HOSVD-based Limited Feedback and Precoding Design for Massive MIMO Systems

Yu Zhu, Yafei Tian, Chenyang Yang
School of Electronics and Information Engineering, Beihang University, Beijing 100191, China
Email: {yuzhu, ytian, cyyang}@buaa.edu.cn

***Abstract*—Massive multiple-input multiple-output (MIMO) systems hold the potential to achieve high spectral efficiencies in future wireless systems. As the number of antennas of the base station (BS) increases, downlink channel quantization for frequency division duplex (FDD) based systems becomes difficult, and most of the works on massive MIMO assume time division duplexing (TDD) to sidestep these challenges by employing channel reciprocity. In this paper, we propose to apply the higher-order singular value decomposition (HOSVD) of the MIMO channel matrix and design a new limited feedback and precoding scheme, which aims at solving the problem of high dimensionality led by large antenna array. The proposed scheme can reduce the complexity of channel quantization, the number of bits for feedback and the computational complexity of precoding matrix. Simulation results demonstrate that the performance of the proposed scheme is nearly the same with the singular value decomposition (SVD) based scheme in single-user MIMO and block diagonalization zero-forcing scheme in multi-user MIMO when using quantized channel feedback.**

***Index Terms*—CSI quantization, limited feedback, massive MIMO, precoding, tensor decomposition.**

I. Introduction

Massive multiple-input multiple-output (MIMO) systems are promising to meet the ever-growing data demand in 5G cellular networks [1]. In contrast to time division duplexing (TDD), which can obtain downlink channel state information (CSI) by channel estimation in uplink, frequency division duplex (FDD) systems are challenging to implement with massive MIMO because most previous solutions for channel estimation and quantization become impractical as the number of antennas increases. However, most current cellular networks are based on FDD. Considering the backward compatibility, it is important to solve the CSI estimation and feedback problem for FDD-based massive MIMO systems.

In massive MIMO systems, with a large number of antennas installed within a limited space at the base station (BS), the channels are likely to be spatially correlated. The spatial correlation can be exploited to reduce the feedback overhead [2]. There are some existing works dedicated to the dimension reduction in the limited feedback. The singular value decomposition (SVD) is employed to alleviate the dimension problem by focusing on the dominant eigenvectors of the spatial covariance matrix [3]. Considering the limitation of physical size, placing the massive antennas in two-dimensional (2D) grid is an effective way for its commercialization, which is termed as "three-dimensional (3D) MIMO" [4]. Inspired by 3D-MIMO system with uniform planar array (UPA) antennas, some works have considered to deal with horizontal and vertical directions independently. [5] has revealed that the 3D correlation matrix can be well approximated by a Kronecker production of azimuth and elevation correlations. This approximation makes the Kronecker-product codebook possible, which is an efficient way of quantizing the channel with UPA antennas. [6] has shown that the Kronecker-product codebook can quantize the channel vector without any assumption on the decomposition of the channel correlation matrix.

When we consider the scene where the transmitter is equipped with a UPA and the receiver is also equipped with multiple antennas, it's natural for us to think of using a tensor to describe the channel. Then we can introduce some existing operations of tensors, such as CANDECOMP/PARAFAC decomposition and Tucker decomposition for dimensionality reduction. We are used to using a vector or a matrix to describe the channel. A vector and a matrix are actually a first-order and a second-order tensor, respectively. Hence tensor is a more general expression and the unfolding operations of tensors make it easy to convert from tensors to matrices. In [7], Tucker decomposition is applied in the rotation matrix required by the rotated codebook to avoid the problem of high dimensionality in 3D-MIMO systems.

However, most previous works have focused on multiple-input single-output (MISO) only. In this paper, we consider downlink MIMO systems where users are equipped with multiple antennas and can receive multiple data streams simultaneously. We propose to apply the Tucker decomposition of the MIMO channel matrix and design a new limited feedback and precoding scheme. We can compute the Tucker decomposition through higher-order singular value decomposition (HOSVD), hence we call the proposed algorithm "HOSVD-based limited feedback and precoding". The proposed scheme can reduce the complexity of channel quantization, the number of bits for feedback and the computational complexity of precoding matrix.

The paper is organized as follows. We explain the system model and some basic knowledge of tensor in Section II. In Section III, we propose the HOSVD-based limited feedback and precoding. Simulations are shown in Section IV, and conclusions follow in Section V.

Notations: Scalars are denoted by lowercase letters, vectors by lowercase boldface letters, matrices by uppercase boldface letters, and higher-order tensors by calligraphic letters. $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ are respectively the transpose, Hermitian and conjugate operation. The Kronecker, outer, and n-mode products are denoted by the symbols $\otimes$, $\circ$, and $\times_n$, respectively. The Frobenius norm, statistical expectation are denoted by $\|\cdot\|$ and $\mathbb{E}\left[\cdot\right]$, respectively.

