

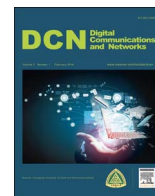
HOSTED BY



ELSEVIER

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.elsevier.com/locate/dcan

Rank-defective millimeter-wave channel estimation based on subspace-compressive sensing[☆]

Majid Shakhsi Dastgahian*, Hossein Khoshbin

Department of Electrical Engineering, Ferdowsi University of Mashhad, Iran

ARTICLE INFO

Keywords:

Millimeter wave communications
Sparse channel estimation
Rank-defective
Subspace enhancement
Multiple measurement vectors (MMV)

ABSTRACT

Millimeter-wave communication (mmWC) is considered as one of the pioneer candidates for 5G indoor and outdoor systems in E-band. To subdue the channel propagation characteristics in this band, high dimensional antenna arrays need to be deployed at both the base station (BS) and mobile sets (MS). Unlike the conventional MIMO systems, Millimeter-wave (mmW) systems lay away to employ the power predatory equipment such as ADC or RF chain in each branch of MIMO system because of hardware constraints. Such systems leverage to the hybrid precoding (combining) architecture for downlink deployment. Because there is a large array at the transceiver, it is impossible to estimate the channel by conventional methods. This paper develops a new algorithm to estimate the mmW channel by exploiting the sparse nature of the channel. The main contribution is the representation of a sparse channel model and the exploitation of a modified approach based on Multiple Measurement Vector (MMV) greedy sparse framework and subspace method of Multiple Signal Classification (MUSIC) which work together to recover the indices of non-zero elements of an unknown channel matrix when the rank of the channel matrix is defected. In practical rank-defective channels, MUSIC fails, and we need to propose new extended MUSIC approaches based on subspace enhancement to compensate the limitation of MUSIC. Simulation results indicate that our proposed extended MUSIC algorithms will have proper performances and moderate computational speeds, and that they are even able to work in channels with an unknown sparsity level.

1. Introduction

Thanks to the tremendous growing demand for data rates in cellular networks, it seems to be essential to shift the operating frequency of cellular systems from the conventional microwave spectrum to promising E-band Millimeter-wave spectrum (30–300 GHz) for indoor and even outdoor applications [1,2]. Free license and wideband spectrum in millimeter wave along with using of the array with massive number of antenna make a plentiful combination for fifth generation (5G) [3]. Fortunately, the very small wavelengths of mmW signals (between 1–10 mm) make it possible to pack a miniaturized large number of antennas into transceivers thereby providing high beamforming gains that can compensate severe path loss caused by shadowing phenomena because of encountering higher frequency to

environmental obstructions such as oxygen absorb; humidity fades and reflective outdoor materials [4]. Propagation measurements in urban environments show that the mmW channel is sparse in angular domain meaning only a few scattering clusters [5,6]. For resolving the problems related to sparsity, compressed sensing (CS) literature has been studied extensively in recent years [7]. The main goal of CS approaches is trying to recover a sparse signal successfully from a few linear measurements. Furthermore, to overcome the poor propagation characteristics in this frequency band and boost the traveling range of waves, adaptive beamforming may make systems less vulnerable to unfavorable shadowing effects. However, in the mmW system, it is unfeasible to dedicate one complete Radio Frequency (RF) chain and one high-resolution analog to digital converter (ADC) or digital to analog converter (DAC) to each branch of antenna due to a high cost

Peer review under responsibility of Chongqing University of Posts and Telecommunications.

* Corresponding author.

E-mail addresses: majid.shakhsidastgahian@stu.um.ac.ir (M.S. Dastgahian), khoshbin@um.ac.ir (H. Khoshbin).

<http://dx.doi.org/10.1016/j.dcan.2016.10.006>

Received 15 April 2016; Received in revised form 17 October 2016; Accepted 24 October 2016

Available online xxxx

2352-8648/ © 2016 Chongqing University of Posts and Telecommunications. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by/4.0/>).

Please cite this article as: Shakhsi Dastgahian, M., Digital Communications and Networks (2016), <http://dx.doi.org/10.1016/j.dcan.2016.10.006>

and power consumption of these components. For this reason and for supporting capability of multiplexing several data streams and also achieving more accurate beamforming gain, a hybrid architecture has been proposed in [8,9] such that the processing is performed in analog stage with the number of RF chains much lower than the number of antennas and baseband stage. Baseband precoder (combiner) is used for correction of limitation of analog RF section. In [8,9], the sparse nature of the poor scattering mmW channel is exploited to develop low-complexity hybrid beamforming. Design of hybrid beamformers has been investigated for other architectures in [10,11]. For instance [10], the past proposed a scheme based on subspace estimation rather than estimation of the whole channel by utilizing the concept of the reciprocity of the channel in TDD MIMO systems. In most of the previous works, estimation of the channel or the beamforming has been based on the sparse recovery problem in compressive sensing for single measurement vector [12]. Convex optimization problem such as LASSO or some simple and fast suboptimal greedy algorithms such as orthogonal matching pursuit (OMP) can be leveraged to resolve the sparse problems [7].

In this paper, we consider a hybrid beamforming model for downlink single-user mmW systems. We assume to have a constant sensing matrix in several times of training mode as well as to know the geometry of the arrays in source and destination. Thus, we utilize the multiple measurement vectors (MMV) for sparse millimeter channel and propose different approaches for solving the channel estimation problem. We can consider the various greedy MMV algorithms where are mostly an extension of single measurement vector (SMV) prescriptions such as Simultaneous OMP (SOMP), Simultaneous Iterative Hard Thresholding (SIHT), Simultaneous Hard Thresholding Pursuit (SHTP) [13]. The main contribution of the paper is the development of the MUSIC based methods rather than the existing simultaneous algorithms for solving the joint sparse channel recovery. In practical rank-defective and noisy channels, MUSIC fails, and we need to propose some new approaches based on subspace enhancement to compensate the limitation of MUSIC [14]. Rank-deficiency may occur due to shortage of snapshots number than the sparsity level value, or correlation between sources or multipath propagation which is usual in millimeter wave communications.

The reminder of this paper is organized as follows. In Section 2, the system and channel models will be introduced. In Section 3, we will take the advantages of the sparse nature of the mmW channel and formalize its estimation as a compressive sensing problem. In Section 4, we first give MUSIC algorithm for estimating the subspace and solving joint recovery problem and then introduce the subspace enhancement approach for noisy and rank-defective channel. Afterwards, a sparsity level-blind algorithm based on the MUSIC and conventional greedy support recovery methods will be described. Performance comparison and conclusions are presented in Sections 5 and 6 respectively.

We use the following notations throughout this paper. The bold upper-case letters denote matrices, and bold lower-cases represent vectors. Furthermore, $\|\mathbf{A}\|_F$ and $\|\mathbf{a}\|$ stand for Frobenius norm of a matrix and Euclidian norm of a vector, whereas $\mathbf{A}^T, \mathbf{A}^H, \mathbf{A}^*$ and \mathbf{A}^\dagger are its transpose, conjugate transpose (Hermitian), conjugate and Moore-Penrose pseudo-inverse, accordingly. The terms of $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of \mathbf{A} and \mathbf{B} . The term of $\mathbf{A} \in \mathbb{C}^{i \times j}$ means that the matrix \mathbf{A} is a complex matrix with dimension of $i \times j$. We use $[\cdot]$ to denote expectation operator. $\text{vec}(\mathbf{A})$ means arranging all of columns of matrix \mathbf{A} in a column vector.

The j th column and j th row of the matrix \mathbf{A} is demonstrated by $\mathbf{A}_{\cdot j}$ and $\mathbf{A}_{(j)}$. For an arbitrary set of χ , the sub matrix of \mathbf{A}_χ is composed of selected columns of \mathbf{A} by entries of χ . For an arbitrary subspace S ,

matrix \mathbf{P}_S means the orthogonal projection onto the range space of S and \mathbf{P}_S^\perp indicates the orthogonal projection onto orthogonal complement of S^\perp .

2. System model

Assuming a single-user downlink Millimeter-wave MIMO system with N_t transmitter antennas at the base station (BS) and N_r receiver antennas at the Mobile station (MS). Each side is equipped with $N_{t_f}^r$ and $N_{r_f}^r$ Radio-Frequency (RF) chains. N_s data streams are considered separately to be sent into a sparse channel. In the proposed model, the number of component of the transmitter array is more than the receiver. Furthermore, the number of RF chains satisfies $N_s \leq N_{t_f}^r \leq N_{r_f}^r \leq \min(N_t, N_r)$. Fig. 1 depicts a hybrid single user MIMO mmW transceiver with spatial multiplexing gain and phase shifter as an analog beamformer.

