Novel Convolutional Restricted Boltzmann Machine manifold learning inspired dynamic user clustering hybrid precoding for millimeter-wave massive multiple-input multiple-output systems

Xiaoping Zhou, Haichao Liu, Bin Wang, Qian Zhang and Yang Wang

Abstract
Millimeter-wave massive multiple-input multiple-output is a key technology in 5G communication system. In particular, the hybrid precoding method has the advantages of being power efficient and less expensive than the full-digital precoding method, so it has attracted more and more attention. The effectiveness of this method in simple systems has been well verified, but its performance is still unknown due to many problems in real communication such as interference from other users and base stations, and users are constantly on the move. In this article, we propose a dynamic user clustering hybrid precoding method in the high-dimensional millimeter-wave multiple-input multiple-output system, which uses low-dimensional manifolds to avoid complicated calculations when there are many antennas. We model each user set as a novel Convolutional Restricted Boltzmann Machine manifold, and the problem is transformed into cluster-oriented multi-manifold learning. The novel Convolutional Restricted Boltzmann Machine manifold learning seeks to learn embedded low-dimensional manifolds through manifold learning in the face of user mobility in clusters. Through proper user clustering, the hybrid precoding is investigated for the sum-rate maximization problem by manifold quasi-conjugate gradient methods. This algorithm avoids the traditional method of processing high-dimensional channel parameters, achieves a high signal-to-noise ratio, and reduces computational complexity. The simulation result table shows that this method can get almost the best summation rate and higher spectral efficiency compared with the traditional method.

Keywords
mmWave massive multiple-input multiple-output, manifold analysis, hybrid precoding, user clustering

Introduction
Millimeter-wave (mmWave) massive MIMO (multiple-input multiple-output) technology gradually becomes a key technology in 5G communication due to its rich spectrum resources.1–3 Due to the high carrier frequency, mmWave signal suffers from high propagation loss so that large-scale antenna arrays are leveraged for path compensation.4 However, in massive MIMO system, the number of antennas at the transmitter and...
receiver is very large, configuring a radio frequency (RF) chain for each antenna in the traditional all-digital solution requires a lot of hardware cost and causes the loss of power. In response to this problem, a hybrid scheme has emerged, considering the reduction of hardware requirements in spectrum efficiency (SE) and energy efficiency (EE). However, the hybrid precoding scheme in wideband channels is currently a difficult problem to solve.

How to obtain the optimal precoding matrix is the key issue of hybrid precoding. For the case of large-scale antennas in mmWave communication, large-scale matrix calculations are usually required. The difficulty of hybrid precoding is to reduce the complexity of the above situation. Some advanced beam-space-based hybrid precoding algorithms have been studied. Previous investigations make full use of the sparsity of the beam space channel according to the sparse signal processing scheme. In the literature, the problem is transformed into finding the optimal precoder with hybrid structure, and an algorithm based on the basis tracking method is proposed. A hybrid precoding scheme designed is proposed according to the Orthogonal Match Pursuit (OMP) algorithm in the literature, which can make full use of channel sparsity. In the multi-user scenario, the low-complexity multi-user hybrid precoding of the mmWave system is studied. A Kronecker decomposition hybrid beamforming (KDBH) method for multi-cell multi-user massive MIMO system based on sparse propagation path is proposed.

However, the resolution of the beam space is not infinite. Due to the existence of power leakage, the sparse channel is non-ideal, and there are many possible non-zero terms. Some papers consider hybrid precoding of interfering mmWave channels. Dealing with interference is very challenging, because the number of antennas is large, and the high-complexity precoding matrix is difficult to implement. To address the high interference problem, a closed-form broadband hybrid precoding scheme was proposed in the literature. An analytical framework of hybrid beamforming (AFHB) in multi-cell mmWave systems was proposed. A combination of analog and digital beamforming is adopted. The former is based on a phase shifter, and the latter is based on a regularized zero-forcing method.

Recently, scholars have proposed the manifold learning in mmWave massive MIMO systems. Yu et al. proposed a manifold optimization (MO)-based hybrid precoding algorithm with lower complexity. To replace the range of the constant envelope with a circular manifold, Chen proposed a Riemannian conjugate gradient manifold algorithm. In Mai et al., a Riemannian vector perturbation manifold for a multi-user massive MIMO system was studied, in which the RF-baseband hybrid precoding was jointly arranged. A Riemannian trust-region Newton manifold (RTRNM) showed an improved method of beamforming in multi-cluster scenarios. The optimization beamforming is utilized to mitigate inter-cell interference by dividing multi-users into multi-clusters with spatial correlation. However, multi-user high-dimensional channels are not mapped into low-dimensional subspaces to achieve dimensionality reduction. Learning a form of a double digital beamforming schemes optimizes the network resource allocation in massive MIMO networks. Moe Thet et al. analyze by greedy algorithm how fast-moving users in static and time-varying user clustering are executed according to the system sum-rate. The manifold learning algorithm is used to reduce the multi-user high-dimensional channels. It reduces the computational complexity while mitigating inter-cell interference-based fully digital beamforming. It focuses on the local linear spatial structure between user channels, and ignores the global spatial characteristics. And it is not possible to quickly analyze the global and local correlations between user channels in the case of moving users.

The traditional user precoding methods are not applicable to multiple users, although they optimize precoding using channel sparsity. In the multi-user scenario, the traditional method suffers from high precoding complexity, high channel dimensionality, and does not consider user mobility. Therefore, it is necessary to propose new algorithms that take into account these important problems in practical communication.

In this article, we propose a low-complexity hybrid precoding algorithm for dynamic user clustering in mmWave massive MIMO systems. Specifically, a large-scale number of antennas embedded in a low-dimensional subspace. The mmWave channel measurement results show that the mmWave has a diffuse scattering phenomenon on the surface of the rough scatterer, and the scattering range will increase as the wavelength decreases. For scenarios where users are dense, when there is not enough space between users, diffuse scattering may cause adjacent users to receive signals of the same path. Therefore, it causes serious inter-user interference. Our goal is to design a mixed precoding matrix, so they manage intra-cell and inter-cell interference requires a lower channel knowledge, and can be used to achieve low-complexity mixed analog/digital architecture, that is, compared with a small number of RF chains, the number of antennas. In order to solve the set classification problem, a manifold discriminant analysis (MDA) is proposed. Set by each user is modeled as a manifold, we will issue expressed as clustering for multi-manifold learning. The manifold discriminative learning seeks to learn the embedding low-dimensional manifolds, wherein manifolds with different user cluster label better separation of high-dimensional partial space of each flow channel in the shape of the correlation is enhanced. Learning by discriminant manifold, the majority of high-dimensional
mapping of the channel to a low-dimensional manifold, it is possible to fully utilize the potential of the high-dimensional channel spatial correlation. By transforming the non-linear problems of high-dimensional channels into global non-linearities and local non-linearities, the purpose of dimensionality reduction is achieved. In low-dimensional manifolds, the intra-cluster channels become more clustered and the separability of embedded features is enhanced. Facing the situation that users will move in different clusters, we introduce the novel Convolutional Restricted Boltzmann Machine (NCRBM) framework into our stream shape learning. By continuously updating the stream shape discriminations in clusters, we get the best state of stream shape learning for users in clusters. Through proper user clustering, the hybrid precoding is investigated for the sum-rate maximization problem by manifold quasi-conjugate gradient methods. To enhance the spectral efficiency of the system, the design of each cluster analog RF precoder should balance the optimizing self-transmission and the interference. The digital precoding matrix is obtained by Karush Kuhn Tucker (KKT). Compared with the traditional method, the proposed method does not require the solution of large-scale channel parameters, and can achieve a high signal-to-noise ratio (SNR) while reducing the computational complexity. The results show that the algorithm can obtain close to the optimal sum-rate and quite high spectral efficiency.

The rest of this article is as follows. Section “System model and channel model” introduces the system model and channel model. Section “User clustering hybrid precoding scheme” introduces the algorithm for dimensionality reduction and the hybrid precoding algorithm in multi-user high-dimensional channel scenarios. Section “Simulation results” presents the simulation results. Section “Conclusion” part summarizes this article.

