Impulsive Noise Recovery and Elimination: A Sparse Machine Learning Based Approach

Sicong Liu, *Member, IEEE*, Liang Xiao, *Senior Member, IEEE*, Lianfen Huang, and Xianbin Wang, *Fellow, IEEE*

***Abstract*—The performance of orthogonal frequency division multiplexing (OFDM) based wireless vehicular communication systems is faced with the great challenge of impulsive noise (IN), which could limit the application of OFDM in ultra-reliable low-latency communication scenarios. In this paper, the challenge of IN elimination for OFDM-based wireless systems is efficiently overcome by the proposed sparse learning algorithms and probabilistic framework inspired by the emerging machine learning theories. For the first time, the sparse machine learning theory is introduced to IN recovery and elimination. Exploiting the measurement vector of IN observed from the reserved null sub-carriers as the input, a novel sparse machine learning based algorithm of sparse cross-entropy minimization is proposed, in which the probability distribution of the IN support is iteratively updated by minimizing the loss function, i.e. the cross-entropy. The proposed algorithm is able to effectively and efficiently learn the sparse pattern and converge to the accurate distribution of IN support. To facilitate an accelerated and even more efficient learning process, regularization is imposed on the loss function by adding a weighting parameter in favor of the accurate distribution. The computer simulation results confirm that the proposed scheme outperforms conventional methods while utilizing fewer spectrum resources over wireless vehicular channels.**

***Index Terms*—Impulsive noise, sparse machine learning, cross-entropy minimization, orthogonal frequency division multiplexing.**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) is one of the most popular broadband technologies currently deployed due to its high spectral efficiency and robustness in frequency-selective multipath channels. Due to these advantages, OFDM has been widely applied in various wireless or wired vehicular related communications systems to support high-speed and broadband transmission, including the wireless access in vehicular environments specified in the IEEE 802.11p Wireless Access in Vehicular Environments (WAVE) standard [1].

Apart from the advantages, an important detrimental characteristic of OFDM-based vehicular communications systems is its vulnerability to the impulsive noise (IN), which is both non-Gaussian and non-stationary and prevailing stochastically in the wireless or wired vehicular channels [2], [3]. In many scenarios, IN could be generated by ignition in vehicles, the switches of electric devices, and strong bursty radio frequency emission, which might cause transmission errors and loss of the data frames of interest [4]. For instance, ignition sparks and engine rotation in vehicles might generate impulsive noise, which is a very common RF contamination in vehicular channels [5], [6]. Due to its impulse nature in time domain, the spectrum of IN can be extremely wide. It is therefore very difficult to separate or eliminate IN from the OFDM data block because all the OFDM sub-carriers are contaminated by the time-domain bursty IN, especially in the presence of intensive IN with high power [7]. The ultra-wide coherence bandwidth of IN in frequency domain and its similar statistics to the OFDM signal in time domain make the separation of IN and OFDM signals extremely difficult. This could cause even severe degradation to the OFDM-based vehicular communications performance in the form of increased symbol error rate/bit error rate and increased transmission latency in case of packet loss [8], [9]. Hence, it is crucial and essential to mitigate IN to support ultra reliable low latency communication (URLLC) services and vehicular ad hoc networks (VANETs) in the next-generation vehicular networking and communications scenarios.

There have been some conventional methods aiming at mitigating the IN impacts. A heavily investigated approach is applying nonlinear operations to the receiver, including clipping, blanking and their combination, to suppress or null out the received samples with the power beyond some threshold [10]. However, these conventional methods tried to suppress the noise power, nonlinearly exclude or clip the noise contaminated signal instead of estimating and eliminating the IN, so useful data might be lost, especially for IN with large power.

Recently, the sparse recovery methods based on the theory of compressed sensing (CS) [11], [12], have been investigated to estimate the IN [13]–[15]. In the related studies, the IN is regarded as a sparse signal in the time-domain due to its bursty nature, and estimated using CS-based algorithms, such as the sparse convex optimization (SCO) algorithm in [14] and

Manuscript received September 7, 2018; revised November 20, 2018; accepted December 15, 2018. Date of publication January 9, 2019; date of current version March 14, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61871339, 61671396, 61731012, and 91638204, in part by the Key Laboratory of Digital Fujian on IoT Communication, Architecture and Safety Technology under Grant 2010499, and in part by the Open Research Fund of National Mobile Communications Research Laboratory Southeast University under Grant 2018D08. The review of this paper was coordinated by Prof. G. Gui. *(Corresponding author: Sicong Liu.)*

S. Liu, L. Xiao, and L. Huang are with the Department of Communication Engineering and the Key Laboratory of Digital Fujian on IoT Communication, Architecture and Security Technology, Xiamen University, Xiamen 361005, China (e-mail: liusc@xmu.edu.cn; adalittlel@gmail.com; lfhuang@xmu.edu.cn).

X. Wang is with the Department of Electrical and Computer Engineering, University of Western Ontario, London, ON N6A 5B9, Canada (e-mail: xianbin.wang@uwo.ca).

Digital Object Identifier 10.1109/TVT.2019.2891617

the greedy algorithms of sparsity adaptive matching pursuit (SAMP) in [16] and priori aided SAMP (PA-SAMP) in [15]. Although superior performance over the conventional passive methods is achieved, the spectral efficiency of the CS-based methods could still be improved since many OFDM sub-carriers are reserved for measurements. Besides, large background noise and sparsity level also cause inaccuracy of IN estimation. Another sparse recovery theory of sparse Bayesian learning was utilized to estimate the asynchronous IN [17] or the narrowband internet-of-things (NB-IoT) interference [18], which exploited the sparsity of the noise and estimate it using Bayesian inference to improve the robustness against large background noise.

Different from the existing methods mentioned above, the emerging and powerful machine learning theory and techniques have drawn tremendous research attention recently [19]–[22]. It can be a great inspiration in order to find out a both efficient and reliable method of sparse recovery, or IN recovery [23]. In machine learning applications, utilizing cross-entropy (CE) as the loss function to train deep neural networks have solved many different learning tasks in various areas, such as communications [24] and network security [25], etc. Some recent studies exploited the CE method to solve combinatorial optimization problems, leading to superior performance than the brute-force approach [26]. However, existing CE methods are not designed for sparse approximation. Consequently, the state-of-art research on IN recovery based on sparse machine learning and cross-entropy is quite insufficient in literature. To fill this gap and improve the robustness, efficiency and accuracy of state-of-the-arts, in this paper, we formulated the sparse machine learning inspired probabilistic framework of IN recovery, and proposed the algorithm of sparse CE minimization (SCEM) to learn the locations of IN. By introducing and properly utilizing the sparse machine learning methodology in IN recovery, the unknown IN can be learnt efficiently and accurately in different severe conditions. The main contributions of this paper are twofold as follows:

- The sparse machine learning theory is introduced to the area of IN recovery for the first time. A novel probabilistic framework of sparse machine learning is formulated for IN recovery in OFDM-based vehicular communications, which improves the spectral efficiency and accuracy of the conventional and CS-based methods.
- A sparse machine learning inspired algorithm, i.e., SCEM, is proposed for IN recovery, which significantly develops the CE method to be capable of supporting sparse recovery. By imposing regularization on the loss function, an enhanced algorithm of regularized SCEM (RSCEM) is further developed to improve the convergence rate and recovery accuracy.

The rest of this paper is organized as follows: the statistical models of IN and the signal model in wireless vehicular communication are introduced in Section II. Section III presents the proposed probabilistic framework formulation and the proposed algorithms of sparse learning for IN recovery, constituting the main contribution of this paper. The simulation results are reported in Section IV with detailed discussions, followed by the conclusions.

Notation: Matrices and column vectors are denoted by boldface letters; $(\cdot)^{\dagger}$ and $(\cdot)^{H}$ denote the pseudo-inversion operation and conjugate transpose; $\|\cdot\|_r$ represents the ℓ_r-norm operation; $|\Pi|$ denotes the cardinality of the set Π; $\mathbf{v}|_{\Pi}$ denotes the entries of the vector $\mathbf{v}$ in the set of Π; Π^c denotes the complementary set of Π; $\mathbf{A}_{\Pi}$ represents the sub-matrix comprised of the Π columns of the matrix $\mathbf{A}$.

