

# An improved reconstruction algorithm based on multi-user detection for uplink grant-free NOMA

Chengyan Hou<sup>1</sup>, Xiaohui Li<sup>1</sup>

<sup>1</sup>School of Electronics and Information Engineering, Anhui University, Hefei ,230000, P.R. China

**Abstract.** For the traditional orthogonal matching pursuit(OMP) algorithm in multi-user detection(MUD) for uplink grant-free NOMA, here is a poor BER performance, so in this paper we propose an temporal-correlation orthogonal matching pursuit algorithm(TOMP) to realize multi-user detection. The core idea of the TOMP is to use the time correlation of the active user sets to achieve user activity and data detection in a number of continuous time slots. We use the estimated active user set in the current time slot as a priori information to estimate the active user sets for the next slot. By maintaining the active user set  $\hat{T}^l$  of size  $K$  ( $K$  is the number of users),but modified in each iteration. Specifically, active user set is believed to be reliable in one iteration but shown error in another iteration, can be added to the set path delay  $\hat{T}^l$  or removed from it. Theoretical analysis of the improved algorithm provide a guarantee that the multi-user can be successfully detected with a high probability. The simulation results show that the proposed scheme can achieve better bit error rate (BER) performance in the uplink grant-free NOMA system.

## 1 Introduction

With the rapid development of mobile communication, the spectrum resources become increasingly scarce, in order to meet the needs of future 5G in spectrum efficiency and the number of connections; Non-orthogonal multiple access (NOMA) as a candidate for 5G technology has been widely concerned [1].The future 5G wireless communication network will meet the wireless networking needs of a variety for network equipment, especially for the scene of 5G Massive Machine Type Communication(MMTC), the sporadic overload systems are some small transmission packets, continue along by orthogonal multiple access (OMA) will be bound to cause the waste of resources and excessive scheduling overhead, therefore, in order to

avoid the waste of resources, we need to study the new non-orthogonal multiple access(NOMA) technology to improve the efficiency of small packet transmission.The existing multi-user detection of NOMA system is based on the premise that all users are active, however, not all users are active at the same time in the system, a large number of statistical results show that even in the busy time, active users generally do not exceed 10% of the total users, which means that although there are massive user connections in 5G, access to the system is only a small part of the user at the same time, so active users are sparse [2].The base station can detect the sparse active users and data simultaneously by means of joint detection. Since the base station does not need to know the active user status information, the user does not need

<sup>a</sup> Corresponding author: houchengyan520@163.com

to interact with the base station in advance, therefore, grant free transmission is highly expected in uplink NOMA systems, which greatly reduces signaling overhead and delay, and realizes user overload. It meets the requirements of future 5G mass connection, high spectrum utilization and low latency.

## 2 Compressive sensing

The premise of compressive sensing (CS) [3] is that the signal is sparse, but, most of the natural signal  $f \in R^N$  is not sparse in the time domain, but  $f$  can be sparse by a sparse basis  $\Psi$  transform, where  $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_N)$ , so the signal  $f$  can be expressed as:

$$f = \sum_{i=1}^N \Psi_i x_i = \Psi \mathbf{x} \quad (1)$$

The sparse vector  $\mathbf{x}$  is  $k$ -order sparse, so  $\|\mathbf{x}\|_0 \leq k$ , indicates that the number of nonzero values are equal to  $k$  ( $k < N$ ). If we apply a corresponding linear transformation to the coefficient  $\mathbf{x}$  with measuring matrix  $\Phi: M \times N$  ( $M \ll N$ ) which is not related to the sparse base  $\Psi$ , get measurements  $\mathbf{y}: M \times 1$ :

$$\mathbf{y} = \Phi \mathbf{f} = \Phi \Psi \mathbf{x} = \mathbf{A} \mathbf{x} \quad (2)$$