Notations

Upper and lower-case boldface letters represent the matrices and the vectors, respectively. $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^*$, $tr(\cdot)$, and $\|\cdot\|_F$ are the Hermitian transpose, inverse, transpose, complex conjugate, trace, and the Frobenius norm of a matrix, respectively. $E(\cdot)$ is the expectation. $diag(\cdot)$ denotes the diagonal matrix. $|G|$ is the cardinality of the set $G$. $\otimes$ indicates the Kronecker product. $CN(0,\sigma^2)$ represents the zero-mean complex Gaussian distribution with zero-mean and the variance $\sigma^2$. $\text{span}(Y)$ denotes the subspace spanned by the column vectors of $Y$. $\nabla(\cdot)$ indicates the gradient. Finally, $I_N$ denotes the $N \times N$ identity matrix.

System model and channel model

System model

We consider a hybrid mmWave massive MIMO system model consisting of B cells. We assume that a base station (BS) equipped with $N_t$ antenna and $N_{RF}$ RF chains ($N_t \geq N_{RF} \geq K$) serves $K$ single-antenna users, as shown in Figure 1.

![Figure 1. Hybrid mmWave massive MIMO system model.](image-url)
in Figure 1. To manage the interference and improve
the data rate for users, the users are partitioned into \( L \)
clusters \( G_1, \ldots, G_L \) with \( g_i = |G_i|, \sum_{i=1}^L g_i = K, \) and
\( G_i \cap G_{i'} = \emptyset, \forall i \neq i' \). \( G_i \) is \( i \)th cluster, where
\( i = 1, \ldots, L \). The sets \( \{G_1, \ldots, G_L\} \) are all user clusters.

Let \( u_{b,i,k}, k = 1, \ldots, g_i \) denote the \( k \)th user of \( G_i \) in
the \( b \)th cell \((b = 1, 2, \ldots, B) \). Hybrid precoding consists of two parts: baseband domain digital precoding and
radio frequency domain analog precoding. In the downlink
channel, the transmitted symbol first passes through the
digital precoder, and the generated signal is fed back to the
radio frequency chain. The output of the RF chain is analog
precoded and then sent to the base station. For a simple
precoding matrix that adaptively steers an
additive Gaussian noise.

The undirected similarity graph of multi-users is rep-
resented by the graph embedding method. To represent
each user set as a manifold, the user channel charac-
teristic graphs \((\{h_{i,k}, m_{k,}\})_{j=1}^{L}\) are constructed, as shown in
Figure 2.

\( \mathbf{m}_{k,j}(0) \) represents the intra-cluster channel weight
function between users \( k \) and user \( j \). \( m_{k,j} \) represents
the inter-cluster channel weight function between users
(2)

where \( h_{i,k} \) is the channel matrix of the user in cluster can be expressed as

\[
h_{i,k} = \sum_{l=1}^{N_l} a_{i,l} a_{l,T} a_{i,l}\]

(3)

where \( a(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j2\pi/\lambda d_{ULA} \sin(\theta)}, \ldots, e^{j(N-1)2\pi/\lambda d_{ULA} \sin(\theta)}\right]^T \)

\( \lambda \) is the wavelength, and \( d_{ULA} \) denotes the spacing
between antennas. Due to the limited space during scat-
tering of during mmWave propagation, the mmWave
massive MIMO channel \( h_{i,k} \) is low-rank. Therefore, we
can use a limited number of RF chains to obtain near-
optimal throughput.

User clustering hybrid precoding scheme

Our goal is to design a hybrid precoding matrix. Therefore,
we must first deal with intra-cluster, inter-
cluster, and inter-cell interference with less known
channel knowledge, and second, we need to use little
RF chains to complete the hybrid analog/digital archi-
tecture, avoiding the high complexity of traditional
methods. Next, we propose a hybrid precoding method
based on manifold learning to achieve the above goals.

NCRBM manifold learning for user clusters

With the increase of antennas and users in the
mmWave massive MIMO system, inter-cell and intra-
cell directional interference will occur during signal
transmission. The high-dimensional channel matrix
requires high-complexity hybrid analog/digital archi-
tectures. By modeling each user set as a manifold, we
formulate the problem as clustering-oriented manifold
discriminative learning.

The undirected similarity graph of multi-users is re-
presented by the graph embedding method. To represent
each user set as a manifold, the user channel charac-
teristic graphs \((\{h_{i,k}, m_{k,}\})_{j=1}^{L}\) are constructed, as shown in
Figure 2.

\( m_{k,j} \) represents the channel vector between the BS
and user \( u_{b,i,k} \). \( x_{b,i,k} \) represents the transmit signal of user
\( u_{b,i,k} \). \( n_{b,i,k} \sim CN(0, \sigma^2) \) is the spatially white
additive Gaussian noise. \( F_{b,i,k} \in C^{N \times n_{RF,i}} \) is the analog
precoding matrix that adaptively steers an
nRF,i-dimensional RF beamspace for the coverage of
\( G_i \) with \( n_{RF,i} \approx g_i \).

\( W_{b,i,k} \in C^{n_{RF,i}} \) is the digital precoding matrix. \( C \)
is the set of complex numbers. \( h_{b,i,k}' = F_{b,i,k} W_{b,i,k} x_{b,i,k} \) are
the desired signal. \( \sum_{k'=1,k \neq k}^{G_i} h_{b,i,k}' F_{b,i,k} W_{b,i,k} x_{b,i,k} \) are
the intra-cluster interference. \( \sum_{l'=1,l \neq l}^{L} h_{b,i,l}' F_{b,i,l} W_{b,l,i} x_{b,i,l} \) are the inter-cluster interference. \( \sum_{b'=1,b \neq b}^{B} h_{b,i,b'} W_{b,i,b'} x_{b,i,b} \) are the inter-cell interference. The
result of the hybrid method is more accurate than the
statistical method, and the method can get faster and
more generalized results, but it cannot provide enough
accuracy in modeling the intra-cluster angle, which is
necessary for beamforming and inter-cluster interference
optimization.\(^{28,38}\)

Channel model

In order to take advantage of the unique spatial
selectivity or scattering characteristics of the mmWave
massive MIMO channel, this article adopts the
Saleh–Valenzuela (SV) model,\(^{35}\) where the channel
matrix of the user in cluster can be expressed as

\[
\mathbf{h}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j2\pi/\lambda d_{ULA} \sin(\theta)}, \ldots, e^{j(N-1)2\pi/\lambda d_{ULA} \sin(\theta)}\right]^T
\]
The weight function \( m_{j,k,j} \) of the intra-cluster is defined as follows

\[
\begin{align*}
0 < m_{j,k,j} &\leq 1, & k,j \text{ in the intra-cluster} \\
0 &\text{ otherwise}
\end{align*}
\]  \( \text{(4)} \)

The weight function \( m_{z,k,j} \) of the inter-cluster is defined as follows

\[
\begin{align*}
0 < m_{z,k,j} &\leq 1, & k,j \text{ in the inter-cluster} \\
0 &\text{ otherwise}
\end{align*}
\]  \( \text{(5)} \)

The weight functions of the intra-cluster show that when users \( k \) and \( j \) are the same cluster, the weight is larger; when users \( k \) and \( j \) are the different cluster, the weight is 0. The weight functions of the inter-cluster show that when users \( k \) and \( j \) are different cluster, the weight is larger; when users \( k \) and \( j \) are the same cluster, the weight is 0. The manifold discriminative learning seek to learn the embedding low-dimensional manifolds, where manifolds with labels of different user groups can be separated more conveniently, and the local spatial correlation of the high-dimensional channels within each manifold is enhanced. Some existing manifold learning algorithms, such as Locally Linear Embedding (LLE),\(^{39}\) cannot retain the complete global non-linear channel structure of user clusters.

Figure 2. User cluster undirected characteristic graph.

Figure 3. Schematic diagram of dimension reduction.

