^k) \ge q(\boldsymbol{x}_i)$ and $q(\boldsymbol{x}_i|\boldsymbol{x}_i^k)|_{\boldsymbol{x}_i = \boldsymbol{x}_i^k} = q(\boldsymbol{x}_i)|_{\boldsymbol{x}_i = \boldsymbol{x}_i^k}$, where $q(\boldsymbol{x}_i)$ is defined in (40). A simple approach is to perform a second-order Taylor series expansion of the objective function $q(\boldsymbol{x}_i)$ around the point $\boldsymbol{x}_i^k$, and construct a separable surrogate upper bound of the objective function $q(\boldsymbol{x}_i|\boldsymbol{x}_i^k)$. As a consequence, we obtain a simple iterative algorithm to perform symbol estimation summarized in Table 2. For simplicity, the details are shown in Appendix VI. It should be noted that the last $L_\text{b} - L_\text{i}$ estimated symbols are abandoned because they are not reliable due to limited observation information. But they will be estimated again in the $(i+1)$-th sub-block thanks to the overlapping structure of data sub-blocks.

Since soft decision decoder is adopted, the LLR of the output of the symbol detection module is needed. With the Gaussian distribution assumption, we model the optimized symbols as

$$\hat{\boldsymbol{x}} = g_x \boldsymbol{x} + \boldsymbol{v}_x \tag{41}$$

where g_x is the gain coefficient and $\boldsymbol{v}_x$ is assumed to follow $\mathcal{CN}(\boldsymbol{0}, \sigma_{v_x}^2 \mathbf{I})$. Since the true symbol vector $\boldsymbol{x}$ is not available,

TABLE 2. The MM algorithm for symbol estimation of the i-th sub-block.

Input: $\boldsymbol{y}_\text{r}$ defined in (58), $\boldsymbol{H}_{i-1}$, the number of maximum MM iterations K.
Initialize: $\boldsymbol{x}_i^0 = \boldsymbol{0} \in \mathbb{C}^{L_\text{i}}$, calculate $\lambda_{\max}$ which is the maximum eigenvalue of $\boldsymbol{H}_{i-1}^\text{H} \boldsymbol{H}_{i-1}$.
For $k = 0 : K$
Calculating the optimal solution $\boldsymbol{x}_\text{o}^k$ defined in (61).
For $n = 1 : L_\text{i}$
If $x_\text{o}^k(n)$ is within the constraint $\mathcal{A}$
$x_i^{k+1} = x_\text{o}^k(n)$.
Else
x_i^{k+1} is the point in the boundary of constraint $\mathcal{A}$ which is closest to $x_\text{o}^k(n)$.
End
End
End

it is approximately replaced by hard decision of $\hat{\boldsymbol{x}}$, denoted by $\tilde{\boldsymbol{x}}$. The ML estimates of the gain and variance are

$$\hat{g}_x = \frac{\tilde{\boldsymbol{x}}^\text{H} \hat{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}^\text{H} \tilde{\boldsymbol{x}}}, \tag{42}$$
$$\hat{\sigma}_{v_x}^2 = \frac{1}{N_\text{d}} \|\hat{\boldsymbol{x}} - \hat{g}_x \tilde{\boldsymbol{x}}\|^2. \tag{43}$$

Therefore, the LLR can be calculated according to [4].

C. ESTIMATION OVER THE CHANNEL

In this subsection, we fix the impulsive noise $\boldsymbol{o}$ and the symbol $\boldsymbol{x}$ and estimate the channel $\boldsymbol{h}$.

Define

$$\boldsymbol{X}_i = \begin{bmatrix} x_i(1) & 0 & \cdots & 0 \\ x_i(2) & x_i(1) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ x_i(L_\text{i}) & x_i(L_\text{i} - 1) & \cdots & x_i(L_\text{i} - L_\text{c} + 1) \end{bmatrix} \in \mathbb{C}^{L_\text{i} \times L_\text{c}}. \tag{44}$$

Dropping the irrelevant terms of $\boldsymbol{h}_i$ in the joint optimization problem (17) leads to the optimization subproblem at the i-th sub-block, i.e.,

$$\min_{\boldsymbol{h}_i, \alpha_i} \sigma_v^{-2} \|\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{h}_i - \boldsymbol{o}_i\|^2 + \sigma_w^{-2} \|\boldsymbol{h}_i - \alpha_i \boldsymbol{h}_{i-1}\|^2. \tag{45}$$

To ease the computation complexity of the optimization problem, $\boldsymbol{h}_i$ and α_i are optimized alternatively. Obviously, the optimization of $\boldsymbol{h}_i$ and α_i leads to least square solutions, i.e.,

$$\boldsymbol{h}_i = \left(\sigma_v^{-2} \boldsymbol{X}_i^\text{H} \boldsymbol{X}_i + \sigma_w^{-2} \mathbf{I}\right)^{-1} \cdot \left(\sigma_v^{-2} \boldsymbol{X}_i^\text{H} (\boldsymbol{y}_i - \boldsymbol{o}_i) + \sigma_w^{-2} \alpha_i \boldsymbol{h}_{i-1}\right), \tag{46}$$
$$\alpha_i = \frac{\boldsymbol{h}_{i-1}^\text{H} \boldsymbol{h}_i}{\|\boldsymbol{h}_{i-1}\|^2}. \tag{47}$$

where α_i in (46) is initialized by α_{i-1} when optimizing the channel $\boldsymbol{h}_i$ based on the assumption of slowly time-varying channel.

