Progress In Electromagnetics Research M, Vol. 68, 163–171, 2018

P-OMP-IR Algorithm for Hybrid Precoding in Millimeter Wave MIMO Systems

Ruiyan Du[1], Fulai Liu[1, *], Xinwei Wang[2], Qingping Zhou[3], and Xiaoyu Bai[2]

Abstract—This paper presents a P-OMP-IR algorithm for the hybrid precoding problem in millimeter wave (mm-Wave) multiple-input multiple-output (MIMO) systems. In the proposed approach, the digital precoding matrix is updated via the orthogonal matching pursuit (OMP) method, and the analog precoding matrix is refined column by column using the dominant singular value and corresponding singular vectors of a residual matrix successively. During the refining phase of the analog precoding matrix, an extended power method is designed to calculate the dominant singular value and the corresponding left and right singular vectors, which is able to reduce the computational complexity significantly. Simulation results show that the proposed algorithm can not only reduce the residual of the hybrid precoder effectively, but also improve the spectral efficiency consistently.

1. INTRODUCTION

Millimeter wave (mm-Wave) multiple-input multiple-output (MIMO) system is emerging as a promising technology for the next generation communication [1]. Although mm-Wave MIMO systems experience higher path loss than existing cellular systems, they can take advantage of the large bandwidth in mm-Wave spectrum and employ large-scale antenna arrays that are packed in a very small area due to short wavelength. With large number of antenna elements at the transmitter/receiver, mm-Wave MIMO array can provide significant beamforming gains to combat path loss [2].

Hybrid precoding architecture has attracted wide attention recently, which only requires a small number of radio frequency (RF) chains between the low-dimensional digital precoder and high-dimensional analog precoder [3]. Several methods of hybrid precoder are conceived in [4–6], including the optimal analog beamforming of clustered subarrays [4], minimum-mean-square-error (MMSE)-based analog/digital beamforming [5], and joint transmit/receive beamforming [6]. The convex quadratic programming method and least squares method are utilized to obtain the analog precoding matrix and digital precoding matrix in [7]. A restricted set of dominant candidate directions is leveraged to find the analog precoding matrix in [8]. A successive interference cancelation-based hybrid precoding scheme for subarray structures is proposed in [9]. The advantages of directional beamforming are studied by considering two-path mm-Wave channels in [10]. In [11, 12], the hybrid precoding problem is reformulated as a sparse signal reconstruction problem and solved via the orthogonal matching pursuit (OMP) method. The sparse hybrid precoding algorithm can approach the full digital MIMO processor by iteratively selecting a beamforming vector from the set of array response vectors; however, due to the limited size of the set of array response vectors, the sparse hybrid precoding algorithm still inevitably leads to performance losses.

In this paper, a P-OMP-IR algorithm is proposed to further improve the performance of the sparse hybrid precoding. The hybrid precoding matrix of the proposed algorithm is initialized through the OMP algorithm. Then, each column vector of the analog precoding matrix is refined by using the

Received 10 April 2018, Accepted 9 May 2018, Scheduled 16 May 2018

* Corresponding author: Fulai Liu (fulailiu@126.com).

[1] Engineer Optimization & Smart Antenna Institute, Northeastern University at Qinhuangdao, China. [2] School of Computer Science and Engineering, Northeastern University, Shenyang, China. [3] Tangshan Normal University, Tangshan, China.

dominant singular vectors of a residual matrix. Afterwards, with the refined analog precoding matrix, the digital precoding matrix is updated via the OMP method again. Finally, by alternatively iterating the above stages, the performance of the hybrid precoder is improved gradually. It is noteworthy that the proposed algorithm depends on the dominant singular vectors of a residual matrix. In order to reduce the computational complexity, an extended power method is proposed to derive the dominant left and right singular vectors simultaneously.

The remaining part of this paper is organized as follows. The system model is introduced in Section 2. Section 3 presents details of the proposed algorithm. Simulations are given to demonstrate the performance of the proposed method in Section 4. Section 5 concludes the paper.

Notations: $\boldsymbol{A}$ stands for a matrix. $\boldsymbol{a}$ represents a vector. $\boldsymbol{A}_{(m,n)}$ denotes the element of $\boldsymbol{A}$ corresponding to the mth row and nth column. $\boldsymbol{A}_{(:,n)}$ represents the nth column of $\boldsymbol{A}$. $\boldsymbol{A}_{(:,1:n)}$ is the first n columns of $\boldsymbol{A}$. $\boldsymbol{A}_{(n,:)}$ represents the nth row of $\boldsymbol{A}$. $(\boldsymbol{A})^T$, $(\boldsymbol{A})^H$, $(\boldsymbol{A})^{-1}$ and $(\boldsymbol{A})^\dagger$ denote the transpose, the conjugate transpose, the inverse and the pseudoinverse of $\boldsymbol{A}$, respectively. $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. $|\boldsymbol{A}|$, $\|\boldsymbol{A}\|_2$ and $\|\boldsymbol{A}\|_F$ represent the determinant, the 2-norm and the Frobenius norm of $\boldsymbol{A}$, respectively. $\mathcal{CN}(a, b)$ is a complex Gaussian distribution with mean a and variance b. $\mathrm{diag}(\boldsymbol{A})$ is a vector which consists of the diagonal elements of $\boldsymbol{A}$. $\mathrm{diag}\{a_1, \cdots, a_N\}$ is a diagonal matrix with the entries in $\{a_1, \cdots, a_N\}$ on its diagonal. $\mathrm{rank}(\boldsymbol{A})$ is the rank of $\boldsymbol{A}$. $\mathbb{E}[\cdot]$ denotes the expectation operation.

