Two Stage Detection for Uplink Massive MIMO MU-SCMA Systems

Cuitao Zhu(✉), Ning Wei, Zhongjie Li, and Hanxin Wang

Hubei Key Laboratory of Intelligent Wireless Communication, College of Electronics and Information Engineering, South-Central University for Nationalities, Wuhan, China
Cuitaozhu@mail.scuec.edu.cn

Abstract. In this paper, we propose a two stage multiuser detection scheme: a linear pre-filtering and iteration removal based message passing algorithm (RM-MPA). As the first stage of the proposed detection, a linear pre-filtering based on Richardson method is proposed to avoid the complicated matrix inversion in an iterative way. Meanwhile, we also present a sub-optimum relaxation parameter to Richardson for lower-complexity. Then the RM-MPA is used for multiuser decoding, which compared the decoding advantages of users and sorted users according to decoding advantages. After the each iteration, the users with higher decoding advantages directly are decoded and removed. The removed users do not participate in the subsequent iterations, therefore, the complexity of subsequent iterations decrease gradually. Simulation results show that the proposed two stages multiuser detection can significantly reduce the computational complexity with better symbol error rate performance.

Keywords: Massive MIMO · Non-orthogonal multiple access
Message passing algorithm

1 Introduction

Future fifth generation (5G) wireless networks are expected to support massive number of connected devices, non-orthogonal multiple access (NOMA) has attracted much attention as an enabling technology [1]. SCMA, as one of the competitive NOMA schemes, has been proposed to address the above requirement. To further improve the spectral efficiency, SCMA can be combined with Massive MIMO technology [2]. However, low complexity and efficient multiuser (MU) detection is one of the vital issues for combining SCMA with Massive MIMO in 5G, which need to be further addressed.

Due to the sparse structure of SCMA spreading signature, several iterative multiuser detection schemes for uplink SCMA systems [3, 4], which are based on message passing algorithm (MPA), were proposed to efficiently approximate to the maximum a

This work was partially supported by the National Natural Science Foundation of China under grant No. 61671483 and the Natural Science Foundation of Hubei province under grant No. 2016CFA089.

X. Liu et al. (Eds.): ChinaCom 2018, LNICST 262, pp. 414–424, 2019.
https://doi.org/10.1007/978-3-030-06161-6_41

posteriori (MAP). Message passing schedule strategy plays an important role in MPA-based detection schemes, which influences the convergence rate. References [5–7] proposed various strategies to further reduce the complexity of the MPA detector. Reference [8] proposed a resource-selection based MPA detector, in which resources with well-conditioned channels are selected to perform the jointly Gaussian algorithm in order to achieve satisfactory performance with low complexity. However, the complexity of the improved methods based on MPA is still high for SCMA with massive MIMO.

To address this issue, we propose an iteration removal based message passing algorithm with linear pre-filtering, which included the sequentially operating two stages. The first is a linear pre-filtering based on Richardson method to avoid the complicated matrix inversion in each iteration. We also propose a suboptimum relaxation parameter to Richardson for lower-complexity. After linear pre-filtering, RM-MPA was used for multiuser decoding based on the filtered signals. We propose a simple and novel method to compare the decoding advantages of users and sort the users according to decoding advantages. Then, n users with higher decoding advantages were directly decoded and removed after each iteration. Moreover, the removed users did not participate in the subsequent iteration, so the complexity of subsequent iterations decreased gradually.

2 System Model

For an uplink multiuser massive MIMO-NOMA system, in which a single base station with a uniform linear array of N_r antennas serves J users with single-antenna, J users share K physical resource elements, the j-th user transmits binary bits are mapped into a K dimensional complex codeword x_j, $X_j = \left[x_{1,j}, x_{2,j}, \ldots, x_{K,j}\right]^T$, for an SCMA encoder, it selected from the corresponding SCMA codebook $\boldsymbol{\chi}_j$. We assume that the power of a codeword is normalized to be 1 for all users, i.e., $\left\|x_j\right\|^2=1$. The system model is depicted in Fig. 1.

