doi:10.3772/j.issn.1006-6748.2021.01.002

Dynamic multi-user detection scheme based on CVA-SSAOMP algorithm in uplink grant-free NOMA①

Xu Lei (徐 磊)②, Tao Shangjin, Bai Shichao, Zhang Jian, Fang Hongyu, Li Xiaohui
(School of Electronic and Information Engineering, Anhui University, Hefei 230039, P. R. China)

Abstract

In the uplink grant-free non-orthogonal multiple access (NOMA) scenario, since the active user at the sender has a structured sparsity transmission characteristic, the compressive sensing recovery algorithm is initially applied to the joint detection of the active user and the transmitted data. However, the existing compressed sensing recovery algorithms with unknown sparsity often require noise power or signal-to-noise ratio (SNR) as the priori conditions, which greatly reduces the algorithm adaptability in multi-user detection. Therefore, an algorithm based on cross validation aided structured sparsity adaptive orthogonal matching pursuit (CVA-SSAOMP) is proposed to realize multi-user detection in dynamic change communication scenario of channel state information (CSI). The proposed algorithm transforms the structured sparsity model into a block sparse model, and without the priori conditions above, the cross validation method in the field of statistics and machine learning is used to adaptively estimate the sparsity of active user through the residual update of cross validation. The simulation results show that, compared with the traditional orthogonal matching pursuit (OMP) algorithm, subspace pursuit (SP) algorithm and cross validation aided block sparsity adaptive subspace pursuit (CVA-BSASP) algorithm, the proposed algorithm can effectively improve the accurate estimation of the sparsity of active user and the performance of system bit error ratio (BER), and has the advantage of low-complexity.

Key words: non-orthogonal multiple access (NOMA), multi-user detection, cross validation, structured sparsity (SP), orthogonal matching pursuit (OMP)

0 Introduction

Multiple access technology has been regarded as one of the key technologies of each generation of wireless communication system. Specially, orthogonal multiple access (OMA) is used in current 4G systems, and that is orthogonal frequency division multiple access (OFDMA). In OMA, the number of supported users is strictly limited by the available orthogonal resources, which is difficult to meet the requirements of large-scale connections in the future communication systems[1]. In order to meet this challenge, non-orthogonal multiple access (NOMA) has been actively studied[1,2], which can achieve system overload through non-orthogonal resource allocation. The uplink transmission is scheduled by the base station (BS) in the request permission process, there will be a large number of transmission delay and signaling overhead problems. In the large-scale connections, these problems becomes more worse or even unacceptable.

In order to solve the problems above, the grant-free transmission is expected in the uplink NOMA system, where users can send data randomly and BS does not know which users are active, so the active status of users must be detected. Because of the characteristics of sporadic communication in the Internet of Things (IoT), the active user presents sparsity, and the compressive sensing (CS) theory[3] can be introduced to recover the sparse signals. Therefore, the multi-user detection (MUD) problem can be converted into a sparse signal recovery problem. However, in these CS-based MUD schemes, signal detection is typically detected independently at different time slots, where the correlation of active user at different time slots is not considered, and the number of active user must be the same at several consecutive time slots. Therefore, Ref. [4] proposed the joint signal detection at several consecutive time slots in a frame, which can improve the system multi-user detection performance by utilizing

① Supported by the National Natural Science Foundation of China (No. 62001001).
② To whom correspondence should be addressed. E-mail: xulei@ahu.edu.cn
Received on Feb. 12, 2020

frame-by-frame sparsity. Ref. [5] proposed a multi-user detection iterative support detection (ISD) algorithm. Based on the multi-user ISD algorithm, Ref. [6] proposed the structured iterative support detection (SISD) algorithm, which uses the structured sparsity of active user to jointly detect the number of active user and the transmitted data in the uplink grant-free NOMA system. An approximate message passing and expectation maximization joint (Joint-EM-AMP) detection algorithm was proposed in Ref. [7]. Based on Ref. [6], Ref. [8] focused on the impact of the active user number and the system overload rate on the multi-user detection performance of the system, and made some corresponding in-depth analysis. Ref. [9] effectively combined cross validation method in statistics and machine learning with the estimation of sparsity in the compressed sensing recovery algorithm. A cross validation aided block sparsity adaptive subspace pursuit (CVA-BSASP) algorithm was proposed in Ref. [10]. The CVA-BSASP algorithm realizes the adaptive sparsity estimation of active user by using the cross validation method in the statistics and the machine learning when the noise power or the signal-to-noise ratio is not required as a priori condition. However, CVA-BSASP algorithm assumes that the channel state information (CSI) of multiple time slots in a frame remains unchanged, which is not in accordance with the actual communication scenario. Moreover, when the maximum sparsity is satisfied, the iteration is stopped, so a large number of unnecessary iterations need to be continued after the algorithm estimates the actual sparsity of active user. This not only greatly increases the complexity of the algorithm, but also reduces the adaptability of the algorithm.