II. System Model

A. Statistical Model of IN

In the OFDM-based vehicular communication system contaminated by IN, the IN vector corresponding to the i-th OFDM symbol is denoted by $\boldsymbol{\xi}_i = [\xi_{i,0}, \xi_{i,1}, \ldots, \xi_{i,N-1}]^T$ of length N. Sparsity is the important and intrinsic property of IN, meaning that the number of nonzero entries is sufficiently small compared with that of the signal dimension. The support is defined as

$$\Pi = \{j \,|\xi_{i,j} \neq 0\,,\ j = 0, 1, \ldots, N-1\}, \tag{1}$$

where the sparsity level $K = |\Pi|$ denotes the number of nonzero entries. Without loss of generality, the support Π can be distributed randomly and all the entries might be zero or nonzero, so there is no a priori constraint of the support distribution. The statistical characteristics of the IN model, including the distribution of the instantaneous amplitude, as well as the probabilistic distribution of the arrival of IN, have already been empirically modeled in literature. Among the most commonly adopted models of instantaneous amplitude are the Gaussian mixture model [27] and the Middleton's Class A model [28].

The Gaussian mixture distribution are widely adopted for the time-domain asynchronous IN [27]. The probability density function (PDF) of the Gaussian mixture model is represented as

$$p_Z(z) = \sum_{j=1}^{J_{\mathrm{m}}} \beta_j \cdot g_j(z), \tag{2}$$

where J_{m} is the number of Gaussian components, β_j is the mixture coefficient of the Gaussian variable, and $g_j(z)$ is the PDF of a Gaussian variable with zero mean and variance of σ_j^2.

The Middleton's Class A model, which is also a typical statistical model of IN [28], can be generated by the special case of Gaussian mixture when the parameters β_j and σ_j^2 in (2) satisfy

$$\beta_j = e^{-A} A^j / j!, \tag{3}$$

$$\sigma_j^2 = (j/A + \omega)/(1 + \omega), \tag{4}$$

$$J_{\mathrm{m}} \to \infty, \tag{5}$$

where the parameters A and ω denote the overlapping factor and the background-to-impulsive-noise power ratio, respectively. The Middleton's Class A model is also a typical statistical model of IN [28], which is often adopted to describe the IN occurring in both wired and wireless channels.

Usually, the intensity of IN with respect to background noise is indicated by interference-to-noise ratio (INR) γ, which is

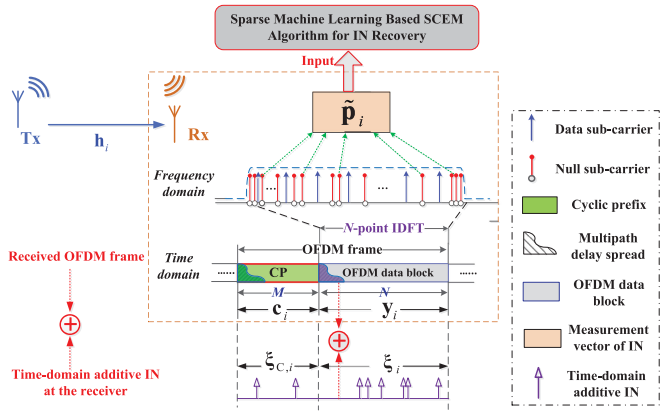


Fig. 1. Time-frequency OFDM frame structure exploited by the sparse machine learning based algorithms for IN recovery and cancelation in the wireless vehicular communication system.

represented by

$$\gamma = \frac{1}{\sigma^2}\mathbb{E}\left\{\frac{1}{K}\sum_{j\in\Pi}|\xi_{i,j}|^2\right\}, \tag{6}$$

where σ^2 is the variance of the background additive white Gaussian noise (AWGN).

The arrival of the IN samples can be described by a Poisson point process (PPP) [29], in which the number of the IN burst samples per second is a Poisson random variable Λ with the probability given by

$$P(\Lambda = n) = \frac{\lambda^n e^{-\lambda}}{n!}, \tag{7}$$

where λ is the Poisson parameter denoting the rate of IN arrival. Without loss of generality, the Middleton's Class A model is adopted in this paper as the distribution of the instantaneous amplitude, and the Poisson point process is adopted as the probability distribution of the IN burst arrival.

B. Signal Model of OFDM-Based Vehicular Communications

The IEEE 802.11p WAVE standard specified a typical OFDM-based broadband wireless vehicular transmission system [1], where the signal frame structure in the presence of IN at the receiver is illustrated in Fig. 1. In the time domain, the i-th transmitted frame consists of the OFDM block $\mathbf{x}_i = \left[x_{i,0}, x_{i,1}, \ldots, x_{i,N-1}\right]^T$ of length N and its cyclic prefix (CP) $\mathbf{c}_i$ of length-M. $\mathbf{x}_i$ is the inverse discrete Fourier transform (IDFT) of the corresponding frequency-domain data block $\tilde{\mathbf{X}}_i$ in the N OFDM sub-carriers, which contains some reserved null sub-carriers whose indices are denoted by the set Ω. The number of null sub-carriers is $R = |\Omega|$. Null sub-carriers are specified in vehicular communication systems exploited as the reserved tones or virtual sub-carrier masks [1], and can be utilized as the measurement of the IN, which will be explained in detail in the proposed algorithms in Section III. The transmitted OFDM frame then passes through the wireless vehicular communication channel with the channel impulsive response (CIR) $\mathbf{h}_i = [h_{i,0}, h_{i,1}, \ldots, h_{i,N-1}]^T$.

At the receiver, the received i-th time-domain OFDM data block $\mathbf{y}_i$ is represented as

$$\mathbf{y}_i = [y_{i,0}, y_{i,1}, \ldots, y_{i,N-1}]^T = \mathbf{h}_i \odot \mathbf{x}_i + \boldsymbol{\xi}_i + \mathbf{w}_i, \tag{8}$$

where $\boldsymbol{\xi}_i$ and $\mathbf{w}_i$ denote the length-N vector of time-domain IN additive to the received OFDM data block, and the corresponding time-domain AWGN with zero mean and variance of σ^2, respectively. The operator $\odot$ denotes circular convolution. Then, transforming the received signal to the frequency domain using discrete Fourier transform (DFT) yields

$$\tilde{\mathbf{Y}}_i = \mathbf{F}_N \mathbf{y}_i = \mathbf{H}_i \tilde{\mathbf{X}}_i + \mathbf{F}_N \boldsymbol{\xi}_i + \tilde{\mathbf{w}}_i, \tag{9}$$

where $\mathbf{F}_N$ denotes the N-point DFT matrix with its entry $(\mathbf{F}_N)_{m,n}$ being $\exp(-j2\pi mn/N)/\sqrt{N}$, and $\tilde{\mathbf{w}}_i$ denotes the frequency-domain AWGN vector. The channel frequency response (CFR) is denoted by the $N \times N$ matrix $\mathbf{H}_i = \mathrm{diag}\{\mathbf{F}_N \mathbf{h}_i\}$. As is described previously, the R sub-carriers of $\tilde{\mathbf{X}}_i$ corresponding to the set Ω are reserved to zero at the transmitter, thus they can be selected out of $\tilde{\mathbf{Y}}_i$ using a selection matrix $\mathbf{S}_R$ to formulate the measurement vector of IN represented by

$$\tilde{\mathbf{p}}_i = \mathbf{S}_R \tilde{\mathbf{Y}}_i = \mathbf{S}_R \mathbf{H}_i \tilde{\mathbf{X}}_i + \mathbf{F}_R \boldsymbol{\xi}_i + \mathbf{S}_R \tilde{\mathbf{w}}_i = \mathbf{0} + \mathbf{F}_R \boldsymbol{\xi}_i + \tilde{\mathbf{w}}_{R,i}, \tag{10}$$