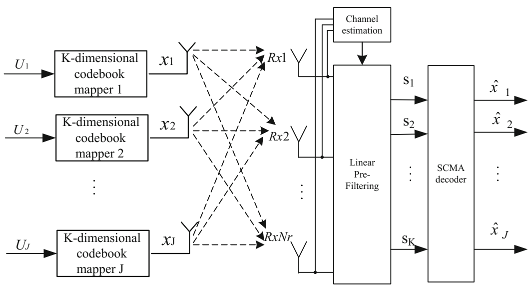

Fig. 1. System model

Since the SCMA codewords have sparse nature, the structure of SCMA encoder can be interpreted by an indicator matrix $F = [f_1, f_2, \ldots, f_J]$, in which $f_j = [f_{1,j}, f_{2,j}, \ldots, f_{K,j}]^T$ is the encoder indicator vector of user j, if $f_{k,j} = 1$, indicates k - th resource occupied by user j. Let d_f denotes the number of nonzero entries in each row of matrix F, i.e. the number of users sharing same resource, d_j denotes the number of nonzero entries in each column of matrix F, which corresponds to the number of resources allocated to the j-th user. In addition, let $\varsigma_j = \{k | f_{k,j} = 1\}$ be the set of resource, which indicates the resources allocated to the j - th user, and $\zeta_k = \{j | f_{k,j} = 1\}$ be the set of users, which indicates the users sharing the k-th resource, respectively. The K-dimensional codewords are transmitted to the receiver through K resources. The received signal at the k - th resource $y_k \in C^{N_r \times 1}$ can be expressed as follows:

$$y_k = H_k x_{k,[d_f]} + n_k, \quad k = 1, 2, \ldots, K \tag{1}$$

Where,

$$y_k = \left[y_k^1, y_k^2, \ldots, y_k^{N_r}\right]^T \tag{2}$$

$$H_k = \begin{bmatrix} h_{k,j_1}^1 & h_{k,j_2}^1 & \cdots & h_{k,j_{df}}^1 \\ h_{k,j_1}^2 & h_{k,j_2}^2 & \cdots & h_{k,j_{df}}^2 \\ & & \cdots & \\ h_{k,j_1}^{N_r} & h_{k,j_2}^{N_r} & \cdots & h_{k,j_{df}}^{N_r} \end{bmatrix} \tag{3}$$

$$x_{k,[df]} = \left[x_{k,j_1}, x_{k,j_2}, \ldots, x_{k,j_{df}}\right]^T \tag{4}$$

$$n_k = \left[n_k^1, n_k^2, \ldots, n_k^{N_r}\right]^T \tag{5}$$

In (2), the element y_k^p denotes the received signal by the p - th antenna from the k - th resource. In (3), the element $h_{k,j}^p$ is the channel gain between user j and the p - th antenna at the k - th resource, In (4), $x_{k,[d_f]}$ indicates superimposing users codeword at the k - th resource, and $n_k \sim CN(0, \sigma^2 I)$ denotes the Gaussian noise.

3 The Proposed Two Stage Detector

3.1 Low-Complexity Linear Pre-Filtering

Theoretical results for massive MIMO systems have shown that linear pre-filtering, such as matched filter (MF), and minimum mean square error (MMSE) filter, are able to achieve near-optimal performance in the large-antenna limit. However, these linear pre-filters involve troublesome matrix inversion. For this reason, we propose an approach based on Richardson method for a linear pre-filtering, and present a simple approach to determine the suboptimum relaxation parameter.

Let W_k be a linear filtering weighting matrix of size $N_r \times d_f, 1 \leq k \leq K$, where the entries depend on the perfectly known channel gains at the receiver. The weighting matrix of MMSE defined as follow:

$$W_k = \left(H_k^H H_k + \frac{d_f}{SNR} I_{d_f} \right)^{-1} H_k^H \tag{6}$$

In (6), I_{d_f} is an identity matrix of size $d_f \times d_f$. MMSE based linear filtering matrix was applied to (1), filtered signal model can be derived as

$$s_k = \left(H_k^H H_k + \frac{d_f}{SNR} I_{d_f} \right)^{-1} H_k^H y_k = P^{-1} \hat{y}_k \tag{7}$$

where, S_k is the filtered signal from k - th resource, $P = \left(H_k^H H_k + \frac{d_f}{SNR} I_{d_f} \right)$ is Hermitian positive definite matrix. Owing to the direct computation of the inverse P^{-1} requires high complexity, so we propose the Richardson method [9] to efficiently solve (7) to reduce the complexity. The method using the Richardson method to solve (7) is given by

$$s_k^{(i+1)} = s_k^{(i)} + \gamma \left(\widehat{y}_k - P s_k^{(i)} \right) \tag{8}$$

Where, $\mathrm{s}_\mathrm{k}^{(\mathrm{i})}$ denotes the solutions at the i - th iteration, γ is the relaxation parameter, which plays an important role in the convergence and convergence rate.