In this paper, the CVA-SSAOMP algorithm is proposed combined with the actual communication scenario in which the intra-frame multiple time slots channel state information (CSI) dynamically changes, and based on the structured sparsity model of the active user at the transmitting end. This algorithm transforms the structured sparsity model into a block sparse structured model equivalently. For the situation that the sparsity of active user is unknown, the cross validation method is used to realize the adaptive estimating the sparsity of active user. When the cross validation method estimates the actual sparsity of active user, the corresponding cross validation residual is minimum. This feature can be used as the condition for early termination of the iteration. In addition, combined with the low-complexity advantages of OMP algorithm, the overall complexity of multi-user algorithm can be reduced, and the adaptability of the algorithm can be further improved. Simulation results show that the proposed algorithm achieves good multi-user detection performance.

Notation: uppercase and lowercase bold letters represent matrices and vectors respectively. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{\dagger}$ and $\|\cdot\|_p$ respectively represent transpose, conjugate transpose, matrix pseudo-inverse and norm. $\mathrm{supp}(\cdot)$ denotes a support set.

1 System model

A classic uplink NOMA system framework with one BS and K users[6] is considered, where both the BS and the users are equipped with a single antenna. The transmission symbol x_k of user k is modulated onto a spreading sequence s_k of length N. Considering the overload of $N < K$ system, the overload rate is K/N, and the number of users can be greater than the length of the spreading sequence. After that, the signals from all active user are superimposed, and then transmitted through N orthogonal OFDM subcarriers. The received signal at BS can be expressed as

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{G}_k \boldsymbol{s}_k x_k + \boldsymbol{v} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v} \tag{1}$$

where, $\boldsymbol{y} = [y_1, y_2, \cdots, y_N]^{\mathrm{T}}$ denotes the received signal on the N OFDM subcarriers. $\boldsymbol{G}_k = diag(\boldsymbol{g}_k)$, $\boldsymbol{g}_k = [g_{1,k}, g_{2,k}, \cdots, g_{N,k}]^{\mathrm{T}}$ denotes the channel gain of the corresponding user k on the N OFDM subcarriers, obeying the complex Gaussian distribution with the mean of 0 and the unit variance. $\boldsymbol{s}_k = [s_{1,k}, s_{2,k}, \cdots, s_{N,k}]^{\mathrm{T}}$ is the spreading sequence of the k-th user, and the length is N. Where $\boldsymbol{H} = [\boldsymbol{G}_1\boldsymbol{s}_1, \boldsymbol{G}_2\boldsymbol{s}_2, \cdots, \boldsymbol{G}_K\boldsymbol{s}_K]$ is the equivalent channel matrix and the value of its elements in the n-th row and k-th columns is $h_{n,k} = g_{n,k}s_{n,k}$. $\boldsymbol{x} = [x_1, x_2, \cdots, x_K]^{\mathrm{T}}$ is the signal transmitted by all K users. $\boldsymbol{v} = [v_1, v_2, \cdots, v_N]^{\mathrm{T}}$ denotes the Gaussian noise vector on N OFDM subcarriers.

It is assumed that the active user and the inactive user are completely synchronized in the entire intra-frame data, in which case the active user set of consecutive time slots in one frame remains unchanged. Therefore, the frame structured sparsity model of the active user can be obtained[4]. Defining $\boldsymbol{x}^{[j]} = [x_1, x_2, \cdots, x_K]^{\mathrm{T}}$ as the signal transmitted by all K users at j-th time slots, the active user support set of the j-th time slots is $\mathrm{supp}(\boldsymbol{x}^{[j]})$, then there are

$$\mathrm{supp}(\boldsymbol{x}^{[1]}) = \mathrm{supp}(\boldsymbol{x}^{[2]}) = \cdots = \mathrm{supp}(\boldsymbol{x}^{[j]}) \tag{2}$$

The assumed $\boldsymbol{x}^{[j]}$ maximum level of sparsity (i. e., the number of non-zero elements of $\boldsymbol{x}^{[j]}$) is S, then the support set $\boldsymbol{x}^{[j]}$ is defined as

$$\Gamma^{[j]} = \{k: k \in \{1, 2, \cdots, K\}, \boldsymbol{x}_k^{[j]} \neq 0\} \tag{3}$$

this denotes an index set of non-zero elements in $\boldsymbol{x}^{[j]}$. According to the statistics data of mobile traffic[11], even during busy hours, the number S of active user is usually much smaller than the number K of all potential users.