We propose to perform the manifold discriminative learning for global dimensionality reduction. The high-dimensional channels are mapped in the low-dimensional manifolds, as shown in Figure 3. In order to reveal the potential non-linear manifold structure of high-dimensional channels, intra-cluster graph and inter-cluster graph are constructed using the label information of user characteristics. In addition, it can make the low-dimensional channels more clustered, and enhance the separability of embedded low-dimensional channels. The radio frequency eigen-beamformer is
considered to be the best solution for user group transmission. The channel eigenvector learning corresponding to the maximum eigenvalue is taken as the spatial direction. In theory, the main direction learned is the beamforming. Multi-users of the same cluster have highly correlated transmission paths. We seek to learn a generic mapping \( A \) that is defined as

\[
h'_k = A^T h_k
\]

where \( A \) is projection matrix, \( h'_k \) is the \( k \)th user low-dimensional mapping of the high-dimensional channels \( h_k \). The original high-dimensional channels \( h_k \) can be transformed into a low-dimensional channel \( h'_k \). The relative spatial relationship of neighboring users in high-dimensional channels remains unchanged in low-dimensional mappings. In order to maintain the manifold structure of the high-dimensional channels, the optimization problem is the projection direction of manifold, that is, \( \forall h_i, h_j (k \neq j) \) of the intra-cluster, the target function of the intra-cluster can be obtained as

\[
\max_A \sum_{k,j} (h'_k - h'_j)^2 m_{\xi, k,j}
\]

The projection can maximize the use of all users in the cluster of the intra-cluster as equation (8), where \( S_{\xi, local} = H(D_k - M_k)H^T \) is the local manifold structure of the intra-cluster, and \( D_k \) is the diagonal matrix and \( D_k = \sum_{k \neq j} M_{\xi}(k,j) \)

\[
\begin{align*}
\max_A & \sum_{k,j} (h'_k - h'_j)^2 m_{\xi, k,j} \\
= & \max_A \frac{1}{2} \sum_{k,j} (A^T h_k - A^T h_j)^2 m_{\xi, k,j} \\
= & \max_A A^T H D_k H^T A - A^T H M_k H^T A \\
= & \max_A A^T S_{\xi, local} A
\end{align*}
\]

According to the SV model, \( R_{i,k} = E[h_i,k h_i^H] \) is the covariance matrix of the \( k \)th user in the \( i \)th cluster. The transmission covariance matrix of users in a cluster is the same, so, \( R_{i,k} \), that is

\[
R_{i,k} = U_{i,k} A_{i,k} U_{i,k}^H
\]

where \( U_{i,k} \in \mathbb{C}^{N \times r} \) is a matrix of eigenvectors corresponding to \( r_i (r_i = N_i) \) non-zero eigenvalues of \( R_{i,k} \). \( A_{i,k} \) is the diagonal matrix whose elements are the non-zero eigenvalues of \( R_{i,k} \in \mathbb{C}^{N \times N} \) corresponding to the non-zero eigenvalues, satisfying \( A_k \in \mathbb{C}^{r \times r} \). Since users in the same user cluster have similar spatial correlations, they have similar local scattering, \( R_i = R_{i,k}, \forall k \in G_i \). Measure of similarity between the user and the similarity criterion is a function of the distance function coefficients. Since \( \text{span}(U) = UU^T \), the similarity measurement function between any two users based on the distance of subspace projection matrix can be expressed as

\[
d_{pm}(U_k U_k^T, V_i V_i^T) = \frac{1}{2} \|U_k U_k^T - V_i V_i^T\|^2_F
\]

\[
= \frac{1}{2} \text{tr} \left( \psi_k \psi_k^T \right)
\]

where \( U_k \) is the eigenvectors matrix of \( R_k \) in any cluster, that is, \( R_k = U_k A_k U_k^H \), and \( V_i \) is the eigenvectors matrix of the \( i \)th cluster center \( R_i \). \( \psi_{k,i} = U_k U_k^T - V_i V_i^T \) is the symmetric positive semi-definite matrix that needs to be learned. The global manifold structure \( S_{\xi, global} \) of intra-cluster is measured as

\[
S_{\xi, global} = \sum_{i=1}^L \sum_{k \in G_i} \frac{1}{2} \text{tr} \left( \psi_{k,i} \psi_{k,i}^T \right)
\]

To effectively utilize the global characteristics and local manifold structure of intra-cluster channels, we can get the intra-cluster dispersion \( \eta_\xi \) by combining equations (9) and (11)

\[
\eta_\xi = \nu S_{\xi, global} + (1 - \nu) S_{\xi, local}
\]

where \( \nu \) is the constants.

The weight functions \( m_{\xi, k,j} \) of the intra-cluster can be obtained as

\[
m_{\xi, k,j} = \exp \left( -\frac{d_{k,j}}{s'} \right)
\]

where \( s' \) is the constants, and \( d_{k,j} \) is the similarity measurement function between user \( k \) and user \( j \).

In order to maintain the manifold structure of the inter-cluster user channels, the optimization problem is the projection direction of manifold, that is, \( \forall h_i, h_j (k \neq j) \) of the inter-cluster, the objective function of the inter-cluster can be obtained as

\[
\max_A \sum_{k,j} (h'_k - h'_j)^2 m_{\xi, k,j}
\]

Therefore, the projection can maximize the use of all users of the inter-cluster, that is

\[
\begin{align*}
\max_A & \sum_{k,j} (h'_k - h'_j)^2 m_{\xi, k,j} \\
= & \frac{1}{2} \max_A \sum_{k,j} (A^T h_k - A^T h_j)^2 m_{\xi, k,j} \\
= & \max_A A^T H D_k H^T A - A^T H M_k H^T A \\
= & \max_A A^T S_{\xi, local} A
\end{align*}
\]
where $S_{\xi, local} = H(D_{\xi} - M_{\xi})H^T$ is the local manifold structure of the intra-cluster, and $D_{\xi}$ is the diagonal matrix and $D_{\xi} = \sum_{k \neq j} M_{\xi}(k, j)$. The global inter-cluster $S_{\xi, global}$ is measured as

$$S_{\xi, global} = \sum_{i = 1}^{L} \sum_{k \neq j} \frac{1}{\sqrt{2}} K - g_{i, j} \text{tr}(\psi_{\xi, i} \psi_{\xi, j}^T)$$  \hspace{1cm} (16)$$

To effectively utilize the global characteristics and local manifold structure of inter-cluster channels, we can get the inter-cluster dispersion $\eta_{\xi}$ by combining equations (15) and (16)

$$\eta_{\xi} = \varphi S_{\xi, global} + (1 - \varphi)S_{\xi, local}$$ \hspace{1cm} (17)

where $\varphi$ is the constants.

The weight functions $m_{\xi, k, j}$ of the inter-cluster can be obtained as

$$m_{\xi, k, j} = \exp\left(-\frac{s''}{d_{k, j}}\right)$$ \hspace{1cm} (18)

where $s''$ is the constants.

After getting the intra-cluster dispersion $\eta_{\xi}$ and the inter-cluster dispersion $\eta_{\xi}$, the mobility of the user causes these two values to change continuously. To cope with this problem, we introduce the NCRBM model \(^40\) to acquire optimal value. The NCRBM is a multi-layer network constructed by Convolutional Restricted Boltzmann Machines (CRBMs) stacked on top of each other. NCRBM is an extension of the standard Restricted Boltzmann Machine (RBM) that has inherited all its properties but with faster operation. NCRBM optimize the intra-cluster dispersion value over time. In $i$th($i = 1, ..., L$) cluster of users, let us denote the intra-cluster dispersion of $\mu$ time interval by $\{\eta_{\xi, k, j}^{L, i_{t}}\}_{t = 0}^{L - 1}$, where $\eta_{\xi} = \{\eta_{\xi, 0}, \eta_{\xi, 0} + \Delta t, ..., \eta_{\xi, 0} + \mu \Delta t\}$ is a set of intra-cluster dispersions belonging to $i$th cluster. And $t_0$ is the start time, and $\Delta t$ is the time interval. In addition, we define $\tilde{\eta}_{\xi}$ as the optimal intra-cluster dispersion after optimization. The optimal value is computed by averaging across manifold structure feature over all the intra-cluster dispersions belonging to set $\eta_{\xi}$. An optimal value is a mean representation of a specific cluster in manifold structure feature space and must be calculated for each cluster separately. The energy of the joint configuration $(\eta_{\xi}, \tilde{\eta}_{\xi})$ of the input intra-cluster dispersion and optimized units for an NCRBM with real-valued input units can be defined as follows

$$E(\eta_{\xi}, \tilde{\eta}_{\xi}) = \frac{1}{2} \sum_{u, v = 1}^{\mu} (\eta_{\xi})_{u, v}^2 - \sum_{q = 1}^{Q} \sum_{m, n = 1}^{\mu} (\eta_{\xi})_{m, n}^q (w^q \ast \eta_{\xi})_{m, n} - \sum_{q = 1}^{Q} b_q \sum_{m, n = 1}^{\mu} (\eta_{\xi})_{m, n}^q - c \sum_{u, v = 1}^{\mu} (\eta_{\xi})_{u, v}$$ \hspace{1cm} (19)

where * stands for the convolution; $w^q, (\eta_{\xi})_{m, n}^q$, and $(\eta_{\xi})_{u, v}$ denote the manifold structure feature detectors horizontally and vertically in filter $q$, the optimized unit on location $(m, n)$, and the input intra-cluster dispersion unit on location $(u, v)$, respectively. Also, $b_q$ is the shared bias among all units in feature map $q$, and $c$ is the bias value for input intra-cluster dispersion units.