The downlink signal at the receiver side before a baseband filter is given by,

$$\mathbf{x}_r = \mathbf{F}_r^H \mathbf{H} \mathbf{F}_t \mathbf{G}_t \mathbf{t} + \mathbf{F}_r^H \mathbf{n} \quad (1)$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the complex sparse channel assumed to be slowly block-fading, $\mathbf{F}_t \in \mathbb{C}^{N_t \times N_{t_f}^r}$ is the analog (RF) precoder, $\mathbf{G}_t \in \mathbb{C}^{N_{t_f}^r \times N_s}$ is the baseband precoder, $\mathbf{t} \in \mathbb{C}^{N_s \times 1}$ is transmitting signal vector with covariance matrix $\mathbb{E}[\mathbf{t}\mathbf{t}^H] = (\mathbf{P}_t/N_s)\mathbf{I}_{N_s}$ and $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the additive Gaussian noise at the receiver with $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \sigma_n^2 \mathbf{I}_{N_r}$. Similarly, $\mathbf{F}_r \in \mathbb{C}^{N_r \times N_{r_f}^r}$ and $\mathbf{G}_r \in \mathbb{C}^{N_{r_f}^r \times N_r}$ are the RF band and baseband combiners, respectively. The received signal, after filtering, is given by,

$$\mathbf{y}_r = \mathbf{G}_r^H \mathbf{x}_r = \mathbf{G}_r^H \mathbf{F}_r^H \mathbf{H} \mathbf{F}_t \mathbf{G}_t \mathbf{t} + \mathbf{G}_r^H \mathbf{F}_r^H \mathbf{n} = \mathbf{C}^H \mathbf{H} \mathbf{P} \mathbf{t} + \mathbf{C}^H \mathbf{n} \quad (2)$$

where \mathbf{C} is defined $\mathbf{F}_r \mathbf{G}_r \in \mathbb{C}^{N_r \times N_r}$. In Fig. 1, the block of phase shifters as analog precoder/combiner can be chosen from predefined codebooks or from random matrices with stochastic phases and constant amplitudes. Thus, a possible value set for μ th phase shifter and ν th RF chain in the matrix \mathbf{F} is $[\mathbf{F}]_{\mu,\nu} = 1/\sqrt{N_t} e^{j\vartheta_{\mu,\nu}}$ where $\vartheta_{\mu,\nu}$ as an M-bits quantized angle is chosen from uniform distribution in range of $[0, 2\pi)$. The total power constraint is compelled by normalizing \mathbf{G}_t such that $\|\mathbf{F}_t \mathbf{G}_t\|_F^2 = N_s$.

Millimeter-wave channel in outdoor environment is limited by a few numbers of propagation paths [5,6]. The statistical model for mmW channel is unsuitable due to poor scattering nature. Based on the parametric physical model of channel with L scatterers and assumption that each scatterer contributes a single propagation path between the BS and MS, the nonlinear channel \mathbf{H} in spatial angles (but linear in the path gains) can be indicated as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^L \beta_l \mathbf{v}_r(\theta_l) \mathbf{v}_t^H(\varphi_l) = \sqrt{\frac{N_t N_r}{L}} \mathbf{V}_r(\theta) \boldsymbol{\beta} \mathbf{V}_t^H(\varphi) \quad (3)$$

where $\boldsymbol{\beta} = \text{diag}([\beta_1, \beta_2, \dots, \beta_L]) \in \mathbb{C}^{L \times L}$ is the L dimensional propagation path gain diagonal matrix with independently and identically distributed complex Gaussian diagonal entries with zero mean and variance $1/L$. The $\mathbf{V}_r(\theta) \in \mathbb{C}^{N_r \times L}$ and $\mathbf{V}_t(\varphi) \in \mathbb{C}^{N_t \times L}$ in (3) represent array response matrices at the BS and MS, respectively. Such matrices are given by

$$\mathbf{V}_r(\theta) = [\mathbf{v}_r(\theta_1), \mathbf{v}_r(\theta_2), \dots, \mathbf{v}_r(\theta_L)] / \sqrt{N_r} \quad (4)$$

$$\mathbf{V}_t(\varphi) = [\mathbf{v}_t(\varphi_1), \mathbf{v}_t(\varphi_2), \dots, \mathbf{v}_t(\varphi_L)] / \sqrt{N_t} \quad (5)$$

The terms of φ_i and θ_i denote the Angle of Departure (AoD) and Angle of Arrival (AoA) of i th the independent path from L total paths. By assuming a uniform linear arrays (ULA) model [15], $\mathbf{v}_r(\theta_i)$ and $\mathbf{v}_t(\varphi_i)$ can be defined as

$$\mathbf{v}_r(\theta_i) = \left[1, e^{-j\frac{2\pi}{\lambda} d \sin(\theta_i)}, \dots, e^{-j\frac{2\pi}{\lambda} (N_r-1) d \sin(\theta_i)} \right]^T / \sqrt{N_r} \quad (6)$$

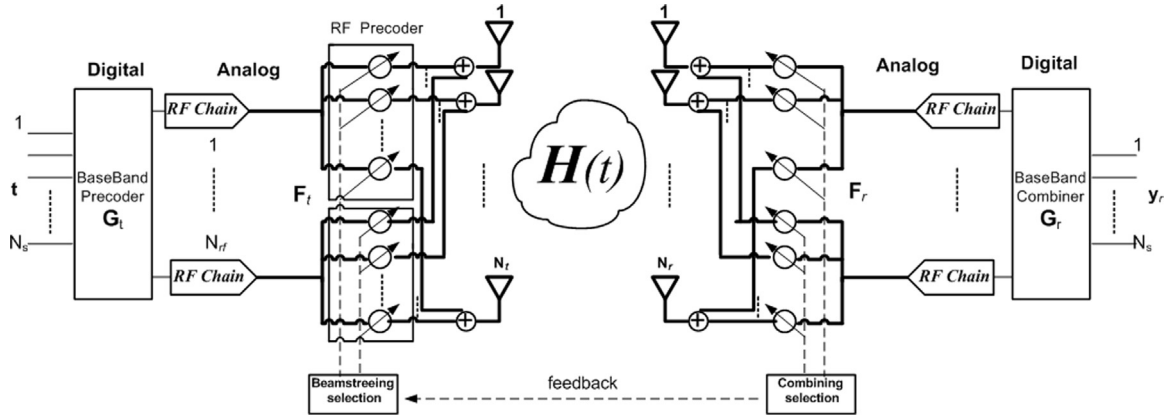


Fig. 1. Hybrid Model of millimeter wave channel.

$$\mathbf{v}_t(\varphi_l) = \left[1, e^{-j\frac{2\pi}{\lambda}d \sin(\varphi_l)}, \dots, e^{-j\frac{2\pi}{\lambda}(N_t-1)d \sin(\varphi_l)} \right]^T / \sqrt{N_t} \quad (7)$$

where λ is the signal wavelength and d is the inter-antenna distance such that it sets to $\lambda/2$ at both the BS and MS. It is assumed that $\mathbf{v}_r(\theta_l)$ and $\mathbf{v}_t(\varphi_l)$ are quasi-static at the receiver and transmitter, which means that $\mathbf{v}_r(\theta_l)$ and $\mathbf{v}_t(\varphi_l)$ vary slowly, and they can be well estimated at both sides of the transceiver.

3. Problem formulation

In this section, we take the advantages of the sparse nature of the mmW channel and formulize channel estimation model as a compressive sensing problem. Authors in [9,16] attempted to obtain training analog vectors from a hierarchical multi-level, but here these vectors are considered as random vectors with fluctuating phase by using one RF chain in training step. As a result, the received signal can be written as

$$\mathbf{v}_k = \mathbf{f}_{r,k}^H \mathbf{H} \mathbf{f}_{t,k} \mathbf{t}_k + \mathbf{f}_{r,k}^H \mathbf{n}_k \quad (8)$$

where \mathbf{v}_k and \mathbf{t}_k are received and transmitted symbol, $\mathbf{f}_{r,k}$ and $\mathbf{f}_{t,k}$ are training analog beamformer at the MS and BS and \mathbf{n}_k is additive received vector noise at the k th time-slot. For representing the sparse characteristics of the channel, we can apply lemma $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ to (8) from [17]. Thus we can rewrite (8) as

$$\mathbf{v}_k = (\mathbf{f}_{t,k}^T \otimes \mathbf{f}_{r,k}^H) \text{vec}(\mathbf{H}) \mathbf{t}_k + \mathbf{f}_{r,k}^H \mathbf{n}_k \quad (9)$$

It can be assumed that BS sends the same symbol \mathbf{t}_k in separate M time slots with distinctive precoder vectors and also MS received in M different combiners. With such an assumption, MS stacks the M measurements in a vector as

$$\mathbf{y} = \mathbf{\Theta} \mathbf{h} \mathbf{t} + \boldsymbol{\zeta} \quad (10)$$

where, $\mathbf{y} = [v_1, \dots, v_M]^T$, $\mathbf{\Theta} = [(\mathbf{f}_{t,1}^T \otimes \mathbf{f}_{r,1}^H), \dots, (\mathbf{f}_{t,M}^T \otimes \mathbf{f}_{r,M}^H)]^T$, $\mathbf{h} = \text{vec}(\mathbf{H})$ and $\boldsymbol{\zeta} = [\mathbf{f}_{r,1}^H \mathbf{n}_1, \dots, \mathbf{f}_{r,M}^H \mathbf{n}_M]^T$.