where $\mathbf{S}_R$ is an $R \times N$ selection matrix consisting of the corresponding R rows from the R rows out of the $N \times N$ identity matrix $\mathbf{I}_N$ indicated by the set Ω. Equation (10) holds since the entries of $\tilde{\mathbf{X}}_i$ corresponding to the set Ω are reserved to zero, so $\mathbf{S}_R \mathbf{H}_i \tilde{\mathbf{X}}_i = \mathbf{0}$. The frequency-domain length-R vector $\tilde{\mathbf{p}}_i = [\tilde{p}_{i,0}, \tilde{p}_{i,1}, \ldots, \tilde{p}_{i,R-1}]^T$ is the measurement vector of the IN at the null sub-carriers, and $\tilde{\mathbf{w}}_{R,i}$ denotes the corresponding length-R AWGN vector. The matrix $\mathbf{F}_R$ denotes the $R \times N$ partial DFT matrix consisting of the R rows of $\mathbf{F}_N$ indicated by the set Ω, which is given by

$$\mathbf{F}_R = \frac{1}{\sqrt{N}}[\boldsymbol{\chi}_0 \quad \boldsymbol{\chi}_1 \cdots \boldsymbol{\chi}_{N-1}], \tag{11}$$

where the vector $\boldsymbol{\chi}_m$ is composed of the entries $\exp(-j2\pi mk/N), k \in \Omega, m = 0, \ldots, N-1$. The matrix $\mathbf{F}_R$ can be regarded as the *observation matrix* utilized for sparse machine learning, where an observation matrix is similar to that in linear sparse inverse problems exploited for measuring data from the unknown vector.

Consequently, the received data in the null sub-carriers set Ω can be rewritten briefly as

$$\tilde{\mathbf{p}}_i = \mathbf{F}_R \boldsymbol{\xi}_i + \tilde{\mathbf{w}}_{R,i}. \tag{12}$$

III. Probabilistic Framework Formulation and Algorithms of Sparse Learning for IN Estimation

A. Probabilistic Framework Formulation of Sparse Learning for IN Estimation

Since the unknown IN to be recovered is a sparse vector in the time domain, it is most important to reconstruct the support, i.e., the indices set of nonzero entries, of IN. Assuming that the sparsity level of IN is no more than K, to ensure that the IN is

sparse, the length-N IN vector $\boldsymbol{\xi}_i$ to be recovered should satisfy

$$\|\boldsymbol{\xi}_i\|_0 \leq K \tag{13}$$

where the ℓ_0-norm operation $\|\cdot\|_0$ calculates the number of nonzero entries of the vector inside. Then, according to the system model analysis in Section II-B and equation (12), the optimal estimated IN $\hat{\boldsymbol{\xi}}_i^*$ that should be recovered is the one that generates the minimum residue error norm for the frequency-domain measurement vector $\tilde{\mathbf{p}}_i$ at the null sub-carriers set Ω. Hence, to recover the optimal IN $\hat{\boldsymbol{\xi}}_i^*$ is to solve the optimization problem given by

$$\hat{\boldsymbol{\xi}}_i^* = \arg\min_{\boldsymbol{\xi}_i} \|\tilde{\mathbf{p}}_i - \mathbf{F}_R\boldsymbol{\xi}_i\|_2, \text{s.t.} \|\boldsymbol{\xi}_i\|_0 \leq K, \tag{14}$$

and thus the residue error norm r is defined as

$$r = \|\tilde{\mathbf{p}}_i - \mathbf{F}_R\boldsymbol{\xi}_i\|_2. \tag{15}$$

Since the ℓ_0-norm constraint is non-convex, the problem (14) is intractable in the conventional signal processing perspective. Specifically, it is a sparse combinatorial optimization problem due to the sparse constraint. If denoting the set of all possible sparse vectors that satisfy the ℓ_0-norm constraint (13) by Ξ, which is given by

$$\Xi = \left\{\boldsymbol{\xi}_i \in \mathbb{C}^N \,\middle|\, \|\boldsymbol{\xi}_i\|_0 \leq K\right\}, \tag{16}$$

and the cardinality of the solution space Ξ is given by

$$|\Xi| = \mathrm{C}_N^K 2^K = \frac{N! \cdot 2^K}{(N-K)!K!}. \tag{17}$$

It can be noted from (17) that the possible entries of the solution space is exponentially and combinatorially increasing with the parameters.

Some sparse approximation methods, including the popular CS-based theory, have been exploited to relax the non-convex optimization problem to a tractable one in literature. For instance, the non-convex ℓ_0-norm constraint in (14) can be relaxed to the convex ℓ_1-norm minimization problem [11] as

$$\arg\min_{\boldsymbol{\xi}_i} \|\boldsymbol{\xi}_i\|_1, \text{s.t.} \|\tilde{\mathbf{p}}_i - \mathbf{F}_R\boldsymbol{\xi}_i\|_2 \leq \epsilon, \tag{18}$$

where ϵ denotes the error norm bound due to the background AWGN noise $\tilde{\mathbf{w}}_{R,i}$, and thus convex programming can be exploited to solve it [30]. However, the performance of the CS-based methods is much related with the restricted isometry property (RIP) of the observation matrix [11], [31]. Besides, performance degradation could be caused due to intensive background noise, and the spectral efficiency could still be improved because many null sub-carriers needs to be reserved [15].

To overcome the difficulties of the state-of-the-art methods, we propose a different probabilistic framework of sparse machine learning based on sparse CE minimization for IN recovery. It is able to efficiently solve the non-convex sparse combinatorial optimization problem (14) without prior requirements for the observation matrix $\mathbf{F}_R$ compared with classical CS-based methods, and more spectrum-efficient by reducing the requirement of null sub-carriers as validated by simulation results. The proposed sparse learning based algorithm significantly develops and improves the conventional CE method [26], and the unknown sparse signal can be accurately recovered, as described in detail in the next sub-section.

Algorithm 1: (*SCEM*): Sparse Cross-Entropy Minimization for Sparse Machine Learning Based IN Recovery.

Input:

1) Measurement vector $\tilde{\mathbf{p}}_i$
2) Observation matrix $\boldsymbol{\Psi} = \mathbf{F}_R$
3) Threshold for residue error norm ϵ
4) Candidate supports number N_c, favorable supports number N_f, maximum iteration number $I_{\max}$

Initialization:

$\mathbf{q}^{(0)} \leftarrow \frac{1}{2} \cdot \overrightarrow{\mathbf{1}}_{N\times 1}$ (initial probability distribution of the IN support)

$k \leftarrow 0$ (iteration count number)

Iterations:

repeat

Randomly generate N_c *candidate supports* $\{\Pi_j^{(k)}\}_{j=1}^{N_c}$ based on the current support distribution $\mathbf{q}^{(k)}$, where each candidate support is generated in a recursive way s.t. $|\Pi_j^{(k)}| \leq K, j = 1, \ldots N_c$

Compute the corresponding IN vectors $\{\hat{\boldsymbol{\xi}}_j^{(k)}\}_{j=1}^{N_c}$, s.t. $\hat{\boldsymbol{\xi}}_j^{(k)}\Big|_{\Pi_j^{(k)}} \leftarrow \boldsymbol{\Psi}_{\Pi_j^{(k)}}^{\dagger}\tilde{\mathbf{p}}_i$, $\hat{\boldsymbol{\xi}}_j^{(k)}\Big|_{\Pi_j^{(k)c}} \leftarrow \mathbf{0}$

Calculate the corresponding residue error norms $r_j^{(k)} = \left\|\tilde{\mathbf{p}}_i - \mathbf{F}_R\hat{\boldsymbol{\xi}}_j^{(k)}\right\|_2, j = 1, \ldots N_c$

Sort $\{r_j^{(k)}\}_{j=1}^{N_c}$ in the ascending order as $r_{[1]}^{(k)} \leq r_{[2]}^{(k)} \leq \cdots \leq r_{[N_c]}^{(k)}$

Select the N_f smallest residue error norms $\{r_{[j]}^{(k)}\}_{j=1}^{N_f}$, and set the corresponding supports $\{\Pi_{[j]}^{(k)}\}_{j=1}^{N_f}$ as the *favorable supports*

Update the probability distribution of IN support to $\mathbf{q}^{(k+1)}$ by minimizing the CE based on (22)