II. System Model and Tensor Prerequisites

A. System model

We consider a massive MIMO downlink system where a BS services K users. The BS is equipped with N_t antennas and the user k is equipped with $N_{r,k}$ antennas. The total number of antennas at all users is $N_r = \sum_{k=1}^{K} N_{r,k}$. The transmit signal of user k is denoted by $\mathbf{s}_k \in \mathbb{C}^{L_k}$, where L_k is the number of streams. $\mathbf{P}_k \in \mathbb{C}^{N_t \times L_k}$ denotes the precoding matrix of user k. The channel matrix from the BS to the user k is denoted by $\mathbf{H}_k \in \mathbb{C}^{N_{r,k} \times N_t}$. The received signal of user k can be written as

$$\mathbf{y}_k = \sum_{i=1}^{K} \mathbf{H}_k \mathbf{P}_i \mathbf{s}_i + \mathbf{n}_k \tag{1}$$

where $\mathbf{n}_k \in \mathbb{C}^{N_{r,k}}$ is the additive complex Gaussian noise.

B. Channel model

Consider a clustered channel model [8]. The channel matrix is defined as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{il} \mathbf{a}_r\left(\phi_{il}^r, \theta_{il}^r\right) \mathbf{a}_t\left(\phi_{il}^t, \theta_{il}^t\right)^H \tag{2}$$

where N_{cl} and N_{ray} denote the number of clusters and the number of rays in each cluster, and α_{il} denotes the gain of the lth ray in the ith cluster. We assume that α_{il} is independent and identically distributed (i.i.d) and follows the distribution $\mathcal{CN}\left(0, \sigma_{\alpha,i}^2\right)$ and $\sigma_{\alpha,i}^2$ is the average power of the ith cluster. $\mathbf{a}_r\left(\phi_{il}^r, \theta_{il}^r\right)$ and $\mathbf{a}_t\left(\phi_{il}^t, \theta_{il}^t\right)$ represent the receive and transmit array response vectors respectively, where ϕ_{il}^t and θ_{il}^t denote azimuth and elevation angles of departure (AODs), ϕ_{il}^r and θ_{il}^r denote azimuth and elevation angles of arrival (AOAs). We assume that the antenna arrays in the transmitter and receiver are uniform linear array (ULA) or UPA. The antenna spacing is denoted by d. $N_t = N_{tv} \times N_{th}$ is the number of transmit antennas and $N_r = N_{rv} \times N_{rh}$ is the number of receive antennas, where N_{xv} and N_{xh} are the number of rows and columns, respectively, $x \in \{t, r\}$. If $N_{xv} = 1$ or $N_{xh} = 1$, the antenna array is ULA. Otherwise, the antenna array is UPA. Therefore, the ith element of array response vector can be written as

$$\mathbf{a}_x\left(\phi, \theta\right)_i = \frac{1}{\sqrt{N_x}} \exp\{j\frac{2\pi}{\lambda} d\left(p \sin\phi^x \sin\theta^x + q \cos\theta^x\right)\} \tag{3}$$

where λ is the signal wavelength, $0 \le p < N_{xh}$ and $0 \le q < N_{xv}$ are the antenna indices.

C. Tensor prerequisites

A tensor is a multidimensional array. The order of a tensor is the number of dimensions. For instance, $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is an Nth order tensor whose elements are denoted by $x_{i_1 i_2 \cdots i_N} = [\mathcal{X}]_{i_1 i_2 \cdots i_N}$ where $i_n \in \{1, \ldots, I_n\}$, $n = 1, 2, \ldots, N$. Fibers are the higher order analogue of matrix rows and columns. A fiber is defined by fixing every index but one. The mode-n unfolding of $\mathcal{X}$ is denoted by $\mathbf{X}_n$ and arranges the mode-n fibers to be the columns of the matrix.

The n-mode product of $\mathcal{X}$ with a matrix $\mathbf{U} \in \mathbb{C}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n \mathbf{U}$ and is of size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$, whose element is

$$\left(\mathcal{X} \times_n \mathbf{U}\right)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} u_{j i_n} \tag{4}$$

The tucker decomposition is a form of higher-order principal component analysis. It decomposes a tensor into a core tensor multiplied by a matrix along each mode, which is

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} \tag{5}$$

$$= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} g_{i_1 i_2 \cdots i_N} \mathbf{a}_{i_1}^{(1)} \circ \cdots \circ \mathbf{a}_{i_N}^{(N)} \tag{6}$$

where $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times I_n}$, $n \in \{1, 2, \cdots, N\}$, are the factor matrices and can be thought of as the principal components in each mode. $\mathbf{a}_{i_n}^{(n)}$ is the i_nth column in the $\mathbf{A}^{(n)}$. The tensor $\mathcal{G} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is the core tensor and the entry $g_{i_1 i_2 \cdots i_N}$ shows the level of interaction between the different components $\mathbf{a}_{i_1}^{(1)}, \cdots, \mathbf{a}_{i_N}^{(N)}$ [9].