Using the linear model of virtual channel described in [18], physical channel \mathbf{H} in fixed virtual transmit and receive directions can be modelled as

$$\mathbf{H} = \sum_{m=1}^{N_r} \sum_{n=1}^{N_t} \mathbf{H}_{v,m,n} \mathbf{v}_r(\theta_m) \mathbf{v}_t^H(\varphi_n) = \mathbf{U}_r \mathbf{H}_v \mathbf{U}_t^H \quad (11)$$

$\mathbf{U}_r = [\mathbf{v}_r(\hat{\theta}_1), \mathbf{v}_r(\hat{\theta}_2), \dots, \mathbf{v}_r(\hat{\theta}_{N_r})]$ is an $N_r \times N_r$ array response matrix similar to (3). But, instead of spatial frequencies $(2\pi/\lambda)d \sin(\theta_l)$, $l = 1, \dots, L$, we

substitute the virtual spatial frequencies $2\pi k/N_r$, $k = 1, 2, \dots, N_r$. Similarly, $\mathbf{U}_t = [\mathbf{v}_t(\hat{\varphi}_1), \mathbf{v}_t(\hat{\varphi}_2), \dots, \mathbf{v}_t(\hat{\varphi}_{N_t})]$ is an $N_t \times N_t$ array response matrix with virtual spatial frequencies $2\pi i/N_t$, $i = 1, 2, \dots, N_t$. Thanks to these spatial virtual directions, the matrices \mathbf{U}_r and \mathbf{U}_t are full-rank DFT matrices. Therefore, \mathbf{H}_v is unitarily equivalent to \mathbf{H} such that it captures all of channel information. Matrix of $\mathbf{H}_v \in \mathbb{C}^{N_r \times N_t}$ represents the virtual complex channel matrix and is not generally diagonal.

In practice, however, the virtual channel matrix \mathbf{H}_v is generally non-sparse due to a mismatch between the scattering angles and uniform fixed angles' grid. To force the matrix \mathbf{H}_v to be sparser [19], has proposed, "aperture shaping". For simplicity, in this paper we assume that \mathbf{H}_v is approximately sparse.

By vectorization of the channel matrix in (11), we have

$$\mathbf{h} = \text{vec}(\mathbf{H}) = (\mathbf{U}_t^T \otimes \mathbf{U}_r) \text{vec}(\mathbf{H}_v) \quad (12)$$

$$\mathbf{h} = (\mathbf{U}_t^T \otimes \mathbf{U}_r) \mathbf{h}_v = \mathbf{W} \mathbf{h}_v \quad (13)$$

where $\mathbf{W} \in \mathbb{C}^{N_r N_t \times N_r N_t}$ is defined as a complex dictionary matrix of the channel and $\mathbf{h}_v \in \mathbb{C}^{N_r N_t \times 1}$ represents a sparse vector with L non zero entries as $L \ll N_r N_t$.

Replacing (13) in the stacked measurement vector in (10) and assuming $\mathbf{t} = 1$, we can write,

$$\mathbf{y} = \mathbf{\Theta} \mathbf{W} \mathbf{h}_v + \boldsymbol{\zeta} \quad (14)$$

$$\mathbf{y} = \mathbf{\Psi} \mathbf{h}_v + \boldsymbol{\zeta} \quad (15)$$

where $\mathbf{\Psi} \in \mathbb{C}^{M \times N_r N_t}$ is sensing matrix with the constraint of $M < N_r N_t$. Eq. (15) can be seen as a single measurement vector (SMV) compressive sensing problem due to L -level sparse in \mathbf{h}_v . The sparsity level of a vector is defined as the number of nonzero components of that vector. The support set of \mathbf{h}_v , with notation of χ , is considered as the set of the indices at which it is nonzero. Unknown vector \mathbf{h}_v is recovered as a unique L -sparse solution of a noiseless model of (15) if and only if $2L$ columns of $\mathbf{\Psi}$ are linearly independent [20]. Reconstructing \mathbf{h}_v in the noiseless case can be easily computed as

$$\mathbf{h}_v = \mathbf{\Psi}_{\chi}^{\dagger} \mathbf{y} \quad (16)$$

where $\mathbf{\Psi}_{\chi}^{\dagger} \in \mathbb{C}^{L \times M}$ denotes the Moore-Penrose pseudo inverse of the sub matrix $\mathbf{\Psi}_{\chi}$. Some of the important greedy techniques such as orthogonal Matching Pursuit (OMP) and its derivation, Hard Iterative Thresholding (IHT) and its extensions have been proposed to resolve the SMV sparse problems [21]. When the SNR is very low, which is a usual case at mmW systems, we need to enhance the number of measurements comparable to the dimension of unknown sparse vector. To prevent large stacking of the measurements, exploiting Multiple

Measurement Vector (MMV) is proposed. A common sensing matrix Ψ is utilized to obtain different measurement vectors of multiple realizations of virtual vector channel. Rather than the recovery of the K unknown vectors separately, it is possible to recover all vectors simultaneously by finding the row support of the unknown \mathbf{H}_v from the matrix formulation as

$$\mathbf{Y} = \Psi \mathbf{H}_v + \mathbf{E} \quad (17)$$

where $\mathbf{H}_v = [\mathbf{h}_{v,1}, \dots, \mathbf{h}_{v,K}] \in \mathbb{C}^{N_r \times N_t \times K}$, $\mathbf{E} = [\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_K] \in \mathbb{C}^{M \times K}$ and thus $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K] \in \mathbb{C}^{M \times K}$. When predominant nonzero entries of $\mathbf{h}_{v,k}$ are shared in the same locations, MMV algorithms can lead to computational speed promotion [22]. In this paper, we assume that \mathbf{H}_v has L sparse rows.

The joint sparse non-convex recovery problem, which attempts to regain the unknown matrix with no more than L nonzero rows, is expressed as

$$\min_{\mathbf{H}_v} \|\mathbf{Y} - \Psi \mathbf{H}_v\|_F \text{ subject to } \|\mathbf{H}_v\|_{r,0} \leq L \quad (18)$$

where $\|\cdot\|_{r,0}$ stands for the number of nonzero rows of the matrix. Some of the famous greedy algorithms for solving the full-rank MMV problem are included of SOMP [23], SIHT, and SHTP [13]. Nevertheless, these procedures are incapable of extract the rank information in order to improve the recovery ability in the worst-case ($\text{rank}(\mathbf{H}_v) < L$). In contrast to high computational complexity of rank-blind methods, MUSIC as a rank-aware approach, provides guaranteed recovery in the full row rank cases with the mild complexity. However, once the row rank is not complete, i.e., $\text{rank}(\mathbf{H}_{v,\chi}) < L$, MUSIC does not operate well and needs to propose a modified version of MUSIC. Rank-aware OMP (RA-OMP), a modified version of SOMP, is an algorithm to improve the rank-defective case by inferring the rank information, but not fully rank aware [24]. A modified version of RA-OMP is Rank Aware Order Recursive Matching Pursuit (RA-ORMP) proposed in [25].

4. Subspace enhancement methods

One of the main disadvantages of the MUSIC technique is a drop of its operation under the condition of rank-deficiency or under ill-conditioning. When the number of snapshots is smaller than the sparsity level L , then no more than K rows can be linearly independent, and the nonzero rows of unknown matrix turned into rank defective. Correlation between sources or multi-path propagation is another reason that caused to rank deficiency. For compensating these limitations, one can use a greedy selection algorithm to find $s = L - r$ atoms of the dictionary (or equivalently columns) and then applies MUSIC to an enhanced subspace to recognize the rest of supports. Rank estimation based on observation data matrix is Minimum Description Length (MDL) proposed in [26]. Unlike the mentioned approach in [27], MDL does not require any subjective threshold setting. Motivated by rank-deficiency in channel, we investigate rank-aware algorithms to improve estimation of an unknown rank-defective Millimeter-Wave channel matrix.

4.1. MUSIC as a rank-aware method

The range of the arbitrary matrix of \mathbf{A} is defined as the space spanned by set of all possible linear combination columns of \mathbf{A} and is denoted by $\Omega(\mathbf{A})$. If these columns are linearly independent, we called them *basic columns* and they are a *basis* for $\Omega(\mathbf{A})$. Thus, signal subspace onto (17) is defined by

$$S \triangleq \Omega(\Psi \mathbf{H}_v) = \Omega(\Psi_\chi \mathbf{H}_{v,\chi}) \quad (19)$$

$\Omega(\Psi_\chi \mathbf{H}_{v,\chi})$ agrees with $\Omega(\Psi_\chi)$ when $\mathbf{H}_{v,\chi}$ is a full row rank matrix. In practice, estimated signal subspace, i.e., \hat{S} , is provided from the EVD or SVD on numerous snapshot in matrix $\mathbf{Y} = \Psi \mathbf{H}_v + \mathbf{E}$ while exact subspace obtained by $\mathbf{Y}\mathbf{Y}^H/K$ when $K \rightarrow \infty$. If we assume that $\Psi_\chi \mathbf{H}_{v,\chi}$ is a full row rank matrix, and SNR is high sufficiently, then range of $\Psi_\chi \mathbf{H}_{v,\chi}$ is equivalent to range of \mathbf{Y} . Consequently, one can find the value of elements of support by projecting the columns of Ψ into orthogonal subspace of $\Omega(\Psi_\chi \mathbf{H}_{v,\chi})$ [28].