It should be noted that the noise variance σ_v^2 and σ_w^2 are assumed known for simplificity. σ_v^2 can be estimated based on training symbols. σ_w^2 can be empirically determined based on

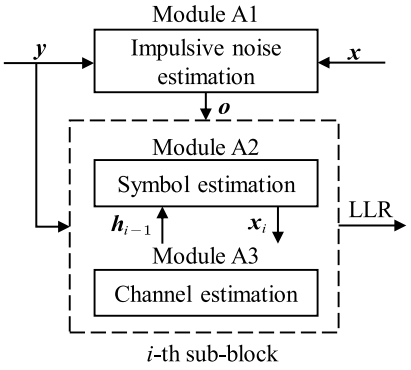

FIGURE 3. The block diagram of module A.

historical experimental data or be refined on-line according to the estimated channel.

Subsection III-A, III-B and III-C present the proposed algorithm of the joint estimation of the impulsive noise, the symbol and the channel, which correspond to module A in Fig. 2. To better illustrate the algorithm, we summarize it in Fig. 3. The impulsive noise estimation is performed with received signal $\boldsymbol{y}$ and decided symbol $\boldsymbol{x}$ from the decoder. The estimated impulsive noise $\boldsymbol{o}$ is feed to the symbol and channel estimation module which is performed sub-block by sub-block. The symbol estimation is performed with $\boldsymbol{y}$ and the estimated channel of the last sub-block $\boldsymbol{h}_{i-1}$. The estimate $\boldsymbol{x}_i$ is feed to the channel estimation module. When the whole block of the estimation of the symbol is completed, the LLR of the estimated symbol is calculated and feed to the soft decision decoder module.

D. DECODER

The decoder receives the soft information from module A. After soft decision decoding, the estimated symbols are refined and feed to module A. The iteration between module A and the decoder continues until error-free or the number of iterations reaches the maximum.

It should be noted that in turbo equalization (TEQ) structure, the soft output of the decoder is feed back to the symbol detection module as a prior to obtain better detection performance. However, in this paper, this potential has not been utilized and it will be further studied in the future.

IV. SIMULATION

In this section, the performance of the proposed JCDI algorithm is demonstrated numerically. The real and imaginary parts of the channel taps are drawn i.i.d. from Gaussian distribution and the length of the channel taps is 10. The QPSK signal undergoes the channel which is then corrupted by the additive white Gaussian noise and the impulsive noise. For comparison, the RLS-DFE receiver, the zero-forcing (ZF) receiver, the linear minimum mean square error (LMMSE) receiver and the joint channel estimation and data decoding (JCD) receiver are also implemented. The details of these algorithms are summarized as follows:

TABLE 3. Parameters setting of simulation.

Parameter	Value
Number of training symbols	511
Number of information bits	3840
LDPC code rate	$1/2$
LDPC symbol node degree	3
L_{c}	10
L_{i}	480
L_{b}	489
A	$1.1 \times \sqrt{2}$
λ_1	5
λ_2	0.01
ϵ	200
Number of S-FISTA iterations	200
Number of MM iterations	20
σ_w^2	0.1
Number of TEQ iterations	1
Number of decoding iterations	5
Number of MC trials	100

1) RLS-DFE: The decision feedback equalization (DFE) structure is adopted where the backward filter is used to cancel the causal ISI. The length of the forward filter and the backward filter are 10 and 9, respectively. The decision delay is 9. The forward filter and the backward filter are updated by the RLS algorithm. The forgetting factor is 0.999. The forward filter, the backward filters and the data symbol are updated symbol by symbol. The soft information of the decided symbols output by the RLS-DFE is feed to the LDPC decoder, whose output is then feed into the RLS-DFE to achieve turbo iteration.
2) ZF/LMMSE: The ZF/LMMSE is almost the same as that of JCDI except that the symbol estimation submodule of the MM algorithm is replaced by the ZF/LMMSE submodule. Note that the ZF/LMMSE has closed form solution for symbol estimation and takes the impulsive noise into consideration.
3) JCD: The JCD is similar to the JCDI except that the impulsive noise estimation submodule is excluded. Thus it could be expected that the JCD works more poorly than the JCDI when there is the impulsive noise.

The parameters setting are shown in Table 3. The irregular LDPC code is adopted whose check matrix is generated by progressive edge growth algorithm. Belief propagation algorithm is adopted for decoding. The size of parity check matrix is 3840×7680. The spectral efficiency is

$$\frac{7680 - 3840}{511 + 3840} = 0.883\ b/s/Hz. \tag{48}$$

Time-invariant channel is adopted in simulation.

The impulsive noise is generated based on mixture Gaussian model [23]. To illustrate the impact of the impulsive noise intuitively, an example of the received signal with and without impulsive noise is given in Fig. 4 and 5, respectively, where the signal-to-impulsive-noise ratio (SINR) is -15 dB. Here the SINR is defined as the ratio of the signal variance to the impulsive noise variance. The amplitude of the impulsive

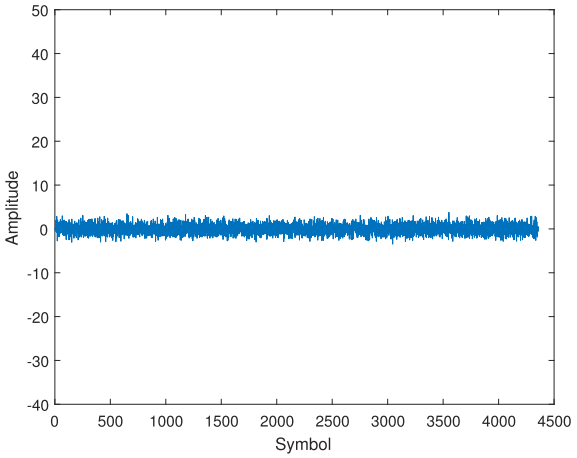

FIGURE 4. An example of the real part of the simulated received signal without impulsive noise.