III. HOSVD-based Limited Feedback and Precoding

In this section, we first introduce a SVD-based precoder, which is used for comparison. Then we introduce the HOSVD-based limited feedback and precoding and employ it for multiuser MIMO (MU-MIMO).

A. SVD-based precoder

The SVD of the channel matrix is defined as

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \tag{7}$$

where $\mathbf{U}$ is the left singular matrix, $\mathbf{\Sigma}$ is the diagonal matrix consists of the singular values in a decreasing order, and $\mathbf{V}$ is the right singular matrix. If the number of streams is L, the SVD-based precoder is the leading L columns of $\mathbf{V}$, weighted by water-filling on the corresponding singular values.

B. HOSVD-based scheme

1) Tucker decomposition of channel matrix: For convenience, we consider a single user first. The channel matrix is denoted by $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$. We can define a $(M+1)$th order tensor $\mathcal{H} \in \mathbb{C}^{N_r \times \prod_{m=1}^{M} N_{t,m}}$ and $\mathbf{H}$ is the mode-1 unfolding of $\mathcal{H}$, i.e., $\mathbf{H} = \mathbf{H}_{(1)}$, where $N_t = \prod_{m=1}^{M} N_{t,m}$. The Tucker decomposition of $\mathcal{H}$ can be written as

$$\mathcal{H} = \mathcal{G} \times_1 \mathbf{B} \times_2 \mathbf{A}^{(1)} \times_3 \mathbf{A}^{(2)} \cdots \times_{M+1} \mathbf{A}^{(M)} \tag{8}$$

$$= \sum_{i=1}^{N_r} \sum_{j_1=1}^{N_{t,1}} \cdots \sum_{j_M=1}^{N_{t,M}} g_{ij_1 \cdots j_M} \mathbf{b}_i \circ \mathbf{a}_{j_1}^{(1)} \circ \cdots \circ \mathbf{a}_{j_M}^{(M)} \tag{9}$$

where $\mathcal{G} \in \mathbb{C}^{N_r \times \prod_{m=1}^{M} N_{t,m}}$ is the core tensor, $\mathbf{B} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{A}^{(m)} \in \mathbb{C}^{N_{t,m} \times N_{t,m}}$, $m \in \{1, 2, \cdots, M\}$, are all unitary matrices. $\mathbf{b}_i$ denotes the ith column of $\mathbf{B}$ and $\mathbf{a}_{j_m}^{(m)}$ denotes the j_mth column of $\mathbf{A}^{(m)}$. The mode-1 unfolding of $\mathcal{H}$ can be written as

$$\mathbf{H} = \mathbf{B}\mathbf{G}_{(1)} \left(\mathbf{A}^{(M)} \otimes \mathbf{A}^{(M-1)} \otimes \cdots \mathbf{A}^{(1)} \right)^T \tag{10}$$

$$= \sum_{i=1}^{N_r} \sum_{j_1=1}^{N_{t,1}} \cdots \sum_{j_M=1}^{N_{t,M}} g_{ij_1 \cdots j_M} \mathbf{b}_i \left(\mathbf{a}_{j_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{j_1}^{(1)} \right)^T \tag{11}$$

where $\mathbf{G}_{(1)}$ is the mode-1 unfolding of $\mathcal{G}$.

We can compute the Tucker decompositon of $\mathcal{H}$ through HOSVD, in which we set $\mathbf{B}$ as the left singular matrix of $\mathbf{H}_{(1)}$ and $\mathbf{A}^{(m)}$ as the left singular matrix of $\mathbf{H}_{(m+1)}$, $m \in \{1, 2, \cdots, M\}$. Additionally, we can compute $\mathcal{G}$ by

$$\mathcal{G} = \mathcal{H} \times_1 \mathbf{B}^H \times_2 \left(\mathbf{A}^{(1)} \right)^H \times_3 \left(\mathbf{A}^{(2)} \right)^H \cdots \times_{M+1} \left(\mathbf{A}^{(M)} \right)^H \tag{12}$$

Here, we need to compute $\mathbf{A}^{(m)}$, $m \in \{1, 2, \cdots, M\}$. This needs M SVD operations. The computational complexity of SVD of $\mathbf{A}^{(m)}$ is $\mathcal{O}\left(N_r N_t N_{t,m}\right)$. Hence the computational complexity of the M SVD operations is about $\mathcal{O}\left(N_r N_t \sum_{m=1}^{M} N_{t,m} \right)$. In the SVD-based algorithm, the computational complexity of SVD of $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is $\mathcal{O}\left(N_r^2 N_t\right)$. Hence the computational complexity of the HOSVD operation is comparable with the SVD-based algorithm.