Remark1. Relying on the MUSIC algorithm, we exploit \mathbf{P}_s as an orthogonal projection matrix onto $\Omega(\Psi_\chi \mathbf{H}_{v,\chi})$ under the conditions of $K \rightarrow \infty$. Thus, for any $k \in \chi$ we have

$$\|\mathbf{P}_s^\perp \Psi_k\| = 0 \quad (20)$$

Based on the above-mentioned explanation and [28], the proof is straightforward and thus omitted.

Additionally, according to (20), $\text{rank}(\mathbf{Q}^H \Psi_\chi) = 0$ where \mathbf{Q} is orthonormal matrix of noise subspace of \mathbf{Y} . However, in practical issue of estimating the rank, due to limitations in SNR or finite number of K , estimation of signal subspace is inaccurate. Thus, Eq. (20) is not satisfied, and we have to minimize $\mathbf{P}_s^\perp \Psi_k$ or equivalently maximize $\mathbf{P}_s \Psi_k$ in MUSIC algorithm. When the rank of $\mathbf{H}_{v,\chi}$, i.e. dimension of signal subspace, is less than the sparsity level then subspace is called proper subspace and MUSIC incapable of completed support recovery. To prohibit the wrong estimating of the support, one can estimate a subspace spanned by $L - r$ columns of Ψ by conventional MMV algorithms like SOMP in a probabilistic way and enhance this subspace to the r -dimensional subspace obtained by MUSIC method deterministically. Because of improper performance of MUSIC method in case of rank-deficiency, it is disregard to explain the algorithm and we refer readers to [25].

4.2. Subspace enhancement MUSIC

Suppose that \mathbf{H}_v has L nonzero rows within support $\chi \subset \{1, \dots, N_t\}$ and also \mathbf{H}_v is rank defective, i.e. $r < L$. Let δ be an arbitrary subset of χ with $L - r$ elements and \hat{S} is estimated subspace of $\Omega(\Psi_\chi \mathbf{H}_{v,\chi})$ obtained by applying the EVD over $\mathbf{Y}\mathbf{Y}^H$. Because of reduction of rank in estimated subspace signal, we are allowed to find an extra subspace to combine it with estimated subspace and compensate incoherence of \hat{S} to $\Omega(\Psi_\chi)$. By approximation from [27],

$$\Omega(\Psi_\chi) \approx \hat{S} + \Omega(\Psi_\delta) \quad (21)$$

Thus, the goal will be to find an enhanced subspace of signal under the condition of linear independency of $L - r$ rows of \mathbf{H}_v [27]. That is, partial support, δ , should be estimated by MMV.

Theory1: Suppose that \bar{S} is an enhanced signal subspace within $\Omega(\Psi_\chi)$. By rewriting (21) as $\bar{S} \approx \hat{S} + \Omega(\Psi_\delta)$, and applying projection operator on both side of it, we have

$$\mathbf{P}_{\bar{S}} = \mathbf{P}_{\Omega(\hat{S} \cup \Psi_\delta)} = \mathbf{P}_{\hat{S}} + (\mathbf{P}_{\hat{S}}^\perp \Psi_\delta) (\mathbf{P}_{\hat{S}}^\perp \Psi_\delta)^\dagger \quad (22)$$

Proof. See Appendix A .

The implementation of (22) seems impractical. We know that the projection onto $\Omega(\hat{S} \cup \Psi_\delta)$ is achieved by applying SVD on matrix $[\mathbf{U}_{\hat{S}}, \Psi_\delta]$ and choosing \mathbf{U}_1 from $[\mathbf{U}_1 | \mathbf{U}_2] \boldsymbol{\Sigma} [\mathbf{V}_1 | \mathbf{V}_2]^H$ [29].

Let δ be an arbitrary subset of χ with cardinality of $L - r$, then $\text{rank}(\hat{\mathbf{Q}}^H \Psi_\delta) = L - r$ where $\hat{\mathbf{Q}}$ is noise subspace resulted by SVD on \mathbf{Y} and consists of orthonormal columns such that $\hat{\mathbf{Q}}^H \mathbf{Y} = \mathbf{0}$. If we suppose the noisy model of $\mathbf{Y} = \Psi \mathbf{H}_v + \mathbf{E}$, then for any $k \in \{1, \dots, N_t\} \setminus \delta$ the term of $\text{rank}(\hat{\mathbf{Q}}^H [\Psi_\delta, \Psi_k]) = L - r$ is satisfied if and only if $k \in \chi$. It

means that Ψ_k is not basic column in augmented matrix of $[\Psi_\delta, \Psi_k]$. Thus, $rank([\Psi_\delta, \Psi_k]) = rank(\Psi_\delta) = L - r$ and consistency is occurred [29]. According to this rule and similar to MUSIC algorithm, it is easy to prove the selection rule to pick up the remained components of support.

Theory2: Suppose that $\delta \subset \chi$ is partial support obtained by one of the MMV algorithms while χ is support set of \mathbf{H}_v . Selection rule for remained entries of support is attained by maximizing $\|\mathbf{P}_{\delta^c}^H \Psi_k\|_2$ where $k \in \{1: N_r N_r\}/\delta$ and $\mathbf{P}_{\delta^c} = \mathbf{P}_{\Omega(\mathbf{U}_\delta, \Psi_\delta)}$.

Proof See Appendix B.

In SOMP as a typical method of MMV and according to [23,30], the selection rule of support is given by $l = \arg \max_{k \in \{1: N_r N_r\}/\delta} \|\mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{Y} \Psi_k\|_2$ for L iteration and $\delta = \{\emptyset\}$ at the first step. The key point of SE-MUSIC for selection rule of δ is the replacement of the stacked data matrix \mathbf{Y} by the estimated orthonormal signal subspace \mathbf{U}_δ . That is, SE-MUSIC algorithm incrementally updates the partially support by the following selection rule,

$$l = \arg \max_{k \in \{1: N_r N_r\}/\delta} \|\mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{U}_\delta \Psi_k\|_2 \quad (23)$$

where δ is $\{\emptyset\}$ at the first iteration. The selection rule in (23) adds a superior member, l , to δ in each step of SOMP. By replacing $orth(\mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{Y})$ instead of $\mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{U}_\delta$ in (23) and exploring all of L support, we can extract RA-OMP from SE-MUSIC. However, rank degeneration problem in a residual matrix of RA-OMP does not happen in SE-MUSIC algorithm when $M = L + 1$.

Motivated by *rank-aware order recursive matching pursuit* (RA-ORMP) algorithm [25,31], we present a generalized Subspace Enhancement Rank Aware MUSIC (GSE-MUSIC) to recover the virtual channel matrix from noisy snapshots. In fact, GSE-MUSIC replaces the snapshot matrix \mathbf{Y} in RA-ORMP by an orthogonal basis matrix for the estimated signal subspace. The strength of GSE-MUSIC rather than the RA-ORMP is rank determination and subspace employment of signal directly in noisy case. However, RA-ORMP can recover support of channel in high SNRs or noiseless cases because the orthonormal basis provided by it includes rank information whereas rank of noisy \mathbf{Y} is full and equal to $\min(M, K)$.

According to (A.2) in Appendix A, $\Omega(\mathbf{P}_{\Psi_\delta}^\perp \mathbf{U}_\delta)$ is enhanced range for signal subspace range. GSE-MUSIC gets an orthogonal basis of estimated signal subspace as an input onto this range and prepares the partially entries of support χ i.e. δ , from the following rule selection.

$$l = \arg \max_{k \in \{1: N_r N_r\}/\delta} \left\| \mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{U}_\delta \Psi_k \right\|_2 / \left\| \mathbf{P}_{\Omega(\Psi_\delta)}^\perp \Psi_k \right\|_2 \quad (24)$$

The key differences between SE-MUSIC and GSE-MUSIC algorithms are orthogonal complementary projecting the dictionary except of partial support atoms, i.e., $\Psi_{k \in \{1, \dots, N_r N_r\}/\delta}$ to the selected atoms, i.e., columns of Ψ_δ , and also renormalize them. Moreover, in SE-MUSIC we only project residual matrix onto the dictionary whereas in GSE-MUSIC, basis of the updated residual matrix $\Xi = \mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{U}_\delta$ is projected.

In Algorithm 1, partial recovery of $L - r$ elements of support χ is established in one of the mentioned approaches. After signal subspace enhancement by partial support, MUSIC provides the rest r entries of support. This algorithm can be considered as a new algorithm that we called Subspace-OMP if we have full recovery of L element and eliminate MUSIC steps. The key note in Subspace-OMP algorithm is sparsity-level blindness.

Algorithm 1. SE and GSE-MUSIC.