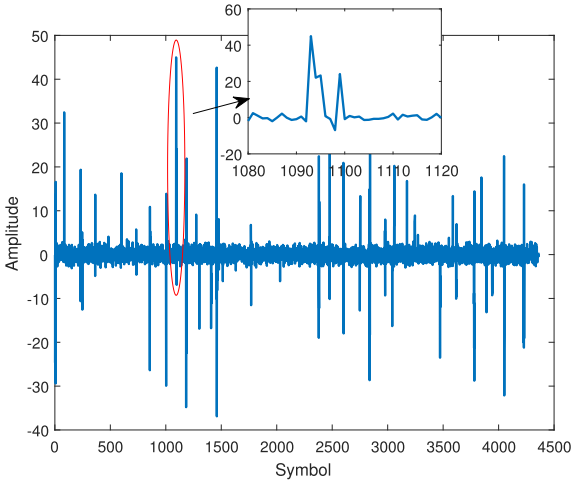

FIGURE 5. An example of the real part of the simulated received signal with impulsive noise. The SINR is −15 dB.

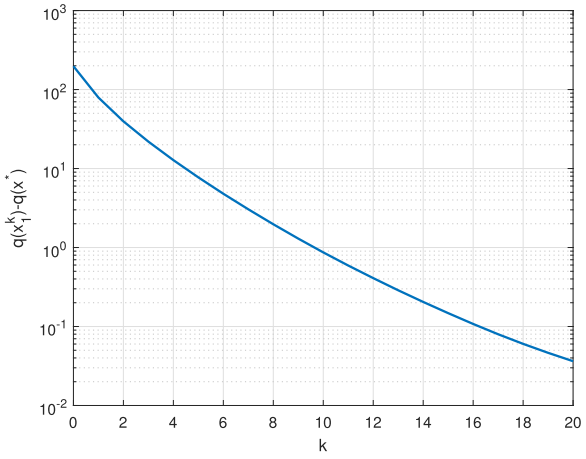

FIGURE 6. Error $q\left(x_i^k\right) - q\left(x^\star\right)$ versus iteration k of the MM algorithm for symbol estimation. The optimal solution $x^\star$ is obtained by CVX software.

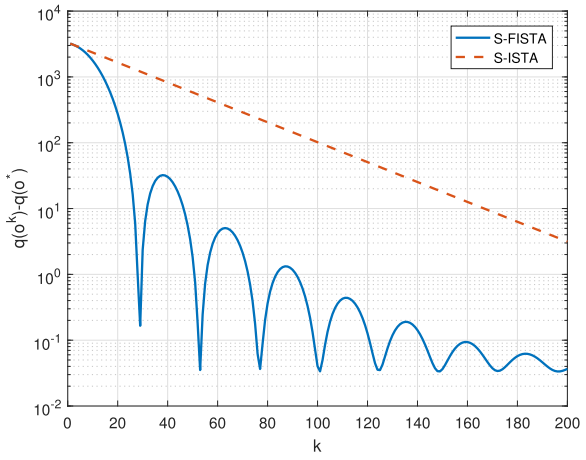

FIGURE 7. Error $q\left(o^k\right) - q\left(o^\star\right)$ versus iteration k of the S-FISTA and the S-ISTA for impulsive noise estimation. The optimal solution $o^\star$ is obtained by CVX software.

noise is significantly larger than that of the communication signals, and the performance of receivers degrades significantly without taking impulsive noise into account. The impulsive noise is distributed in clusters, demonstrating the group sparse property of the impulsive noise.

First, the convergence of the MM algorithm and the proposed S-FISTA algorithm are evaluated by simulation. Fig. 6 and 7 give a typical realization of the error $q(\boldsymbol{x}_i^k) - q(\boldsymbol{x}_i^\star)$ and $q(\boldsymbol{o}^k) - q(\boldsymbol{o}^\star)$ versus the iteration k of the MM algorithm and the proposed S-FISTA algorithm, where the optimal value $q(\boldsymbol{x}_i^\star)$ and $q(\boldsymbol{o}^\star)$ are obtained via the CVX software [28]. The S-FISTA converges faster than the S-ISTA. For the S-FISTA algorithm, the convergence speed is controlled by adaptive step size determined by convergence error ϵ. More convergence error allowed leads to faster convergence.

The normalized mean square error (NMSE) of channel estimation is evaluated. The SINR is set to 0 dB. The channel is randomly generated and fixed for different Monte Carlo (MC) trials. Fig. 8 shows the NMSE of the channel estimation of the proposed JCDI and JCD receiver. It can be seen that the NMSE of JCD decreases as SNR increases, and saturates about −33 dB. The NMSE of JCDI decreases almost linearly as SNR increases and achieves about −46 dB when $SNR = 24$ dB. This demonstrates that incorporating the impulsive noise estimation submodule significantly improves the channel estimation performance.

The coded BER performance of the proposed JCDI is evaluated and results are shown in Fig. 9. Let the SINR be 0 dB. All of the receivers listed are based on TEQ structure. The proposed JCDI receiver has best BER performance, followed by the JCD, LMMSE, RLS-DFE and ZF. Compared to JCD, JCDI estimates the impulsive noise, and achieves the lower coded BER by removing the impulsive noise. Although ZF and LMMSE incorporates the impulsive noise estimation submodule, their coded BERs are high, which are even higher than that of the JCD receiver. The reason is that all these receivers ignore the amplitude constraint of the symbol, leading to poor estimates of the symbols. The symbols are hard to

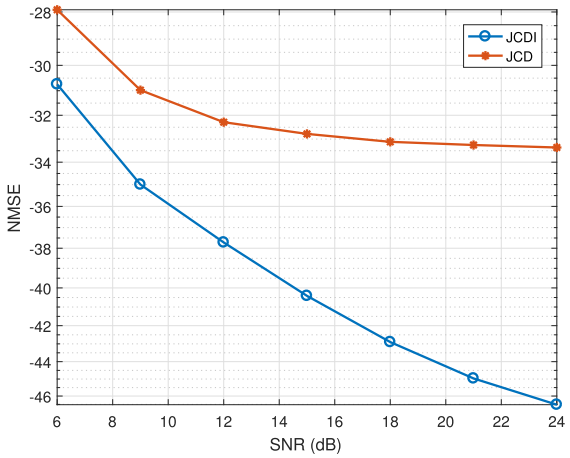

FIGURE 8. The NMSE of channel estimation with the SINR of 0dB. The blue line with cycles denotes the JCDI algorithm where the impulsive noise is considered and the red line with stars denotes the JCD algorithm where the impulsive noise is ignored.