The vector  $\mathbf{x}$  is  $k$ -order sparse, so the initial signal can be accurately recovered from the linear measurement set using the optimal method, the observation matrix  $\Phi$  and the sparse base  $\Psi$  follow the constraint isometric (RIP) condition, so reconstruct the signal  $\mathbf{x}$ , i.e., to find the minimum  $\ell_0$  norm [4]:

$$\hat{\mathbf{x}} = \min_{\mathbf{x} \in R^n} \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{A} \mathbf{x} = \mathbf{y} \quad (3)$$

where  $\mathbf{A} = \Phi \Psi$ ,  $\mathbf{A} \in R^{M \times N}$  ( $M \ll N$ ), for the general  $\mathbf{A}$  and  $\mathbf{y}$ , there are nondeterministic polynomials for the equivalent constraint problem of  $\ell_0$ , so the following method is proposed:

$$\min_{\mathbf{x} \in R^n} \|\mathbf{x}\|_1 \quad s.t. \quad \mathbf{y} = \mathbf{A} \mathbf{x} \quad (\ell_1)$$

The optimal sparse solution requires a very sparse premise, and the perceived matrix  $\mathbf{A}$  must satisfies the incoherent condition, however, in practice, these conditions cannot be satisfied,  $\ell_1$  cannot find the optimal solution, we must turn to other nonconvex

problems, which needs weaker conditions to guarantee the accuracy of the successful recovery [5]:

$$\min_{\mathbf{x} \in R^n} \|\mathbf{x}\|_p \quad s.t. \quad \mathbf{y} = \mathbf{A} \mathbf{x} \quad (\ell_p)$$

where  $0 < p < 1$ , it can be seen from the above formula that when  $p$  is close to zero. The solution of  $\ell_p$  will be closer to the solution of  $\ell_0$ .

## 3 System model

This paper considers a typical uplink NOMA system with one base station (BS) and  $K$  users, here the single antenna scheme is considered without loss of generality. The transmitted symbol  $x_k$  for user  $k$  is modulated onto a spreading sequence  $\mathbf{s}_k$  of length  $N$ , in particular, we consider the case of  $N < K$ , that is, the overloaded system, where the number of users can be larger than the length of spreading sequences; After that, signals from all active users are superimposed, and then are transmitted over  $N$  orthogonal OFDM subcarriers [6]. The received signal on subcarrier  $n$  at the BS is denoted as:

$$y_n = \sum_{k=1}^K g_{nk} s_{nk} x_k + v_n, \quad n = 1, 2, \dots, N \quad (4)$$

where  $s_{nk}$  is the  $n$ th component of spreading sequence  $\mathbf{s}_k$ , and  $v_n$  is the Gaussian noise on subcarrier  $n$  with zero mean and variance  $\sigma^2$ ,  $g_{nk}$  is the channel gain of user  $k$  on the  $n$ th subcarrier, and all of them are independent and identically distributed complex Gaussian variables with zero mean and unit variance [6-8], we combine the received signals over all  $N$  subcarriers, and then the received signal vector  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$  is denoted as:

$$\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{v} \quad (5)$$

where  $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$  denotes the vector of transmitted signal for all  $K$  users,  $\mathbf{H}$  is the measurement matrix of size  $N \times K$ , its value  $h_{nk}$  in the  $n$ th row and the  $k$ th column equals to  $g_{nk} s_{nk}$ ,  $\mathbf{v} = [v_1, v_2, \dots, v_N]^T$  is the noise vector that follows the distribution of  $CN(0, \sigma^2 I_N)$  [8].

## 4 TOMP algorithm

The core idea of the TOMP is to use the time correlation of the active user sets to achieve user activity and data detection in a number of continuous time slots. This is the biggest difference with OMP. The main steps of the algorithm are as follows:

Input:

Measurements :  $\mathbf{Y} = [y^{(1)}, y^{(2)}, \dots, y^{(l)}]$ ;

Measurement matrix:  $\Phi = [\phi_1, \phi_2, \dots, \phi_N]$ ;

Maximum sparsity level:  $K$ .

Initialization:

1) The initial support is set as  $\hat{T}_0 = \phi$ , residual

signal  $\mathbf{R} = \mathbf{Y}$ ;

Iteration:

2) if  $k > K$ , stop, otherwise continue.