2. SYSTEM MODEL

Consider a single-user mm-Wave MIMO system with the hybrid precoder and combiner as shown in Fig. 1, where N_t denotes the number of transmit antennas, N_r the number of receive antennas, and N_s the number of data streams. The number of RF chains is denoted by N_{RF} such that $N_s \leq N_{RF} \leq \min(N_t, N_r)$. Without loss of generality, it is assumed that the transmitter and receiver have the same number of RF chains. $\boldsymbol{F}_{RF} \in \mathbb{C}^{N_t \times N_{RF}}$ represents the analog precoding matrix, and $\boldsymbol{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$ is the digital precoding matrix. The transmitted signal $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = \boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\boldsymbol{s} \tag{1}$$

where $\boldsymbol{s}$ is the $N_s \times 1$ vector of transmitted symbols and $\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^H] = \frac{1}{N_s}\boldsymbol{I}$. The normalized transmit power is imposed by $\|\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\|_F^2 = N_s$. Then the received signal can be written as

$$\boldsymbol{y} = \sqrt{\rho}\boldsymbol{W}_{BB}^H\boldsymbol{W}_{RF}^H\boldsymbol{H}\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\boldsymbol{s} + \boldsymbol{W}_{BB}^H\boldsymbol{W}_{RF}^H\boldsymbol{n} \tag{2}$$

where ρ is the average received power, $\boldsymbol{W}_{RF}$ the $N_r \times N_{RF}$ analog combining matrix, $\boldsymbol{W}_{BB}$ the $N_{RF} \times N_s$ digital combining matrix at the receiver, $\boldsymbol{n}$ the noise drawn from the Gaussian distribution $\mathcal{CN}(0, \sigma_n^2)$, and $\boldsymbol{H}$ the $N_r \times N_t$ channel matrix.

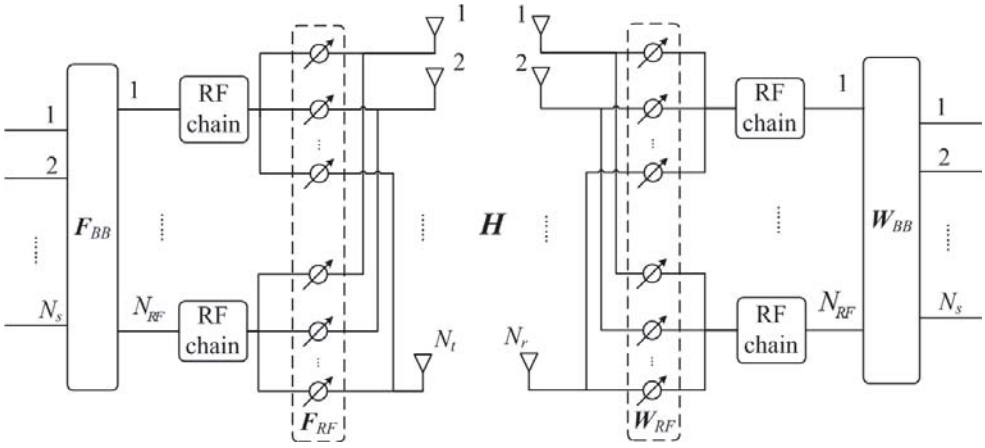

Figure 1. A single-user mm-Wave MIMO system with hybrid analog-digital precoder and combiner.

The narrowband Saleh-Valenzuela clustering channel model is adopted to characterize the mm-Wave propagation environment [13], i.e.,

$$\boldsymbol{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{il} \boldsymbol{a}_r(\phi_{il}^r, \theta_{il}^r) \boldsymbol{a}_t(\phi_{il}^t, \theta_{il}^t)^H \tag{3}$$

where N_{cl} is the number of clusters, N_{ray} the number of the propagation paths in each cluster, α_{il} the gain of the lth ray in the ith cluster distributed as $\mathcal{CN}(0,1)$. $\boldsymbol{a}_t(\phi_{il}^t, \theta_{il}^t)$ and $\boldsymbol{a}_r(\phi_{il}^r, \theta_{il}^r)$ denote the transmit array response vector and receive array response vector, respectively. $(\phi_{il}^r, \theta_{il}^r)$ stands for the azimuth and elevation angles of arriver (AOAs), and $(\phi_{il}^t, \theta_{il}^t)$ denotes the azimuth and elevation angles of departure (AODs). When a uniform planar array (UPA) is considered, the array response vector of the lth ray in the ith cluster can be expressed as [14]

$$\boldsymbol{a}(\phi_{il}, \theta_{il}) = \left[1 \quad \cdots \quad e^{j\frac{2\pi}{\lambda} d(p \sin\phi_{il} \sin\theta_{il} + q\cos\theta_{il})} \quad \cdots \quad e^{j\frac{2\pi}{\lambda} d((M-1)\sin\phi_{il}\sin\theta_{il} + (N-1)\cos\theta_{il})}\right]^T \tag{4}$$

where d denotes the antenna spacing; λ represents the signal wavelength; $p\,(0 \le p \le M)$ and $q\,(0 \le q \le N)$ are the antenna indices in the 2-dimensional plane.

In matrix form, the channel model in Eq. (3) can be rewritten as

$$\boldsymbol{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \boldsymbol{A}_r \boldsymbol{\Lambda} \boldsymbol{A}_t^H \tag{5}$$

where $\boldsymbol{A}_r = [\boldsymbol{a}_r(\phi_{11}^r, \theta_{11}^r), \cdots, \boldsymbol{a}_r(\phi_{N_{cl},N_{ray}}^r, \theta_{N_{cl},N_{ray}}^r)]$ and $\boldsymbol{A}_t = [\boldsymbol{a}_t(\phi^t, \theta_{11}^t), \cdots, \boldsymbol{a}_t(\phi_{N_{cl},N_{ray}}^t, \theta_{N_{cl},N_{ray}}^t)]$ are the array response matrices of transmitter and receiver respectively. $\boldsymbol{\Lambda} = \mathrm{diag}\{\alpha_{11}, \cdots, \alpha_{N_{cl},N_{ray}}\}$ is a diagonal matrix.