$k \leftarrow k + 1$

until $r_{[1]}^{(k-1)} \leq \epsilon$ or $k > I_{\max}$ (halting condition)

Output:

1) Learnt support probability distribution $\hat{\mathbf{q}} = \mathbf{q}^{(k)}$
2) Recovered IN support $\hat{\Pi} = \Pi_{[1]}^{(k-1)}$
3) Recovered sparse IN vector $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_{[1]}^{(k-1)}$

B. Proposed Sparse Machine Learning Inspired Algorithm for IN Recovery: Sparse Cross-Entropy Minimization

Based on the probabilistic framework of sparse learning, the purpose of the SCEM algorithm proposed in this paper is to efficiently solve the sparse combinatorial optimization problem (14) by iteratively minimizing the cross-entropy between the current support distribution and the one minimizing the residue error norm. The pseudo-code of the proposed SCEM algorithm is summarized in Algorithm 1, and the computing flowchart of the

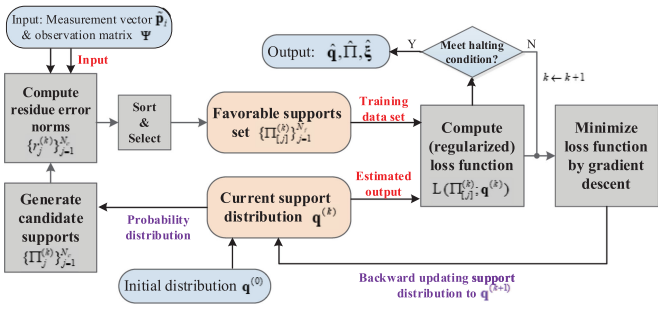


Fig. 2. Computing flowchart of the iterative sparse machine learning based algorithm of SCEM for IN recovery.

essential computing modules, parameters, nodes, and data flows of the algorithm is illustrated in Fig. 2.

It can be observed from Fig. 2 that the proposed sparse machine learning algorithm iteratively learns the probability distribution of the IN support by minimizing the loss function (i.e., the cross-entropy). In each iteration within the algorithm loop, the algorithm generates a set of candidate supports randomly based on the current support distribution $\mathbf{q}^{(k)}$ (initialized by $\mathbf{q}^{(0)}$), and computes the corresponding residue error norms using the measurement vector from the input. After sorting the residue error norms, the set of favorable supports is selected out, which serves as the training data set. Then, the loss function is computed by calculating the cross-entropy between the training data set (obtained from the favorable supports having the minimum residue error norms thus most close to the real IN support) and the estimated output (i.e., the current probability distribution $\mathbf{q}^{(k)}$ of the IN support stored by the algorithm). By minimizing the loss function by gradient descent, the support distribution is backward updated to $\mathbf{q}^{(k+1)}$ for the next iteration. This process will drive the support distribution gradually to be trained towards the one with minimum estimation error. The iterations continue until the halting condition of the algorithm is met, and the output of the algorithm is thus achieved.

The overall structure and explanations of Algorithm 1 are described as follows:

Phase 1- Input: The measurement vector $\tilde{\mathbf{p}}_i$, the observation matrix $\mathbf{\Psi}$, the residue error norm threshold ϵ determined by the power of the background AWGN in (12), and the numbers of candidate supports and favorable supports, i.e., N_c and N_f, are input to the algorithm.

Phase 2 - Initialization: The initial probability distribution of the IN support is set as $\mathbf{q}^{(0)} \leftarrow \frac{1}{2} \cdot \overrightarrow{\mathbf{1}}_{N\times 1}$, where $\mathbf{q}^{(k)} = [q_0^{(k)}, q_1^{(k)} \cdots q_{N-1}^{(k)}]^T$, and $q_n^{(k)}$ denotes the probability that the n-th entry is in the IN support $\Pi^{(k)}$, i.e.,

$$\Pr(n \in \Pi^{(k)}) = q_n^{(k)}, n = 0, \ldots N-1. \tag{19}$$

Since the nonzero entries can be randomly distributed in the support, assuming each entry has an initial probability of 0.5 to be nonzero is rational without loss of generality.

Phase 3 - Main iterations: The main process is composed of multiple iterations, and terminates until the halting condition of the algorithm is met. The main process includes the following steps:

1) Candidate supports generation (Line 4): N_c *candidate supports* $\{\Pi_j^{(k)}\}_{j=1}^{N_c}$ are generated based on the support distribution $\mathbf{q}^{(k)}$. Each candidate support $\Pi_j^{(k)}$ is generated in an efficient and simple recursive manner to obtain a K-sparse support. Let π_l denote the current temporary support in the recursive generation process, where the initial temporary support $\pi_0 = \{0, 1, \ldots, N-1\}$. Then, based on the current temporary support π_l and its corresponding probability $\{q_n^{(k)}\}_{n\in\pi_l}$ derived from the current support distribution $\mathbf{q}^{(k)}$, a sparser temporary support π_{l+1} can be generated by a Bernoulli trial on each entry $n \in \pi_l$ as

$$\pi_{l+1} = \{n | n \in \pi_l, \text{and } f_n^{(\pi_l)} = 1\}, \tag{20}$$

where the $\{0,1\}$-valued parameter $f_n^{(\pi_l)}$ is the outcome of the Bernoulli trial on the entry $n \in \pi_l$ with Bernoulli probability $q_n^{(k)}$. Afterwards, $l \leftarrow l+1$ and keep doing this until $|\pi_l| \le K$, and then the candidate support is set as $\Pi_j^{(k)} = \pi_l$.

2) Computing IN and residue (Line 5-6): the estimated IN vectors $\{\hat{\boldsymbol{\xi}}_j^{(k)}\}_{j=1}^{N_c}$ corresponding to the candidate supports are calculated based on the least squares principle implemented on the candidate supports $\{\Pi_j^{(k)}\}_{j=1}^{N_c}$, and the corresponding residue error norms $\{r_j^{(k)}\}_{j=1}^{N_c}$ are calculated by (15) using the estimated IN vectors.

3) Favorable supports selection (Line 7-8): the candidate supports are sorted by the residue error norms in the ascending order in order to pick out the best N_f candidate supports with smallest estimation error, which is closest to the real IN support and regarded as the favorable supports $\{\Pi_{[j]}^{(k)}\}_{j=1}^{N_f}$. The implicit probability distribution implied by the favorable supports is the training target for the current support distribution $\mathbf{q}^{(k)}$, which is gradually driven towards the ground-truth distribution by iterative minimizing the CE between them.

4) Learning support distribution by minimizing CE (Line 9): The CE is utilized as the loss function $\mathcal{L}(\Pi_{[j]}^{(k)}; \mathbf{q}^{(k)})$ in the perspective of machine learning theory, which is given by

$$\mathcal{L}(\Pi_{[j]}^{(k)}; \mathbf{q}^{(k)}) = -\frac{1}{N_f} \sum_{j=1}^{N_f} \ln \Pr\left(\Pi_{[j]}^{(k)} \middle| \mathbf{q}^{(k)}\right), \tag{21}$$

where $\{-\ln \Pr(\Pi_{[j]}^{(k)} | \mathbf{q}^{(k)})\}$ is the negative logarithm likelihood (NLL) of the favorable support $\Pi_{[j]}^{(k)}$ conditioned on the current probability distribution $\mathbf{q}^{(k)}$. By minimizing the loss function in (21), the current support distribution $\mathbf{q}^{(k)}$ is updated to $\mathbf{q}^{(k+1)}$, which is given by

$$\mathbf{q}^{(k+1)} = \arg\min_{\mathbf{q}^{(k)}} \left\{ -\frac{1}{N_f} \sum_{j=1}^{N_f} \ln \Pr\left(\Pi_{[j]}^{(k)} \middle| \mathbf{q}^{(k)}\right) \right\}, \tag{22}$$