3) The estimated signal:

$$\mathbf{Z} = \Phi^H \mathbf{R} \quad (6)$$

4) User estimated support:

$$\hat{T}_0 = \hat{T}_0 \cup \arg \max_i \lambda_i = \sum_j |Z_{i,j}| \quad (7)$$

5) Residual: The number of elements in the estimated support do not exceed the sparsity level  $K$ :

$$\mathbf{R} = \mathbf{Y} - \Phi_{\hat{T}_0} \Phi_{\hat{T}_0}^+ \mathbf{Y} \quad (8)$$

6) The estimate of the support set at the end of the iteration by  $\hat{T}_0$ , which also serves as the initial estimate support for the following operation:

$$\hat{T}^l = \hat{T}_0 \quad (9)$$

7) Search matching users:  $l = l + 1$ , searching for the most relevant atoms and indexes to join set  $\hat{T}^{l-1}$ , in each iteration, the user that correlates best with the residual signal will be included in the active user set at first. So, the active user set can be calculated by:

$$\tilde{T}^l = \hat{T}^{l-1} \cup \arg \max_i \beta_i = \phi_i^H r^{l-1} \quad (10)$$

where  $r^{l-1}$  is the residual signal.

8) Update: The sparse transmitted signal vector in this  $l$ th time slot can be updated as follows:

$$\tilde{x}^l = \Phi_{\tilde{T}^l}^+ y \quad (11)$$

9) Modify the estimated support: Estimated support set is constantly being modified in each iteration. For example, in an iteration, the active user set is reliable, but the active user set is not accurate at the next iteration, and then it should be removed from the  $\hat{T}^l$ . The estimated support is revised by:

$$\hat{T}^l = \{K \text{ indices with largest magnitude in } \tilde{x}^l\} \quad (12)$$

10) Update the residual signal:

$$r^l = y - \Phi_{\hat{T}^l} \Phi_{\hat{T}^l}^+ y \quad (13)$$

11) When the energy of the residual error signal does not decrease, the iteration will be terminated with the following condition:

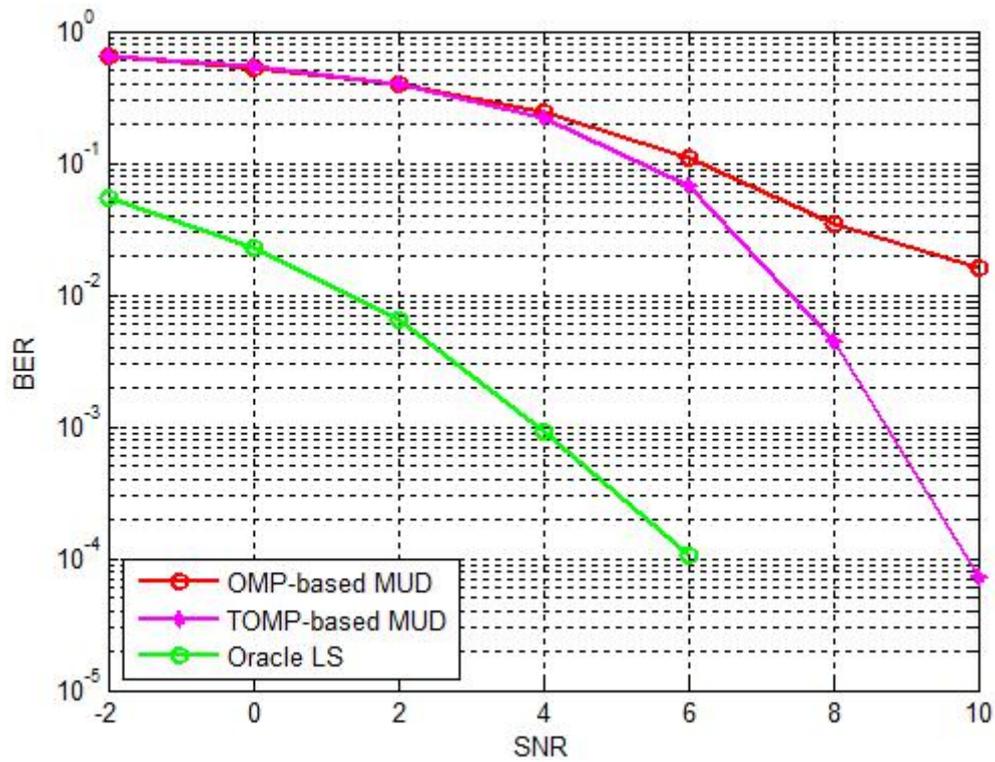
$$\|r^l\|_2 \geq \|r^{l-1}\|_2 \quad (14)$$