When the Gaussian symbols are transmitted throughout the mm-Wave channel, the achieved spectral efficiency R can be given by [15]

$$R = \log_2\left(\left|\boldsymbol{I}_{N_s} + \frac{\rho}{N_s} \boldsymbol{R}_n^{-1} \boldsymbol{W}_{BB}^H \boldsymbol{W}_{RF}^H \boldsymbol{H} \boldsymbol{F}_{RF} \boldsymbol{F}_{BB} \boldsymbol{F}_{BB}^H \boldsymbol{F}_{RF}^H \boldsymbol{H}^H \boldsymbol{W}_{RF} \boldsymbol{W}_{BB}\right|\right) \tag{6}$$

where $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{W}_{BB}^H \boldsymbol{W}_{RF}^H \boldsymbol{W}_{RF} \boldsymbol{W}_{BB}$ is the noise covariance matrix after the receiving processing. Furthermore, all the elements of $\boldsymbol{F}_{RF}$ and $\boldsymbol{W}_{RF}$ should satisfy the unit modulus constraints, namely, $|\boldsymbol{F}_{RF(m,n)}| = |\boldsymbol{W}_{RF(m,n)}| = 1$.

Note that directly maximizing spectral efficiency R requires a joint optimization over the four matrix variables $(\boldsymbol{F}_{RF}, \boldsymbol{F}_{BB}, \boldsymbol{W}_{RF}, \boldsymbol{W}_{BB})$. However, it is intractable to find the global optimal solution of the joint optimization problem. Fortunately, the joint design problem can be approximately separated into two sub-problems [12], that is, the precoding and combining problems that have similar mathematical formulations. This paper will mainly focus on the precoding problem, and the presented algorithm can be used to solve the combining problem equivalently.

The precoder optimization problem can be expressed as [16]

$$\min_{\boldsymbol{F}_{RF}, \boldsymbol{F}_{BB}} \|\boldsymbol{F}_{\mathrm{opt}} - \boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\|_F^2 \qquad \text{s.t.} \quad |\boldsymbol{F}_{RF(m,n)}| = 1, \quad \|\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\|_F^2 = N_s \tag{7}$$

where $\boldsymbol{F}_{\mathrm{opt}}$ is the unconstrained fully digital precoder that can be obtained from the SVD of the channel $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, i.e., $\boldsymbol{F}_{\mathrm{opt}} = \boldsymbol{V}^H$. $|\boldsymbol{F}_{RF(m,n)}| = 1$ is the unit modulus constraint, and $\|\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\|_F^2 = N_s$ denotes the transmit power constraint. Finding the optimal analytical solutions of Eq. (7) is still intractable due to the unit modulus constraint of $\boldsymbol{F}_{RF}$, and Section 3 will further analyze the optimization problem.

3. ALGORITHM FORMULATION

In this section, the proposed P-OMP-IR algorithm is carried out. The OMP-based sparse precoding algorithm is implemented to initialize the hybrid precoding matrix, and then, the RF and digital precoding matrices are refined alternatively by using the dominant singular vectors of a residual matrix and the OMP method, respectively. The specific process is described in the rest of this section.

3.1. Analog Precoding Matrix Refinement

Let's start with the initialization of the hybrid precoding matrix. As stated above, the hybrid precoding matrix is initialized with the OMP-based sparse hybrid algorithm in [12]. Essentially, the above initialization problem can be considered as a sparse signal reconstruction problem, which is solved with sparse signal recovery methods. According to the compressed sensing theory, the dictionary plays an important role in the OMP method, which can be obtained by either an analytical or a learning-based approach. The analytical approach generates the dictionary via a predefined mathematical transform, while for the learning-based approach, the dictionary is adapted from a set of training signals. Learned dictionaries have potential to improve the performance of the sparse signal recovery algorithms (e.g., the OMP algorithm) further, since the learned dictionaries can capture the salient information directly from the training signals [17]. Therefore, in this section, a novel design is proposed to reduce the residual of the hybrid precoder through refining the matrix $\boldsymbol{F}_{RF}$ iteratively. The residual β between the optimal precoder $\boldsymbol{F}_{\text{opt}}$ and the hybrid precoder $\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}$ can be given by

$$\beta = \|\boldsymbol{F}_{\text{opt}} - \boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\|_F^2 = \left\|\boldsymbol{F}_{\text{opt}} - \sum_{i=1}^{N_{RF}} \boldsymbol{f}_{RF}^i \boldsymbol{f}_{BB}^i\right\|_F^2 \tag{8}$$

where $\boldsymbol{f}_{RF}^i$ $(1 \le i \le N_{RF})$ is the ith column vector of matrix $\boldsymbol{F}_{RF}$, and $\boldsymbol{f}_{BB}^i$ $(1 \le i \le N_{RF})$ denotes the ith row vector of matrix $\boldsymbol{F}_{BB}$. In the refinement stage of $\boldsymbol{F}_{RF}$, the column vectors of $\boldsymbol{F}_{RF}$ are modified one by one to reduce the coherence among them [18].

For the i_0th column of $\boldsymbol{F}_{RF}$ and the corresponding i_0th row of $\boldsymbol{F}_{BB}$, the optimization problem can be expressed as

$$\underset{\boldsymbol{f}_{RF}^{i_0}, \boldsymbol{f}_{BB}^{i_0}}{\arg\min} \left\|\boldsymbol{F}_{\text{opt}} - \sum_{i \ne i_0}^{N_{RF}} \boldsymbol{f}_{RF}^i \boldsymbol{f}_{BB}^i - \boldsymbol{f}_{RF}^{i_0} \boldsymbol{f}_{BB}^{i_0}\right\|_F^2. \tag{9}$$