Consider reconstructing J time slots active user transmit signal $[\boldsymbol{x}^{[1]}, \boldsymbol{x}^{[2]}, \cdots, \boldsymbol{x}^{[J]}]$ from J time slots received signal $[\boldsymbol{y}^{[1]}, \boldsymbol{y}^{[2]}, \cdots, \boldsymbol{y}^{[J]}]$, at J consecutive time slots (e.g., $J = 7$ is considered in the LTE-advanced standard[12]), so

$$\boldsymbol{y}^{[j]} = \boldsymbol{H}^{[j]}\boldsymbol{x}^{[j]} + \boldsymbol{v}^{[j]}, \; j = 1, 2, \cdots, J \tag{4}$$

where $\boldsymbol{y}^{[j]} \in \mathbb{C}^{N\times 1}$ is the active user transmit signal at j-th time slots received. $\boldsymbol{H}^{[j]} \in \mathbb{C}^{N\times K}$ is the equivalent channel matrix at j-th time slot, which is obtained by multiplying the channel matrix and the spreading matrix, and the channel matrix dynamically changes at different time slots, which $\boldsymbol{v}^{[j]} \in \mathbb{C}^{N\times 1}$ is the noise vector at j-th time slot.

2 Cross validation aided structured sparsity adaptive orthogonal matching pursuit algorithm

The CVA-SSAOMP algorithm proposed in this paper contains two parts.

Part one is structured sparsity model recombine. The received signal $[\boldsymbol{y}^{[1]}, \boldsymbol{y}^{[2]}, \cdots, \boldsymbol{y}^{[J]}]$ of J time slots in Eq. (4) is connected end to end according to the elements of each row, and then converted into one-dimensional signal. At the same time, the equivalent channel matrix $[\boldsymbol{H}^{[1]}, \boldsymbol{H}^{[2]}, \cdots, \boldsymbol{H}^{[J]}]$, the transmission signal $[\boldsymbol{x}^{[1]}, \boldsymbol{x}^{[2]}, \cdots, \boldsymbol{x}^{[J]}]$, and the noise signal $[\boldsymbol{v}^{[1]}, \boldsymbol{v}^{[2]}, \cdots, \boldsymbol{v}^{[J]}]$ of J time slots are adjusted accordingly.

Part two is signal reconstruction algorithm. Divide the recombined J time slot received signals in part one into data of length N_{cv} and N_e. Data of length N_{cv} is used to adaptively estimate the sparsity of active user, and data of length N_e is used for signal reconstruction. The specific content is as follows.

2.1 Structured sparsity model recombine

Firstly, the structured sparsity model is recombined to convert multiple time slots signal detection into single time slots signal detection. The received signal $[\boldsymbol{y}^{[1]}, \boldsymbol{y}^{[2]}, \cdots, \boldsymbol{y}^{[J]}]$ of the J time slots in one frame is rearranged to obtain a recombination signal $\boldsymbol{Y}$, and the corresponding adjusted equivalent channel matrix $[\boldsymbol{H}^{[1]}, \boldsymbol{H}^{[2]}, \cdots, \boldsymbol{H}^{[J]}]$ is rearranged to obtain $\boldsymbol{P}$, rearranged signal $[\boldsymbol{x}^{[1]}, \boldsymbol{x}^{[2]}, \cdots, \boldsymbol{x}^{[J]}]$ to be recombined to obtain $\boldsymbol{X}$, rearranged noise signal $[\boldsymbol{v}^{[1]}, \boldsymbol{v}^{[2]}, \cdots, \boldsymbol{v}^{[J]}]$ to be recombined to obtain $\boldsymbol{V}$.

The specific recombination method is as follows.

Recombining received signal of J time slots into

$$\boldsymbol{Y} = [y_1^{[1]} \;\; y_1^{[2]} \;\; \cdots \;\; y_1^{[J]} \;\; \cdots \;\; y_N^{[1]} \;\; y_N^{[2]} \;\; \cdots \;\; y_N^{[J]}]^{\mathrm{T}} \tag{5}$$

Recombining the transmit signal of J time slots into

$$\boldsymbol{X} = [x_1^{[1]} \;\; x_1^{[2]} \;\; \cdots \;\; x_1^{[J]} \;\; \cdots \;\; x_K^{[1]} \;\; x_K^{[2]} \;\; \cdots \;\; x_K^{[J]}]^{\mathrm{T}} \tag{6}$$

Recombining the noise signal of J time slots into

$$\boldsymbol{V} = [v_1^{[1]} \;\; v_1^{[2]} \;\; \cdots \;\; v_1^{[J]} \;\; \cdots \;\; v_N^{[1]} \;\; v_N^{[2]} \;\; \cdots \;\; v_N^{[J]}]^{\mathrm{T}} \tag{7}$$