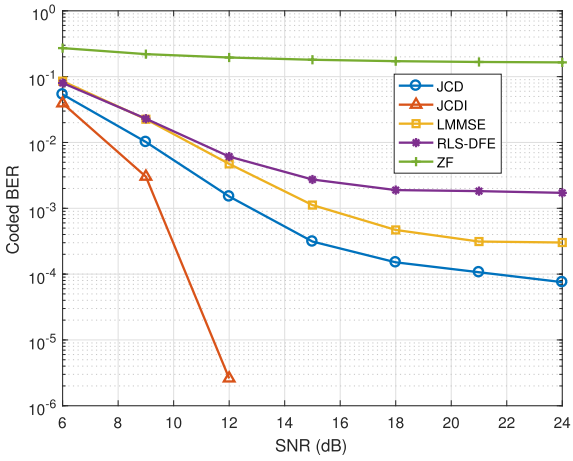

FIGURE 9. The coded BER curve under impulsive noise with the SINR of 0 dB. The blue line with cycle represents the JCD receiver, the red line with triagnle represents the JCDI receiver, yellow line with square represents the LMMSE receiver, the purple line with star represents the RLS-DFE receiver and the green line with plus sign represents the ZF receiver.

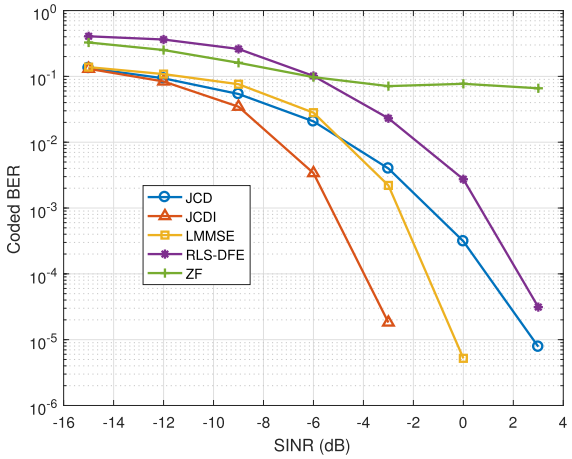

FIGURE 10. The coded BER of the algorithms versus SINR under the SNR of 15 dB. The blue line with cycle represents the JCD receiver, the red line with triagnle represents the JCDI receiver, the yellow line with square represents the LMMSE receiver, the purple line with star represents the RLS-DFE receiver and the green line with plus sign represents the ZF receiver.

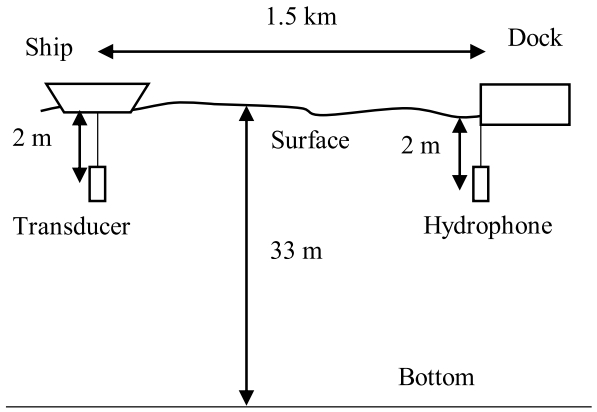

FIGURE 11. The experimental disployment schematic diagram.

be correctly decoded via the LDPC decoder given those poor LLRs of the symbol estimates. The coded BER performance of all the receivers except the JCDI receiver could not improve anymore with the increase of the SNR. This is due to that the impulsive noise is ignored or not perfectly estimated.

Fig. 10 presents the coded BER performance versus the SINR. JCDI performs best, followed by LMMSE, JCD, RLS-DFE and ZF. As the SINR increases, the coded BERs of all algorithms decrease. The results demonstrate that JCDI effectively cancels the effects of the impulsive noise and achieves the best coded BER performance.

V. EXPERIMENT

In this section, underwater acoustic communication experimental data is adopted to verify the proposed receiver.

The experiment was conducted on September 23, 2023 in the Zhoushan Sea, China. The experimental deployment is shown in Fig. 11 where the location between the transmitter and the receiver was about 1.5 km. The water depth was about 33 m. The depth of the transducer and the hydrophone both were about 2 m. The QPSK single carrier signal was transmitted. The information bits are encoded by the irregular LDPC code which is decoded by belief propagation algorithm. The pulse shaping filter adopts a square root raised cosine (SRRC) filter. The signal with bandwidth of 3.6 kHz was transmitted simultaneously over three bands with the carrier frequency of 5.5 kHz, 10 kHz and 15 kHz, respectively. The size of parity check matrix is 3840×7680. The data rate is

$$\frac{7680 - 3840}{(511 + 3840)\frac{1+0.2}{3600}} = 2.648\ kb/s. \tag{49}$$

The rest parameters of the signal and the receiver is listed in Table 4 where the parameters of the JCDI and RLS-DFE algorithms are carefully selected for better performance.

TABLE 4. Signal parameters of experiment.