1. $\delta = \{\emptyset\}$;
2. exploit \mathbf{U}_δ and r (rank) by eigenvalue decomposition and MDL;
3. $for j = 1:L - r$

4. $\mathbf{P}_{\Omega(\Psi_\delta)} = QR\{\Psi_\delta\}$;
5. $\mathbf{P}_{\Omega(\Psi_\delta)}^\perp \widehat{\Sigma} = QR\{(\mathbf{I} - \mathbf{P}_{\Omega(\Psi_\delta)}) \mathbf{U}_\delta\}$;
6. $\mathbf{P}_{\Omega(\Psi_\delta)}^\perp = \mathbf{I} - \mathbf{P}_{\Omega(\Psi_\delta)}$;
7. if SE-MUSIC: $l = \arg \max_{k \in \{1: N_r N_r\}/\delta} \|\mathbf{P}_{\Omega(\Psi_\delta)}^\perp \mathbf{U}_\delta \Psi_k\|_2$
8. if GSE-MUSIC: $l = \arg \max_{k \in \{1: N_r N_r\}/\delta} \left\| \mathbf{P}_{\Omega(\Psi_\delta)}^\perp \widehat{\Sigma} \Psi_k \right\|_2 / \left\| \mathbf{P}_{\Omega(\Psi_\delta)}^\perp \Psi_k \right\|_2$;
9. $\delta = \delta \cup \{l\}$;
10. *end for*;
11. $\mathbf{U}_{aug} = [\mathbf{U}_\delta, \Psi_\delta]$; find enhancing subspace
12. $\mathbf{U}_{aug} = orth(\mathbf{U}_{aug})$; find orthonormal basis
13. $\gamma = \{1: N_r N_r\}/\delta$
14. *for j = 1: length(γ)*
15. $\kappa_j = \|\mathbf{U}_{aug}^H \Psi_j\|_2$
16. *end for*
17. $\chi = \delta \cup \{r \text{ largest basics of } \kappa\}$;
18. $\widehat{\mathbf{H}}_{v,(\chi)} = \Psi_{(\chi)}^\dagger \mathbf{Y}$

4.3. The case of unknown sparsity level

Until now, we assume to know sparsity level of virtual channel as a priori. Although, it is not a deterrent restriction to process of greedy algorithms if value of L is unavailable. One way for estimation of sparsity level is proposed in [32] based on covariance matrix of observation data. However, this method is suitable in full rank channels because calculation of sparsity level is based on Eigenvalues. As a prior knowledge, in most of the greedy algorithms, the stopping criterion is based on the L iterations. Provided that L is unknown, for an intuitive termination criterion, we use angle function idea from [33], a metric for two subspace even with the different dimensions. We have

$$\|\mathbf{P}_{\Omega(\Psi_\beta)}^\perp \mathbf{U}_\delta\| < \lambda \quad (25)$$

where λ is a predefined threshold. In the first iteration, Ψ_β is selected from r columns of the dictionary such that maximum correlation with the signal subspace of \mathbf{U}_δ is occurred. Constraint of (25) is replaced to line 3 from Algorithm 1. If (25) is satisfied, the Algorithm 1 pursues from line 4–17. Afterward, Ψ_β is composed of $r+1$ atoms (columns) such that *one of them* has chosen from support set and the other r atoms are provided by applying the projection of enhanced subspace onto Ψ_β . Then, algorithm goes back to line 3 to check the condition (25) again. The procedure is reiterated until (25) fails.

After estimation of virtual channel and extracting the estimated mmW channel, thanks to the assumption of the Gaussian signalling over the link in (2), we are able to achieve the estimated spectral efficiency as follows,

$$\widehat{R} = \log_2 \left(\mathbf{I}_{N_s} + \frac{P_s}{N_s} \widehat{\mathbf{R}}_n^{-1} \widehat{\mathbf{H}}_{eff} \widehat{\mathbf{H}}_{eff}^H \right) \quad (26)$$

where $\widehat{\mathbf{H}}_{eff} = \widehat{\mathbf{C}}^H \widehat{\mathbf{H}} \widehat{\mathbf{P}}$ is effective estimated channel with $\widehat{\mathbf{C}} = \widehat{\mathbf{F}}_r \widehat{\mathbf{G}}_r$ and $\widehat{\mathbf{P}} = \widehat{\mathbf{F}}_t \widehat{\mathbf{G}}_t$ as estimated combiner and precoder matrices respectively. Furthermore, $\widehat{\mathbf{R}}_n = \sigma_n^2 \widehat{\mathbf{C}} \widehat{\mathbf{C}}^H$ is the covariance matrix of filtered noise matrix. Considering the estimated virtual channel provided by one of the proposed algorithms and using (11), the estimated mmW channel model can be written

$$\widehat{\mathbf{H}} = \mathbf{U}_r \widehat{\mathbf{H}}_{v,(\chi)} \mathbf{U}_t^H \quad (27)$$

By applying the singular value decomposition on $\widehat{\mathbf{H}}$ in (27) and choosing the primary N_s columns of the left-handed unitary matrix, i.e., $\widehat{\mathbf{U}}$, and choosing the first N_s columns of the right-handed unitary

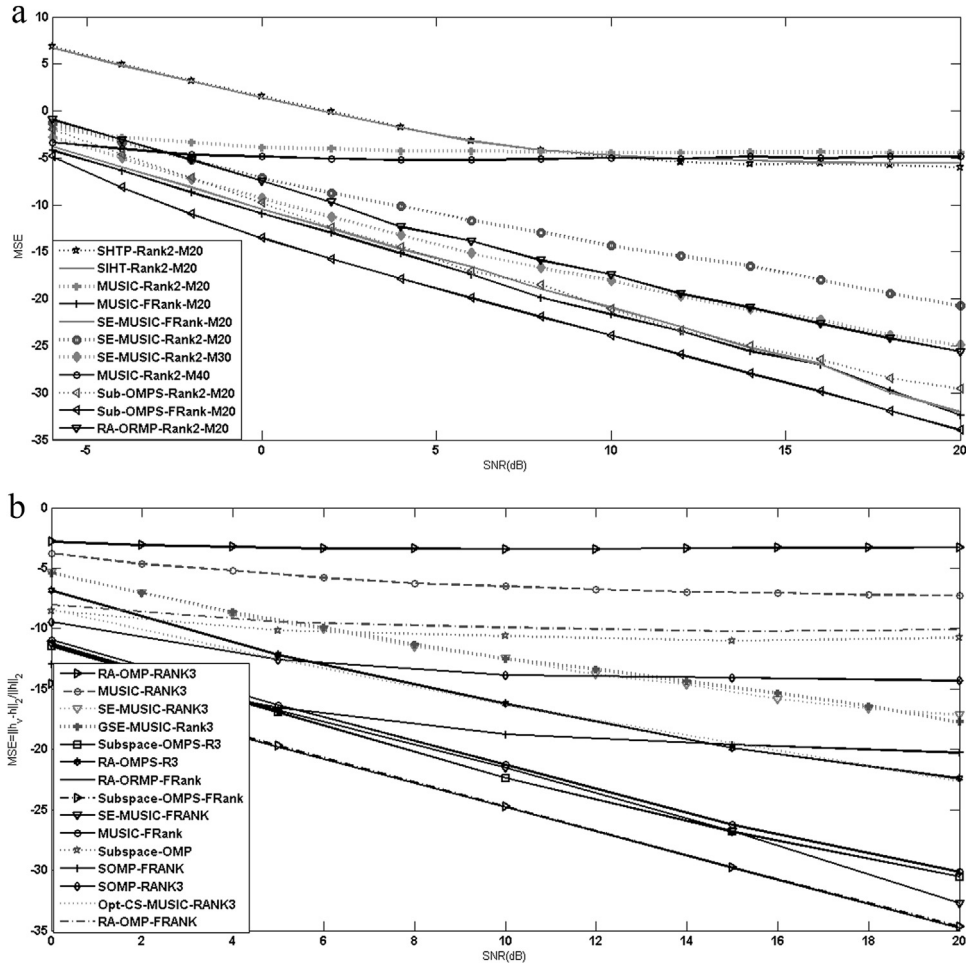


Fig. 2. (a). Comparing subspace-based algorithms in sparsity level 4. (b). MSE of different algorithms when sparsity level is 8 and different rank.

matrix, i.e., $\hat{\mathbf{V}}$, and finally replacing them rather than $\hat{\mathbf{C}}$ and $\hat{\mathbf{P}}$ respectively, we can design $\hat{\mathbf{F}}_r, \hat{\mathbf{G}}_r, \hat{\mathbf{F}}_t$ and $\hat{\mathbf{G}}_t$ by solving the general sparse optimization problem as follows,

$$(\hat{\mathbf{F}}_x, \hat{\mathbf{G}}_x) = \arg \min \|\hat{\mathbf{W}}_x - \hat{\mathbf{F}}_x \hat{\mathbf{G}}_x\|_F,$$

$$s.t. \quad [\hat{\mathbf{F}}_{x,l,l} \text{ and } [\hat{\mathbf{G}}_{x,l,l} \in \{[C_x]_{i,l} | 1 \leq i \leq N_{x,\text{cod}}, l = 1, 2, \dots, N_{x,\text{rf}} \} \quad (28)$$

$$\|\hat{\mathbf{F}}_x \hat{\mathbf{G}}_x\|_F^2 = N_s$$

where the subscript x can be substituted for *transmitter/receiver*, $\hat{\mathbf{W}}_x$ can be replaced by $\hat{\mathbf{U}}$ or $\hat{\mathbf{V}}$ dependent on x and C_x is a general codebook included of quantized steering vectors for transceiver and is chosen from

$$\left\{ \begin{array}{l} \mathbf{v} \left(\frac{k\pi}{N_Q} \right), k = 0, 1, \dots, N_{x,\text{cod}} - 1, N_Q = 2^Q \\ \mathbf{v}_x(\theta) = [1, e^{j\pi \cos(\theta)}, \dots, e^{j\pi(N_x-1)\cos(\theta)}]^T / \sqrt{N_x} \end{array} \right\} \quad (29)$$

Q in (29) indicates the number of bits for controlling the phase shifters and $N_{x,\text{cod}}$ implies the number of steering vectors existing in transmitter/receiver codebook. Optimization problem in (28) is solvable by one of the iterative sparse algorithms.