The network assigns a probability to every possible pair of input and optimized unit through this energy function as follows

$$P(\eta_{\xi}, \tilde{\eta}_{\xi}) = \frac{1}{Z} \exp(-E(\eta_{\xi}, \tilde{\eta}_{\xi}))$$ \hspace{1cm} (20)

where $Z$ is the partition function.

The optimization of the model parameters can be performed by minimizing the following objective function using the Contrastive Divergence

$$\min_{w, \eta, \xi} \left\{ -\sum_{i = t_0}^{t_0 + \mu \Delta t} \log\left(\sum_{i = 0}^{Q} P(\eta_{\xi}, \tilde{\eta}_{\xi})\right) + \lambda_{\beta \text{sparsity}} \sum_{q = 1}^{Q} \left| p - \frac{1}{\mu^2} \sum_{m, n = 1}^{\mu} P(\tilde{\eta}_{\xi})_{m, n} = 1 | \eta_{\xi} \right| \right\}$$ \hspace{1cm} (21)

The second term in equation (21) is the sparsity regularization proposed to prevent the model from being overcomplete. For each intra-cluster dispersion, $\eta_{\xi}$, we have

$$\tilde{\eta}_{\xi} = \frac{1}{Q} \sum_{q = 1}^{Q} P\left(\tilde{\eta}_{\xi}^q\right) = 1 | \eta_{\xi}$$ \hspace{1cm} (22)

where $Q$ is the number of manifold structure features. Then, the optimal value for each cluster can be computed as follows

$$\tilde{\eta}_{\xi} = \frac{1}{Q} \sum_{q = 1}^{Q} \tilde{\eta}_{\xi}^q$$ \hspace{1cm} (23)

NCRBM objective function is comprised of two main parts, for example, generative and optimized parts. Generative objective (the first two terms in equation (24)) is the same as sparsity regularized CRBM, while this way of optimization does not guarantee the best intra-cluster dispersion value. Optimization of the
generative part is performed by minimizing Contrastive Divergence. The second two terms in equation (24) correspond to the optimized function, which can be optimized by following a gradient-based process. Minimizing the following objective function can get the optimal intra-cluster dispersion

$$\min_{w, b, \epsilon} \left\{ -\sum_{i=1}^{L} \log \left( \sum_{q=1}^{Q} P \left( \eta^q_{ki}, \tilde{\eta}^q_{ki} \right) \right) + \lambda_{\text{sparsity}} \sum_{q=1}^{Q} \left| \rho - \frac{1}{\mu^2} \sum_{m,n=1}^{\mu} P \left( \eta^q_{m,n} = 1 \mid \eta^q_{ki} \right) \right|^2 \right. $$

$$+ \lambda_{\text{dis}} \frac{1}{Q} \sum_{r=1}^{Q} \sum_{q=1}^{Q} e_l \left\| P \left( \eta^q_{m,n} = 1 \mid \eta^q_{ki} \right) - M_t \right\|^2_F - \beta_{\text{dis}} \sum_{u,v=1}^{L} \left\| M_u - M_v \right\|^2_F \right\}$$

where $L$ is the number of clusters. In equation (24), the third term tries to move the features of each intra-cluster dispersion in cluster $r$ to its manifold structure $M_r$, while the last term tries to maximize the distance between class-maps $M_u$ and $M_v$ for a better separation between clusters. The weighting parameters $\lambda_{\text{dis}}$ and $\beta_{\text{dis}}$ are used to adjust the amount of contribution of the optimized terms in the overall process.

The gradient of the optimized part can be computed exactly at each iteration. For each cluster, we summed the gradient contributions brought by the two components. Based on the definition in equation (23), update of intra-cluster dispersion in each cluster can be performed after optimizations. The limitation of this approach is that too long-time intervals can cause the obtained intra-cluster dispersion to be inaccurate. So, it needs to be run for several times, which makes the overall computing time of the algorithm increase.

For inter-cluster dispersion, we use a similar approach to obtain the optimized inter-cluster values $\tilde{\eta}_c$. In this way, if the users in a cluster move within a certain time, we use the NCRBM model to obtain optimal intra-cluster dispersion and inter-cluster dispersion while maintaining the manifold structure feature within a certain time. However, the overhead of manifold learning increases due to the constant updating of the dispersion values for intra- and inter-cluster users. This is mainly due to the fact that the introduced NCRBM model requires some iterations.

The discriminative function $J(A)$ is transformed as

$$J(A) = \max_A \frac{A^T \eta_c A}{A^T \tilde{\eta}_c A}$$

$$J(A) = \max_A \frac{A^T (\tilde{\eta}_c \text{whole} + (1 - \tilde{\nu}) \tilde{\eta}_c \text{local}) A}{A^T (\tilde{\eta}_c \text{whole} + (1 - \tilde{\nu}) \tilde{\eta}_c \text{local}) A}$$

s.t. $h_{i,k} = A^T h_{i,k}$

According to equation (27), the low-dimensional mapping of the $k$th user channel matrix $h_{i,k}$ is determined by the projection matrix $A$. The Lagrange multipliers are introduced to transform the original optimization problem into a Lagrangian function problem to find the optimal projection matrix $A$ with intra-cluster and inter-cluster dispersion values.\(^4\) By solving the generalized eigenvalues of the discriminative function, we can obtain the projection matrix $A = [A_1, \ldots, A_n]$, $n$ is the dimensionality reduction of user channel matrix. After user clustering, the channel correlation of users in the same cluster is enhanced.

Then, according to the intra-cluster graph and inter-cluster graph constructed using the label information of user characteristics, the user clusters can be divided more accurately with lower complexity. Based on the maximum and minimum distances and the weighted likelihood similarity criterion, an optimized spatial fuzzy $c$-means clustering algorithm is proposed. The algorithm is an iterative optimization that minimizes the cost function defined as follows

$$J(\mu_{i,k}) = \sum_{k=1}^{K} \sum_{i=1}^{L} \mu_{i,k}^2 d_{i,k}$$

where $d_{i,k} = (1/\sqrt{2}) \text{tr}(\psi_{i,k} A^T \psi_{i,k} A)$ is the similarity measurement function between the $i$th user and the $k$th cluster center. $\mu_{i,k}$ represents the membership function of user $u_{i,k}$ in the $k$th cluster, and $\Xi$ is a constant. The parameter $\Xi$ controls the ambiguity of the result partition; in this article, we set $\Xi = 2$. When assigning high membership values to users $u_{i,k}$ close to the cluster center and assigning low membership values to users $u_{i,k}$ far away from the cluster center, the cost function $J(\mu_{i,k})$ is the smallest. The membership function represents the probability that a user $u_{i,k}$ belongs to a specific cluster. Member functions and cluster center are updated as follows

$$\mu_{i,k} = \frac{1}{\sum_{r=1}^{R} \mu_{i,r}^2 d_{i,k}^{-1/r}}$$

and
\[ V_{i,k} V_{i,k}^T = \frac{\sum_{k=1}^{K} \mu_{i,k}^3 U_{i,k} U_{i,k}^T}{\sum_{k=1}^{K} \mu_{i,k}^3} \]  

where \( d_{i,\rho} = (1/\sqrt{2}) \text{tr}(\psi_i \psi_{i,\rho}^T) \) is the similarity measurement function between the \( i \)th user and the \( \rho \)th cluster center. In summary, to represent each user set as a manifold, the process of clustering-oriented manifold discriminative learning is as follows:

Step 1: construct the user channel characteristic graphs \( \{(h_{i,k}, m_{i,k})\}_{i=1}^{L} \).
Step 2: find the two farthest distances \( U_{i} \) and \( U_{r} \), and their centers as the first user group, that is, \( V_{1}^{(0)} = U_{i}, V_{2}^{(0)} = U_{r} \). The number of the user clusters is \( i = 2 \).
Step 3: from the Euclidean distance criterion \( d_{pm}(U_{i} U_{i}^T, V_{i} V_{i}^T) = (1/\sqrt{2}) \text{tr}(\psi_i \psi_{i,\rho}^T) \), all users are gathered in the \( i \) user cluster.
Step 4: among the \( i \) user groups that have completed clustering, find the weakest similarity point in each user group to obtain the \( i \) user clusters. Calculate the sum distance \( d_{i,k} \) between the user \( k(k = 1, 2, \ldots, K) \), the membership functions \( \mu_{i,k}^{(0)} \), and the center point \( V_{i}^{(0)}(i = 1, 2, \ldots, L) \) of each user cluster in turn.
Step 5: calculate the spatial membership function and update the center point \( V_{i}^{(0)}(i = 1, 2, \ldots, L) \) of each user cluster with

\[ V_{i,k} V_{i,k}^T = \frac{\sum_{k=1}^{K} (\mu_{i,k}^{(0)})^3 U_{i,k} U_{i,k}^T}{\sum_{k=1}^{K} (\mu_{i,k}^{(0)})^3} \]

Then, the maximum value among \( d_{i,k} \) is found. \( V_{i+1} = \arg \max_i d_{i,k} \). All users entering \((i + 1)\) are re-divided into different user clusters.

