K-means clustering algorithm in antenna selection for Massive MIMO

Xiao Xuefang\textsuperscript{1,2} Lei Guowei\textsuperscript{3,4} Lin Yun\textsuperscript{1,2}

\textsuperscript{1}School of optoelectronics and communication engineering, Xiamen university of Technology, Xiamen 361024, China. \textsuperscript{2}Key Laboratory of Communication Network and Information Processing in Fujian Province, Xiamen 361024, China

\textsuperscript{3}School of electronic engineering, Beijing university of posts and telecommunications, Beijing 100876, China. \textsuperscript{4} School of science, Jimei university, Xiamen 361021, China

xfxiao@xmut.edu.cn gwlei@163.com

Abstract. Massive multiple-input multiple-output (MIMO) is considered as the promising technique in next generation of wireless communication system. It has high spectral efficiency and energy efficiency. In addition, it can mitigate the interferences among users as the amount of antennas at base station (BS) grows. But, the high complexity and cost of hardware pose huge challenges to massive MIMO. In future, antenna selection (AS) is still a choice to decrease the burden of BS. In this letter, the AS algorithm based on K-means clustering is proposed to maximizing the capacity with low computational complexity. Finally, the simulated results are presented to validate the proposed method.

1. Introduction

Massive multiple-input multiple-output (MIMO) is regarded as one of the key technologies of the next generation of wireless communication systems. It employs a large number of antennas at the base station (BS) to serve certain number of users. As the number of antennas at BS grows without limit, all of the effects of uncorrelated noise and fast fading disappear [1]. Typically, a base station with a large number of antennas can serve several single-antenna users in the same time-frequency resource. Therefore, it has high spectral efficiency and energy efficiency.

These benefits, however, come at the cost of a dramatic increase in hardware and computational complexity. Actually, each antenna element at BS is connected to a single RF chain, which is composed of an amplifier, an analog-to-digital converter, and mixers. These elements make RF chains generally expensive and particularly power demanding. Recently, some antenna selection (AS) methods in typical MIMO are extended to massive MIMO [2]-[5]. The AS algorithm in [2]-[3] is developed from traditional square maximum-volume submatrices method, in which the number of selected transmit antenna is equal to the number of users. In general, the AS algorithm is given based on channel capacity[4].

Theoretically, the capacity of MIMO channels increases linearly with the minimum number of transmit and receive antennas while assuming that the channel is independently faded and the transmission power is fixed [2]. However, spatial correlation can decrease the capacity of feasible system due to insufficient antenna spacing and scattering [5]. This letter aims at the effects of channel correlation, and investigates the decrease of channel correlation in AS massive MIMO system.
Moreover, the K-means clustering algorithm is proposed in AS massive MIMO system. Simulated results are given finally to validate the proposed method.

2. System model and algorithm design

Consider a massive MIMO system with $N_t$ antennas at BS side and $K$ ($N_t > J$) users with single antenna in a rich-scattering environment where users are randomly located. Since AS is employed at BS generally, we assume $L_t$ antennas selected out of $N_t$ in a downlink situation. Then the signal model can be expressed as

$$r = \sqrt{\rho / L_t} H_d s + w \quad (1)$$

where $r$ is the $J \times 1$ signal vector, $s$ is the $N_t \times 1$ signal vector, $w$ is the added white Gaussian noise, $H_d$ is the $J \times L_t$ channel matrix, and $\rho$ is the average signal-to-noise ratio (SNR) at receiver. The ergodic capacity in downlink massive MIMO is given as [6]

$$C_{DL} = \log_2 \left[ \text{det}(I_{L_t} + \frac{\rho}{L_t} H_d^H H_d) \right] \quad (2)$$

The overall channel $H \in \mathbb{C}^{J \times N_t}$ is assumed to be quasi-static fading and is expressed as

$$H = \left[ H_1, \ldots, H_{N_t} \right]$$

$$= \begin{bmatrix}
    h_{1,1} & \cdots & h_{1,N_t} \\
    \vdots & \ddots & \vdots \\
    h_{J,1} & \cdots & h_{J,N_t}
\end{bmatrix} \quad (3)$$

In (3), $H_i$ ($i=1, \ldots, N_t$) stands for the column vector, i.e. $H_i = [h_{1,i}, h_{2,i}, \ldots, h_{J,i}]'$. Since $H$ can be rewritten as

$$H = R^{1/2} \tilde{H} T^{1/2} \quad (4)$$

where $\tilde{H}$ has i.i.d. circular symmetric complex Gaussian entries, with independent real and imaginary parts, each with variance $1/2$, and $R$ is receive covariance matrix and $T$ is transmit covariance matrix [7].

As correlation matrices $R$ and $T$ can decrease the capacity of massive MIMO system. First of all, we aim to decrease the correlation of matrices. In this paper, we propose the K-means clustering algorithm in AS for massive MIMO. K-means clustering is originally popular for cluster analysis in data mining. It aims to partition observations into $K$ clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. To illustrate this procedure, we assume three users ($J=3$) in downlink massive MIMO, and select four antennas ($L_t=4$) to serve these users. As seen in figure 1, these users construct the Cartesian coordinates with $J$-dimensions. There are $N_t$ star points distributed in this Cartesian coordinates space. Each star point stands for $H_i$ ($i=1, \ldots, N_t$) in (3). To select four antennas from $N_t$ antennas, we can partition these star points into four clusters, e.g. $K=1$, $K=2$, $K=3$, and $K=4$. As shown in (5), the criterion of clustering is according to the dot product between $H_i$ and $H_c$ (the centroid of its cluster).