2) CSI quantization: As is mentioned in Section II-C, the Tucker decomposition is a form of higher-order principal component analysis. $\mathbf{B}$ and $\mathbf{A}_{(m)}$, $m \in \{1, 2, \cdots, M\}$, are the factor matrices and can be thought of as the principal components in each mode. Hence most of the channel power is contained in some of the columns of the factor matrices. Considering (10), if the number of streams is L, we plan to choose the L leading columns of $\left(\mathbf{A}^{(M)} \otimes \mathbf{A}^{(M-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)^*$ to quantize and feedback as CSI. If we denote one of the columns as $\left(\mathbf{a}_{t_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{t_1}^{(1)} \right)^*$ and use it as the precoding vector, then we can obtain the equivalent channel

$$\mathbf{H} \left(\mathbf{a}_{t_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{t_1}^{(1)} \right)^*$$

$$= \sum_{i=1}^{N_r} \sum_{j_1=1}^{N_{t,1}} \cdots \sum_{j_M=1}^{N_{t,M}} g_{ij_1 \cdots j_M} \mathbf{b}_i \left(\mathbf{a}_{j_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{j_1}^{(1)} \right)^T \left(\mathbf{a}_{t_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{t_1}^{(1)} \right)^* \tag{13}$$

$$= \sum_{i=1}^{N_r} \sum_{j_1=1}^{N_{t,1}} \cdots \sum_{j_M=1}^{N_{t,M}} g_{ij_1 \cdots j_M} \mathbf{b}_i \left\{ \left[\left(\mathbf{a}_{j_M}^{(M)} \right)^T \left(\mathbf{a}_{t_M}^{(M)} \right)^* \right] \otimes \cdots \otimes \left[\left(\mathbf{a}_{j_1}^{(1)} \right)^T \left(\mathbf{a}_{t_1}^{(1)} \right)^* \right] \right\} \tag{14}$$

$$= \sum_{i}^{N_r} g_{it_1 \cdots t_M} \mathbf{b}_i \tag{15}$$

We can obtain (13) by replacing $\mathbf{H}$ with (11). If $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are square matrices such that the products $\mathbf{AC}$ and $\mathbf{BD}$ exist, then $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D})$ exists and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$. For this reason, we can obtain (14) from (13). We choose the L leading columns of $\left(\mathbf{A}^{(M)} \otimes \mathbf{A}^{(M-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)^*$ according to the power of each column corresponding equivalent channel. From (15), we can obtain

$$\| \mathbf{H} \left(\mathbf{a}_{t_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{t_1}^{(1)} \right)^* \|^2 = \sum_{i=1}^{N_r} |g_{it_1 \cdots t_M}|^2 \tag{16}$$

Therefore, $\| \mathbf{H} \left(\mathbf{a}_{t_M}^{(M)} \otimes \cdots \otimes \mathbf{a}_{t_1}^{(1)} \right)^* \|$ equals to the norm of the $\left(t_1 + \sum_{u}^{M} (t_u - 1) \prod_{v=1}^{u-1} N_{t,v} \right)$th column of $\mathbf{G}_{(1)}$.

Hence, we need to compute the norm of each column of $\mathbf{G}_{(1)}$ then sort them in decreasing order. The largest L norms are denoted by $\{\lambda_1, \lambda_2, \cdots, \lambda_L\}$, and $\{c^l, l = 1, \cdots, L\}$ are the corresponding column indexes. Additionally, c^l is corresponding to the t_m^lth column of $\mathbf{A}^{(m)}$, $m \in \{1, 2, \cdots, M\}$ and they meet the following equation:

$$c^l = t_1^l + \sum_{u=2}^{M} \left(t_u^l - 1 \right) \prod_{v=1}^{u-1} N_{t,v} \tag{17}$$

To know which columns of $\mathbf{A}^{(m)}$ need to be quantized, we compute t_m^l from (17) as

$$t_1^l = \left(c^l - 1 \right) \% N_{t,1} + 1 \tag{18}$$

$$t_m^l = \frac{\left(c^l - t_1^l - \sum_{u=2}^{m-1} \left(t_u^l - 1 \right) \prod_{v=1}^{u-1} N_{t,v} \right) \% \prod_{v=1}^{m} N_{t,v}}{\prod_{v=1}^{m-1} N_{t,v}} + 1 \tag{19}$$

where $m \in \{2, \cdots, M\}$ and $\%$ is the modulo operator.

We use codebooks to quantize $\mathbf{a}_{t_m^l}^{(m)}$, where $m \in \{1, 2, \cdots, M\}$ and $l \in \{1, 2, \cdots, L\}$. For instance, we can adopt DFT codebooks. For $\mathbf{a}_{t_m^l}^{(m)}$, the length of codeword we need equals to $N_{t,m}$. For the SVD-based algorithm, the length of codeword is N_t. As is mentioned above, $N_t = \prod_{m=1}^{M} N_{t,m}$, hence $N_{t,m} < N_t$ for $m \in \{1, 2, \cdots, M\}$ and the codeword length needed reduces.