5. Simulation results

In this section, we evaluate numerical results of proposed algorithms and compare their performance to conventional MMV and SMV problems. Under adoption of hybrid analog/digital architecture illustrated in Fig. 1, RF phase shifters in analog parts of precoder and combiner are able to be controlled with 8 quantization bits. Similar to [9], the operational carrier frequency of the system is 28 GHz with consideration of bandwidth of 100 MHz. The path-loss exponent is assumed to be $\beta_{\text{loss}} = 3.5$ and the angles of arrival and departure are selected randomly with a uniform distribution from range of $[0, 2\pi]$.

In Fig. 2(a), we consider $N_t = 32$, $N_r = 8$, $L=4$ and $K=4$ but with different values of M . Proposed subspace OMPS is compared to various rank-aware algorithms such as SE-MUSIC, MUSIC, RA-ORMP and rank-blind conventional SHTP, SIHT with full and deficient rank. The figure indicates that the performance of virtual channel estimation is totally improved when M is increasing. The performance of proposed SE-MUSIC with deployments of $M=20$ and decreasing rank to 2 outperforms than the rank-defective MUSIC with $M=40$, conventional rank-blindness SIHT, SHTP. The SE-MUSIC algorithm generally has similar behaviour to MUSIC in full row rank case. The key note in this figure is better performance of proposed Subspace-OMPS in rank-

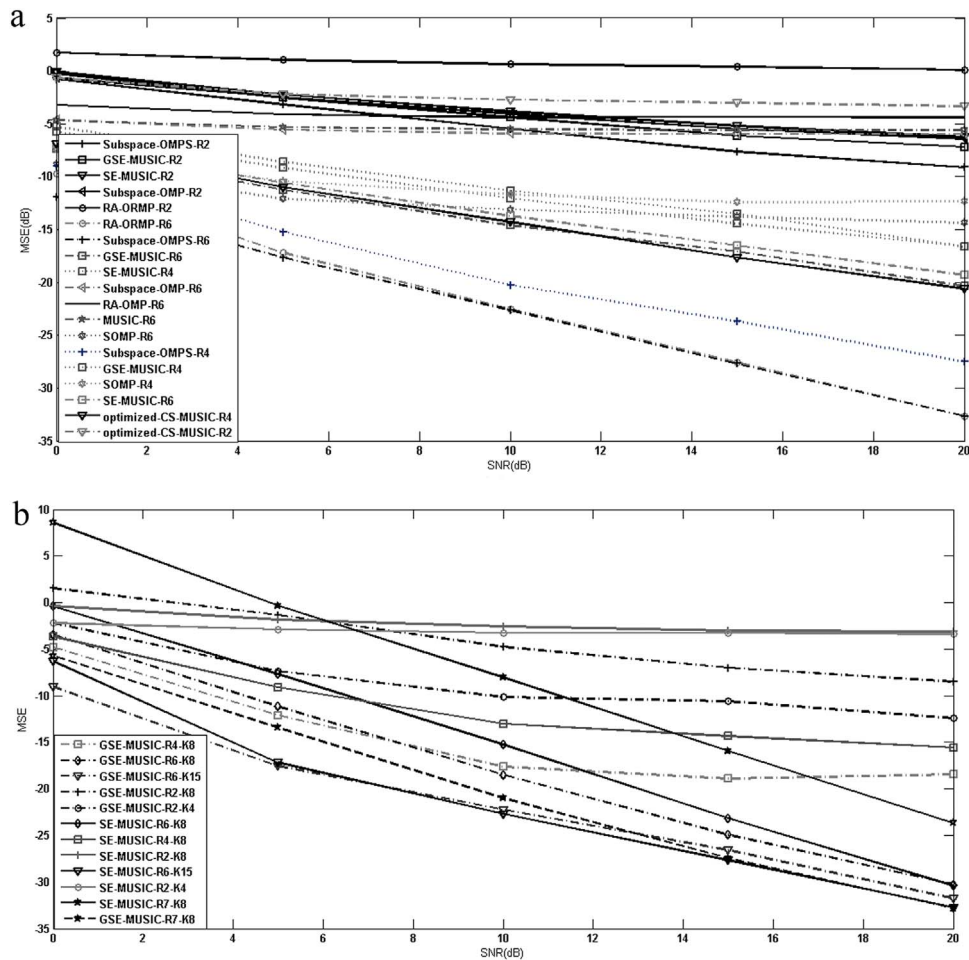


Fig. 3. (a). Impact of growth of rank in rank-aware algorithm when sparsity level is 8. (b). Unknown channel estimation when sparsity level is unknown too.

defective and even in full-rank case rather than the others.

In Fig. 2(b), the sparsity level is 6 with $M = 20$ and $K = 6$. The number of BS and MS antenna are 32 and 8 respectively. First, full row rank and then rank-defective scenarios for rank-blind and rank-aware algorithms are experienced. As shown in the figure, the performance of the MSE in SE-MUSIC and GSE-MUSIC is better than the MUSIC, SOMP and SHTP when SNR is increased. In other words, algorithms based on OMPS approach along with exploiting subspace guarantee recovery of the channel for the full row rank or even rank-defective cases in spacious SNRs.

RA-ORMP and Subspace-OMP in full-rank case are similar but better than MUSIC based algorithms while in the rank-defective case, proposed Subspace-MUSIC outperforms than the others, especially in middle to high SNRs.

In Fig. 3(a), the sparsity level is changed to be 8 with $M = 20$, $K = 8$ and the same antenna size. The goal of this experiment is finding the performance of proposed algorithms on MSE while the rank of channel is decreasing. In additional to the evaluated algorithms, here we compare results to optimized CS-MUSIC proposed in [34]. The performance shows that Optimized-CS-MUSIC behaves similar to SE and GSE-MUSIC.

In Fig. 3(b), the case when the sparsity level is unknown is illustrated for MUSIC algorithms based on subspace enhancement

such as SE-MUSIC and GSE-MUSIC. For this figure, the same system setup of Fig. 3(a) is adopted again, and the MSE achieved by constant size of training vector.

The result shows that when the channel is rank-defective with unknown number of multipath, the MSE of GSE-MUSIC outperforms than SE-MUSIC in a different range of SNRs. Increasing value of K improves MSE in the vast range of SNRs especially in low SNRs when rank is full or even incomplete. Simulation results show that the value of threshold in (25) depends on the rank of the unknown channel matrix. As an empirical rule, we choose the value of $(r/10)$ as a proper threshold.

In Fig. 4(a), the spectral efficiency is represented by the proposed algorithms when desired number of paths, i.e., L equals 8. Algorithms include MUSIC, SE-MUSIC, GSE-MUSIC, RA-ORMP and Subspace-OMP for the rank-defective case of channel are tested for different values of BS and MS antenna number with constant RF chains such that $N_{rf}^t = 10$ and $N_{rf}^r = 6$, and compared with the spectral efficiency of the perfect channel. The values of the M and K are 20 and 8 respectively and channel estimation is performed while $SNR = 10dB$.

The results indicate that spectral efficiency can be achieved using the proposed algorithms based on subspace enhancement despite their low-complexity.