Recombining the equivalent channel matrix of J time slots into

$$\boldsymbol{P} = \begin{pmatrix} h_{1,1}^{[1]} & 0 & 0 & & h_{1,K}^{[1]} & 0 & 0 \\ 0 & \ddots & 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & h_{1,1}^{[J]} & & 0 & 0 & h_{1,K}^{[J]} \\ & \vdots & & \ddots & & \vdots & \\ h_{1,N}^{[1]} & 0 & 0 & & h_{N,K}^{[1]} & 0 & 0 \\ 0 & \ddots & 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & h_{1,N}^{[J]} & & 0 & 0 & h_{N,K}^{[J]} \end{pmatrix} \tag{8}$$

Therefore, after the structured sparsity model is recombined, the mathematical model obtained is expressed as

$$\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{X} + \boldsymbol{V} \tag{9}$$

where $\boldsymbol{Y} \in \mathbb{C}^{NJ\times 1}$ is the received signal of J time slots, $\boldsymbol{P} \in \mathbb{C}^{NJ\times KJ}$ is the equivalent channel matrix of J time slots, $\boldsymbol{X} \in \mathbb{C}^{KJ\times 1}$ is all the data transmitted by the K users at J time slots in one frame, and $\boldsymbol{V} \in \mathbb{C}^{NJ\times 1}$ is a noise vector.

So far, the structured sparsity model has been recombined into a block sparsity structure. The multiple time slots structured sparsity signal reconstruction in Eq. (4) are converted into the reconstruction of the single time slots structured sparsity in Eq. (9), and it provides necessary preliminary preparation for the CVA-SSAOMP algorithm designed below.

2.2 CVA-SSAOMP algorithm

The specific steps of the CVA-SSAOMP algorithm are as follows.

Step 1 Recombine the structured sparsity model to obtain the block sparse structure.

Step 2 Selection of training data and validation data. The block sparse structure is segmented into training data and validation data by cross validation. The received signal $\boldsymbol{Y}$ after recombination takes vector $\boldsymbol{Y}_{cv} \in \mathbb{C}^{N_{cv}\times 1}$ of length N_{cv} as validation data, vector $\boldsymbol{Y}_e \in \mathbb{C}^{N_e\times 1}$ of length $N_e = JN - N_{cv}$ as training data, corresponding extended channel coefficient matrix is divided into $\boldsymbol{P}_{cv} \in \mathbb{C}^{N_{cv}\times JK}$ and $\boldsymbol{P}_e \in \mathbb{C}^{N_e\times JK}$, and noise is divided into $\boldsymbol{V}_e \in \mathbb{C}^{N_e\times 1}$ and $\boldsymbol{V}_{cv} \in \mathbb{C}^{N_{cv}\times 1}$. Eq. (10) can be obtained.

$$\begin{bmatrix}\boldsymbol{Y}_e\\ \boldsymbol{Y}_{cv}\end{bmatrix} = \begin{bmatrix}\boldsymbol{P}_e\\ \boldsymbol{P}_{cv}\end{bmatrix}\boldsymbol{X} + \begin{bmatrix}\boldsymbol{V}_e\\ \boldsymbol{V}_{cv}\end{bmatrix} \tag{10}$$

Step 3 Initialize the iteration parameters. The initial support set Γ is an empty set, and the initial residual $\boldsymbol{r}$ is received signal $\boldsymbol{Y}_e$. The initial sparsity of active user is $S = 1$, and the initial iteration number is $n = 1$.

Step 4 Calculate the correlation matrix and find the support set corresponding to the largest correlation coefficient. The vector $\boldsymbol{w}$, $\boldsymbol{w} \in \mathbb{C}^{JK\times 1}$ is obtained by calculating the inner product of the equivalent channel matrix $\boldsymbol{P}_e$ and the residual signal $\boldsymbol{r}$ of the n-th iteration by Eq. (11). Then divide $\boldsymbol{w}$ into K groups and calculate the energy of each group separately. All atomic indexes corresponding to the largest energy group are added to the support set $\boldsymbol{\Gamma}$.

$$\boldsymbol{w} = \boldsymbol{P}_e^{\mathrm{H}} \boldsymbol{r}^n \tag{11}$$

Step 5 The estimated signal $\hat{\boldsymbol{C}}$ is recovered by the pseudo-inverse matrix of Eq. (12) using the alternative support set $\boldsymbol{\Gamma}$.

$$\hat{\boldsymbol{C}}(\boldsymbol{\Gamma}) = (\boldsymbol{P}_e(\boldsymbol{\Gamma}))^{\dagger}\boldsymbol{Y}_e \tag{12}$$

Step 6 Backtracking the update support set. Using the recovered estimated signal $\hat{\boldsymbol{C}}$, select all the atomic indexes corresponding to the S group with the largest energy to obtain the updated support set Λ.