3) Feedback: As is mentioned above, we use codebooks to quantize $\{\mathbf{a}_{t_1^1}^{(1)}, \cdots, \mathbf{a}_{t_M^1}^{(M)}\}, \cdots, \{\mathbf{a}_{t_1^L}^{(1)}, \cdots, \mathbf{a}_{t_M^L}^{(M)}\}$ then get the index groups $\left(\mathrm{CI}_{t_1^1}^{(1)}, \cdots, \mathrm{CI}_{t_M^1}^{(M)} \right), \cdots, \left(\mathrm{CI}_{t_1^L}^{(1)}, \cdots, \mathrm{CI}_{t_M^L}^{(M)} \right)$, in which $\mathrm{CI}^{(m)}$ denotes the codebook adopted by $\mathbf{A}^{(m)}$, $\mathrm{CI}_{t_m^l}^{(m)}$ denotes the codeword index corresponding to $\mathbf{a}_{t_m^l}^{(m)}$. The user sends the index groups to the BS in the feedback link as CSI.

To explain why the feedback bits reduce, we consider a simple example, where $M = 2$ and $L = 4$. Here we need to quantize $\{\mathbf{a}_{t_1^1}^{(1)}, \mathbf{a}_{t_2^1}^{(2)}\}, \{\mathbf{a}_{t_1^2}^{(1)}, \mathbf{a}_{t_2^2}^{(2)}\}, \{\mathbf{a}_{t_1^3}^{(1)}, \mathbf{a}_{t_2^3}^{(2)}\}, \{\mathbf{a}_{t_1^4}^{(1)}, \mathbf{a}_{t_2^4}^{(2)}\}$.

If $\mathbf{a}_1^{(1)}$ and $\mathbf{a}_2^{(1)}$ are the leading two left singular vectors of $\mathbf{A}^{(1)}$, $\mathbf{a}_1^{(2)}$ and $\mathbf{a}_2^{(2)}$ are the leading two left singular vectors of $\mathbf{A}^{(2)}$, the four columns chosen from $\left(\mathbf{A}^{(2)} \otimes \mathbf{A}^{(1)}\right)^*$ are usually composed of $\left(\mathbf{a}_1^{(2)} \otimes \mathbf{a}_1^{(1)}\right)^*$, $\left(\mathbf{a}_1^{(2)} \otimes \mathbf{a}_2^{(1)}\right)^*$, $\left(\mathbf{a}_2^{(2)} \otimes \mathbf{a}_1^{(1)}\right)^*$ and $\left(\mathbf{a}_2^{(2)} \otimes \mathbf{a}_2^{(1)}\right)^*$. Hence we just need to quantize $\mathbf{a}_1^{(1)}, \mathbf{a}_2^{(1)}, \mathbf{a}_1^{(2)}$ and $\mathbf{a}_2^{(2)}$. This means that there are some overlaps in $\{t_1^1, t_1^2, t_1^3, t_1^4\}$ and $\{t_2^1, t_2^2, t_2^3, t_2^4\}$. By taking advantage of this, the proposed scheme can reduce the feedback bits.

4) Precoding in SU-MIMO: The codewords corresponding to the feedback codeword index groups can be denoted by $\left(\mathbf{cw}_{t_1^1}^{(1)}, \cdots, \mathbf{cw}_{t_M^1}^{(M)}\right), \cdots, \left(\mathbf{cw}_{t_1^L}^{(1)}, \cdots, \mathbf{cw}_{t_M^L}^{(M)}\right)$. Therefore, the precoding matrix can be written as

$$\mathbf{P} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_L]^* \tag{20}$$

where

$$\mathbf{p}_l = \mathbf{cw}_{t_M^l}^{(M)} \otimes \cdots \otimes \mathbf{cw}_{t_1^l}^{(1)} \tag{21}$$

If we use the water-filling to allocate power among data streams, the precoding matrix becomes

$$\mathbf{P} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_L]^* \Lambda^{\frac{1}{2}} \tag{22}$$

where Λ is a diagnal matrix, the optimal power loading coefficients in Λ are found using water-filling on $\{\lambda_1, \lambda_2, \cdots, \lambda_L\}$.

5) Precoding in MU-MIMO: Now, we employ the HOSVD-based precoder in MU-MIMO. The data rate of user k can be written as

$$\mathrm{R}_k = \log_2 |\mathbf{I} + \frac{E_{s,k}\mathbf{H}_k\mathbf{P}_k\mathbf{P}_k^H\mathbf{H}_k^H}{N_0\mathbf{I} + \sum_{j\neq k} E_{s,j}\mathbf{H}_k\mathbf{P}_j\mathbf{P}_j^H\mathbf{H}_k^H}| \tag{23}$$

where $\mathbf{H}_k$ and $\mathbf{P}_k$ denote the channel matrix and precoding matrix of user k respectively, $E_{s,k} = \mathbb{E}\{\mathbf{s}_k\mathbf{s}_k^H\}$, and N_0 is the power of noise. In order to increase R_k, we need to reduce the inter-user interference, i.e., $\sum_{j\neq k} E_{s,j}\mathbf{H}_k\mathbf{P}_j\mathbf{P}_j^H\mathbf{H}_k^H$.