Let $\boldsymbol{G}_{i_0} = \boldsymbol{F}_{\text{opt}} - \sum_{i \ne i_0}^{N_{RF}} \boldsymbol{f}_{RF}^i \boldsymbol{f}_{BB}^i$, the problem in Eq. (9) can be rewritten as

$$\underset{\boldsymbol{f}_{RF}^{i_0}, \boldsymbol{f}_{BB}^{i_0}}{\arg\min} \left\|\boldsymbol{G}_{i_0} - \boldsymbol{f}_{RF}^{i_0} \boldsymbol{f}_{BB}^{i_0}\right\|_F^2. \tag{10}$$

Ignoring the unit modulus constraint temporarily, the optimal solution of the problem in Eq. (10) can be given by the Eckart-Young-Mirsky theorem [19], i.e.,

$$\underset{\boldsymbol{f}_{RF}^{i_0}, \boldsymbol{f}_{BB}^{i_0}}{\arg\min} \left\|\boldsymbol{G}_{i_0} - \boldsymbol{f}_{RF}^{i_0} \boldsymbol{f}_{BB}^{i_0}\right\|_F^2 = \left\|\boldsymbol{G}_{i_0} - \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^H\right\|_F^2 \tag{11}$$

where σ_1 is the largest singular value of $\boldsymbol{G}_{i_0}$, and $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the left and right singular vectors corresponding to σ_1. By Eq. (11), $\boldsymbol{f}_{RF}^{i_0}$ and $\boldsymbol{f}_{BB}^{i_0}$ can be refined by the following equations

$$\boldsymbol{f}_{RF}^{i_0} = \boldsymbol{u}_1, \quad \boldsymbol{f}_{BB}^{i_0} = \sigma_1 \boldsymbol{v}_1^H. \tag{12}$$

After the refinement of all the column vectors of $\boldsymbol{F}_{RF}$ and the row vectors of $\boldsymbol{F}_{BB}$, the constrained RF precoding matrix $\boldsymbol{F}_{RF}$ can be given by $\boldsymbol{F}_{RF(m,n)} = \frac{\boldsymbol{F}_{RF(m,n)}}{|\boldsymbol{F}_{RF(m,n)}|}$, $\forall m, n$ [20].

3.2. Estimation of the Dominant Singular Value and the Corresponding Singular Vectors

From Eq. (12), it is clear that the proposed method depends on the dominant singular value and vectors. In order to circumvent the SVD involved high computational complexity, the power method [21] is extended in this subsection to quickly estimate the dominant singular value and the corresponding singular vectors. The following theorem describes the proposed extended power method in detail.

Theorem 1: Suppose that $\boldsymbol{G}$ is an $m \times n$ matrix with singular values $\sigma_1 > \sigma_2 \geq \ldots \geq \sigma_r > 0$ and dominant singular vectors $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$. For the vector sequence

$$\begin{cases} \boldsymbol{b}_{k-1} = \dfrac{\boldsymbol{G}\boldsymbol{a}_{k-1}}{\|\boldsymbol{G}\boldsymbol{a}_{k-1}\|_2} \\ \boldsymbol{a}_k = \dfrac{\boldsymbol{G}^H\boldsymbol{b}_{k-1}}{\|\boldsymbol{G}^H\boldsymbol{b}_{k-1}\|_2} \end{cases} \qquad k = 1, 2, \cdots, \tag{13}$$

if $\boldsymbol{a}_0^H\boldsymbol{v}_1 = \alpha_1 \neq 0$, then the following statements are true:

$$\text{(a)} \lim_{k\to\infty} \boldsymbol{a}_k = \frac{\alpha_1}{|\alpha_1|}\boldsymbol{v}_1, \quad \text{(b)} \lim_{k\to\infty} \boldsymbol{b}_k = \frac{\alpha_1}{|\alpha_1|}\boldsymbol{u}_1, \quad \text{(c)} \lim_{k\to\infty} \|\boldsymbol{G}\boldsymbol{a}_k\|_2 = \sigma_1.$$

Proof: Let the SVD of $\boldsymbol{G} \in \mathbb{C}^{m\times n}$ be $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where $\boldsymbol{U} = [\boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \cdots \ \ \boldsymbol{u}_m]$, $\boldsymbol{V} = [\boldsymbol{v}_1 \ \ \boldsymbol{v}_2 \ \ \cdots \ \ \boldsymbol{v}_n]$, $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \text{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$. Decomposing $\boldsymbol{a}_0$ into $\boldsymbol{a}_0 = \sum_{i=1}^n \alpha_i\boldsymbol{v}_i$ ($\alpha_1 = \boldsymbol{a}_0^H\boldsymbol{v}_1 \neq 0$), from Eq. (13), $\boldsymbol{a}_k$ and $\boldsymbol{b}_k$ can be written as:

$$\begin{aligned} \boldsymbol{a}_k &= \frac{\boldsymbol{G}^H\boldsymbol{b}_{k-1}}{\|\boldsymbol{G}^H\boldsymbol{b}_{k-1}\|_2} = \frac{\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{a}_{k-1}}{\|\boldsymbol{G}\boldsymbol{a}_{k-1}\|_2\|\frac{\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{a}_{k-1}}{\|\boldsymbol{G}\boldsymbol{a}_{k-1}\|_2}\|_2} = \frac{\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{a}_{k-1}}{\|\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{a}_{k-1}\|_2} = \frac{(\boldsymbol{G}^H\boldsymbol{G})^k\boldsymbol{a}_0}{\|(\boldsymbol{G}^H\boldsymbol{G})^k\boldsymbol{a}_0\|_2} = \frac{(\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^H)^k\boldsymbol{a}_0}{\|(\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^H)^k\boldsymbol{a}_0\|_2} \\ &= \frac{(\sum_{i=1}^r \boldsymbol{v}_i\boldsymbol{\sigma}_i^2\boldsymbol{v}_i^H)^k\boldsymbol{a}_0}{\|(\sum_{i=1}^r \boldsymbol{v}_i\boldsymbol{\sigma}_i^2\boldsymbol{v}_i^H)^k\boldsymbol{a}_0\|_2} = \frac{(\sum_{i=1}^r \boldsymbol{v}_i\boldsymbol{\sigma}_i^{2k}\boldsymbol{v}_i^H)(\sum_{i=1}^n \alpha_i\boldsymbol{v}_i)}{\|(\sum_{i=1}^r \boldsymbol{v}_i\boldsymbol{\sigma}_i^{2k}\boldsymbol{v}_i^H)(\sum_{i=1}^n \alpha_i\boldsymbol{v}_i)\|_2} = \frac{\sum_{i=1}^r \alpha_i\sigma_i^{2k}\boldsymbol{v}_i}{\|\sum_{i=1}^r \alpha_i\sigma_i^{2k}\boldsymbol{v}_i\|_2} \\ &= \frac{\alpha_1\sigma_1^{2k}(\boldsymbol{v}_1 + \sum_{i=2}^r \frac{\alpha_i}{\alpha_1}(\frac{\sigma_i}{\sigma_1})^{2k}\boldsymbol{v}_i)}{\|\alpha_1\sigma_1^{2k}(\boldsymbol{v}_1 + \sum_{i=2}^r \frac{\alpha_i}{\alpha_1}(\frac{\sigma_i}{\sigma_1})^{2k}\boldsymbol{v}_i)\|_2} \end{aligned}$$