3.2 Selection of the Suboptimal Relaxation Parameter

In this section, we will deduce the suboptimum relaxation parameter based on approximation error. By formula (7) and (8), we can have the following equation:

$$\begin{aligned} s_k^{(i+1)} - s_k &= s_k^{(i)} + \gamma \left(\widehat{y}_k - P s_k^{(i)} \right) - s_k \\ &= (I_k - \gamma P) \left(s_k^{(i)} - s_k \right) \\ &= \cdots \\ &= B^{i+1} \left(s_k^{(0)} - s_k \right) \end{aligned} \tag{9}$$

Where, $B = I_K - \gamma P$ is the iteration matrix of Richardson algorithm, from (9) we can derive the approximation error induced by the Richardson algorithm, which can be evaluated as the following:

$$\left\| s_k^{(i+1)} - s_k \right\|_2 = \left\| B^{i+1} \right\|_F \left\| s_k^{(0)} - s_k \right\|_2 \tag{10}$$

From (10), the approximation error induced by Richardson algorithm is mainly affected by iteration matrix B and initial solution $S_k^{(0)}$. The relaxation parameter γ only affected by matrix B, if $\left\|B^{i+1}\right\|_F$ was minimized, the value of γ is optimum. However, the direct computation of $\left\|B^{i+1}\right\|_F$ is complicated, since the iteration matrix B is a random matrix and it is hard to obtain the joint distribution of all the elements. Fortunately, it has been proved in [9] that when i goes infinity, we have the following equation

$$\lim_{i \to \infty} \left\|B^{i+1}\right\|_F^{1/(i+1)} = \rho(B) \tag{11}$$

Where, $\rho(B)$ is the spectral radius of iteration matrix B. From (11), we can regard $\rho(B)$ as the asymptotic convergence rate of Richardson algorithm, and a smaller $\rho(B)$ will lead to a faster convergence rate. Let $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and the smallest eigenvalues of matrix P. According to the definition $B = I_K - \gamma P$, the spectral radius $\rho(B)$ will depend on γ as

$$\rho(B) = \max(|1 - \gamma\lambda_{\max}|\ ,\ |1 - \gamma\lambda_{\min}|) \tag{12}$$

From (12), when $|1 - \gamma\lambda_{\min}| = |1 - \gamma\lambda_{\max}|$, the smallest spectral radius can be achieved. Since γ should also satisfy $0 < \gamma < 2/\lambda_{\max}$, to guarantee the convergence of Richardson algorithm. So, we can conclude that when $\gamma\lambda_{\max} - 1 = 1 - \gamma\lambda_{\min}$, the smallest spectral radius $\rho(B)$ can be achieved and the fastest convergence rate, which means the optimal relaxation parameter γ_{opt} is

$$\gamma_{opt} = \frac{2}{\lambda_{\max} + \lambda_{\min}} \tag{13}$$

From (13), we can see that to obtain the optimal relaxation parameter γ_{opt}, we need to know a priori $\lambda_{\max}$ and $\lambda_{\min}$, which in practice is difficult. Therefore, directly using (13) to determine the optimum relaxation parameter γ_{opt} is very sophisticated in practical massive MIMO systems. To address this issue, we propose a suboptimal relaxation parameter γ'_{opt} with a negligible performance loss.

In massive MIMO systems, the elements of channel matrix $H_k \in C^{N_r \times d_f}$ are independent and identically distributed complex Gaussian random variables, the sample covariance matrix Φ is [10]

$$\Phi = \frac{1}{d_f} \sum_{v=1}^{d_f} h_v h_v^H = \frac{1}{d_f} H_k H_k^H \tag{14}$$

Where, h_v denotes the v column vector of matrix H_k, when $\frac{N_r}{d_f} \to \beta \in (0,\ \infty)$, according to the law of Bai-Yin, the largest and the smallest eigenvalues of matrix Φ meet as

$$\lim_{N_r \to \infty} \lambda'_{\max}(\Phi) = \sigma^2\left(1+\sqrt{\beta}\right)^2 \tag{15}$$

$$\lim_{N_r \to \infty} \lambda'_{\min}(\Phi) = \sigma^2\left(1-\sqrt{\beta}\right)^2 \tag{16}$$