Parameter	Value
Number of training symbols	511
Number of information bits	3840
LDPC code rate	$1/2$
Roll-off factor of SRRC filter	0.2
Span of SRRC filter	40
Sampling frequency	96 kHz
L_{i}	480
L_{b}	499
A	$1.4 \times \sqrt{2}$
λ_1	10
λ_2	1
ϵ	1000
Number of S-FISTA iterations	200
Number of MM iterations	80
σ_w^2	0.001
Forward filter length	50
Backward filter length	5
Decision delay	9
Forgetting factor	0.999
Initial matrix of the RLS algorithm	$10\mathbf{I}$
Proportional factor of PLL	0.001
Integral tracking factor of PLL	0.0001
Number of TEQ iterations	1
Number of decoding iterations	5
Number of MC trials	100

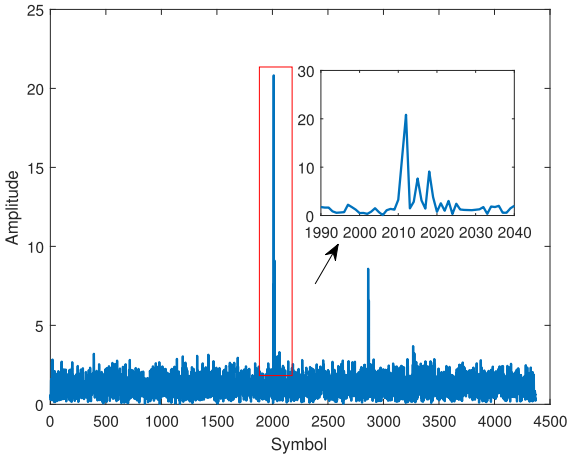

FIGURE 12. The baseband signal of the first packet which is interfered by group sparse impulsive noise.

To combat time-varying channels, the phase-lock loop is adopted in the RLS-DFE receiver.

The impulsive noise is observed in one of the received packets, as shown in Fig. 12 where the sparse and group sparse impulsive noise can be observed.

The coded BER results of the experiment data are listed in Table 5. It shows that the proposed JCDI receiver has better coded BER performance than that of the RLS-DFE receiver. The bit error rate of RLS-DFE decreases as the number of TEQ iterations increases from zero to 1. The BER is non-decreasing as we continue to increase the number of TEQ iterations.

TABLE 5. BER of the experiment data using the proposed JCDI iterative receiver.

Packet	TEQ iteration	Coded BER	
		JCDI	RLS-DFE
1	0	0	0.002865
	1	0	0.001823
	2	0	0.003646
2	0	0	0.000521
	1	0	0.000260
	2	0	0.000260
3	0	0	0
	1	0	0
	2	0	0

VI. CONCLUSION

In this paper, an iterative receiver named JCDI taking impulsive noise into account is proposed. The receiver consists of two main modules, where the first module performs symbol estimation, channel estimation and impulsive noise estimation, and the second module performs the data decoding which utilizes the encoding structure. In particular, an efficient S-FISTA algorithm is proposed to estimate the group-sparse impulsive noise. Numerical simulations and real experiments are conducted to show that the proposed JCDI receiver has better coded BER performance than conventional receiver.

APPENDIX A
THE FUNDAMENTAL DEFINITIONS AND THEOREMS OF DEVELOPING THE SMOOTHED FISTA

To develop the smoothed FISTA, we introduce the following definitions and theorems [27] to make this paper self-contained.

Definition 1: [27, Chapter 5, p. 107] L-smoothness: Let $L \geq 0$. A function $f : \mathbb{E} \to (-\infty, \infty]$ is said to be L-smooth over a set $D \subseteq \mathbb{E}$ if it is differentiable over D and satisfies

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_\star \leq L\|\boldsymbol{x} - \boldsymbol{y}\| \text{ for all } \boldsymbol{x}, \boldsymbol{y} \in D, \tag{50}$$

where $\|\cdot\|_\star$ is the dual norm. The constant L is called the smoothness parameter.

Definition 2: [27, Chapter 10, p. 305] Smoothable functions: A convex function $h : \mathbb{E} \to \mathbb{R}$ is called (α, β)-smoothable ($\alpha, \beta > 0$) if for any $\mu > 0$ there exists a convex differentiable function $h_\mu : \mathbb{E} \to \mathbb{R}$ such that the following holds:

(a) $h_\mu(\boldsymbol{x}) \leq h(\boldsymbol{x}) \leq h_\mu(\boldsymbol{x}) + \beta\mu$ for all $\boldsymbol{x} \in \mathbb{E}$.

(b) h_μ is $\frac{\alpha}{\beta}$-smooth.

The function h_μ is called a $\frac{1}{\mu}$-smooth approximation of h with parameters (α, β).

Theorem 1: [27, Chapter 10, p. 311] Let $h^1, h^2 : \mathbb{E} \to \mathbb{R}$ be convex functions, and let γ_1, γ_2 be nonnegative numbers. Suppose that for a given $\mu > 0$, h_μ^i is a $\frac{1}{\mu}$-smooth approximation of h_i with parameters (α_i, β_i) for $i = 1, 2$. Then $\gamma_1 h_\mu^1 + \gamma_2 h_\mu^2$ is a $\frac{1}{\mu}$-smooth approximation of $\gamma_1 h^1 + \gamma_2 h^2$ with parameters $(\gamma_1\alpha_1 + \gamma_2\alpha_2, \gamma_1\beta_1 + \gamma_2\beta_2)$.

Theorem 2: [27, Chapter 10, p. 311] Let $\mathcal{A}:\mathbb{E}\to\mathbb{V}$ be a linear transformation between the Euclidean spaces $\mathbb{E}$ and $\mathbb{V}$. Let $h:\mathbb{V}\to\mathbb{R}$ be a convex function and define $q(\boldsymbol{x})=h(\mathcal{A}(\boldsymbol{x})+\boldsymbol{b})$ where $\boldsymbol{b}\in\mathbb{V}$. Suppose that for a given $\mu>0$, h_μ is a $\frac{1}{\mu}$-smooth approximation of h with parameters (α,β). Then the function $q_\mu(\boldsymbol{x})=h_\mu(\mathcal{A}(\boldsymbol{x})+\boldsymbol{b})$ is a $\frac{1}{\mu}$-smooth approximation of q with parameters $(\alpha\|\mathcal{A}\|^2,\beta)$.