$$d_j = \langle H_j, H_c \rangle / \sqrt{\langle H_j, H_j \rangle \times \langle H_c, H_c \rangle} \quad (5)$$
As aforementioned, any $H_i$ and $H_j$ from two different clusters should have least correlation. Hereby, the whole process of AS for downlink massive MIMO is given as follows:
(A) Generate $K$ ($K=L_t$) different $H_c$
(B) Calculate the dot products between $H_i$ and $H_c$ based on (5). From these dot products, partition $H_i$ with the largest $d_i$ and group them into four clusters.
(C) Calculate the centroid of each cluster, substitute $H_c$ with new one.
(D) Repeat above steps (A),(B),(C) until the centroid of each cluster changes little.
(E) From each cluster, find the one with largest 2-norm, and construct a new $H_d$ with $K$ submatrices $H_i$.

Furthermore, one way to evaluate the orthogonality of all users is singular value spread of the propagation matrix. The propagation matrix $H_d$ has a singular value decomposition (SVD) [6]

$$H_d=U\Sigma V^H$$

(6)

Where $U$ and $V$ are unitary matrices, and $\Sigma=\text{diag}\{\lambda_1,\ldots,\lambda_J\}$ is the $J\times L_t$ diagonal matrix, and the singular value spread is defined as

$$\kappa = \frac{\max\{\lambda_i\}}{\min\{\lambda_i\}}$$

(7)

Simulated results: We examine the AS algorithms via simulated results. We turn our attentions to ergodic capacity and singular value spread. First, it is assumed that the channel is polluted by additive white Gaussian noise (AWGN).

Fig.2 depicts the ergodic capacity versus SNR. It is further assumed that the number of users is $J=4$, the number of selected antennas at BS is $L_t=4$, and the number of total antennas is $N_t=64$. It is shown that the Rand AS algorithm has the lowest capacity. The K-means AS algorithm performs better than RMV AS algorithm. The optimal AS algorithm performs best due to its exhaustive search.
Fig. 2 Ergodic Capacity versus signal-to-noise ratio (SNR).

With aforementioned assumptions, we evaluate the cumulative distribution functions versus singular value spread. It is shown in Fig. 3 that, singular value spread with K-means AS algorithm is larger than those with RMV and Rand AS algorithms. Moreover, singular value spread with optimal AS algorithm turns to be more stable.

Fig. 3 Cumulative distribution functions versus singular value spread.

Table 1 is given as the comparison of computational complexity for various algorithms. Optimal AS is to find the optimal subset that yields biggest capacity from all possible subsets. The algorithm is exhaustive because there are $|\Theta|=C_{NL}^{L_r} \times C_{NL}^{L_t}$ subsets in all. Rand AS is to choose from the total $C_{NL}^{L_r} \times C_{NL}^{L_t}$ subsets at random. It is a fast algorithm as it does not need any computation. RMV AS requires the inversion and refreshment of matrix, so the complexity is given by $O(NtK(K-J))$. The complexity of K-means AS largely depends on the partition and clustering. The number of iterations is denoted as $T$, which is often small until convergence. Hence, the complexity of K-means AS is given by $O(N, KT)$.

| Table 1: The computational complexity for various algorithms. |
|---------------------------------|-----------------|-----------------|-----------------|
| Optimal                        | Rand            | RMV [3]         | K-means         |
| $C_{NL}^{L_r} \times C_{NL}^{L_t}$ | $O(1)$          | $O(NtK(K-J))$   | $O(N, KT)$      |

3. Conclusion

In this paper, the K-means clustering algorithm has been introduced. Based on K-means clustering, a new AS algorithm has been proposed in AS for downlink massive MIMO. It can be seen from simulated results that the proposed algorithm should have performances close to optimal AS with low computational complexity. The study can also be extended to any number of AS in downlink or uplink massive MIMO system.
Acknowledgments
This work is supported by Capital of Xiamen university of technology under Grant ‘D2018005’, ‘Foundation project of department of education of Fujian Province’ (Grant JAT160260), and ‘Li Shangda Discipline Construction Fund’ (Grant ZC2016008) of Jimei University.

References
[1] Marzetta, T.L. Noncooperative cellular wireless with unlimited numbers of base station antennas[J]. IEEE Trans. Wireless Communications, 2010, 9(11): 3590–3600.
[2] Bing Fang, Zuping Qian, et al. RAISE: A new fast transmit antenna selection algorithm for massive MIMO systems[J]. Wireless personal communications, 2015, 80 (3):1147–1157.
[3] Tang H. and Nie Z.P. RMV Antenna selection algorithm for massive MIMO[J]. IEEE Signal processing letters, 2018, 25(2):239-242.
[4] Gao Y., Vinck H. et al. T.Massive MIMO antenna selection: switching architectures, capacity bounds, and optimal antenna selection algorithms[J].IEEE Signal processing letters, 2018, 66(5):1346-1360.
[5] Al-Hussaibi W.A. and Falah H.A. A closed-form approximation of correlated multiuser MIMO ergodic capacity with antenna selection and imperfect channel estimation[J]. IEEE Transactions on vehicular technology, 2018, 67(6):5515-5519.
[6] Gao X., Edfors O.,et al. Massive MIMO performance evaluation based on measured propagation data[J]. IEEE Transactions on wireless communications, 2015, 14(7):2899-3911.
[7] Hanlen L. and Grant A. Capacity analysis of correlated MIMO channels[J]. IEEE Transactions on information theory, 2012, 58(11):6773-6787.