Where, σ_z^2 is the noise power. In massive MIMO systems, the number of antennas N_r is very large, the smallest and the largest eigenvalues of matrix P will converge to the deterministic values as

$$\lambda_{\max} \approx d_f\left(1+\sqrt{\beta}\right)^2 + \frac{d_f}{SNR} \tag{17}$$

$$\lambda_{\min} \approx d_f\left(1-\sqrt{\beta}\right)^2 + \frac{d_f}{SNR} \tag{18}$$

By replacing (17) and (18) in (13), the suboptimum relaxation parameter γ'_{opt} is given by

$$\gamma'_{opt} = \frac{2}{d_f\left[\left(1+\sqrt{\beta}\right)^2 + \left(1-\sqrt{\beta}\right)^2\right] + 2\frac{d_f}{SNR}} \tag{19}$$

From (19), we can observe that γ'_{opt} depends on the number of receive antennas N_r and d_f, which are deterministic and known after the massive MIMO configuration has been fixed. Thus, we do not need to re-compute γ'_{opt} as H varies. Furthermore, γ'_{opt} does not need to know a priori $\lambda_{\max}$ and $\lambda_{\min}$ in comparison to γ_{opt} given by (13).

4 Low Complexity SCMA Decoder Based on Iteration Removal

4.1 Decoding Advantage Analysis

The procedure of RM-MPA can be explained by the factor graph, which is a bipartite graph including resource nodes r_k and user nodes u_j. In general, MPA consists of the exchange of messages between the nodes of a factor graph. There are two approaches for scheduling messages in MPA, i.e., parallel schedule strategy and serial schedule strategy. In this paper, we tackle the problem of how to schedule message for the serial schedule strategy that results in effectively reduced complexity and the best convergence rate.

Let $I_{r_k \to u_j}(x_j)$ and $I_{u_j \to r_k}(x_j)$ be the message propagated along branch from r_k to u_j, and from u_j to r_k, respectively. In the each iteration, messages are first sent from user nodes to resource nodes. Each resource nodes then computes extrinsic messages and sends back to the user nodes based on the previously received information. These user nodes-to-resource nodes messages will then be used to calculate the new resource nodes-to-user nodes messages in the next iteration. Thus, the two messages generating functions are defined as follows:

$$I^t_{r_k \to u_j}(x_j) = \sum_{\sim x_j} \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left\| y_k - \sum_{m \in \zeta_k} h_{k,m} x_{k,m} \right\|^2 \right) \times \prod_{\substack{l \in \zeta_k/j \\ l<j}} I^t_{u_l \to r_k}(x_l) \prod_{\substack{l \in \zeta_k/j \\ l>j}} I^{t-1}_{u_l \to r_k}(x_l) \right\} \tag{20}$$

$$I^t_{u_j \to r_k}(x_j) = normalize\left(\prod_{m \in \varsigma_j/k} I^t_{r_m \to u_j}(x_j) \right) \tag{21}$$

Where, $l<j$ and $l>j$ represent the user l before and after the user j, respectively. The $\zeta_{k/j}$ denotes to remove the user j from the set ζ_k, t is the iteration index, at the maximum iterations $t_{\max}$, the decoding soft output can be expressed as

$$Q(x_j) = \prod_{k \in \varsigma_j} I^{t_{\max}}_{r_k \to u_j}(x_j) \tag{22}$$

As mentioned above, the computational complexities of the original MPA based on serial schedule strategy, relate to $t_{\max} \cdot KM^{d_f}$. Furthermore, for original MPA, we can see that codewords of all users are decoded together until the maximum iterations $t_{\max}$ is reached. Thus, the schedule strategy for message update has a disadvantage that all users participate in the process of the each iteration, and the parameter d_f is same in the each iteration, this leads to high computational complexity in the each iteration. To tackle this issue, we present a low complexity MPA detector based on iteration removed.