APPENDIX B
THE DERIVATION OF (36)

According to the definition of the proximal minimization and the definition of $g(\check{\boldsymbol{o}})$, i.e., $g(\check{\boldsymbol{o}})=\lambda_1\sum_{n=1}^{L_y}\left\|\check{\boldsymbol{o}}(2n-1:2n)\right\|$, the proximal operator (31) is expanded to

$$\check{\boldsymbol{o}}^{k+1}=\arg\min_{\check{\boldsymbol{o}}^{k+1}}\left\{\frac{1}{\tilde{L}}\lambda_1\sum_{n=1}^{L_y}\left\|\check{\boldsymbol{o}}^{k+1}(2n-1:2n)\right\|+\frac{1}{2}\left\|\check{\boldsymbol{o}}^{k+1}-\left(\check{\boldsymbol{o}}^{k}-\frac{1}{\tilde{L}}\nabla F_\mu\left(\check{\boldsymbol{o}}^{k}\right)\right)\right\|^2\right\}. \tag{51}$$

For ease of derivation, complex variable $o^{k+1}(n)$ defined in (22) is adopted. Define

$$\boldsymbol{a}=\begin{bmatrix}\check{\boldsymbol{a}}(1)+j\check{\boldsymbol{a}}(2)\\ \check{\boldsymbol{a}}(3)+j\check{\boldsymbol{a}}(4)\\ \vdots\\ \check{\boldsymbol{a}}(2L_y-1)+j\check{\boldsymbol{a}}(2L_y)\end{bmatrix}\in\mathbb{C}^{L_y}, \tag{52}$$

where $\check{\boldsymbol{a}}$ is defined in (36). Substituting (22), (36) and (52) into (51) leads to

$$\boldsymbol{o}^{k+1}=\arg\min_{\boldsymbol{o}^{k+1}}\times\left\{\frac{1}{\tilde{L}}\lambda_1\sum_{n=1}^{L_y}\left\|o^{k+1}(n)\right\|+\frac{1}{2}\left\|\boldsymbol{o}^{k+1}-\boldsymbol{a}\right\|^2\right\}. \tag{53}$$

Note that the optimization problem (53) is separable w.r.t. the impulsive noise $o^{k+1}(1),o^{k+1}(2),\cdots,o^{k+1}(L_y)$. Therefore, the optimization problem (53) is simplified to

$$o^{k+1}(n)=\arg\min_{o^{k+1}(n)}\times\left\{\frac{1}{\tilde{L}}\lambda_1\left|o^{k+1}(n)\right|+\frac{1}{2}\left|o^{k+1}(n)-a(n)\right|^2\right\}, \tag{54}$$

for every n. Define the objective function in (54) as $J\left(o^{k+1}(n)\right)$ which can be formulated as

$$\begin{aligned}J\left(o^{k+1}(n)\right)&\propto\frac{1}{\tilde{L}}\lambda_1\left|o^{k+1}(n)\right|+\frac{1}{2}\left|o^{k+1}(n)\right|^2-\Re\left\{o^{k+1}(n)a^*(n)\right\}\\&=\frac{1}{2}\left|o^{k+1}(n)\right|^2+\left(\frac{1}{\tilde{L}}\lambda_1-|a(n)|\cos\right.\\&\quad\left.\times\left(\angle o^{k+1}(n)-\angle a(n)\right)\right)\left|o^{k+1}(n)\right|,\end{aligned} \tag{55}$$

which is a quadratic function of $\left|o^{k+1}(n)\right|$. Therefore, if $|a(n)|\cos\left(\angle o^{k+1}(n)-\angle a(n)\right)>\frac{1}{\tilde{L}}\lambda_1$, the optimal $\left|o^{k+1}(n)\right|$ and $\angle o^{k+1}(n)$ minimizing $J\left(o^{k+1}(n)\right)$ is

$$\left|o^{k+1}(n)\right|=|a(n)|\cos\left(\angle o^{k+1}(n)-\angle a(n)\right)-\frac{1}{\tilde{L}}\lambda_1, \tag{56}$$

$$\angle o^{k+1}(n)=\angle a(n). \tag{57}$$

Otherwise, $\left|o^{k+1}(n)\right|=0$.

APPENDIX C
THE DERIVATION OF THE MM ALGORITHM FOR SOLVING (41)

Define

$$\boldsymbol{y}_{\mathrm{r}}=\boldsymbol{y}_i-\boldsymbol{H}_{i-1,\mathrm{tail}}\boldsymbol{x}_{i-1,\mathrm{tail}}-\boldsymbol{o}_i. \tag{58}$$

Ingoring the constant term σ_v^{-2} which does not affect the optimization result, introduce the initial point $\boldsymbol{x}_i^k$ where the superscript denotes the k-th iteration and expand the objective function in (40)

$$\begin{aligned}&\|\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i\|^2\\&=\|\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k-\boldsymbol{H}_{i-1}\left(\boldsymbol{x}_i-\boldsymbol{x}_i^k\right)\|^2\\&=\left(\boldsymbol{x}_i-\boldsymbol{x}_i^k\right)^{\mathrm{H}}(\boldsymbol{H}_{i-1})^{\mathrm{H}}\boldsymbol{H}_{i-1}\left(\boldsymbol{x}_i-\boldsymbol{x}_i^k\right)\\&\quad-2\Re\left\{\left(\boldsymbol{x}-\boldsymbol{x}_i^k\right)^{\mathrm{H}}(\boldsymbol{H}_{i-1})^{\mathrm{H}}\left(\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k\right)\right\}\\&\quad+\left\|\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k\right\|^2.\end{aligned} \tag{59}$$