Step 6: if \( i = i + 1 \geq L \) is true, perform Step 5; otherwise repeat Step 3.
Step 7: \( \| U_{i} V_{i}^T \|_F^2 \) is calculated, each user is arranged to the user clusters with the largest similarity coefficient.
Step 8: output cluster result, and the number of users in each cluster.
Step 9: calculate the \( m_{i,k,j} \) and \( m_{k,i,j} \) according to equations (13) and (18); construct intra-cluster graph and inter-cluster graph using the label information of user characteristics.
Step 10: calculate the \( S_{k,\text{whole}}, S_{k,\text{local}}, S_{k,\text{local}} \), and \( S_{k,\text{whole}} \) according to equations (9), (11), (15), and (16).

Step 11: calculate the \( \eta_{k} \) and \( \eta_{k} \) according to equations (12) and (17).
Step 12: calculate the optimized \( \tilde{\eta}_{k} \) and \( \hat{\eta}_{k} \) according to equation (23).
Step 13: optimize the discriminative function \( J(A) \) according to equation (27).
Step 14: according to the obtained projection matrix, get the projection in low-dimensional subspace \( h_{i,k} \).

**Manifold discriminative learning for hybrid precoding**

On the basis of manifold discriminative learning for global dimensionality reduction and user clustering, we investigate the sum-rate maximization problem for hybrid precoding. In order to design the precoding matrix \( F_{G_i} W_{G_i} \), the sum-rate maximization of mixed precoding \( g \) is studied, such that they manage intra-cluster interference and inter-cluster interference. To improve the spectral efficiency of the systems, the design of each cluster analog precoding should balance optimizing self-transmission and interference. By representing each user set as a manifold, the received signal of the cluster can be represented as

\[
y'_{G_i} = H_{G_i}^T H_{G_i} W_{G_i} x_{G_i} + n_{G_i} \]

where \( y'_{G_i} = [y'_{G_i,1}, \ldots, y'_{G_i,K}]^T \) represents the received signal, \( H_{G_i} = [H_{G_i,1}, \ldots, H_{G_i,K}] \) represents the channel matrix for the \( i \)th cluster, \( F_{G_i} = [F_{G_i,1}, \ldots, F_{G_i,K}] \) represents the precoding matrix of each cluster, \( W_{G_i} = \text{diag}(W_{G_i,1}, \ldots, W_{G_i,K}) \). \( H_{G_i}^T H_{G_i} W_{G_i} x_{G_i} \) is the desired signal, \( \sum_{k'=1}^{K} H_{G_i,k'}^T F_{G_i,k'} W_{G_i,k'} x_{G_i,k'} \) is the intra-cluster interference, \( \sum_{k'=1, k' \neq k}^{K} H_{G_i,k'}^T F_{G_i,k'} W_{G_i,k} x_{G_i,k} \) is the inter-cluster interference after the low-dimensional mapping. In order to adapt to special scenarios and requirements, the hybrid precoding matrix can be determined by per-cluster processing (PCP). The goal of PCP is to balance the performance and complexity by effectively separating the clusters in the RF beam domain.

In PCP mode, the analog precoding matrix \( F_{G_i} \) of each cluster is calculated according to manifold quasi-conjugate gradient algorithm, while the digital precoding matrix \( W_{G_i} \) is calculated by each user cluster according to their equivalent channel matrix. Let \( H_{G_i}^T = H_{G_i}^T F_{G_i} \) denote the equivalent channel matrix after analog precoding, and it is an approximate block diagonal matrix, which can be expressed as
where $H'_{eq0} = H'_{eq} \cdot H'_{eq}$ represents the diagonal elements of the matrix in equation (32), and off-diagonal elements of the matrix $H'_{eq} H'_{eq} (i \neq i')$ represents the interference channel matrix between user clusters. After analog precoding, the inter-cluster interference is eliminated, that is, $H'_{eq} H'_{eq} \approx 0$. $H'_{eq}$ can be expressed as

$$H'_{eq} = \begin{bmatrix} H'_{eq1} & 0 & \cdots & 0 \\ 0 & H'_{eq2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H'_{eqG} \end{bmatrix}$$

The digital precoding matrix $W'$ is a block diagonal matrix as follows

$$W' = \text{diag}(W'_{G1}, \ldots, W'_{G_L})$$

With scalar equalization $\beta_{Gi}^{-1}$, the signal estimate $\hat{x}_{Gi}$ for $G_i$ can be expressed as

$$\hat{x}_{Gi} = \beta_{Gi}^{-1} \left( H'_{eqi} H'_{eqi} W'_{Gi} x_{Gi} + \sum_{k' = 1, k' \neq i}^{\left| G_i \right|} H'_{eqk'} H'_{eqk'} W'_{Gi,k'} x_{Gi,k'} + \sum_{i' = 1, i' \neq i}^{L} H'_{eqi'} H'_{eqi'} W'_{Gi,i'} x_{Gi,i'} + n_{Gi} \right)$$

where $\beta_{Gi}$ is a scaling equalization that is jointly optimized with the hybrid precoding. The conditional mean square error (MSE) for $G_i$ is

$$\varepsilon(F'_{Gi}, W'_{Gi}, \beta_{Gi}) = E \left[ \| x_{Gi} - \hat{x}_{Gi} \|^2 \right]$$

$$= E \left[ \| x_{Gi} - \beta_{Gi}^{-1} \left( H'_{eqi} H'_{eqi} W'_{Gi} x_{Gi} \right) \|^2 \right] + E \left[ \sum_{k' = 1, k' \neq i}^{\left| G_i \right|} \| \beta_{Gi}^{-1} H'_{eqk'} H'_{eqk'} W'_{Gi,k'} x_{Gi,k'} \|^2 \right] + \sum_{i' = 1, i' \neq i}^{L} \| \beta_{Gi}^{-1} H'_{eqi'} H'_{eqi'} W'_{Gi,i'} x_{Gi,i'} \|^2 + \beta_{Gi}^{-2} n_{Gi}$$

The conditional MSE in equation (31) is simplified as

$$\varepsilon(F'_G, W'_G, \beta_G) = \varepsilon_{G}^{(1)} + \varepsilon_{G}^{(2)}$$

where

$$\varepsilon_{G}^{(1)} = E \left[ \| x_{Gi} - \beta_{Gi}^{-1} \left( H'_{eqi} H'_{eqi} W'_{Gi} x_{Gi} \right) \|^2 \right]$$

$$\varepsilon_{G}^{(2)} = E \left[ \sum_{k', k' \neq i}^{\left| G_i \right|} \left\| \beta_{Gi}^{-1} H'_{eqk'} H'_{eqk'} W'_{Gi,k'} x_{Gi,k'} \right\|^2 \right] + \beta_{Gi}^{-2} n_{Gi}$$

Therefore, the hybrid precoding based on interference leakage is jointly optimized with $F'_G, W'_G$, and $\beta_G$. According to the literature, $W'_G$ can be transformed into $W''_G = \beta_G W'_G$, where $W''_G$ is an unnormalized digital precoding matrix, which can be obtained by KKT conditions as

$$W''_G = \left( H'_{eq0} H'_{eq0} + \gamma_{G}^{-1} I_{G} \right)^{-1} H'_{eq0} H$$

where $\gamma_{G}^{-1}$ is the regularization factor, which depends on noise variance and base station transmit power. $I_{G}$ is

$$I_{G} = \sum_{k' = 1, k' \neq i}^{\left| G_i \right|} H_{Gk', Gk'}^H F_{Gk', Gk'} W_{Gk', Gk'} x_{Gk', k} + \sum_{i' = 1, i' \neq i}^{L} H_{G, i'}^H H_{G, i'} W_{G, i'} x_{G, i'} + n_{Gi}$$