In Fig. 4(b) experiment is repeated for unknown channel state to

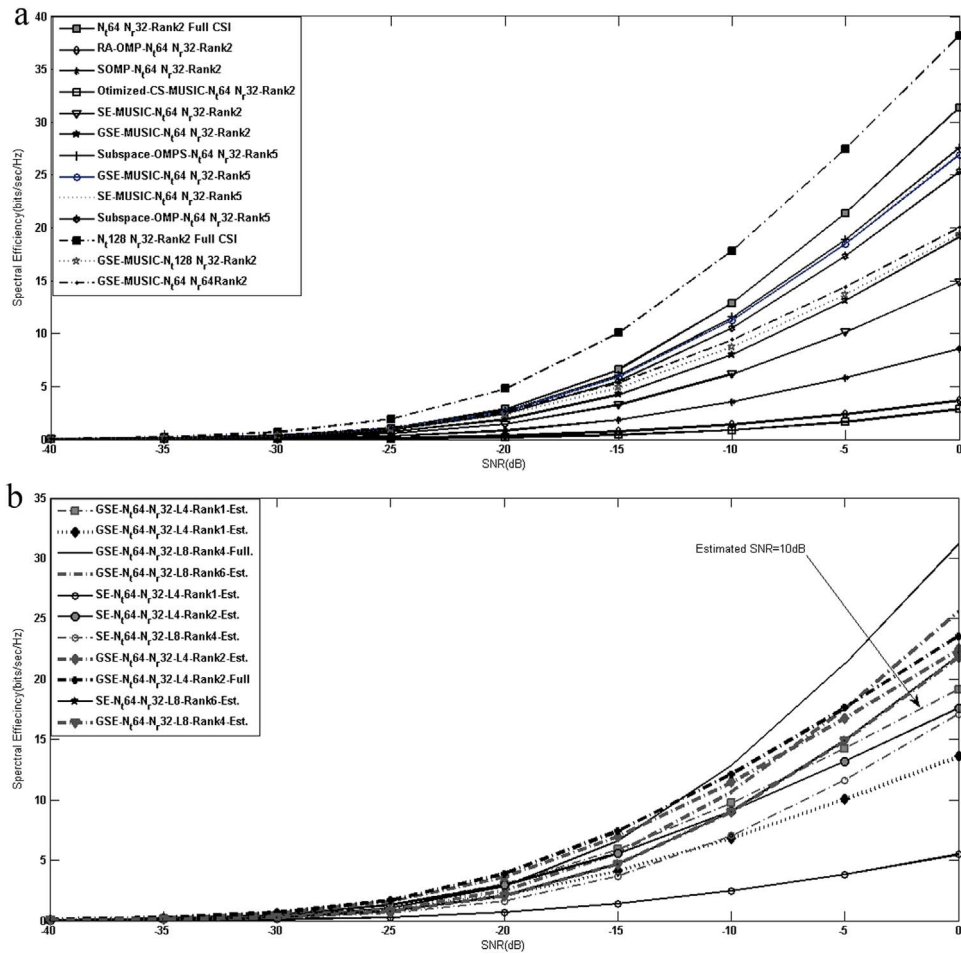


Fig. 4. (a). Spectral efficiency in different rank and size of antenna when sparsity level is 8. (b). Impact of sparsity level and rank on spectral behaviour.

sparsity level of 4 and 8 in SE and GSE-MUSIC while $SNR = 0dB$. Results represent prominence of GSE-MUSIC in different arrangements.

In Fig. 5(a), the impact of RF chain limitations and number of quantized bit of phase shifters is evaluated. Two system models with various suggested algorithms and different quantization bits for the phase shifters are considered, one with 10 RF chains at the BS, 6 RF chains at the MS, and the other with 5 RF chains at the BS, and 3 RF chains at the MS. The other parameters are the same as the last simulations with $L=5$, Rank=3, $N_t = 64, N_r = 32$. Simulation results in $SNR = 20dB$ show that the offered subspace based algorithms can achieve near optimal rates while the sufficient number of RF chains and quantization bits exist.

In Fig. 5(b)-6(a) performance of success rate is attained in terms of SNR and M when various algorithms under rank defective condition and different K are experienced. Number of transmit antenna, receiver antenna and sparsity level set to 32, 8 and 8 respectively. In Fig. 5(b) when $K=8$ and rank changes between 2 and 6, algorithm Subspace-OMP is prior than the others. When rank is increased, success rate improves such that around of $SNR=0$ dB support is recovered com-

pletely. In Fig. 6(a) algorithm RA-ORMP in small M, low rank and middle SNR conditions is unsuccessful while Subspace-OMP and GSE-MUSIC are prosperous. Support recovery in Subspace-OMP when rank is 6 and $M=15$ is completed.

In Fig. 6(b) parameters, SNR, K and M for SE and GSE-MUSIC algorithms are variant while rank is fixed to 5. As shown in this figure, GSE outperforms SE-MUSIC. Additionally, increasing of the value K and SNR each of them solitary makes growth in success rate.

Comparing average computation's time of each iteration in various approaches under same conditions of $N_t = 16N_r = 8$, $L=8$, $K=8$ and Rank=4 is summarized in Table 1. As shown in Table 1, Subspace-MUSIC is faster than the others, but it is blind when sparsity level is unknown. GSE-MUSIC is the next fast algorithm.

6. Conclusion and future works

In this paper, we explored the potential of MUSIC-based and Rank-aware algorithms in rank-defective or ill-conditioned mmW channel estimation while these approaches exploit the sparse nature of the channel with small training overhead. The hybrid architecture is

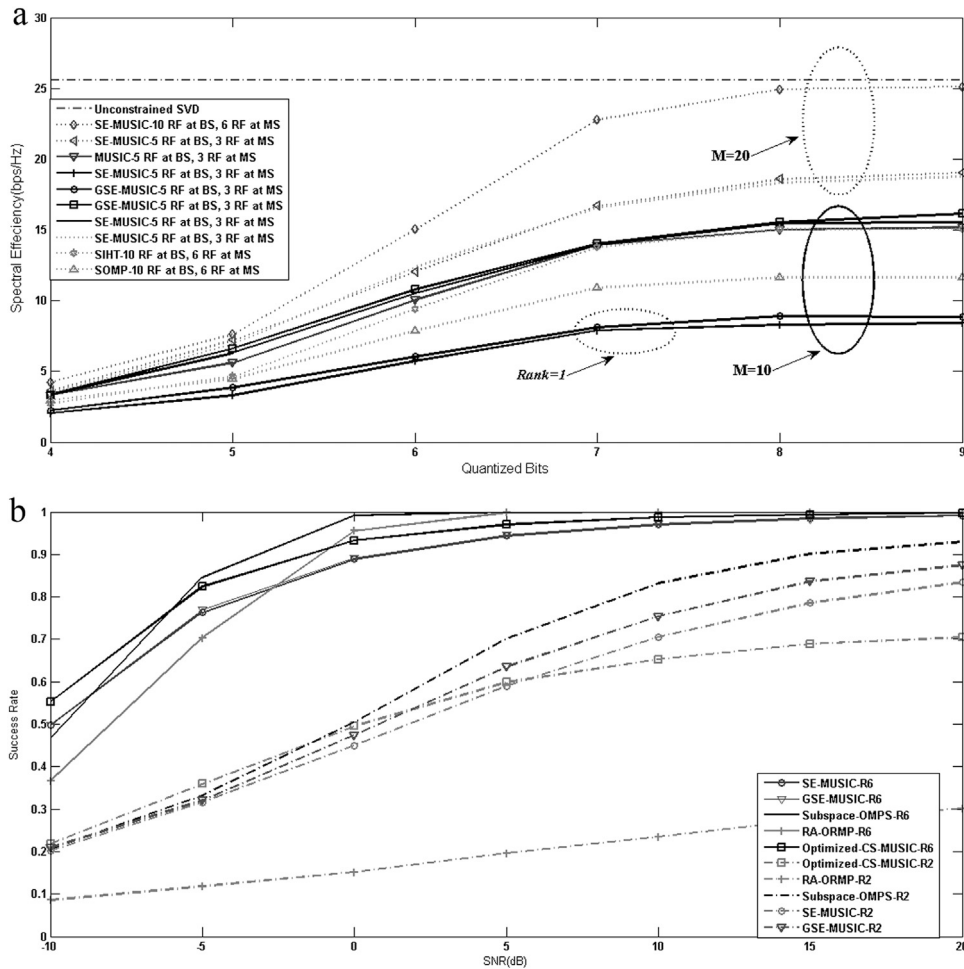


Fig. 5. (a). Impact of number of RF chains and quantized bits on spectral when $N_t=16$, $N_r=4$ and sparsity level 8. (b). Impact of rank on success rate in different SNR when sparsity level and M are 8 and 20 respectively.

composed of analog phase-shifters and digital base-band processor in the transceiver along with the large antenna array and RF chains very smaller than the length of the array, achieving near optimal spectral efficiencies even in rank-imperfect outdoor channels. We first enumerated the conventional MMV algorithms as extended SMV methods for full-rank channel, and then developed subspace enhancement approaches for channel with imperfect rank. Numerical results showed that the proposed rank-aware OMP offers near-optimal solution and achieves better spectral efficiency similar to the fully digital counterparts. We also provided a channel estimation method that can succeed

in the unknown multipath (sparsity) and noisy measurement conditions.

For future work, one can extend rank-defective mmW channel estimation based on Bayesian enhancement approaches, for example mentioned in [35]. It would also be interesting to extract the hybrid precoding/combining of rank-defective multi-user mmW according to some studies such as [36]. Furthermore, it would be arousing to consider the interference-cancellation problem in Multi-User frequency-selective mmW networks.