Step 7 The estimated signal $\hat{\boldsymbol{X}}$ is recovered by the pseudo-inverse matrix of Eq. (13). The backtracking updating support set Λ is used to recover the estimated signal $\hat{\boldsymbol{X}}$ whose sparsity is S, and the estimated signal under this sparsity is stored.

$$\hat{\boldsymbol{X}}(\boldsymbol{\Lambda}) = (\boldsymbol{P}_e(\Lambda))^{\dagger}\boldsymbol{Y}_e \tag{13}$$

Step 8 The estimated signal residual $\boldsymbol{r}$ is updated by the Eq. (14) using the training data.

$$\boldsymbol{r} = \boldsymbol{Y}_e - \boldsymbol{P}_e \times \hat{\boldsymbol{X}} \tag{14}$$

Step 9 If the updated residual value satisfies $\| \boldsymbol{r}^{(n)} < \boldsymbol{r}^{(n-1)} \|_2$, the support set is updated, the number of iterations is increased by one, and the process returns to Step 4 to continue the loop iteration under the current sparsity with the new residual value, otherwise Step 10 is executed.

Step 10 The cross validation residuals $\boldsymbol{c}$ is calculated using the validation data by Eq. (15), the cross validation residual is stored, and $S = S + 1$, returning to Step 4, continuing the iteration with the new sparsity. The iteration is stopped when the cross validation residual is successively smaller than the subsequent four cross validation residuals.

$$\boldsymbol{c} = \boldsymbol{Y}_{cv} - \boldsymbol{P}_{cv} \times \hat{\boldsymbol{X}} \tag{15}$$

Step 11 Find the sparsity corresponding to the minimum value of the cross validation residuals is the actual sparsity of active user, and the estimated signal corresponding to this sparsity is used as the recovery signal.

The proposed algorithm has the following characteristics. (1) Adaptive estimation of sparsity. In many applications, residual-based iterative stop conditions are widely used in compressed sensing recovery algorithms with unknown sparsity. However, the residual margin decreases monotonously, making it difficult to determine the optimal termination point. By contrast, the proposed CVA-SSAOMP algorithm can use cross validation to estimate the sparsity level because the cross validation residual typically has a minimum when the estimated sparsity is equal to the actual sparsity level. (2) Adaptability of algorithm. In the compressed sensing greedy algorithm, the measurement vector and the sparse transform base of the signal are irrelevant, and all the measured values are evenly distributed to the respective observations. Therefore, these observations have equal weights when reconstructing the signal, so partial observations can be used to reconstruct the sparse signal. The proposed CVA-SSAOMP algorithm can extract part of the observations as training data and it is used to reconstruct sparse signal. The remaining observations values are used as cross validation data to replace the prior conditions of noise level or sparsity to achieve the estimation of sparsity. In the algorithm, it is not necessary to set the maximum sparsity as the stop iteration condition, which makes the algorithm more adaptable.

2.3 CVA-SSAOMP algorithm with mean filtering

In Step 11 of the above algorithm, the sparsity corresponding to the minimum value of the cross validation residuals is taken as the actual sparsity of active user. In this process, there will be an error that the sparsity corresponding to the minimum value of the cross validation residuals is not the actual sparsity of active user. This is because the noise affects the accurate estimation of the sparsity of active user, and the error estimation will directly affect the system BER performance. Therefore, the mean filtering method in image processing is used to take the average value of the estimated sparsity of active user, approach the actual sparsity of active user at low SNR, and remove a small number of error at high SNR, thereby improving the system BER performance.

2.4 Algorithm complexity analysis

The analysis of the complexity of the CVA-SSAOMP algorithm is as follows.

(1) The inner product of the $N_e \times KJ$ dimension of the channel matrix and the $N_e \times 1$ dimension residual is first calculated in Step 4, multiplied N_e times, added $N_e - 1$ times, and totaled KJ rows, and the total complexity is $KJ(N_e + N_e - 1)$. Then calculate the energy of the K groups vector, and the computational complexity is $K(2J - 1)$. The two parts of the computational complexity are added as $KJ(N_e + N_e - 1) + K(2J - 1)$.

(2) The signal $\hat{\boldsymbol{C}}$ is estimated by the least squares method in Step 5. When the estimated sparsity of active user is S, the support set Γ has SJ atoms, the dimension of $\boldsymbol{P}_e$ is $N_e \times SJ$, and the complexity of the least squares estimation signal $\hat{\boldsymbol{C}}$ is $4SJN_e$.

(3) The grouping energy of the estimated signal $\hat{\boldsymbol{C}}$ is calculated in Step 6, and the computational complexity is $S(2J - 1)$.