Define the maximum eigenvalue of $\left(\boldsymbol{H}_{i-1}^t\right)^{\mathrm{H}}\boldsymbol{H}_{i-1}^t$ as $\lambda_{\max}$. The surrogate function of (59) can be constructed as

$$\begin{aligned}&\|\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i\|^2\\&\le\lambda_{\max}\|\boldsymbol{x}_i-\boldsymbol{x}_i^k\|^2\\&\quad-2\Re\left\{\left(\boldsymbol{x}_i-\boldsymbol{x}_i^k\right)^{\mathrm{H}}(\boldsymbol{H}_{i-1})^{\mathrm{H}}\left(\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k\right)\right\}\\&\quad+\left\|\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k\right\|^2\\&=\lambda_{\max}\|\boldsymbol{x}_i\|^2-2\lambda_{\max}\Re\left\{\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{x}_i^k\right\}\\&\quad-2\Re\left\{\boldsymbol{x}_i^{\mathrm{H}}(\boldsymbol{H}_{i-1})^{\mathrm{H}}\left(\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k\right)\right\}+const\\&=\lambda_{\max}\|\boldsymbol{x}_i-\boldsymbol{x}_{\mathrm{o}}^k\|^2+const\end{aligned} \tag{60}$$

where

$$\boldsymbol{x}_{\mathrm{o}}^k=\boldsymbol{x}_i^k+\frac{1}{\lambda_{\max}}\left(\boldsymbol{y}_{\mathrm{r}}-\boldsymbol{H}_{i-1}\boldsymbol{x}_i^k\right). \tag{61}$$

In this way, it becomes easier to optimize the new objective function since every variable of vector $\boldsymbol{x}_i^t$ can be optimized individually. Apparently, the optimal solution at the k-th iteration is $\boldsymbol{x}_i=\boldsymbol{x}_{\mathrm{o}}^k$ if the solution is within the constraint. If not, it is not hard to imagine that the optimal solution will

be the point on the boundary of the constraint which is closest to $\boldsymbol{x}_{\mathrm{o}}^{k}$.

Take the constraint

$$\begin{aligned} &|\Re\{x_i(n)\}| \leq A, \\ &|\Im\{x_i(n)\}| \leq A, n = 1, 2, \cdots \end{aligned} \tag{62}$$

as an example where $x_i(n)$ is the n-th element of vector $\boldsymbol{x}_i$. The optimal solution minimizing the approximated objective function indicated by (60) under the constraint (62) is

$$\begin{aligned} x_i^{k+1}(n) = &\operatorname{sign}\left\{\Re\left\{x_{\mathrm{o}}^{k}(n)\right\}\right\} \min\left\{A, \Re\left\{x_{\mathrm{o}}^{k}(n)\right\}\right\} \\ &+ j \cdot \operatorname{sign}\left\{\Im\left\{x_{\mathrm{o}}^{k}(n)\right\}\right\} \min\left\{A, \Im\left\{x_{\mathrm{o}}^{k}(n)\right\}\right\} \end{aligned} \tag{63}$$

where $x_{\mathrm{o}}(n)$ is the n-th element of vector $\boldsymbol{x}_{\mathrm{o}}$.

REFERENCES

[1] J. Zhou, Y. Si, and Y. Chen, "A review of subsea AUV technology," *J. Mar. Sci. Eng.*, vol. 11, no. 6, p. 1119, May 2023.

[2] L. Wan, H. Jia, F. Zhou, M. Muzzammil, T. Li, and Y. Huang, "Fine Doppler scale estimations for an underwater acoustic CP-OFDM system," *Signal Process.*, vol. 170, May 2020, Art. no. 107439.

[3] L. Wan, H. Zhou, X. Xu, Y. Huang, S. Zhou, Z. Shi, and J.-H. Cui, "Adaptive modulation and coding for underwater acoustic OFDM," *IEEE J. Ocean. Eng.*, vol. 40, no. 2, pp. 327–336, Apr. 2015.

[4] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[5] X. Tu, X. Xu, and A. Song, "Frequency-domain decision feedback equalization for single-carrier transmissions in fast time-varying underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 46, no. 2, pp. 704–716, Apr. 2021.

[6] Q. Guo, D. Huang, S. Nordholm, J. Xi, and Y. Yu, "Iterative frequency domain equalization with generalized approximate message passing," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 559–562, Jun. 2013.

[7] M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Phase-coherent digital communications for underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 19, no. 1, pp. 100–111, Jan. 1994.

[8] Y. Tsuda and T. Shimamura, "An improved NLMS algorithm for channel equalization," in *Proc. ISCAS*, Phoenix-Scottsdale, AZ, USA, 2002, pp. 1–4.

[9] J. Xi, S. Yan, L. Xu, and C. Hou, "Sparsity-aware adaptive turbo equalization for underwater acoustic communications in the Mariana trench," *IEEE J. Ocean. Eng.*, vol. 46, no. 1, pp. 338–351, Jan. 2021.

[10] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.

[11] L. Wan, J. Zhu, E. Cheng, and Z. Xu, "Joint CFO, gridless channel estimation and data detection for underwater acoustic OFDM systems," *IEEE J. Ocean. Eng.*, vol. 47, no. 4, pp. 1215–1230, Oct. 2022.

[12] G. Yang, Q. Guo, H. Ding, Q. Yan, and D. D. Huang, "Joint message-passing-based bidirectional channel estimation and equalization with superimposed training for underwater acoustic communications," *IEEE J. Ocean. Eng.*, vol. 46, no. 4, pp. 1463–1476, Oct. 2021.

[13] Q. Guo, L. Ping, and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4340–4349, Aug. 2009.

[14] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Efficient estimation and prediction for sparse time-varying underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 1112–1125, Jul. 2020.

[15] C. M. Anil, "Underwater acoustic communications in warm shallow water channels," Ph.D. dissertation, Dept. Elect. Comput. Eng., NUS, Singapore, 2006.