Let a $\{0,1\}$-valued length-N vector $\mathbf{f}_{[j]}$ denote the favorable support $\Pi_{[j]}^{(k)}$, where its n-th entry $f_{[j],n} = (\mathbf{f}_{[j]})_n$ satisfies

$$f_{[j],n} = \begin{cases} 1, & n \in \Pi_{[j]}^{(k)} \\ 0, & n \notin \Pi_{[j]}^{(k)} \end{cases} \tag{23}$$

Then the conditional probability $\Pr(\Pi_{[j]}^{(k)}|\mathbf{q}^{(k)})$ in the CE in (22) is given by

$$\Pr\left(\Pi_{[j]}^{(k)}|\mathbf{q}^{(k)}\right) = \Pr\left(\mathbf{f}_{[j]}|\mathbf{q}^{(k)}\right), \tag{24}$$

where $f_{[j],n}$ is a Bernoulli random variable given by

$$\Pr\left(f_{[j],n}=1\right) = q_n^{(k)},\ \Pr(f_{[j],n}=0) = 1-q_n^{(k)}. \tag{25}$$

Thus, one can derive that

$$\Pr(\mathbf{f}_{[j]}|\mathbf{q}^{(k)}) = \prod_{n=0}^{N-1}\left(q_n^{(k)}\right)^{f_{[j],n}}\left(1-q_n^{(k)}\right)^{1-f_{[j],n}}. \tag{26}$$

By substituting (26) into (22), the first derivative of the CE with respect to $q_n^{(k)}$ can be derived as

$$\begin{aligned}
&\frac{\partial}{\partial q_n^{(k)}}\left\{-\frac{1}{N_f}\sum_{j=1}^{N_f}\ln\Pr\left(\Pi_{[j]}^{(k)}\Big|\,\mathbf{q}^{(k)}\right)\right\}\\
&=\frac{\partial}{\partial q_n^{(k)}}\left\{-\frac{1}{N_f}\sum_{j=1}^{N_f}\Big[f_{[j],n}\ln q_n^{(k)}+(1-f_{[j],n})\ln(1-q_n^{(k)})\Big]\right\}\\
&=-\frac{1}{N_f}\sum_{j=1}^{N_f}\left[\frac{f_{[j],n}}{q_n^{(k)}}-\frac{1-f_{[j],n}}{1-q_n^{(k)}}\right].
\end{aligned} \tag{27}$$

To minimize the CE, the first derivative (27) is set to zero, so the updated support distribution $\mathbf{q}^{(k+1)}$ can be learnt by

$$q_n^{(k+1)} = \frac{1}{N_f}\sum_{j=1}^{N_f} f_{[j],n},\ n=0,1,\ldots,N-1. \tag{28}$$

5) Iteration switching (Line 10-11): if the halting condition is satisfied when $r_{[1]}^{(k-1)} \le \epsilon$ or $k > I_{\max}$, the algorithm ends. Otherwise, the algorithm goes into the next iteration.

Phase 4 - Output: The output of the algorithm includes the learnt support probability distribution $\hat{\mathbf{q}} = \mathbf{q}^{(k)}$, the recovered IN support $\hat{\Pi} = \Pi_{[1]}^{(k-1)}$, and the recovered sparse IN vector $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_{[1]}^{(k-1)}$, which obtains the solution of the sparse combinatorial optimization problem (14) as $\hat{\boldsymbol{\xi}}_i^* = \hat{\boldsymbol{\xi}}$. Then, the recovered IN signal $\hat{\boldsymbol{\xi}}_i^*$ can be canceled out from the received OFDM data block $\mathbf{y}_i$ in (8) before the subsequent demapping and decoding procedures to eliminate the impacts of IN.

C. Enhanced Sparse Machine Learning Based Algorithm for IN Recovery: Regularized SCEM

In the proposed SCEM algorithm where the CE plays the important role of loss function, each NLL corresponding to each favorable support $\Pi_{[j]}^{(k)}$ has an average contribution to the CE given in (22), so the favorable supports with different residue error norms contribute the same to the loss function. In fact, different supports should reflect different contributions on the loss function so as to encourage the algorithm to learn the support with less error. Out of this insight, an enhanced sparse learning algorithm of RSCEM is proposed, in which the loss function (21) is regularized by multiplying with the weighting parameter $\lambda_{[j]}$ to generate the regularized loss function $\mathcal{L}_{\text{reg}}(\Pi_{[j]}^{(k)};\mathbf{q}^{(k)})$ given by

$$\mathcal{L}_{\text{reg}}\left(\Pi_{[j]}^{(k)};\mathbf{q}^{(k)}\right) = -\frac{1}{N_f}\sum_{j=1}^{N_f}\lambda_{[j]}\ln\Pr\left(\Pi_{[j]}^{(k)}\Big|\,\mathbf{q}^{(k)}\right), \tag{29}$$

where the regularization weighting parameter $\lambda_{[j]}$ is given by

$$\lambda_{[j]} = \frac{\overline{r}^{(k)}}{r_{[j]}^{(k)}},\ j=1,2,\ldots,N_f, \tag{30}$$

where $\overline{r}^{(k)}$ is the average residue error norm over the favorable supports given by

$$\overline{r}^{(k)} = \frac{1}{N_f}\sum_{j=1}^{N_f} r_{[j]}^{(k)}. \tag{31}$$

Note that a smaller residue error norm $r_{[j]}^{(k)}$ leads to a larger weighting parameter $\lambda_{[j]}$ in (30). Hence, the NLL corresponding to a more accurate support will have a larger contribution to the regularized loss function in (29), which will drive the support distribution $\mathbf{q}^{(k)}$ to converge to the ground-truth support more accurately and more efficiently. The pseudo-code of RSCEM is thus similar to that of SCEM given in Algorithm 1 except for the procedure of minimizing the loss function in Line 9, where the regularized loss function is now adopted to update the distribution as given by

$$\mathbf{q}^{(k+1)} = \arg\min_{\mathbf{q}^{(k)}} -\frac{1}{N_f}\sum_{j=1}^{N_f}\lambda_{[j]}\ln\Pr\left(\Pi_{[j]}^{(k)}\Big|\,\mathbf{q}^{(k)}\right). \tag{32}$$

To calculate the minimum regularized loss function in (32), we inherit the same notation as in the previous sub-section, i.e. the Bernoulli vector $\mathbf{f}_{[j]}$ in (23) denoting the favorable support $\Pi_{[j]}^{(k)}$. Through similar deduction from (23) to (26), and substituting (26) into (32), we can obtain the first derivative of the regularized loss function with respect to $q_n^{(k)}$, represented as

$$\begin{aligned}
&\frac{\partial}{\partial q_n^{(k)}}\left\{-\frac{1}{N_f}\sum_{j=1}^{N_f}\lambda_{[j]}\ln\Pr\left(\Pi_{[j]}^{(k)}\Big|\,\mathbf{q}^{(k)}\right)\right\}\\
&\quad=\frac{\partial}{\partial q_n^{(k)}}\left\{-\frac{1}{N_f}\sum_{j=1}^{N_f}\lambda_{[j]}\Big[f_{[j],n}\ln q_n^{(k)}\right.\\
&\qquad\left.+(1-f_{[j],n})\ln(1-q_n^{(k)})\Big]\right\}\\
&\quad=-\frac{1}{N_f}\sum_{j=1}^{N_f}\lambda_{[j]}\left[\frac{f_{[j],n}}{q_n^{(k)}}-\frac{1-f_{[j],n}}{1-q_n^{(k)}}\right].
\end{aligned} \tag{33}$$

Setting the first derivative of the regularized loss function given in (33) to zero, the regularized loss function can be minimized, yielding the updated support probability distribution

$\mathbf{q}^{(k+1)}$ given by

$$q_n^{(k+1)} = \frac{\sum_{j=1}^{N_f} \lambda_{[j]} f_{[j],n}}{\sum_{j=1}^{N_f} \lambda_{[j]}},\ n = 0, 1, \ldots, N-1. \qquad (34)$$

Comparing (34) with (28), it can be observed that, for the algorithm of SCEM, all the entries $\{f_{[j],n}\}_{j=1}^{N_f}$ have the same contribution to the updating of $q_n^{(k+1)}$ in (28), so the different accuracy among favorable supports are not taken into consideration. On the other hand, for the enhanced RSCEM algorithm, a more accurate support $\Pi_{[j]}^{(k)}$ will impose a larger weighting parameter $\lambda_{[j]}$ on and have a larger contribution to the updating of $q_n^{(k+1)}$ as implied by (34). In fact, (28) can be regarded as a special case of (34) when $\lambda_{[j]} = 1,\ j = 1, 2, \ldots, N_f$. Consequently, it can be derived that the enhanced RSCEM algorithm will learn the ground-true support distribution more accurately and more efficiently than SCEM, which is also validated in the simulation results in the next section.