Owing to the proper property of message passing serial schedule strategy, RM-MPA based on serial schedule strategy can decode and remove the users with higher decoding advantages after the first iteration. Furthermore, when the resources are handled later, the users have the higher reliability of the message carried by them. Thus, later users occupy resources, which have the higher reliability of the decoding soft output in the each iteration. So, there is an inherent decoding advantage between users in RM-MPA based multiuser detection schemes. Let $e^t_{k,j}(x_j)$ denotes the number of external message used by the user j, which occupied the k-th resource update message at the t iterations. That can be defined as follow

$$e^t_{k,j}(x_j) = \left\{ n = \sum f_{l,j'} | f_{k,j'} = 1 \& f_{l,j'} = 1\ ,\ j' \neq j, l<k \right\} \tag{23}$$

From (23), the user j utilizes the total number of external messages produced in this iteration at the t iterations.

$$e_j^t(x_j) = \sum_{k \in \varsigma_j} e_{k,j}^t(x_j) \tag{24}$$

From (24), the bigger $e_j^t(x_j)$, the more number of external messages utilized by user j in this iteration. Then, user j has higher decoding reliability of after this iteration, and has higher decoding advantage to other users. According to the sequence of resources being processed in the RM-MPA, we will define an advantage level for each resource. The higher level, the more decoding advantage, decoding reliability is greater. For instance, resource k, the corresponding advantage level is l_k, then, the decoding advantage level of user j defined as

$$a_j = \sum_{k \in \varsigma_j} f_{k,j} \cdot l_k \tag{25}$$

According to (25), users sorted by decoding advantage in descending order, then the order of decoding reliability of users can be obtained after the each iteration, and decode and remove the decoding reliable users, then do next iteration.

4.2 Iteration Removed Strategy

(1) Iteration Removed Strategies Based on Decoding Advantages of Users

As mentioned above, greater decoding advantage of users, the higher decoding reliability in RM-MPA-based detection schemes. Therefore, users can be sorted according to decoding advantage from high to low. After each iteration, the n users with higher decoding advantages were directly decoded and removed, and these users no longer to participate in the next iteration. Thus, after the each iteration the number of users is cut down n, $1 \leq n \leq K/N$. After $t_{\max}$ iterations, no matter how much remaining user, all decoded together. Thus, it can be seen that more users removed, the lower complexity of decoding algorithm after the each iteration.

(2) Iteration Removed Strategies Based on Orthogonal User Grouped

In the system with SCMA, if each user occupies a data stream layer, then there are maximum users C_K^N, meanwhile, there are maximum C_K^N indicator vectors f_j in the indicator matrix F. If indicator vectors are grouped according to mutually orthogonal, all indicator vectors can be divided into $C_K^N/(K/N) = C_{K-1}^{N-1}$ groups. In this paper, we assume that K can be divisible by N. There are K/N indicator vectors in each group, and the indicator vector f_j of users is mutually orthogonal in a group. If there are same number of users in each group, and users occupy all resources, this grouping is called an orthogonal complete user group, otherwise, called an orthogonal non-complete.

On former orthogonal users grouped, iteration decoding and removal based on Orthogonal users grouped describe as follow: optionally selecting a group of orthogonal user after the each iteration, the users in a group are directly decoded and removed, and these users no longer to participate in the next iteration. To an orthogonal complete user group, selected randomly an orthogonal users group, then the users are decoded and removed.

5 Numerical Results

This section presents the performances of the proposed linear pre-filtering and multiuser decoding schemes based on decoding advantages of users (RM-MPA-URS) and Orthogonal user grouped (RM-MPA-OUGR) for uplink massive MIMO MU-SCMA systems.

Figure 2 shows the Symbol Error Rate (SER) performance of the MMSE linear pre-filtering based on Richardson algorithm and Neumann series, and exact matrix inversion method over Rayleigh fading channels. The SCMA system is with 16-QAM, $\beta = 8$, $N_r = 64$, $d_f = 8$.

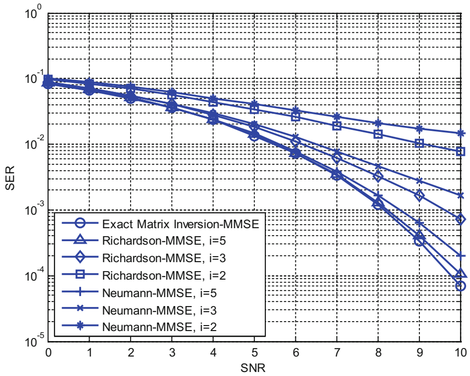

Fig. 2. SER performance comparison

We can see from the Fig. 2 that the SER performance of the MMSE linear pre-filtering based on Richardson algorithm outperforms Neumann series with 5 iterations, 2 iterations and 3 iterations, and the more advantage with increase of SNR. On the other hand, the SER performance of the MMSE linear pre-filtering based on Richardson algorithm and exact matrix inversion method have become much closer with the increase of iterations, however, the computational cost of the exact matrix inversion method is higher than that in Richardson algorithm.