[16] X. Tu, X. Xu, and A. Song, "Statistical analysis and hybrid modeling of high-frequency underwater acoustic channels affected by wind-driven surface waves," *J. Acoust. Soc. Amer.*, vol. 151, no. 5, pp. 3266–3279, May 2022.

[17] P. Gendron and B. Nandram, "Modeling heavy-tailed correlated noise with wavelet packet basis functions," *J. Stat. Planning Inference*, vol. 112, nos. 1–2, pp. 99–114, Mar. 2003.

[18] Y. Hou, R. Liu, B. Dai, and L. Zhao, "Joint channel estimation and LDPC decoding over time-varying impulsive noise channels," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2376–2383, Jun. 2018.

[19] D. Li, Y. Wu, M. Zhu, and J. Tao, "An enhanced iterative receiver based on vector approximate message passing for deep-sea vertical underwater acoustic communications," *J. Acoust. Soc. Amer.*, vol. 149, no. 3, pp. 1549–1558, Mar. 2021.

[20] J. Wolf, "Redundancy, the discrete Fourier transform, and impulse noise cancellation," *IEEE Trans. Commun.*, vol. COM-31, no. 3, pp. 458–461, Mar. 1983.

[21] F. Abdelkefi, P. Duhamel, and F. Alberge, "Impulsive noise cancellation in multicarrier transmission," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 94–106, Jun. 2005.

[22] F. Abdelkefi, P. Duhamel, and F. Alberge, "A necessary condition on the location of pilot tones for maximizing the correction capacity in OFDM systems," *IEEE Trans. Commun.*, vol. 55, no. 2, pp. 356–366, Feb. 2007.

[23] M. Nassar, P. Schniter, and B. L. Evans, "A factor graph approach to joint OFDM channel estimation and decoding in impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1576–1589, Mar. 2014.

[24] S. V. Zhidkov, "Analysis and comparison of several simple impulsive noise mitigation schemes for OFDM receivers," *IEEE Trans. Commun.*, vol. 56, no. 1, pp. 5–9, Jan. 2008.

[25] Y.-N. Tian, X. Han, J.-W. Yin, Q.-Y. Liu, and L. Li, "Group sparse underwater acoustic channel estimation with impulsive noise: Simulation results based on Arctic ice cracking noise," *J. Acoust. Soc. Amer.*, vol. 146, no. 4, pp. 2482–2491, Oct. 2019.

[26] J. Lin, M. Nassar, and B. L. Evans, "Impulsive noise mitigation in powerline communications using sparse Bayesian learning," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1172–1183, Jul. 2013.

[27] A. Beck, *First-order Methods in Optimization*. Philadelphia, PA, USA: SIAM, 2017.

[28] M. Grant and S. Boyd. *CVX: MATLAB Software for Disciplined Convex Programming*. Accessed: Feb. 16, 2024. [Online]. Available: http://cvxr.com/cvx/

MINHAO ZHANG received the B.S. degree in ocean engineering and technology from Zhejiang University, Hangzhou, China, in 2016. He is currently pursuing the Ph.D. degree with the Ocean College, Zhejiang University, Zhoushan, China.

His research interest includes underwater acoustic communications.

JIANG ZHU (Member, IEEE) received the B.E. degree in electronic engineering from Harbin Engineering University, Harbin, China, in 2011, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2016.

He was a Visiting Student with Lehigh University, in 2015. He joined the Ocean College, Zhejiang University, in 2016, where he is currently an Associate Professor. His current research interests include statistical signal processing and compressed sensing.

YIFAN WANG received the B.S. degree in underwater acoustic engineering from Harbin Engineering University, Harbin, China, in 2021. He is currently pursuing the Ph.D. degree with the Ocean College, Zhejiang University, Zhoushan, China.

His research interests include underwater acoustic communication and direction of arrival estimation.

YANBING FU received the B.S. degree from the College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, in 2023. He is currently pursuing the M.S. degree in electronic information engineering with the Ocean College, Zhejiang University, Zhoushan, China.

His research interests include underwater acoustic communications and signal processing.

YAN WEI (Member, IEEE) received the B.Sc. and M.Sc. degrees in naval architecture and ocean engineering from Wuhan University of Technology, Wuhan, China, in 2004 and 2007, respectively, and the Ph.D. degree from the Department of Maritime and Transportation Technology, Delft University, Delft, The Netherlands, in 2012.

Since 2012, she has been with the Ocean College, Zhejiang University, where she is currently an Associate Professor. Her current research interests include the flow of noise and its influence on underwater acoustic communications, ocean observatory networking, and cross-medium communication.

XINGBIN TU (Member, IEEE) received the B.S. degree in marine technology and the Ph.D. degree in marine physics from the College of Ocean and Earth Science, Xiamen University, Xiamen, China, in 2012 and 2020, respectively.

In 2014, he enrolled from the master's program into the doctoral program. From 2016 to 2018, he was a Visiting Scholar with The University of Alabama, Tuscaloosa, AL, USA. From 2020 to 2022, he was a Postdoctoral Fellow with the Ocean College, Zhejiang University, Zhoushan, China, where he has been a Research Associate with the Institute of Ocean Sensing and Networks, since 2022. His research interests include underwater acoustic communications and signal processing.

Dr. Tu has been a member of the Technical Program Committee for the International Workshop/Conference on Underwater Networks and Systems (WUWNet), since 2021.

FENGZHONG QU (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2002 and 2005, respectively, and the Ph.D. degree in electrical and computer engineering from the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, in 2009.

From 2009 to 2010, he was an Adjunct Research Scholar with the Department of Electrical and Computer Engineering, University of Florida. Since 2011, he has been with the Ocean College, Zhejiang University, Zhoushan, China, where he is currently a Professor and the Chair of the Institute of Ocean Sensing and Networking. His current research interests include underwater acoustic communication and networks, wireless communications, signal processing, and intelligent transportation systems.