(4) The least squares estimated signal $\hat{\boldsymbol{X}}$ is calculated in Step 7. At this time, the support set has SJ atoms, the dimension of $\boldsymbol{P}_e$ is $N_e \times SJ$, and the complexity of the least squares estimated signal $\hat{\boldsymbol{X}}$ is $4SJN_e$.

(5) Calculate the estimated signal residual complexity is $2SJN_e$ in Step 8.

(6) Calculate the complexity of the cross validation residuals is $2SJN_{cv}$ in Step 10.

Adding the computational complexity of the above parts, removing the smaller term, the single iteration computational complexity of the CVA-SSAOMP algorithm is $O(2JN_eK + 8JSN_e + 2J^2SN + JK)$. The upper limit of the number of iterations of the CVA-BSASP algorithm is $L_{CVA\text{-}BSASP} \leqslant \bar{S}\min\left(\frac{-\log\rho_{\min}}{-\log C_K} + 1, \frac{1.5\bar{S}}{-\log C_K}\right)$ in Ref. [10], where $\bar{S} = 2S$ is the maximum sparsity of active user set in the algorithm. Thus, the upper limit of the number of iterations of the CVA-SSAOMP algorithm can be expressed as $L_{CVA\text{-}SSAOMP} \leqslant (S + 4)\min\left(\frac{-\log\rho_{\min}}{-\log C_K} + 1, \frac{1.5S + 6}{-\log C_K}\right)$. This shows that $L_{CVA\text{-}SSAOMP} < L_{CVA\text{-}BSASP}$. The algorithm complexity of CVA-SSAOMP is $O(L_{CVA\text{-}SSAOMP} \cdot (2JN_eK + 8JSN_e + 2J^2SN + JK))$.

In the multi-user detection of J time slots in one frame, the complexity of the OMP algorithm is $O(2JSNK + 3JS^2N)$, the complexity of the SP algorithm is $O(2JSNK + 6JS^2N)$, and the complexity of the CVA-BSASP algorithm is $O(L_{CVA\text{-}BSASP} \cdot (2JN_eK + 12JSN_e + 2J^2SN + JK))$, where $N_e = 450$. The complexity of the CVA-SSAOMP algorithm is $O(L_{CVA\text{-}SSAOMP} \cdot (2JN_eK + 8JSN_e + 2J^2SN + JK))$. The calculation is known $O(2JN_eK + 8JSN_e + 2J^2SN + JK) < O(2JN_eK + 12JSN_e + 2J^2SN + JK)$, and $L_{CVA\text{-}SSAOMP} < L_{CVA\text{-}BSASP}$. Therefore, the complexity of the CVA-SSAOM algorithm proposed in this paper is far less than the complexity of the CVA-BSASP algorithm proposed in Ref. [10].

3 Simulation results and analysis

Monte Carlo method is used for system simulation, and the main simulation parameters in this paper are listed in Table 1.

Table 1 Simulation parameters

Parameters	Value
Number of users (K)	200
Number of subcarriers (N)	100
Number of active user (S)	20
Number of time slots (J)	7
Amount of cross validation data (N_{cv})	[140 : 20 : 360]

The curve of system BER performance among different amount of cross validation data is shown in Fig. 1. In which, the amount of cross validation data is a period from 140 to 360, the sparsity of active user (i. e., the actual number of active user) is $S = 20$, and SNR = 8 dB. It can be seen from Fig. 1 that when the amount of cross validation data is $N_{cv} = 250$, the system BER performance is optimal. When the amount of cross validation data $N_{cv} < 250$, the increase in the amount of cross validation data makes the estimation accuracy rate of active user increase, thereby improving the system BER performance. When the amount of cross validation data $N_{cv} > 250$, the further increase in

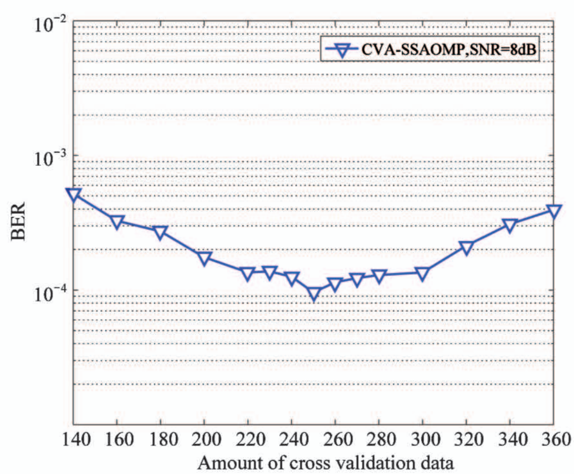

Fig. 1 The curve of system BER performance among different amount of cross validation data

the amount of cross validation data reduces the amount of data used to reconstruct the signal, resulting in a decrease in system BER performance.