$$\begin{aligned} \boldsymbol{b}_k &= \frac{\boldsymbol{G}\boldsymbol{a}_k}{\|\boldsymbol{G}\boldsymbol{a}_k\|_2} = \frac{\boldsymbol{G}\boldsymbol{G}^H\boldsymbol{b}_{k-1}}{\|\boldsymbol{G}^H\boldsymbol{b}_{k-1}\|_2\|\frac{\boldsymbol{G}\boldsymbol{G}^H\boldsymbol{b}_{k-1}}{\|\boldsymbol{G}^H\boldsymbol{b}_{k-1}\|_2}\|_2} = \frac{\boldsymbol{G}\boldsymbol{G}^H\boldsymbol{b}_{k-1}}{\|\boldsymbol{G}\boldsymbol{G}^H\boldsymbol{b}_{k-1}\|_2} = \frac{(\boldsymbol{G}\boldsymbol{G}^H)^k\boldsymbol{b}_0}{\|(\boldsymbol{G}\boldsymbol{G}^H)^k\boldsymbol{b}_0\|_2} = \frac{(\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^H)^k\boldsymbol{b}_0}{\|(\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^H)^k\boldsymbol{b}_0\|_2} \\ &= \frac{(\sum_{i=1}^r \boldsymbol{u}_i\boldsymbol{\sigma}_i^2\boldsymbol{u}_i^H)^k\boldsymbol{G}\boldsymbol{a}_0}{\|(\sum_{i=1}^r \boldsymbol{u}_i\boldsymbol{\sigma}_i^2\boldsymbol{u}_i^H)^k\boldsymbol{G}\boldsymbol{a}_0\|_2} = \frac{(\sum_{i=1}^r \boldsymbol{u}_i\boldsymbol{\sigma}_i^{2k}\boldsymbol{u}_i^H)(\sum_{i=1}^r \alpha_i\sigma_i\boldsymbol{u}_i)}{\|(\sum_{i=1}^r \boldsymbol{u}_i\boldsymbol{\sigma}_i^{2k}\boldsymbol{u}_i^H)(\sum_{i=1}^r \alpha_i\sigma_i\boldsymbol{u}_i)\|_2} = \frac{\sum_{i=1}^r \alpha_i\sigma_i^{2k+1}\boldsymbol{u}_i}{\|\sum_{i=1}^r \alpha_i\sigma_i^{2k+1}\boldsymbol{u}_i\|_2} \\ &= \frac{\alpha_1\sigma_1^{2k+1}(\boldsymbol{u}_1 + \sum_{i=2}^r \frac{\alpha_i}{\alpha_1}(\frac{\sigma_i}{\sigma_1})^{2k+1}\boldsymbol{u}_i)}{\|\alpha_1\sigma_1^{2k+1}(\boldsymbol{u}_1 + \sum_{i=2}^r \frac{\alpha_i}{\alpha_1}(\frac{\sigma_i}{\sigma_1})^{2k+1}\boldsymbol{u}_i)\|_2} \end{aligned}$$

It is well known that $\lim_{k\to\infty}\left(\frac{\sigma_i}{\sigma_1}\right)^k = 0$, therefore,

$$\lim_{k\to\infty} \boldsymbol{a}_k = \lim_{k\to\infty} \frac{\alpha_1\sigma_1^{2k}\boldsymbol{v}_1}{\|\alpha_1\sigma_1^{2k}\boldsymbol{v}_1\|_2} = \frac{\alpha_1}{|\alpha_1|}\boldsymbol{v}_1, \qquad \lim_{k\to\infty} \boldsymbol{b}_k = \lim_{k\to\infty} \frac{\alpha_1\sigma_1^{2k+1}\boldsymbol{u}_1}{\|\alpha_1\sigma_1^{2k+1}\boldsymbol{u}_1\|_2} = \frac{\alpha_1}{|\alpha_1|}\boldsymbol{u}_1 \tag{14}$$

According to Eq. (14), it is clear that

$$\lim_{k\to\infty} ||\boldsymbol{G}\boldsymbol{a}_k||_2 = ||\boldsymbol{G}\boldsymbol{v}_1||_2 = ||\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{v}_1||_2 = \left\|\left(\sum_{i=1}^r \boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^H\right)\boldsymbol{v}_1\right\|_2 = ||\sigma_1\boldsymbol{u}_1||_2 = \sigma_1.$$

This concludes the proof.

According to Theorem 1, the extend power algorithm is given in Table 1, where η is the terminal threshold.