We assume that each user in the MU-MIMO system adopts the HOSVD-based scheme. Since most of the channel power is contained in the directions of the precoding matrix's columns, we can choose users whose precoding matrices are approximately orthogonal to reduce the inter-user interference.

If the users apply DFT codebooks in a limited feedback system, the BS will receive the codeword index groups sent from different users. The index groups sent from user k are denoted by $\left(\mathrm{CI}_{t_1^1}^{k,(1)}, \cdots, \mathrm{CI}_{t_M^1}^{k,(M)}\right), \cdots, \left(\mathrm{CI}_{t_1^{L_k}}^{k,(1)}, \cdots, \mathrm{CI}_{t_M^{L_k}}^{k,(M)}\right)$. Then choosing users whose precoding matrices are approximately orthogonal means that choosing users whose index groups do not overlap. More specifically, the index groups of user k_1 and k_2 do not overlap means that each of the L_{k_2} index groups of user k_2 is different from all of the L_{k_1} index groups of user k_1. The ath index group of user k_1 and the bth index group of user k_2 are different means that the corresponding elements in $\left(\mathrm{CI}_{t_1^a}^{k_1,(1)}, \cdots, \mathrm{CI}_{t_M^a}^{k_1,(M)}\right)$ and $\left(\mathrm{CI}_{t_1^b}^{k_2,(1)}, \cdots, \mathrm{CI}_{t_M^b}^{k_2,(M)}\right)$ are different at least in one element. This guarantees that $\mathbf{a}_{t_M^a}^{k_1,(M)} \otimes \cdots \otimes \mathbf{a}_{t_1^a}^{k_1,(1)}$ and $\mathbf{a}_{t_M^b}^{k_2,(M)} \otimes \cdots \otimes \mathbf{a}_{t_1^b}^{k_2,(1)}$ are orthogonal.

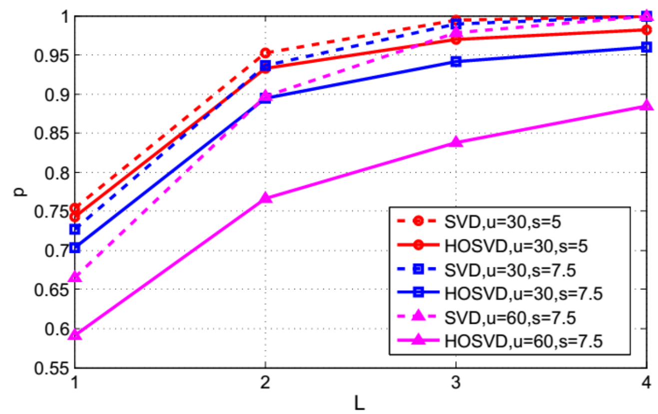

Fig. 1. The percentage of the channel power contained in the directions of the column vectors of the precoding matrices generated by SVD and HOSVD based algorithms as a function of the number of streams L.

The precoding matrix used by the BS can be written as

$$\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_K]\Lambda^{\frac{1}{2}} \tag{24}$$

where $\mathbf{P}_k$ is the precoding matrix of user k and the diagnal elements of Λ are the power loading coefficients.

Block diagonalization zero-forcing (BD-ZF) [10] is a generalization of channel inversion techniques when there are multiple antennas at each receiver. If we apply BD-ZF at BS, it will need two SVD operations for each user. The first SVD operation is to find the null space of the matrix formed by stacking all the other users' channel matrices together. The second SVD operation is to find the solution that maximizes sum capacity for the system under the zero-interference constraint. The proposed HOSVD-based scheme does not need any SVD operation and only need some simple matrix multiplication operations. Hence the computational complexity at BS will reduce.

Remarks: We describe the channel using a $(M+1)$th order tensor above. M is a variable parameter. When we choose the value of M, we need to consider multiple factors such as the form and size of the transmit antenna array, the length of codewords available, the performance expected, the degree of spatially correlation of the channel and so on. For example, the transmitter is equipped with a 8×8 UPA and the length of codewords is 8, hence we can set $M = 2$ and $N_{t,1} = N_{t,2} = 8$. If a higher sum rate is required, we can choose a smaller M. If a fewer feedback bits is required or the degree of spatially correlation of the channel is high, we can choose a larger M.