The variation of cross validation residuals with the estimated sparsity of active user is shown in Fig. 2. In which, the amount of cross validation data is N_{cv} = 250, the sparsity of active user (i. e., the actual number of active user) is $S = 20$, and the maximum sparsity of active user (i. e., iterative stop sparsity) is set to 40. It can be seen from Fig. 2 that when the cross validation residual takes the minimum value, the estimated sparsity of active user is the actual sparsity. When the estimated sparsity of active user is greater than the actual sparsity, the cross validation residual shows a linear growth trend. Based on the above characteristics, after estimating the actual sparsity of active user, the proposed algorithm does not need to continue to perform a large number of unnecessary iterations.

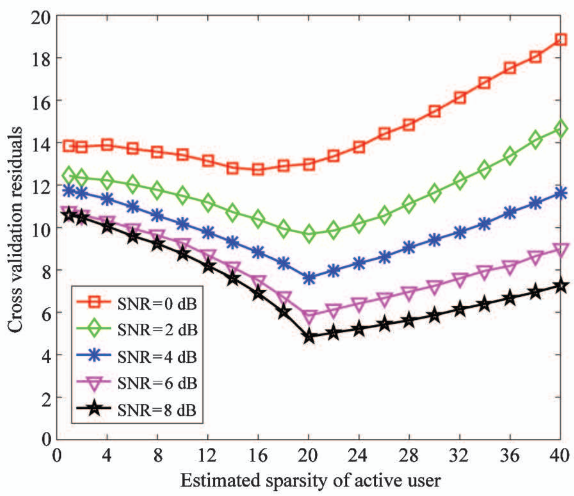

Fig. 2 The variation of cross validation residuals with the estimated sparsity of active user

The estimated sparsity accuracy curves of active user under different SNR are shown in Fig. 3. In which, the amount of cross validation data is N_{cv} = 250, the actual sparsity of active user is 20, and the estimated sparsity accuracies of CVA-SSAOMP algorithm with mean filtering, CVA-SSAOMP algorithm and CVA-BSASP algorithm are compared. It can be seen from Fig. 3 that the estimated sparsity accuracy of CVA-SSAOMP algorithm is significantly higher than that of CVA-BSASP algorithm. When the SNR is very low, the estimated sparsity accuracy of each algorithm is very low. In particular, the CVA-SSAOMP algorithm with mean filtering has the lowest accuracy. When the SNR is gradually increased, the estimated sparsity of each algorithm is gradually improved. Compared with the other algorithms, the CVA-SSAOMP algorithm with mean filtering can achieve accurate sparsity estimation of active user at lower SNR. This is because that, when the SNR is gradually increased, the affect of the noise on the estimation error is gradually reduced, the estimated sparsity accuracy can be improved by cross validation with the limited amount of data and the mean filtering algorithm plays a positive role quickly, which can further improve the estimated sparsity accuracy of active user.

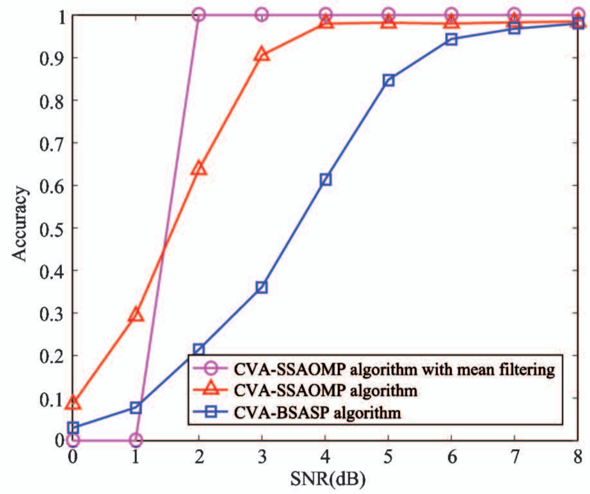

Fig. 3 The estimated sparsity accuracy curves of active user under different SNR

The comparison of system BER performance among different algorithms is shown in Fig. 4. In which, the amount of cross validation data is N_{cv} = 250 and the actual sparsity of active user is 20, and the system BER performances of the proposed CVA-SSAOMP algorithm with mean filtering, CVA-SSAOMP algorithm and traditional OMP algorithm, SP algorithm, CVA-BSASP algorithm proposed in Ref. [10] are compared. It can be seen from Fig. 4 that the system BER performance of CVA-SSAOMP algorithm with mean filtering and that of CVA-SSAOMP algorithm are better than other algorithms. When the SNR is very low, based on the characteristics of the estimated sparsity of active user, the system BER performance of the CVA-SSAOMP algorithm with mean filtering is slightly lower than that of the CVA-SSAOMP algorithm. When the SNR is gradually increased, the system BER performance of the CVA-SSAOMP algorithm with mean filtering is gradually better than that of the CVA-SSAOMP algorithm. When the SNR is 6dB, compared with traditional OMP algorithm, SP algorithm, and the CVA-BSASP algorithm proposed in Ref. [10], the BER performance of proposed CVA-SSAOMP algorithm is better than that of the other three algorithms, and is improved by 98.1%, 97%, and 55.6% respectively. The BER performance of CVA-SSAOMP algorithm with mean filtering is improved by about 25% compared with the CVA-SSAOMP algorithm.