The optimal value given in Ayach et al. is $\gamma_{G}^{-1} = P_{tol}/K\sigma^2$. $P_{tol}$ is the total power of the transmitted signal. The optimal scaling factor $\beta_G$ can be obtained from the base station transmission power with $tr(F'_G W'_G H F'_G H) \leq P_{tol}$ as

$$\beta_G = \frac{P_{tol}}{\sqrt{\sum_{i = 1}^{L} tr(F'_G W'_G W_{G, G_i} H F'_G H)}}$$

Accordingly, equation (38) can be expressed as
We solve the following objective function with the
strained optimization problem, which can be solved
tion, but this problem is equivalent to the uncon-
However, the above function is not a convex optimiza-
the users in the cluster
metrical and discriminant embedding space
radio frequency precoding matrix
From the above analysis, it is essentially to find a
After simple mathematical derivation, equation (39)
can be expressed as
From the above analysis, it is essentially to find a
radio frequency precoding matrix \( F_{c} \), so that the objective function obtains a minimum value of \( \hat{c}_{G_{t}}^{(1)} \) and \( \hat{c}_{G_{t}}^{(2)} \). However, the above function is not a convex optimization, but this problem is equivalent to the unconstrained optimization problem, which can be solved using the manifold optimization method.
To obtain a minimum value, we discover both geometrical and discriminant embedding space \( \Psi_{L} \) for each cluster \( L \) in the radio frequency precoding matrix \( F_{c} \).
Following some simple algebraic steps, equation (45) can be reduced to

\[
S_{\phi} = \frac{1}{2} \sum_{f_i} \left\| \Psi_L^T F_{f_i} - \Psi_L^T F_{f_i} \right\|^2 B_{\phi,f_i,f_i}^L \\
= \sum_{f_i} \Psi_L^T F_{f_i} D_{\phi,f_i}^L (F_{f_i})^T \Psi_L \\
- \sum_{f_i} \Psi_L^T F_{f_i} B_{\phi,f_i,f_i}^L (F_{f_i})^T \Psi_L \\
= \Psi_L^T \tilde{F}^* (D_{\phi} - B_{\phi}) (\tilde{F}^*)^T \Psi_L \\
= \Psi_L^T \Omega_{\phi} (\tilde{F}^*)^T \Psi_L
\]

where \( D_{\phi}^L \) a diagonal matrix with \( D_{\phi}^L(f_i,f_i) = \sum_{f_j} B_{\phi,f_i,f_j}^L(f_j,f_j) \), \( \Omega_{\phi} = (D_{\phi} - B_{\phi})^L \) is the Laplacian matrix of the \( B_{\phi,f_i,f_j}^L \), and \( f^* = \{1, 2, \ldots, G_t\} \).

Similarly, equation (46) can be simplified as

\[
S_{\phi} = \frac{1}{2} \sum_{f_i} \left\| \Psi_L^T F_{f_i} - \Psi_L^T F_{f_i} \right\|^2 B_{\phi,f_i,f_i}^L \\
= \Psi_L^T \tilde{F}^* (D_{\phi} - B_{\phi}) (\tilde{F}^*)^T \Psi_L \\
= \Psi_L^T \Omega_{\phi} (\tilde{F}^*)^T \Psi_L
\]

Thus, the discriminant embedding space \( \Psi_L \) can be derived by maximizing the following objective function

\[
J(\Psi_L) = \frac{\tau(S_{\phi})}{\tau(S_{\phi})} = \frac{\tau(S_{\phi})}{\tau(S_{\phi})}
\]

This is equivalent to find the largest \( \theta \) eigenvalue of the following generalized eigenvalue problem

\[
\tilde{F}^* \Omega_{\phi} (\tilde{F}^*)^T v = \alpha \tilde{F}^* \Omega_{\phi} (\tilde{F}^*)^T v
\]

Let the largest \( \theta \) eigenvalues \( \{\alpha_1, \alpha_2, \ldots, \alpha_\theta\} \) that correspond to the largest \( \theta \) eigenvectors \( \{v_1, v_2, \ldots, v_\theta\} \) be the solutions of equation (52), which are chosen as the discriminant embedding space \( \Psi_L \), where

\[
\alpha_1 \geq \ldots \geq \alpha_2 \geq \alpha_\theta.
\]

Once we have obtain the discriminant embedding space \( \Psi_L \), the most distinguishing precedings are preserved in the most suitable projection space. In the low-dimensional discriminative embedding space, neighboring users of the same cluster approach each other, while preventing users of other clusters from entering the neighborhood.

Local discriminant matrix is defined as follows

\[
G_{local} = \frac{\alpha'}{1 + \alpha'} \Omega_{\phi}
\]

where \( \alpha' \) is a scalar parameter. The local discriminant matrix is used as the input of kernel function.

In the general case, the distance in a Grassmann manifold (a particular class of manifolds) is the length of the shortest geodesic connecting two users in lineal subspaces \( U_{f_i} \) and \( U_{f_j} \), this is

\[
d_G(U_{f_i}, U_{f_j}) = ||\Theta||_2
\]

where \( \Theta = [\tau_1, \tau_2, \ldots, \tau_\theta] \) is the principal angles \( 0 \leq \tau_1 \leq \tau_2 \leq \ldots \leq \tau_\theta \leq \pi/2 \).

Having defined projection distance over Grassmann manifold, the distance from user \( f_i \) to \( f_j \) in the discriminant embedding space \( \Psi_L \) also needs to be taken into account. To cover the intrinsic linear subspace, the user-to-user distance is denoted by

\[
\frac{(F'_{f_i})^T \Psi_L (\Psi_L)^T F'_{f_j})^2}{(F'_{f_i})^T \Psi_L (\Psi_L)^T (F'_{f_j})} = \frac{(F'_{f_i})^T \Psi_L (\Psi_L)^T (F'_{f_j})}{(F'_{f_i})^T \Psi_L (\Psi_L)^T (F'_{f_j})} + \alpha ||U_{f_i} - U_{f_j} U_{f_j}^T U_{f_j}||_F
\]

When the two distances are formalized as above, we arrive at the following form of user-to-user distance metric

\[
d(f_i, f_j, \Psi_L) = (1 - \alpha) \frac{(F'_{f_i})^T \Psi_L (\Psi_L)^T F'_{f_j})^2}{(F'_{f_i})^T \Psi_L (\Psi_L)^T (F'_{f_j})} \frac{\alpha ||U_{f_i} - U_{f_j} U_{f_j}^T U_{f_j}||_F}{(F'_{f_i})^T \Psi_L (\Psi_L)^T (F'_{f_j})} + \alpha ||U_{f_i} - U_{f_j} U_{f_j}^T U_{f_j}||_F^2
\]

In equation (55), the former describes how far away the origins of the two coordinate systems \( U_{f_i} \) and \( U_{f_j} \) in the discriminant embedding space \( \Psi_L \), whereas the latter reflects the correlation between two orthonormal basis matrices.

The local discriminant matrix \( G_{local} \) is chosen as the input of Kernel function

\[
G_{\phi_{f_i}} = \left\{ \begin{array}{ll}
\exp \left( -\sum_{t=1}^{\theta} \left( \alpha_{t, \theta} + \sum_{t=1}^{\theta} \beta_{t, \theta} U_{f_i} \right) - \left( \alpha_{t, \theta} + \sum_{t=1}^{\theta} \beta_{t, \theta} U_{f_j} \right) \right)_{G_{local}} \\
0, \quad \text{otherwise}
\end{array} \right.
\]

where \( t \) is the parameter of the kernel function, \( \alpha_{t, \theta} \) and \( \alpha_{t, \theta} \) are the mean vectors of the manifolds where the users are located, and \( \beta_{t, \theta} \) are the vector of free parameters of \( F'_{f_i} \) and \( F'_{f_j} \).