IV. Simulation Results

In this section, we evaluate the performance of our proposed scheme. We assume that the BS is equipped with a UPA with 8×8 antenna elements. Each user is equipped with a 2×2 UPA. The antenna spacing is equal to half of the signal wavelength. The channel parameters are given by $N_{cl} = 8$, $N_{ray} = 10$ and $\sigma_{\alpha,i}^2 = 1$. The azimuth and elevation AoDs and AoAs

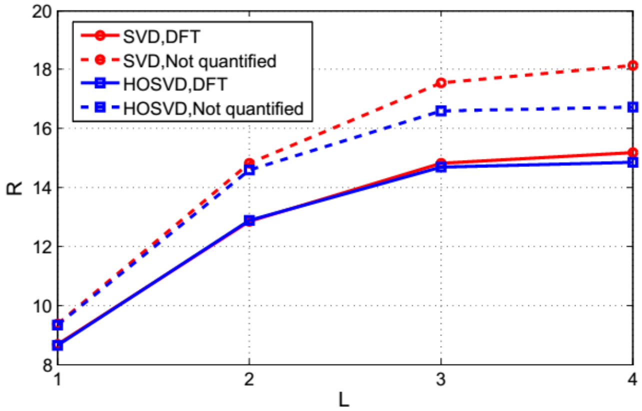

Fig. 2. The data rate achieved by SVD and HOSVD based schemes for a single user as a function of the number of streams using unquantized and DFT codebooks.

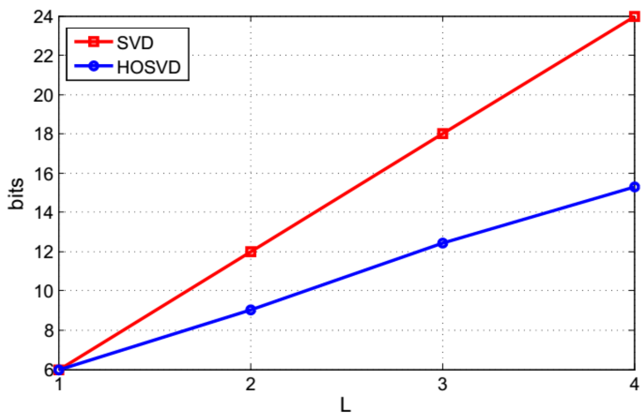

Fig. 3. The average feedback bits of SVD and HOSVD based schemes using DFT codebooks.

follow the Laplacian distribution with uniformly distributed mean angles. Specifically, the central AODs of each cluster in azimuth and elevation follow the uniform distribution, the range of which is u degrees. The central AOAs of each cluster in azimuth and elevation follow the uniform distribution varying from 0° to 360°. The angular spread of rays in each cluster is s degrees. The parameters in HOSVD-based scheme are given by $M = 2$, $N_{t,1} = N_{t,2} = 8$.

First, we simulate a single user case. We consider perfect CSI and use the SVD-based scheme as a comparison. Fig 1 shows the percentage of the channel power contained in the directions of the precoding matrices' column vectors, i.e., $p = \frac{\sum_{l=1}^{L} \lambda_l^2}{\mathrm{tr}\{\mathbf{H}\mathbf{H}^H\}}$, as a function of the number of streams L. A larger p means that more channel power is utilized. It also means that the potential interference that a user may cause to others is smaller in a MU-MIMO system. We can see that when u and s are small, the value of p achieved by HOSVD is close to SVD. When u and s increase, the gap between the two algorithms widens.

Fig 2 compares the data rate R as a function of L using unquantized and quantized codebooks, respectively. The SNR is 30dB, $u = 30$, $s = 7.5$. We choose the power loading coefficients by water-filling. For quantized codebooks, we use 6 bits DFT codebook for SVD precoder and 3 bits DFT codebook for HOSVD precoder. We can see that for unquantized codebooks, the perfomance of HOSVD is worse than SVD. However, when using DFT codebooks, the performance gap narrows, even nearly disappers. Fig 3 compares the average feedback bits of the two schemes using DFT codebooks. When transmitting multiple streams, the average feedback bits of HOSVD is less than SVD. When using quantized codebooks, the codeword length needed by SVD is limited by the number of transmit antennas, while for HOSVD, we can change the codeword length by adjusting M and $N_{t,m}$, $m \in \{1, 2, \cdots, M\}$.

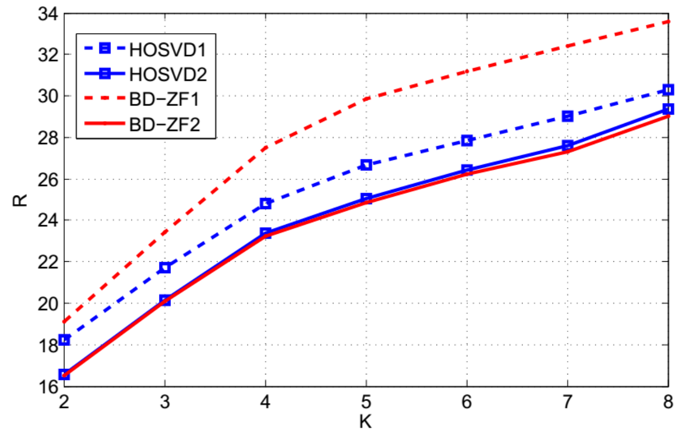

Fig. 4. The sum rate as a function of the number of users in a MU-MIMO system with $L = 4$.