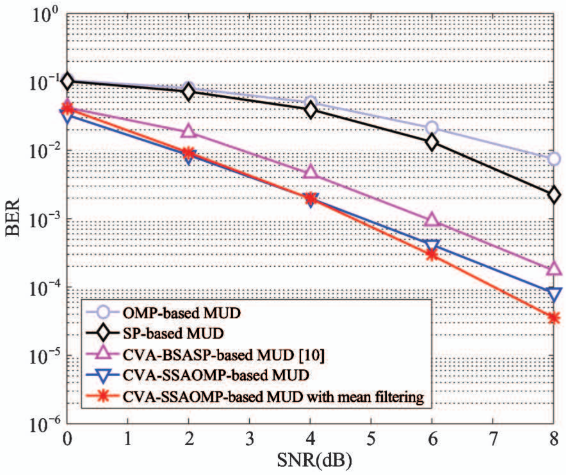

Fig. 4 The comparison of system BER performance among different algorithms

4 Conclusion

This paper proposes a CVA-SSAOMP algorithm for the structured sparsity model of active user in the uplink grant-free NOMA scenario. In this algorithm, based on the feature that the cross validation residual takes the minimum estimated sparsity of active user as the actual sparsity, the iterative stop condition is set. Combined with the low-complexity advantage of the OMP algorithm, the complexity of the proposed algorithm is effectively reduced. The introduction of low-complexity mean filtering further improves the accuracy of the estimated sparsity. Compared with traditional OMP algorithm, SP algorithm and CVA-BSASP algorithm, the proposed algorithm can effectively improve the BER performance.

References

[1] Dai L L, Wang B C, Yuan Y F, et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends[J]. *IEEE Communications Magazine*, 2015, 53(9):74-81

[2] Ding Z G, Dai L L, Poor H V. MIMO-NOMA design for small packet transmission in the Internet of Things[J]. *IEEE Access*, 2016, 4:1393-1405

[3] Shim B, Song B. Multiuser detection via compressive sensing[J]. *IEEE Communications Letters*, 2012, 16(7):972-974

[4] Abebe A T, Kang C G. Iterative order recursive least square estimation for exploiting frame-wise sparsity in compressive sensing-based MTC[J]. *IEEE Communications Letters*, 2016, 20(5):1018-1021

[5] Wang Y L, Yin W T. Sparse signal reconstruction via iterative support detection[J]. *SIAM Journal on Imaging Sciences*, 2010, 3(3):462-491

[6] Wang B C, Dai L L, Mir T, et al. Joint user activity and data detection based on structured compressive sensing for NOMA[J]. *IEEE Communications Letters*, 2016, 20(7):1473-1476

[7] Wei C, Liu H P, Zhang Z C, et al. Approximate message passing-based joint user activity and data detection for NOMA[J]. *IEEE Communications Letters*, 2017, 21(3):640-643

[8] Zhao X J, Zhang A H, Yang S Y, et al. Multi-user detection of NOMA system based on structured compressive sensing[J]. *Modern Electronic Technology*, 2018, 41(5):1-4

[9] Zhang J Y, Chen L M, Boufounos P T, et al. On the theoretical analysis of cross validation in commpressive sensing[C] // Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 2014: 3370-3374

[10] Du Y, Cheng C, Dong B H, et al. Block-compressed-sensing-based multiuser detection for uplink grant-free NOMA systems[C] // Proceedings of the 2018 IEEE International Conference on Communications, Kansas City, USA, 2018: 7894-7909

[11] Hong J P, Choi W, Rao B D. Sparsity controlled random multiple access with compressed sensing [J]. *IEEE Transactions on Wireless Communications*, 2015, 14(2): 998-1010

[12] European Telecommunications Standards Institude. 3GPP TS-36. 211 Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 12)[S]. Nice, French: European Telecommunications Standards Institude, 2016

Xu Lei, born in 1984. He received his Ph. D degree in Electromagnetic Field and Microwave Technology from Anhui University in 2012. He received his M. S. degree in Communication and Information System from Anhui University in 2009. He also received his B. S. degree in Electronic Science and Technology from Huaibei Normal University in 2006. Since 2012, he has been a lecturer in School of Electronics and Information Engineering of Anhui University. His research interests include wireless relay communications, non-orthogonal multiple access, and mmWave massive MIMO.