D. Computational Complexity Analysis

The computational complexity of the proposed algorithms and the state-of-the-art ones are theoretically analyzed and compared in this sub-section.

For the proposed algorithms, considering the complexity of each iteration of SCEM in Algorithm 1: Line 4 (generating N_c candidate supports) - $\mathcal{O}(N_c)$; Line 5 (calculating N_c IN vectors) - $\mathcal{O}\left(N_c RK^2\right)$; Line 6 (calculating N_c residue error norms) - $\mathcal{O}(N_c RK)$; Line 7–8 (sorting N_c residue error norms and selecting N_f smallest ones) - $\mathcal{O}(N_c \log N_c)$; Line 9 (updating the IN support distribution) - $\mathcal{O}(NN_f)$. Therefore, summing them together, the total complexity of each iteration of SCEM is $\mathcal{O}\left(N_c RK^2 + NN_f\right)$. Similarly, since RSCEM only involves the calculation of N_f weighting parameters in (30) with the complexity of $\mathcal{O}(N_f)$, one can derive that the total complexity of each iteration of RSCEM is also $\mathcal{O}\left(N_c RK^2 + NN_f\right)$. Then, considering the maximum iteration number $I_{\max}$, the total complexity of the SCEM and RSCEM algorithms are $\mathcal{O}\left(I_{\max}(N_c RK^2 + NN_f)\right)$. Since the IN is sparse, the sparsity level $K \ll N$. Moreover, it is not necessary for R, $I_{\max}$ and N_c to be very large, either, which will be verified in the simulation results in the next section.

Hence, it can be concluded that the computational complexity of the proposed algorithms is acceptable, and comparable to the complexity of the existing CS-based algorithm, i.e., $\mathcal{O}\left(NRK^2\right)$ as derived in [15].

Furthermore, considering the total number of iterations $I_{\max}$ required to guarantee accurate recovery, it is difficult to derive a theoretical closed-form expression of the iteration number for the proposed machine learning based algorithms. The halting condition that the residue error norm should be less than the given threshold ϵ might be reached at any iteration, which is irrelevant to the parameters of K, R, N, etc. Besides, the convergence rate of learning is closely related with the initial status $\mathbf{q}^{(0)}$ and the form of loss function. However, one can effectively evaluate the average number of iterations through the numerical simulation results in the next section, which shows that the total iteration number does not linearly increase with the sparsity level K and $I_{\max}$ does not have to be very large, so the total complexity is acceptable.

TABLE I
PARAMETER PROFILE OF ITU-R VEHICULAR B MULTIPATH CHANNEL

Tap	Relative delay (ns)	Average power (dB)
1	0	-2.5
2	300	0
3	8 900	-12.8
4	12 900	-10.0
5	17 100	-25.2
6	20 000	-16.0

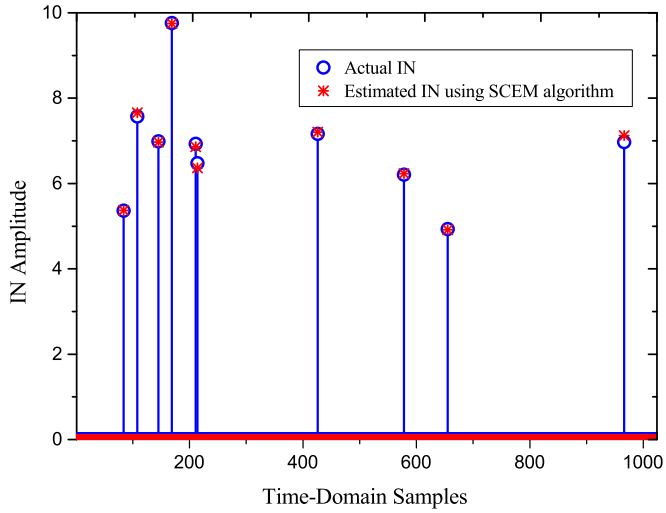


Fig. 3. Graphical visualization of one realization of the IN recovery using the proposed SCEM algorithm.

IV. Simulation Results and Discussions

The performance of the proposed sparse learning method of IN recovery in wireless vehicular communication systems is evaluated through simulations. The simulation setup is basically configured in a typical wireless vehicular transmission scenario. The OFDM sub-carrier number $N = 1024$ and the number of null sub-carriers $R = 128$, and the length of CP is $M = 128$. The modulation scheme of 16QAM and the low density parity check (LDPC) code with code length of 1944 bits and code rate of 0.5 as specified in [1] are adopted. The typical 6-tap multipath channel model called Vehicular-B specified by ITU-R in low-speed vehicular scenario with the relative receiver velocity of 30 km/h is adopted, whose parameters are listed in Table I [32]. The Middleton's Class A model for the IN amplitude distribution and the PPP model for the IN arrival distribution are adopted, with the parameters configured as $A = 0.15, \omega = 0.02, \lambda = 50/\text{sec}$. In the following simulations results, unless explicitly stated, the sparsity level of IN $K = 10$, the INR $\gamma = 30$ dB, and the parameters $N_c = 70, N_f = 15$, and $I_{\max} = 15$ are configured for the proposed SCEM and RSCEM algorithms.

The performance of one realization of the IN recovery using the proposed algorithm of SCEM with $K = 10$ is depicted in Fig. 3. The measurement vector serving as the input of SCEM in Algorithm 1 is obtained from the R null sub-carriers. Based on the sparse learning iterations driven by minimizing

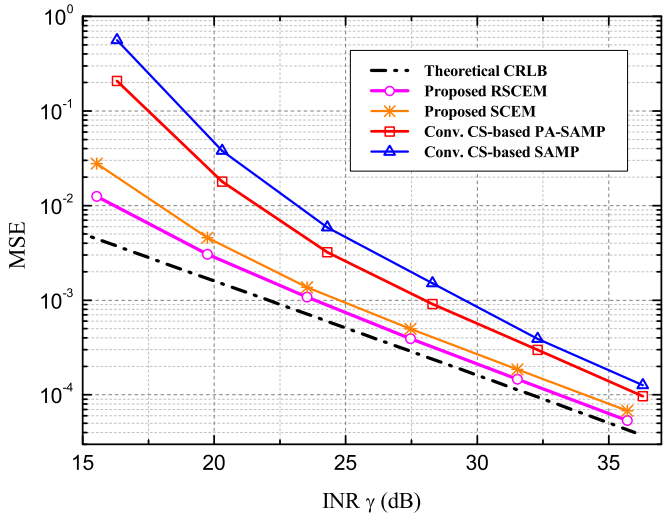


Fig. 4. MSE performance of the IN reconstruction using the proposed algorithms of SCEM and RSCEM compared with state-of-art CS-based methods.

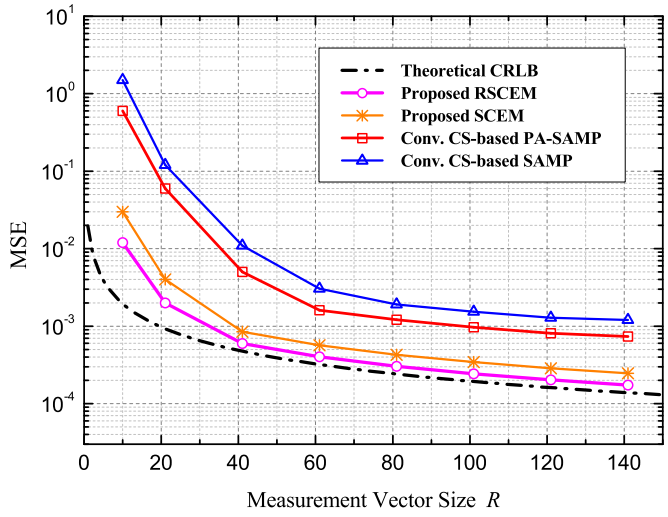


Fig. 5. MSE performance of the IN recovery using the proposed and CS-based algorithms against the measurement vector length R.

the cross-entropy, the accurate support distribution and the IN vector are recovered using the proposed algorithm of SCEM. It is observed from Fig. 3 that the IN estimation matches the actual IN very well.