By minimizing the objective function below, we can find a suitable mapping on which the manifolds belonging to the same subspace can be closer and the manifolds in different subspaces can be further apart.
The above formula can be simplified to
\[
\Gamma^T D_{\text{kernel}} (D - G_{f_i f_i}) D^T_{\text{kernel}} \Gamma = 1
\] (59)

where
\[
D_{\text{kernel}} = \sum_{f_i} \left\| \left( \alpha_{f_i} + \sum_{f_j} \beta_{f_j} U_{f_j} \right) - \left( \alpha_{f_i} + \sum_{f_j} \beta_{f_j} U_{f_j} \right) \right\|_2^2
\] (60)

D is a diagonal matrix, this is, \(D(f_i, f_j) = \sum_{f_k} G_{f_k f_k}, D\) establishes a natural metric between different subspaces. To limit \(z\) on a fixed scale, we add the following constraint
\[
z D z^T = \Gamma^T D_{\text{kernel}} D^T_{\text{kernel}} \Gamma = 1
\] (61)

Therefore, the objective function with constraint is obtained as follows
\[
\Gamma = \arg \min_{\Gamma} \Gamma^T D_{\text{kernel}} (D - G_{f_i f_i}) D^T_{\text{kernel}} \Gamma
\]
\[
s.t. \quad \Gamma^T D_{\text{kernel}} D^T_{\text{kernel}} \Gamma = 1 \quad (62)
\]

Let \(J(\Gamma)\) represent the objective function. The hybrid precoding optimization problem based on interference leakage under orthogonal constraints is
\[
\arg \min_{\Gamma} J(\Gamma) = \sum_{i=1}^{L} \ell_{G_i}^{(1)} + \sum_{i=1}^{L} \ell_{G_i}^{(2)} \quad (63)
\]
\[
s.t. \quad \Gamma^T D_{\text{kernel}} D^T_{\text{kernel}} \Gamma = 1
\]

Finally, we can obtain the optimal \(\Gamma\) by minimizing eigenvalue solution to the generalized eigenvalue problem.

Therefore, solving the objective function can be transformed into a convex optimization problem. The optimal radio frequency precoding matrix \(F'_{G_i}\) is found to obtain a minimum value of \(\ell_{G_i}^{(1)}\) and \(\ell_{G_i}^{(2)}\).

The manifold algorithm to find the optimal radio frequency precoding matrix \(F'_{G_i}\) is as follows:

**Step 1:** initialize the analog precoding matrix \(F'_{G_i,0}\), error threshold \(\varepsilon \in (0, 1)\), the discriminant embedding space \(\Psi_{L}\).

**Step 2:** the learning discriminant embedding space \(\Psi_{L}\) can be derived according to equations (45)-(51).

**Step 3:** find the largest \(\vartheta\) eigenvalue according to equation (52).

**Step 4:** compute local discriminant matrix \(G_{\text{local}}\) according to equation (53).

**Step 5:** compute subspace-to-subspace distance according to equation (56).

**Step 6:** update \(\Gamma\) to subject to \(\Gamma^T D_{\text{kernel}} D^T_{\text{kernel}} \Gamma = 1\), and find the optimal radio frequency precoding matrix \(F'_{G_i}\) according to equation (63). Update the analog precoding matrix until convergence to satisfy the error threshold condition, the algorithm ends.

For the intra-cluster, there is correlation between the channels between users. Users in non-adjacent clusters are much smaller than users in adjacent clusters, but the interference is equal. So, the impact of remote user clusters on users in the cluster can be ignored. The SNR of a user cluster \(G_i\) in \(b\)th cell is as follows

\[
\text{SNR}_{G_i} = \frac{|H_{G_i,0} H_{G_i}^* W_{G_i}^*|^2 P_{G_i}}{\sum_{k' = 1, k' \neq k}^{[G_i]} |IN_{G_i, k'}|^2 P_{G_i} + \sum_{r = 1, r \neq i}^L |IN_{G_r}|^2 P_{G_r} + \sigma_G^2}
\] (64)

where \(IN_{G_i, k'} = H_{G_i, k'.k} H_{G_i}^* W_{G_i}^*\), \(IN_{G_r} = H_{G_r,0} H_{G_r}^* W_{G_r}^*\), and \(P_{G_i}\) are the transmit power of the \(G_i\)th cluster, and \(P_{G_i, k'}\) and \(P_{G_r}\) are the transmit power of the \(k'\)th user in the \(i\)th cluster and the transmit power of the \(G_r\)th cluster, respectively.

The capacity of mmWave massive MIMO system can be expressed as

\[
\text{SUM} = \sum_{G_i - G_i} \log_2 (1 + \text{SNR}_{G_i})
\] (65)

Equation (65) can be written as

\[
\text{SUM} = \sum_{G_i \neq G_i} \log_2 \left( 1 + \frac{|H_{G_i,0} H_{G_i}^* W_{G_i}^*|^2 P_{G_i}}{\sum_{k' = 1, k' \neq k}^{[G_i]} |IN_{G_i, k'}|^2 P_{G_i} + \sum_{r = 1, r \neq i}^L |IN_{G_r}|^2 P_{G_r} + \sigma_G^2} \right)
\] (66)
In this section, we will study the SE and bit error rate (BER) performance of the proposed hybrid precoder. We compare the method in this article with several traditional methods, that is, OMP, KDHB, AFHB, MO, and RTRNM. We also consider the performance of this method for non-mobile users and mobile users. The basic simulation parameters are as follows.

The carrier frequency is 60 GHz. The AoAs and AoDs are uniformly distributed in $[0, 2\pi]$, and a common Angle Spread (AS) $\Delta = 8$. The complex gain of each path obeys the distribution $CN(0, 1)$. The ULA is adopted in simulations. Among them, there are many overlapping parts between the channel power azimuth spectrum, which causes inter-cluster interference.

Figure 4 shows the differences of sum-rate performance with traditional schemes in the mmWave massive MIMO system of hybrid precoding. Let $N_t = 128, N_R = 32$, and $K = 32$. In Figure 4, we can see that the proposed scheme can achieve a very high sum-rate than other existing precodings against different SNR. The reason for this situation is that traditional methods have not solved the problem of non-linear resolution of multi-user high-dimensional channels. In our proposed scheme, each user set is represented by manifold, and the cluster-oriented multi-manifold learning scheme is used to solve the above problems. Aiming at the high-density hotspot scene of the cell, the geometric model of clustering users is studied. This method can better solve the interference within and between clusters. Through user clustering and hybrid precoding, the achievable summation rate of the mmWave massive MIMO system is improved.

Figure 5 shows the differences of the BER performance of different hybrid precoding schemes, where the channel parameters in the single-cell scenario are the same as above. From Figure 5, in the case of different SNRs, a conclusion similar to Figure 4 can be drawn. We can also see that the proposed-based manifold discriminative learning scheme achieve a better BER performance than other schemes. The proposed scheme improves beamspace resolution and reduces the influence of power leakage on beamspace channel.

Figure 6 compares the average sum-rate of this article and the traditional scheme, and other existing precodings with different numbers of users. We set $n_{RF,i} = g_i$ in each cluster. It can be seen from Figure 6 that this method is superior to other methods. Due to the continuous increase in the number of users, the traditional solution does not consider the problem of non-linear resolution of multi-user high-dimensional channels. This scheme can well solve the interference in clusters, between clusters and between cells. The
solution proposed in this article can significantly improve the average and rate of mmWave massive MIMO systems.

Figure 7 shows the change trend of the system average SNR as the SNR changes. Figure 8 shows the average SE when the BS antenna changes. We can learn from the figure that the proposed method achieves an average SE that is significantly higher than other traditional schemes. From Figure 7 that through the proposed manifold learning scheme, each user and its adjacent high-dimensional channels are in a global and local non-linear neighborhood. Aiming at the high-density hotspot scene of the community, the geometric model of clustering users is studied. The proposed scheme manages the multi-user and inter-cell interference and improves the data rate for cell-edge users. From Figure 8, the proposed method can effectively and extensively use antennas in multiple low-dimensional manifolds.

Conclusion

A hybrid precoding scheme for user clustering is proposed, which can solve the problem of large-scale mmWave MIMO in multiple low-dimensional manifolds, avoiding the high-dimensional complex operations of traditional schemes. For the BS, mmWave massive MIMO obtains a low-dimensional learning channel matrix by manifold. Then user clustering hybrid precoding is studied for the transmitted signal based on the low-dimensional channel matrix. The manifold discriminative learning seek to learn the embedding low-dimensional subspace, where manifolds with different user group labels are easier to distinguish, and the local spatial correlation of high-dimensional channels in each manifold is enhanced. Through proper user clustering, the hybrid precoding is investigated for the sum-rate maximization problem by manifold quasi-conjugate gradient methods. The simulation results show that the method has good robustness on the basis of reducing the computational complexity of the mmWave massive MIMO system.