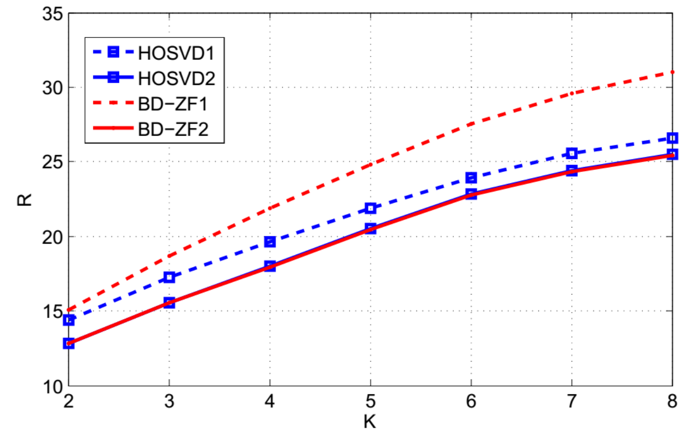

Fig. 5. The sum rate as a function of the number of users in a MU-MIMO system with $L = 2$.

Fig 4 and Fig 5 compare the sum rate R as a function of the number of users K in a MU-MIMO system. We simulate the proposed HOSVD-based scheme and BD-ZF scheme. "BD-ZF1" and "BD-ZF2" refer to the BD-ZF scheme using reconstructed channel with exact and quantized CSI feedback, respectively. The reconstructed channel matrix of a single user can be written as

$$\tilde{\mathbf{H}} = \Sigma[\tilde{\mathbf{h}}_1, \cdots, \tilde{\mathbf{h}}_L]^T \tag{25}$$

$$\tilde{\mathbf{h}}_l = \mathbf{a}_{t_M^l}^{(M)} \otimes \cdots \otimes \mathbf{a}_{t_1^l}^{(1)} \tag{26}$$

where Σ is a diagonal matrix whose diagonal elements are $\{\lambda_1, \lambda_2, \cdots, \lambda_L\}$. For quantized CSI feedback, $\mathbf{a}_{t_m^l}^{(m)}$ is replaced by $\mathbf{cw}_{t_m^l}^{(m)}$. We choose K users from 100 users located randomly using the scheduling method described in Section III-B5. The reconstructed channel matrix of the K users can be written as

$$\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1, \cdots, \tilde{\mathbf{H}}_K]^T \tag{27}$$

where $\tilde{\mathbf{H}}_k$ is the reconstructed channel matrix of user k. "HOSVD1" and "HOSVD2" refer to the HOSVD-based scheme using exact and quantized CSI feedback. We assume the numbers of each user's streams are identical and equal to L. L equals to 4 and 2 for Fig 4 and Fig 5 respectively. We set parameters as $u = 30$, $s = 7.5$, the cell radius is $250m$, the SNR of users in cell edge is $5\mathrm{dB}$. We use DFT codebooks to quantize CSI, and use the power loading coefficients by waterfilling among all streams of the chosen users. We can see that the performance of the proposed scheme with quantized CSI, i.e., "HOSVD2", and the BD-ZF scheme using reconstructed channel with quantized CSI, i.e., "BD-ZF2", are nearly the same.

V. Conclusions

We have designed a novel limited feedback and precoding scheme for massive MIMO systems by introducing HOSVD. The scheme can be employed in MU-MIMO easily through a simple user scheduling method. The proposed scheme can reduce the length of codewords hence is able to reduce the complexity of channel quantization and the number of bits for feedback. It can also reduce the computational complexity of precoding matrix at BS. Simulation results demonstrated that the performance of the proposed scheme is nearly the same with the SVD-based scheme in SU-MIMO and BD-ZF scheme in MU-MIMO using quantized CSI feedback.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing —the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[3] J. Y. Ko and Y. H. Lee, "Adaptive beamforming with dimension reduction in spatially correlated MISO channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 4998–5002, Oct. 2009.

[4] G. Liu, X. Hou, F. Wang, J. Jin, H. Tong, and Y. Huang, "Achieving 3D-MIMO with massive antennas from theory to practice with evaluation and field trial results," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–10, 2016.

[5] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proceedings of IEEE International Conference on Communications*, Jun. 2014, pp. 5865–5870.

[6] J. Choi, K. Lee, D. J. Love, T. Kim, and R. W. Heath, "Advanced limited feedback designs for FD-MIMO using uniform planar arrays," in *Proceedings of IEEE Global Communications Conference*, Dec. 2015, pp. 1–6.

[7] F. Yuan, "Tucker decomposition for rotated codebook in 3D MIMO system under spatially correlated channel," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, p. 1, 2015.

[8] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[9] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[10] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.