The mean square error (MSE) performances of the proposed algorithms and the state-of-art CS-based algorithms (SAMP [16] and PA-SAMP [15]) for IN recovery in the wireless vehicular communication system are shown in Fig. 4. The theoretical Cramer-Rao lower bound (CRLB) $2\sigma^2 \cdot (N \cdot K/R)$ [33] is illustrated as benchmark. The measurement length is set as $R = 128$. It can be observed that the proposed RSCEM and SCEM algorithms achieve the MSE of 10^{-3} at the INR of 23.7 dB and 24.8 dB, respectively, which outperforms conventional PA-SAMP and SAMP algorithms by approximately 3.5 dB and 5.0 dB, respectively. It is noted from Fig. 4 that the MSE of the proposed SCEM and RSCEM algorithms are asymptotically approaching the theoretical CRLB with the increase of the INR, which verifies the high accuracy of the proposed sparse learning method for IN recovery. It is also shown by Fig. 4 that the proposed enhanced algorithm RSCEM enjoys a further INR gain of about 0.8 dB over the SCEM algorithm, which proves that the accuracy of the RSCEM algorithm can be further improved by imposing regularization on the loss function.

Moreover, the MSE performance of IN recovery for different schemes with respect to the measurement length R is reported in Fig. 5, where the INR is set as $\gamma = 30$ dB. it is shown in Fig. 5 that the MSE of the proposed algorithms decreases fast with the increase of the measurement vector length R, i.e. the number of null sub-carriers utilized for IN measurement, whereas the decrease of the MSE of the existing CS-based methods is much slower. At the MSE of 10^{-3}, the proposed RSCEM and SCEM algorithms costs only $R = 32$ and 38 null sub-carriers, respectively, while CS-based algorithms cost more than 100 null sub-carriers. Hence, it can be concluded that the introduction of sparse machine learning will greatly reduce the amount of measurement data, i.e. spectrum resource required for accurate recovery, achieving higher spectral efficiency than conventional counterparts.

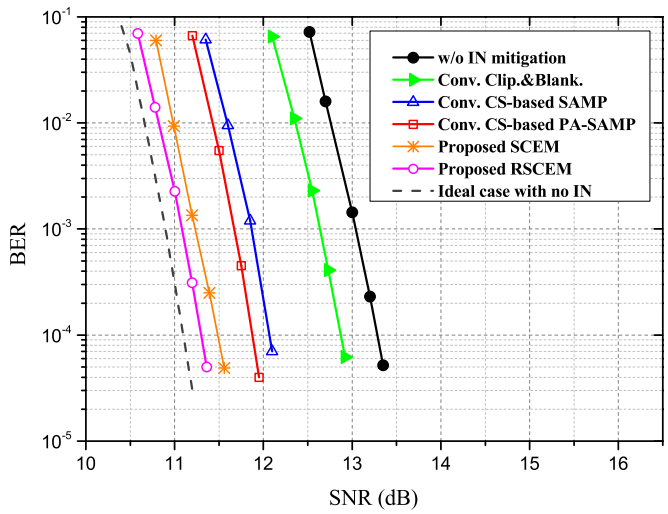


Fig. 6. BER performance of different schemes for IN mitigation under the wireless vehicular transmission channel.

The bit error rate (BER) performance of different IN mitigation schemes under the wireless vehicular transmission channel is illustrated in Fig. 6, including the proposed sparse learning based algorithms of SCEM and RSCEM, as well as the conventional non-CS-based method (clipping and blanking [10]) and the state-of-art CS-based methods. The worst case without IN mitigation and the ideal case with no IN are also depicted as benchmarks. It can be found that at the target BER of 10^{-4}, the proposed sparse learning based algorithms outperform the state-of-art CS-based methods, the conventional method of clipping and blanking, and the worst case without IN mitigation by approximately 0.6 dB, 1.5 dB, and 1.9 dB, respectively. It is also shown that the proposed enhanced algorithm of RSCEM using regularization of the loss function can further improve the performance of the proposed SCEM algorithm by 0.2 dB. Furthermore, the gap between the proposed RSCEM algorithm and the ideal case with no IN is only about 0.2 dB, validating the effectiveness of the proposed scheme.

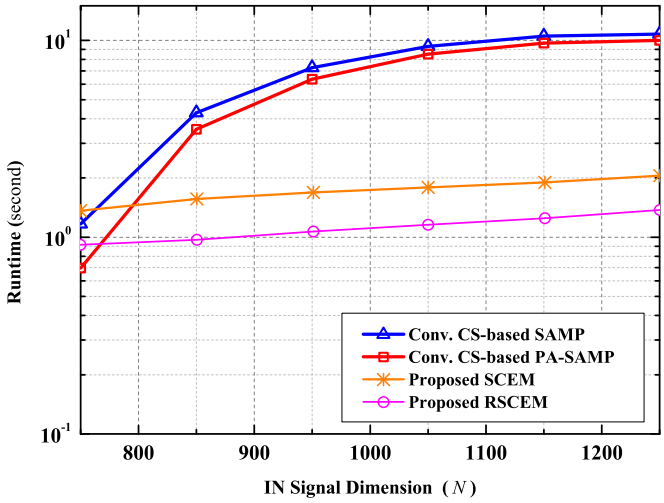


Fig. 7. Comparison of the computational complexity of the proposed and CS-based algorithms with respect to IN dimension N.

TABLE II
THE AVERAGE NUMBER OF ITERATIONS OF DIFFERENT IN RECOVERY ALGORITHMS AGAINST SPARSITY LEVEL

Sparsity Level	Conventional PA-SAMP	Conventional SAMP	Proposed SCEM	Proposed RSCEM
10	7.5	9.8	14.3	11.2
20	16.4	20.1	14.9	11.8
30	25.8	29.6	15.6	12.7

The runtime, as a transparent indicator of the convergence rate, i.e. iteration number, and computational complexity, is simulated with respect to the dimension of the unknown sparse IN vector N, as shown in Fig. 7. The average runtime of an algorithm of interest is computed using the average of 10^3 runs. Note that for fair comparison, the runtime of different schemes is compared in the condition that the MSE $\leq 10^{-3}$ is reached when the algorithms converge. It can be observed from Fig. 7 that the runtime of the conventional CS-based algorithms increases approximately linearly with the IN dimension N, validating the theoretical complexity $\mathcal{O}\left(NRK^2\right)$. On the other hand, it can be observed from Fig. 7 that the runtime of the proposed RSCEM and SCEM algorithms increases much slower with N than that of CS-based algorithms, which is also consistent with the theoretical complexity $\mathcal{O}\left(I_{\max}(N_c RK^2 + NN_f)\right)$. It is shown that the convergence of the proposed algorithms is faster than that of the CS-based ones when $N \geq 800$, and the gain grows larger with the increase of N. It is also verified that the enhanced algorithm RSCEM is more complexity-efficient and converges faster than SCEM.

For both the conventional CS-based and the proposed algorithms, the number of iterations against the sparsity level K is investigated through numerical experiments, as shown in Table II. For fair comparison, the number of iterations of different schemes is also compared in the condition that the MSE $\leq 10^{-3}$ is reached when the algorithms converge. It is indicated that the number of iterations for the CS-based algorithms increase approximately linearly with the sparsity level, whereas the iteration number of the proposed algorithms almost keep invariant, which is consistent with the analysis. Besides, it is proved from the results that the enhanced algorithm RSCEM has fewer iterations than SCEM and thus converges faster. It is also verified that setting the maximum iteration number as $I_{\max} = 15$ is sufficient for accurate recovery.