More realistic precoding for MIMO is expected in the future research. In popular learning after user clustering we just use the traditional method, the choice of user clustering center is crucial, if the user is in high speed of movement will make our method inaccurate. So, more advanced research is needed. In addition, our current method is applied in hybrid precoding of full connections; precoding research is still needed for sub-connections as well as dynamic connections.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by the Shanghai Capacity Building Projects in Local Institutions under Grant 19070502900.

ORCID iDs

Xiaoping Zhou https://orcid.org/0000-0002-4051-9558
Yang Wang https://orcid.org/0000-0001-8100-9194

References

1. Wang X, Jia M, Guo Q, et al. Joint power, original bandwidth, and detected hole bandwidth allocation for
2. Wang P, Li Y, Song L, et al. Multi-gigabit millimeter wave wireless communications for 5G: from fixed access to cellular networks. *IEEE Commun Mag* 2015; 53(1): 168–178.

3. Rangan S, Rappaport TS and Erkip E. Millimeter-wave cellular wireless networks: potentials and challenges. *P IEEE* 2014; 102(3): 366–385.

4. Vlachos E, Alexandropoulos GC and Thompson J. Massive MIMO channel estimation for millimeter wave systems via matrix completion. *IEEE Signal Proc Let* 2018; 25(11): 1675–1679.

5. Sun Y and Qi C. Weighted sum-rate maximization for analog beamforming and combining in millimeter wave massive MIMO communications. *IEEE Commun Lett* 2017; 21(8): 1883–1886.

6. Sohrabi F and Yu W. Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays. *IEEE J Sel Area Comm* 2017; 35(7): 1432–1443.

7. Liu A, Lau VKN and Zhao M. Stochastic successive convex optimization for two-timescale hybrid precoding in massive MIMO. *IEEE J Sel Top Signa* 2018; 12(3): 432–444.

8. He S, Qi C, Wu Y, et al. Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems. *IEEE Access* 2016; 4: 9895–9905.

9. Alexandropoulos GC and Chouvardas S. Low complexity channel estimation for millimeter wave systems with hybrid A/D antenna processing. In: 2016 *IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, 4–8 December 2016, pp.1–6. New York: IEEE.

10. Han S, I C, Xu Z, et al. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Commun Mag* 2015; 53(1): 186–194.

11. Yu X, Zhang J and Letaief KB. A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems. *IEEE J Sel Top Signa* 2018; 12(2): 282–297.

12. Ayach OE, Rajagopal S, Abu-Surra S, et al. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE T Wirel Commun* 2014; 13(3): 1499–1513.

13. Huang C, Liu L, Yuen C, et al. A LSE and sparse message passing-based channel estimation for mmWave MIMO systems. In: 2016 *IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, 4–8 December 2016, pp.1–6. New York: IEEE.

14. Gao Z, Dai L, Han S, et al. Compressive sensing techniques for next-generation wireless communications. *IEEE Wirel Commun* 2018; 25(3): 144–153.

15. Chen C, Tsai C, Liu Y, et al. Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems. *IEEE T Signal Proces* 2017; 65(6): 1412–1424.

16. Huang Y, Zhang J and Xiao M. Constant envelope hybrid precoding for directional millimeter-wave communications. *IEEE J Sel Area Comm* 2018; 36(4): 845–859.

17. Zhu G, Huang K, Lau VKN, et al. Hybrid beamforming via the Kronecker decomposition for the millimeter-wave massive MIMO systems. *IEEE J Sel Area Comm* 2017; 35(9): 2097–2114.

18. He S, Wang J, Huang Y, et al. Codebook-based hybrid precoding for millimeter wave multiuser systems. *IEEE T Signal Proces* 2017; 65(20): 5289–5304.

19. Kim M and Lee YH. MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels. *IEEE T Veh Technol* 2015; 64(6): 2714–2720.

20. Mir T, Zain Siddiqi M, Mir U, et al. Machine learning inspired hybrid precoding for wideband millimeter-wave massive MIMO systems. *IEEE Access* 2019; 7: 62852–62864.

21. Zhang J, Huang Y, Wang J, et al. Hybrid precoding for wideband millimeter-wave systems with finite resolution phase shifters. *IEEE T Veh Technol* 2018; 67(11): 11285–11290.

22. Park S, Alkhateeb A and Heath RW. Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems. *IEEE T Wirel Commun* 2017; 16(5): 2907–2920.

23. Li H, Li M and Liu Q. Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems. *IEEE T Commun* 2020; 68(1): 602–614.

24. Jiang J, Yuan Y and Zhen L. Multi-user hybrid precoding for dynamic subarrays in mmWave massive MIMO systems. *IEEE Access* 2019; 7: 101718–101728.

25. Sun S, Rappaport TS, Shafi M, et al. Analytical framework of hybrid beamforming in multi-cell millimeter-wave systems. *IEEE T Wirel Commun* 2018; 17(11): 7528–7543.

26. Yu X, Shen J, Zhang J, et al. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems. *IEEE J Sel Top Signa* 2016; 10(3): 485–500.

27. Chen J. Low-PAPR precoding design for massive multiuser MIMO systems via Riemannian manifold optimization. *IEEE Commun Lett* 2017; 21(4): 945–948.

28. Mai R, Le-Ngoc T and Nguyen DHN. Two-timescale hybrid RF-baseband precoding with MMSE-VP for multi-user massive MIMO broadcast channels. *IEEE T Wirel Commun* 2018; 17(7): 4462–4476.

29. Lin T, Cong J, Zhu Y, et al. Hybrid beamforming for millimeter wave systems using the MMSE criterion. *IEEE T Commun* 2019; 67(5): 3693–3708.

30. Zhou X, Wang P, Yang Z, et al. A manifold learning performance analysis of user scheduling in massive MIMO with fast moving users. In: 2019 *IEEE 30th annual interna- tional symposium on personal, indoor and mobile radio communications (PIMRC)*, Istanbul, 8–11 September 2019, pp.1–6. New York: IEEE.

31. Salous S, Espositi VD, Fuschini F, et al. Millimeter-wave propagation: characterization and modeling toward fifth-generation systems [Wireless Corner]. *IEEE Antenn Propag* M 2016; 58(6): 115–127.

32. Feng J, Wang J, Zhang H, et al. Fault diagnosis method of joint fisher discriminant analysis based on the local and global manifold learning and its kernel version. *IEEE T Autom Sci Eng* 2016; 13(1): 122–133.
34. Sun Y, Gao Z, Wang H, et al. Principal component analysis-based broadband hybrid precoding for millimeter-wave massive MIMO systems. *IEEE T Wirel Commun* 2020; 19(10): 6331–6346.

35. Li Y, Cao G and Cao W. LMDAPNet: a novel manifold-based deep learning network. *IEEE Access* 2020; 8: 65938–65946.

36. Selvan SE, Amato U, Gallivan KA, et al. Descent algorithms on oblique manifold for source-adaptive ICA contrast. *IEEE T Neural Net Lear* 2012; 23(12): 1930–1947.

37. Lu S, Hong M and Wang Z. A nonconvex splitting method for symmetric nonnegative matrix factorization: convergence analysis and optimality. *IEEE T Signal Proc* 2017; 65(12): 3120–3135.

38. Pascual-Garcia J, Molina-Garcia-Pardo J, Martinez-Inglés M, et al. On the importance of diffuse scattering model parameterization in indoor wireless channels at mm-Wave frequencies. *IEEE Access* 2016; 4: 688–701.

39. Meng D, Cao G, Cao W, et al. Supervised feature learning network based on the improved LLE for face recognition. In: 2016 international conference on audio, language and image processing (ICALIP), Shanghai, China, 11–12 July 2016, pp.306–311. New York: IEEE.

40. Van Tulder G and de Bruijne M. Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines. *IEEE T Med Imaging* 2016; 35(5): 1262–1272.

41. Ruan H, Xiao P, Xiao L, et al. Joint iterative optimization-based low-complexity adaptive hybrid beamforming for massive MU-MIMO systems. *IEEE T Commun* 2021; 69(3): 1707–1722.

42. Belkin M and Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: *Proceedings of the advances in neural information processing systems*, 2002, pp.585–591. https://proceedings.neurips.cc/paper/2001/hash/f106b7f99d2cb30c3db1c3cc0fde9c9f-cb-Abstract.html