12) Output: After calculating the active user set in each time slot, the sparse transmission signal vector can be restored in the  $l$ th time.

$$\hat{x}^{(l)} = \Phi_{\hat{T}^l}^+ y^{(l)} \quad (15)$$

## 5 Simulation results

The simulation platform is the MATLAB R2014a version. Let's assume that the number of active users are already known at the BS, with QPSK modulation; Especially, the prerequisites for the simulation system are that the number of users is  $K = 200$ , active users is  $k = 20$  and orthogonal subcarriers is  $N = 100$ , so the overloading factor is 200%. The simulation results are shown in the following Fig.1 and Table.1:



**Fig.1.** Comparison of performance between BER and SNR in uplink grant-free NOMA system

**Table 1.** Experimental data

BER	SNR(dB)		
	OMP-based MUD	TOMP-based MUD	Oracle ls
$10^{-1}$	—	5.4	6
$10^{-2}$	10.4	7.4	1.4
$10^{-4}$	—	9.8	6

In summary, it can be clearly seen from the above Fig.1 and Table.1 that when the SNR is less than 4 dB, the BER performance of the multi-user detection based on the traditional OMP algorithm is approximately the

same as the BER performance of the improved algorithm. However, when SNR increases gradually, the BER performance of the multi-user detection based on the improved algorithm is superior to the traditional OMP algorithm. For example, when SNR is  $10^{-2}$ , the multi-user detection based on the improved algorithm is 3 dB SNR gains over the traditional OMP. This is due to the temporal correlation of active user sets in several continuous time slots. In addition, we can also see that, compared to the oracle LS algorithm, when the SNR is  $10^{-4}$ , the TOMP-based MUD suffers from about 4dB performance loss compared to the Oracle LS. This is due to the uncertainty support estimation of the greedy recovery algorithm in CS when the SNR is lower.

## 6 Conclusions

In this paper, a novel optimization algorithm is proposed based on the MUD scheme of traditional CS classical orthogonal matching pursuit (OMP) in NOMA systems. Simulation results show that the proposed scheme can achieve better BER performance in uplink grant-free NOMA systems, a novel solution for MUD is provided.

## References

1. L. Dai, B. Wang, Y. Yuan, S. Han, Z. Wang. *Nonorthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends*, IEEE Commun. Mag., **53**, (2015), 74-81.
2. Z. Ding, L. Dai, H. V. Poor. *MIMO-NOMA design for small packet transmission in the Internet of Things*, IEEE Access, **4**, (2016), 1393-1405.
3. A. T. Abebe, C. G. Kang. *Iterative order recursive least square estimation for exploiting frame-wise sparsity in compressive sensing based MTC*, IEEE Commun. Lett., **20**, (2016), 1018-1021.
4. E. J. Candes, T. Tao. *Near optimal signal recovery from random projections: Universal encoding strategies*, IEEE Transactions on information theory, **52**, (2006), 5406-5425.
5. L. Gan. *Block compressed sensing of natural images*, Proceedings of 15th International Conference on Digital Signal Processing, (2007), 403-406.
6. B. Wang, L. Dai, T. Mir, et al, *Dynamic compressive sensing based multi-user detection for uplink grant-free NOMA*, IEEE Communications Letters, **20**, (2016), 2320-2323.
7. B. Shim, B. Song, *Multiuser detection via compressive sensing*, IEEE Commun. Lett., **16**, (2012), 972-974.
8. J. P. Hong, W. Choi, B. D. Rao. *Sparsity controlled random multiple access with compressed sensing*, IEEE Trans. Wireless Commun., **14**, (2015), 998-1010.