V. CONCLUSIONS

A novel sparse machine learning based probabilistic framework of IN recovery is formulated for reliable transmission of OFDM-based wireless vehicular communication systems. The original non-convex sparse combinatorial optimization problem of IN recovery is efficiently and accurately solved by the proposed sparse learning algorithm of SCEM, which iteratively learns the probability distribution of the IN support by minimizing the loss function of cross-entropy. By imposing regularization on the loss function, the enhanced algorithm of RSCEM is proposed to further improve the convergence rate and accuracy. It is verified by theoretical analysis and numerical simulation results that the proposed algorithms outperform conventional ones in spectrum efficiency, estimation accuracy and complexity. Moreover, the proposed scheme can also be widely applied in other vehicular related and OFDM-based broadband communication systems contaminated by IN to improve the system performance in senarios like URLLC.

REFERENCES

[1] *IEEE Standard for Information Technology—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments*, IEEE Standard 802.11p, Jul. 2010.

[2] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Trans. Electromagn. Compat.*, vol. 44, no. 1, pp. 249–258, Feb. 2002.

[3] H. Suraweera and J. Armstrong, "Noise bucket effect for impulse noise in OFDM," *Electron. Lett.*, vol. 40, no. 18, pp. 1156–1157, Sep. 2004.

[4] A. Nasri and R. Schober, "Robust L_p-norm decoding for BICM-based secondary user systems," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3084–3090, Nov. 2010.

[5] D. Thompson and J. Dixon, "Vehicle noise," in *Advanced Applications in Acoustics, Noise and Vibration*. London, U.K.: Spon Press, 2004, pp. 236–291.

[6] A. B. Vallejo-Mora, J. J. Sanchez-Martinez, F. J. Canete, J. A. Cortes, and L. Diez, "Characterization and evaluation of in-vehicle power line channels," in *Proc. IEEE Global Telecommun. Conf.*, 2010, pp. 1–5.

[7] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 145–147, Feb. 1996.

[8] F. Juwono, Q. Guo, D. Huang, and K. P. Wong, "Deep clipping for impulsive noise mitigation in OFDM-based power-line communications," *IEEE Trans. Power Del.*, vol. 29, no. 3, pp. 1335–1343, Jun. 2014.

[9] R. Shepherd, J. Gaddie, and D. Nielson, "New techniques for suppression of automobile ignition noise," *IEEE Trans. Veh. Technol.*, vol. VT-25, no. 1, pp. 2–12, Feb. 1976.

[10] S. 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.

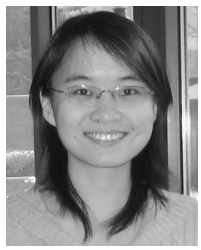
[11] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[12] W. Ding, F. Yang, C. Pan, L. Dai, and J. Song, "Compressive sensing based channel estimation for OFDM systems under long delay channels," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 313–321, Jun. 2014.

[13] S. Liu, F. Yang, X. Wang, J. Song, and Z. Han, "Structured-compressed-sensing-based impulsive noise cancelation for MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6921–6931, Aug. 2017.

[14] T. Naffouri, A. Quadeer, and G. Caire, "Impulse noise estimation and removal for OFDM systems," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 976–989, Mar. 2014.
[15] S. Liu, F. Yang, W. Ding, and J. Song, "Double kill: Compressive-sensing-based narrow-band interference and impulsive noise mitigation for vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5099–5109, Jul. 2016.
[16] T. Do, G. Lu, N. Nguyen, and T. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2008, pp. 581–587.
[17] 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.
[18] S. Liu, F. Yang, J. Song, and Z. Han, "Block sparse Bayesian learning-based NB-IoT interference elimination in LTE-advanced systems," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4559–4571, Oct. 2017.
[19] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.
[20] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surv. Tut.*, vol. 19, no. 4, pp. 2432–2455, Oct.–Dec. 2017.
[21] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, and M. Guizani, "Security in mobile edge caching with reinforcement learning," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 116–122, Jun. 2018.
[22] N. Kato *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017.
[23] Y. Li, X. Cheng, and G. Gui, "Co-robust-ADMM-net: Joint ADMM framework and DNN for robust sparse composite regularization," *IEEE Access*, vol. 6, pp. 47943–47952, 2018.
[24] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
[25] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?" *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 41–49, Sep. 2018.
[26] D. P. Kroese, R. Y. Rubinstein, I. Cohen, S. Porotsky, and T. Taimre, "Cross-entropy method," in *Encyclopedia of Operations Research and Management Science*. New York, NY, USA: Springer, 2013, pp. 326–333.
[27] M. Nassar, K. Gulati, Y. Mortazavi, and B. Evans, "Statistical modeling of asynchronous impulsive noise in powerline communication networks," in *Proc. IEEE Global Telecommun. Conf.*, Houston, TX, USA, Dec. 2011, pp. 1–6.
[28] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106–127, Aug. 1977.
[29] N. Andreadou and F. Pavlidou, "Modeling the noise on the OFDM powerline communications system," *IEEE Trans. Power Del.*, vol. 25, no. 1, pp. 150–157, Jan. 2010.
[30] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[31] Z. Han, H. Li, and W. Yin, *Compressive Sensing for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[32] "Guideline for evaluation of radio transmission technology for IMT-2000," Recommendation ITU-R M.1225, 1997.
[33] S. Liu, F. Yang, and J. Song, "Narrowband interference cancelation based on priori aided compressive sensing for DTMB systems," *IEEE Trans. Broadcast.*, vol. 61, no. 1, pp. 66–74, Mar. 2015.

Sicong Liu (S'15–M'17) received the B.S.E. and Ph.D. degrees (with the highest Hons.) in electronic engineering from Tsinghua University, Beijing, China, in 2012 and 2017, respectively. In 2010, he was a Visiting Scholar with City University of Hong Kong. He was a Senior Research Engineer with Huawei Technologies before joining, in 2018, Xiamen University, Xiamen, China, where he is currently an Assistant Professor. His current research interests lie in sparse signal processing, machine learning, and wireless communications. He was the Editor, Track Chair, or TPC Member of several IEEE and other academic journals and conferences.



Liang Xiao (M'09–SM'13) received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 2009. She is currently a Professor with the Department of Communication Engineering, Xiamen University, Xiamen, China. Her current research interests include smart grids, network security, and wireless communications.

Lianfen Huang received the B.S. degree in radio physics and Ph.D. degree in communication engineering from Xiamen University, Xiamen, China, in 1984 and 2008, respectively. She was a Visiting Scholar with Tsinghua University, Beijing, China, in 1997, and with The Chinese University of Hong Kong, Hong Kong, in 2012. She is currently a Professor of Communication Engineering with Xiamen University. Her research interests include wireless communications, wireless networks, and signal processing.

Xianbin Wang (S'98–M'99–SM'06–F'17) received the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2001.

He is currently a Professor and Tier 1 Canada Research Chair with Western University, London, ON, Canada. Prior to joining Western University, he was with the Communications Research Centre Canada (CRC) as a Research Scientist/Senior Research Scientist between July 2002 and December 2007. From January 2001 to July 2002, he was a System Designer with STMicroelectronics. He has more than 350 peer-reviewed journal and conference papers, in addition to 29 granted and pending patents and several standard contributions. His current research interests include 5G technologies, Internet of Things, communications security, machine learning, and locationing technologies.

Dr. Wang is a Fellow of Canadian Academy of Engineering, and an IEEE Distinguished Lecturer. He is the recipient of many awards and recognitions, including Canada Research Chair, CRC President's Excellence Award, Canadian Federal Government Public Service Award, Ontario Early Researcher Award, and six IEEE Best Paper Awards. He is currently an Editor/Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON BROADCASTING, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was also an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS between 2007 and 2011, and the IEEE WIRELESS COMMUNICATIONS LETTERS between 2011 and 2016. He was involved in many IEEE conferences including GLOBECOM, ICC, VTC, PIMRC, WCNC and CWIT, in different roles such as Symposium Chair, Tutorial Instructor, Track Chair, Session Chair, and TPC Co-Chair.