3.3. Summary of the P-OMP-IR Algorithm

Based on the above discussion, the pseudo-code of the proposed hybrid P-OMP-IR precoding algorithm can be summarized in Table 2. The proposed algorithm starts by selecting parts of columns from the array response matrix $\boldsymbol{A}_t$ with the OMP algorithm (Step 1–Step 7) to form the initial hybrid precoding matrix $\boldsymbol{F}_{RF}^{(0)}$ and $\boldsymbol{F}_{BB}^{(0)}$.

Table 1. The extended power algorithm.

Input: $\boldsymbol{G}$, $\boldsymbol{a}_0$, η
Initializations: $k=0$
repeat
1: $k=k+1$
2: $\boldsymbol{b}_{k-1}=\frac{\boldsymbol{G}\boldsymbol{a}_{k-1}}{\|\boldsymbol{G}\boldsymbol{a}_{k-1}\|_2}$
3: $\boldsymbol{a}_k=\frac{\boldsymbol{G}^H\boldsymbol{b}_{k-1}}{\|\boldsymbol{G}^H\boldsymbol{b}_{k-1}\|_2}$
until $\|\boldsymbol{a}_k-\boldsymbol{a}_{k-1}\|_2\le\eta$
4: $\sigma_1=\|\boldsymbol{G}\boldsymbol{a}_k\|_2$
Output: $\boldsymbol{u}_1=\boldsymbol{b}_k$, $\boldsymbol{v}_1=\boldsymbol{a}_k$, σ_1

Table 2. P-OMP-IR algorithm.

Input: $\boldsymbol{F}_{\text{opt}}$, N_{RF}, $\boldsymbol{A}_t$, ε
Initializations: $\boldsymbol{F}_{\text{res}}=\boldsymbol{F}_{\text{opt}}$, $\boldsymbol{F}_{RF}^{(0)}=\boldsymbol{A}_t$, $t=0$
repeat
1: $\boldsymbol{F}_{RF}=\text{EmptyMatrix}$
for $j=1\to N_{RF}$ **do**
2: $\boldsymbol{\Phi}=\boldsymbol{F}_{RF}^{(t)H}\boldsymbol{F}_{\text{res}}$
3: $k=\arg\max(\text{diag}(\boldsymbol{\Phi}\boldsymbol{\Phi}^H))$
4: $\boldsymbol{F}_{RF}=\left[\boldsymbol{F}_{RF}\middle\vert\boldsymbol{F}_{RF(:,k)}^{(t)}\right]$
5: $\boldsymbol{F}_{BB}^{(t)}=(\boldsymbol{F}_{RF}^H\boldsymbol{F}_{RF})^{-1}\boldsymbol{F}_{RF}^H\boldsymbol{F}_{\text{opt}}$
6: $\boldsymbol{F}_{\text{res}}=\frac{\boldsymbol{F}_{\text{opt}}-\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}^{(t)}}{\|\boldsymbol{F}_{\text{opt}}-\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}^{(t)}\|_F}$
end for
7: $\boldsymbol{F}_{RF}^{(t)}=\boldsymbol{F}_{RF}$
8: $\delta^t=\|\boldsymbol{F}_{\text{opt}}-\boldsymbol{F}_{RF}^{(t)}\boldsymbol{F}_{BB}^{(t)}\|_F$
Refinement of analog precoder
for $i=1\to N_{RF}$ **do**
9: $\boldsymbol{G}_i=\boldsymbol{F}_{\text{opt}}-\sum_{n\neq i}^{N_{RF}}\boldsymbol{F}_{RF(:,n)}^{(t)}\boldsymbol{F}_{BB(n,:)}^{(t)}$
10: Compute $\boldsymbol{u}_1$, $\boldsymbol{v}_1$ and σ_1 of $\boldsymbol{G}_i$ by the extended power method
11: $\boldsymbol{F}_{RF(:,i)}^{(t)}=\boldsymbol{u}_1, \boldsymbol{F}_{BB(i,:)}^{(t)}=\sigma_1\boldsymbol{v}_1^H$
end for
12: $t=t+1$
13: $\boldsymbol{F}_{RF(m,n)}^{(t)}=\frac{\boldsymbol{F}_{RF(m,n)}^{(t-1)}}{\lvert\boldsymbol{F}_{RF(m,n)}^{(t-1)}\rvert},\forall m,n$
until $(\delta^{t-1}-\delta^t)\le\varepsilon$
14: $\boldsymbol{F}_{BB}=\sqrt{N_s}\frac{\boldsymbol{F}_{BB}^{(t)}}{\|\boldsymbol{F}_{RF}^{(t)}\boldsymbol{F}_{BB}^{(t)}\|_F}$
Output: $\boldsymbol{F}_{RF}=\boldsymbol{F}_{RF}^{(t)}$, $\boldsymbol{F}_{BB}$

Afterwards, the hybrid precoder is refined by using the extended power method in Step 9–Step 13. Then, the updated matrix $\boldsymbol{F}_{RF}^{(t)}$ is used as a new input to the OMP algorithm for the next iteration. The process continues until $(\delta^{t-1}-\delta^t)\le\varepsilon$, where ε denotes the predefined threshold. Finally, the transmitter power constraint is considered in Step 14.

4. SIMULATION RESULTS

In this section, several simulation results are presented to evaluate the performance of the proposed hybrid precoding algorithm for a single-user 144×36 mm-Wave MIMO system.