Figure 3 shows the SER performance of the two stages detections: MMSE based on exact matrix inversion method with PMPA, and MMSE based on Richardson with PMPA, and MMSE based on Richardson with RM-MPA. The SCMA system with 16-QAM, $N_r = 64$, $J = 48$, over Rayleigh fading channels.

We can see from the Fig. 3 that the SER performance of the exact matrix inversion-MMSE-PMPA is the best with 10 iterations. However, the exact matrix inversion-

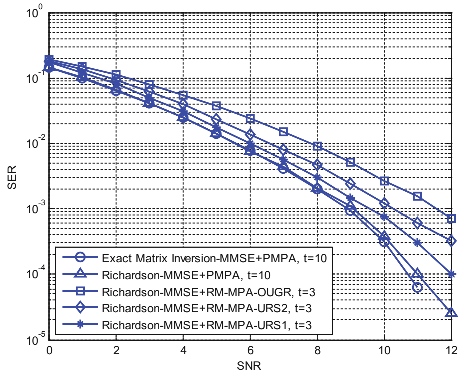

Fig. 3. SER performance comparison of two stages detections

MMSE-PMPA requires higher computational costs. The SER performance of Richardson-MMSE-PMPA is lower than that obtained by the exact matrix inversion-MMSE-PMPA. Meanwhile, for the SER performance comparison of Richardson-MMSE-RM-MPA with two removed strategies with 10 iterations, Fig. 3 shows the SER performance of the Richardson-MMSE-RM-MPA-URS is the best. However, Richardson-MMSE-RM-MPA-OUGR requires the lowest computational cost.

6 Conclusions

In this paper, we proposed a novel Richardson-MMSE-RM-MPA detector for uplink SCMA systems with massive MIMO. In order to detect the transmitted signals of users with lower computational cost, the proposed detection sequentially employs a linear pre-filtering and iteration removal based on Message Passing Algorithm. Simulation results show that the Richardson-MMSE-RM-MPA-OUGR detector can observably reduce the computational complexity with the performance degraded unnoticeably.

References

1. Dai, L., Wang, B., Yuan, Y., Han, S., Chih-Lin, I., Wang, Z.: Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. IEEE Commun. Mag. **53**(9), 74–81 (2015)
2. Liu, T.T., Li, X.M., Qiu, L.: Capacity for downlink massive MIMO MU-SCMA system. In: *Proceedings of the* IEEE WCSP, pp. 1–5. Nanjing, China (2015)
3. Du, Y., Dong, B.H., Chen, Z., Fang, J., Wang, X.J.: A fast convergence multiuser detection scheme for uplink SCMA systems. IEEE Wirel. Commun. Lett. **5**(4), 388–391 (2016)
4. Du, Y., et al.: Low-complexity detector in sparse code multiple access systems. IEEE Commun. Lett. **20**(9), 1812–1815 (2016)
5. Du, Y., Dong, B., Chen, Z., Fang, J., Yang, L.: Shuffled multiuser detection schemes for uplink sparse code multiple access systems. IEEE Commun. Lett. **20**(6), 1231–1234 (2016)

6. Du, Y., Dong, B., Chen, Z., Fang, J., Gao, P., Liu, Z.: Low-complexity detector in sparse code multiple access sytems. IEEE Commun. Lett. **20**(9), 1812–1815 (2016)
7. Wei, F., Chen, W.: Low complexity iIterative receiver design for sparse code multiple access. IEEE Trans. Commun. **65**(2), 621–634 (2016)
8. Tian, L., Zhao, M., Zhong, J., Wen, L.: Resource-selection based low complexity detector for uplink SCMA systems with multiple antennas. IEEE Wirel. Commun. Lett. (2017)
9. Bjorck, A.: Numerical Methods in Matrix Computations. Texts in Applied Mathematics (2015)
10. Bai Z.D.: Methodologies in spectral analysis of large dimensional random matrices, a review [m]// Advances In Statistics. (2015)