The mm-Wave propagation channel is modeled by $L = 50$ paths which are equally divided into 5 clusters $\mathcal{C}_i$ $(i = 1, 2, \cdots, 5)$, and each cluster contains 10 rays. The average azimuth and elevation AODs of each cluster, i.e., $\phi^t_{\mathcal{C}_i} = \frac{1}{10}\sum_{l\in\mathcal{C}_i}\phi^t_l$, $\theta^t_{\mathcal{C}_i} = \frac{1}{10}\sum_{l\in\mathcal{C}_i}\theta^t_l$ $(i = 1, 2, \cdots, 5)$, distribute uniformly in $(0, 2\pi)$. The azimuth and elevation AODs of rays in each cluster are drawn from Laplace distribution, i.e., $\phi^t_{l,l\in\mathcal{C}_i} \sim \mathcal{L}(\mu^t_{\phi,\mathcal{C}_i}, b^t_{\phi,\mathcal{C}_i})$, $\theta^t_{l,l\in\mathcal{C}_i} \sim \mathcal{L}(\mu^t_{\theta,\mathcal{C}_i}, b^t_{\theta,\mathcal{C}_i})$, where the location parameters $\mu^t_{\phi,\mathcal{C}_i} = \phi^t_{\mathcal{C}_i}$, $\mu^t_{\theta,\mathcal{C}_i} = \theta^t_{\mathcal{C}_i}$, and the scale parameters $b^t_{\phi,\mathcal{C}_i}$ and $b^t_{\theta,\mathcal{C}_i}$ are set as 10°. The statistic properties of azimuth and elevation AOAs are the same as the azimuth and elevation AODs. All the results are averaged over 1000 random channel realizations.

Figure 2 compares the proposed algorithm with the OMP-based sparse precoding algorithm in [12], the successive refinement (SR)-OMP-based sparse precoding algorithm in [22], the PE-AltMin algorithm in [16] and the optimal full digital precoding when SNR = 0 dB, and $N_s = 4$. It is clear that the proposed algorithm achieves better performance than other hybrid precoders, which implies that the proposed refinement process is effective for reducing the residual of the hybrid precoder. Besides, more RF chains can reduce the performance gap between the proposed algorithm and the full digital precoder remarkably, since more degrees of freedom can be used as the the number of RF chains increases.

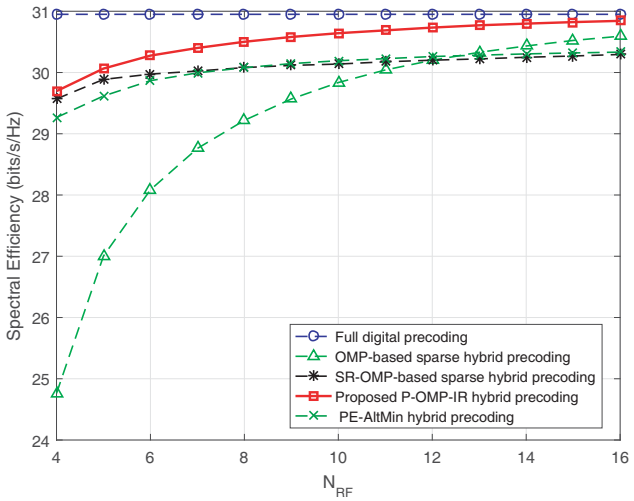

Figure 2. Spectral efficiencies given by different algorithms as functions of the number of RF chains N_{RF} when SNR = 0 dB and $N_s = 4$.

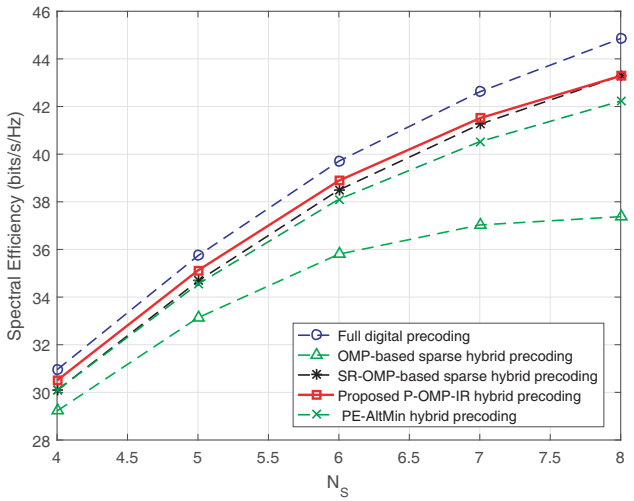

Figure 3. Spectral efficiencies given by different algorithms as functions of the number of data streams N_s when SNR = 0 dB and $N_{RF} = 8$.

Figure 3 shows spectral efficiency curves with different numbers of data streams N_s, when SNR = 0 dB and $N_{RF} = 8$. Obviously, the advantage of the proposed algorithm is more significant when N_s is smaller. For the case $N_s = N_{RF} = 8$, the spectral efficiencies given by the proposed algorithm and the SR-OMP-based sparse precoding algorithm are almost the same, and the performance deterioration of the proposed precoding is more severe than the full digital precoder. Therefore, for fixed number of RF chains, transmitting relatively less data streams could reduce the performance loss of the hybrid precoder effectively.

Figure 4 gives the residuals of the proposed approach for different numbers of iterations when SNR = 0 dB, $N_s = 4$, and the number of RF chains varies from 4 to 8. Explicitly, the residuals decrease as the iteration increases for different RF chains, which suggests that the proposed algorithm is convergent under normal cases. Furthermore, the residuals of 8 RF chains are much smaller than that of 4 RF chains. This difference is caused by the increasing degrees of freedom of the hybrid precoder.

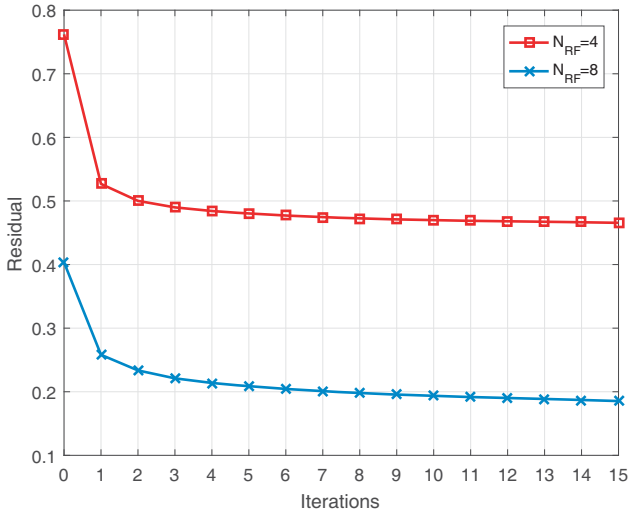

Figure 4. Residuals of the proposed hybrid precoding algorithm as functions of the iterations.

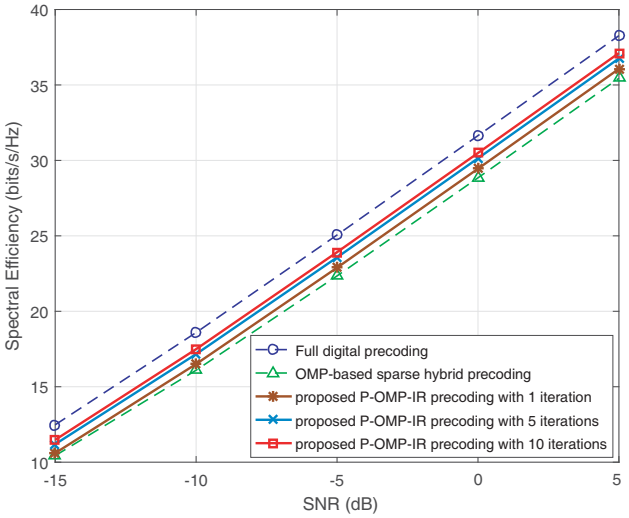

Figure 5. Spectral efficiencies given by different algorithms as functions of the SNR and iterations when $N_s = 4$ and $N_{RF} = 6$.

Figure 5 gives the spectral efficiency curves of the proposed method with different iterations when $N_{RF} = 6$ and $N_s = 4$. The optimal full-digital precoding and OMP-based sparse precoding serve as benchmarks. It can be seen that the spectral efficiency given by the proposed algorithm is improved gradually as the iterations increase. When $t = 5$, the proposed algorithm can provide about 1 dB SNR gain compared with the OMP-based sparse precoding algorithm. However, if iterations increase continuously, the improvements will be marginal, which means that the proposed algorithm can approach a steady solution with a few iterations.

5. CONCLUSION

In this paper, a hybrid precoding algorithm is proposed for mm-Wave MIMO systems. In the presented approach, the OMP technique is used to initialize the hybrid precoder, then the digital precoding matrix and RF precoding matrix are alternatively refined by the OMP method and the dominant singular vectors given by the extend power method. Simulation results show that the proposed algorithm can not only approach a steady solution with a few iterations, but also offer higher spectral efficiencies than existing sparse precoding algorithms.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of Hebei Province (No. F2016501139), the Fundamental Research Funds for the Central Universities under Grant (Grant No. N172302002, Grant No. N162304002), and the National Natural Science Foundation of China (Grant No. 61501102, Grant No. 61473066).

REFERENCES

1. Kutty, S. and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, Vol. 18, No. 2, 949–973, 2016.
2. Heath, R. W., N. Gonzalez-Prelcic, Jr., S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, Vol. 10, No. 3, 436–452, 2016.

3. Han, S., I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, Vol. 53, No. 1, 186–194, 2015.
4. Kanatas, A. G., "A receive antenna subarray formation algorithm for MIMO systems," *IEEE Commun. Lett.*, Vol. 11, No. 5, 1396–1398, 2007.
5. Venkateswaran, V. and A. J. Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, Vol. 58, No. 8, 4131–4143, 2010.
6. Nsenga, J., A. Bourdoux, W. V. Thillo, V. Ramon, and F. Horlin, "Joint Tx/Rx analog linear transformation for maximizing the capacity at 60 GHz," *IEEE Int. Conf. Commun.*, 1–5, 2011.
7. Ni, W., X. Dong, and W. S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, Vol. 65, No. 15, 3922–3933, 2017.
8. Singh, J. and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, Vol. 14, No. 5, 2670–2683, 2015.
9. Dai, L., X. Gao, J. Quan, S. Han, and I. Chih-Lin, "Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems," *Proc. IEEE Int. Conf. Commun.*, 1334–1339, 2015.
10. Raghavan, V., S. Subramanian, J. Cezanne, and A. Sampath, "Directional beamforming for millimeter-wave MIMO systems," *IEEE Global Commun. Conf.*, 1–7, 2015.
11. Alkhateeb, A., O. E. Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," *Inf. Theory Appl. Workshop*, 1–5, 2013.
12. Ayach, O. E., S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, Vol. 13, No. 3, 1499–1513, 2014.
13. Rusu, C., R. Mendez-Rial, N. Gonzalez-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, Vol. 15, No. 12, 8380–8393, 2016.
14. Balanis, C., *Antenna Theory*, Wiley, 1997.
15. Alkhateeb, A., O. E. Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," *IEEE Inf. Theory Appl. Workshop*, 1–5, 2013.
16. Yu, X., J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, Vol. 10, No. 3, 485–500, 2016.
17. Wei, D., T. Xu, and W. Wang, "Simultaneous codeword optimization (SimCO) for dictionary update and learning," *IEEE Trans. Signal Process.*, Vol. 60, No. 12, 6340–6353, 2011.
18. Aharon, M., M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, Vol. 54, No. 11, 4311–4322, 2006.
19. Wei, M., "Perturbation theory for the Eckart-Young-Mirsky theorem and the constrained total least squares problem," *Linear Algebra Appl.*, Vol. 280, No. 2, 267–287, 1998.
20. Alkhateeb, A. and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, Vol. 64, No. 5, 1801–1818, 2016.
21. Liu, F. L., R. Y. Du, J. P. Guo, and S. M. Guo, "P-GLRT algorithm for cooperative spectrum sensing," *Wireless Personal Commun.*, Vol. 81, No. 3, 1079–1089, 2015.
22. Mirza, J., B. Ali, S. S. Naqvi, and S. Saleem, "Hybrid precoding via successive refinement for millimeter wave MIMO communication systems," *IEEE Commun. Lett.*, Vol. 21, No. 5, 991–994, 2017.