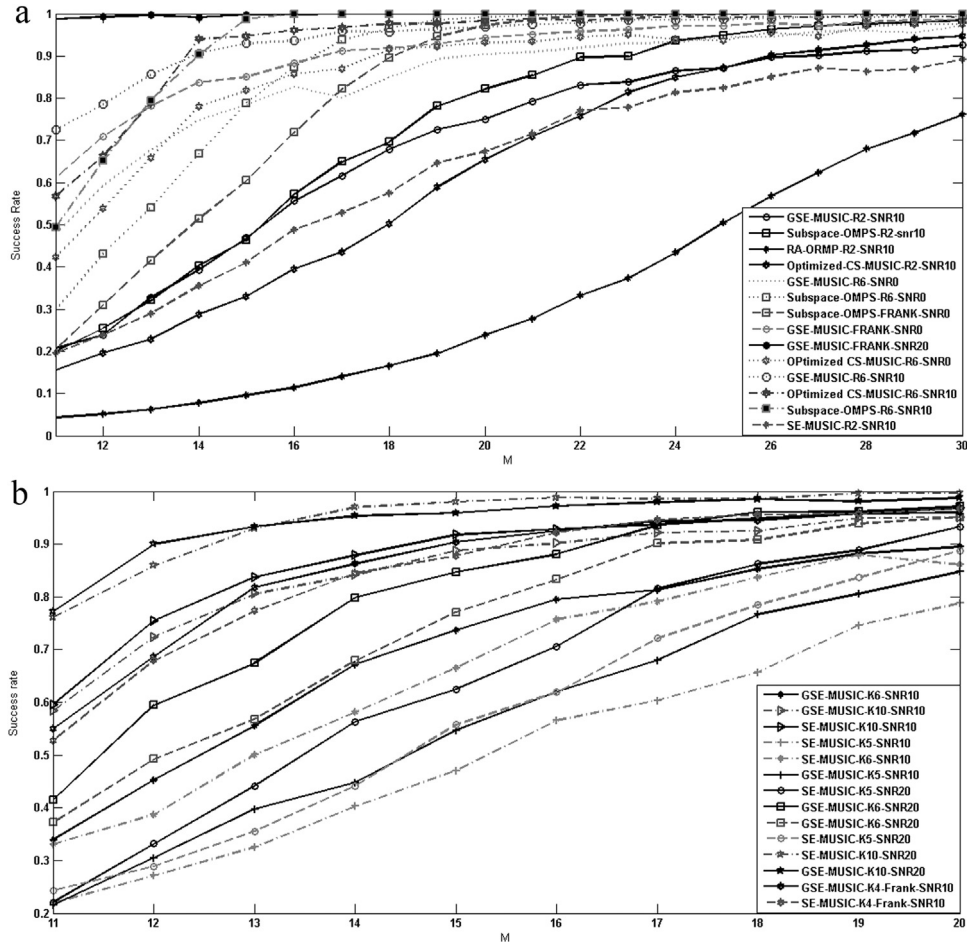


Fig. 6. (a). Impact of rank on success rate in terms of M when sparsity level and K are both 8. (b). Impact of K on success rate when SNR is 0 dB in terms of M when sparsity level and rank are 8 and 5 respectively.

Table 1

Average Elapsed Time for one iteration of different proposed algorithms.

Algorithm	Subspace-OMPS	GSE-MUSIC	SE-MUSIC	Optimized CS-MUSIC	SOMP
Elapsed Time(sec)	0.04888	0.0537	0.0664	0.0945	0.0736

Appendix A. Proof of Theorem 1

By assuming $\mathbf{U}_{\bar{\gamma}} = \text{orth}[\mathbf{U}_{\bar{\gamma}}, \mathbf{\Psi}_{\bar{\delta}}]$ as an enhanced basis matrix with L columns within $\Omega(\mathbf{\Psi}_{\bar{\gamma}})$ such that $\mathbf{U}_{\bar{\gamma}}$ as is a $r \times r$ estimated signal subspace within $\Omega(\mathbf{\Psi}_{\bar{\gamma}}) = \Omega(\mathbf{\Psi}_{\bar{\gamma}} \mathbf{H}_{v,(\gamma)})$ we have

$$\Omega(\mathbf{U}_{\bar{\gamma}}) = \Omega[\mathbf{U}_{\bar{\gamma}}, \mathbf{\Psi}_{\bar{\delta}}] \quad (\text{A.1})$$

On the other hand, $\Omega(\mathbf{U}_{\bar{\gamma}})^{\perp} = \Omega(\mathbf{Q}_{\bar{\gamma}})$ where $\mathbf{Q}_{\bar{\gamma}}$ is an estimation of noise subspace of $\mathbf{Y}\mathbf{Y}^H/\mathbf{K}$. By applying the projection update rule [37] on (A.1) we have

$$\Omega(\mathbf{U}_{\bar{\gamma}}) = \Omega(\mathbf{U}_{\bar{\gamma}}) + \Omega(\mathbf{P}_{\mathbf{U}_{\bar{\gamma}}}^{\perp} \mathbf{\Psi}_{\bar{\delta}}) = \Omega(\mathbf{\Psi}_{\bar{\delta}}) + \Omega(\mathbf{P}_{\mathbf{\Psi}_{\bar{\delta}}}^{\perp} \mathbf{U}_{\bar{\gamma}}) \quad (\text{A.2})$$

Since $\mathbf{P}_{\mathbf{Q}_{\bar{\gamma}}}$ is equivalent to $\mathbf{I} - \mathbf{P}_{\mathbf{U}_{\bar{\gamma}}} = \mathbf{P}_{\mathbf{U}_{\bar{\gamma}}}^{\perp}$ then we have,

$$\Omega(\mathbf{U}_{\bar{\gamma}}) = \Omega(\mathbf{U}_{\bar{\gamma}}) + \Omega(\mathbf{P}_{\mathbf{Q}_{\bar{\gamma}}} \mathbf{\Psi}_{\bar{\delta}}) \quad (\text{A.3})$$

By applying the projection operator in both of equation (A.2) and considering $\bar{\gamma}$ and $\hat{\delta}$ instead of $\mathbf{U}_{\bar{\gamma}}$ and $\mathbf{U}_{\bar{\gamma}}$ respectively and knowing that $\mathbf{P}_{\mathbf{G}} = \mathbf{G}\mathbf{G}^{\dagger}$ for an arbitrary matrix of \mathbf{G} [38], we have

$$\mathbf{P}_{\bar{\gamma}} = \mathbf{P}_{\Omega[\mathbf{U}_{\bar{\gamma}} \mathbf{\Psi}_{\bar{\delta}}]} = \mathbf{P}_{\Omega(\mathbf{U}_{\bar{\gamma}})} + \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\mathbf{U}_{\bar{\gamma}})}^{\perp} \mathbf{\Psi}_{\bar{\delta}})} \quad (\text{A.4})$$

$$= \mathbf{P}_{\bar{\gamma}} + (\mathbf{P}_{\bar{\gamma}}^{\perp} \mathbf{\Psi}_{\bar{\delta}}) (\mathbf{P}_{\bar{\gamma}}^{\perp} \mathbf{\Psi}_{\bar{\delta}})^{\dagger}$$

In practical issues, one can calculate $\mathbf{P}_{\bar{\gamma}}$ by considering of the unitary matrix part of QR decomposition on $[\mathbf{U}_{\bar{\gamma}} \mathbf{\Psi}_{\bar{\delta}}]$ since the columns of orthogonal projector matrix can be obtained from any set of orthonormal vectors onto $\Omega[\mathbf{U}_{\bar{\gamma}}, \mathbf{\Psi}_{\bar{\delta}}]$.

Appendix B. Proof of Theory 2

If $\Psi_{:,k}$ is combination of columns from matrix Ψ_δ then,

$$\text{rank}(\mathbf{Q}^H[\Psi_\delta, \Psi_{:,k}]) = L - r \quad (\text{B.1})$$

Lemma B.1. Let \mathbf{A} be any $m \times n$ matrix and r be any positive integer such that $1 \leq r \leq n$. \mathbf{A} has rank of r if and only if there is a $r \times r$ sub matrix of \mathbf{A} with nonzero determinant, while every $s \times s$ sub matrix of \mathbf{A} has zero determinant for $s \geq r$.

If we suppose that $\mathbf{A} = \mathbf{Q}^H[\Psi_\delta, \Psi_{:,k}]$ with a dimension of $(M - r) \times (L - r + 1)$ then $\mathbf{A}^H \mathbf{A}$ is a $(L - r + 1) \times (L - r + 1)$ matrix. Since $\text{rank}(\mathbf{A}) = L - r$, then sub matrix $\mathbf{A}^H \mathbf{A}$ has zero determinant according to the Lemma B.1. So we have

$$|\mathbf{Q}^H[\Psi_\delta, \Psi_{:,k}]^H \mathbf{Q}^H[\Psi_\delta, \Psi_{:,k}]| = 0 \quad (\text{B.2})$$

where $|\cdot|$ indicates the determinant of a matrix. We can rewrite (B.2) as

$$\begin{aligned} & \left| \begin{bmatrix} \Psi_\delta^H \\ \Psi_{:,k}^H \end{bmatrix} \mathbf{Q} \mathbf{Q}^H [\Psi_\delta, \Psi_{:,k}] \right| = \\ & \left| \begin{bmatrix} \Psi_\delta^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta & \Psi_\delta^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_{:,k} \\ \Psi_{:,k}^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta & \Psi_{:,k}^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_{:,k} \end{bmatrix} \right| = 0 \end{aligned} \quad (\text{B.3})$$

By using lemma $\left| \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \right| = |\mathbf{A}| |\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B}|$ from [29] and knowing that $|\Psi_\delta^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta| > 0$ because of $\text{rank}(\mathbf{Q}^H \Psi_\delta) = L - r$ and a little manipulation in (B.3) we have,

$$\Psi_{:,k}^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_{:,k} - \Psi_{:,k}^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta (\Psi_\delta^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta)^{-1} \Psi_\delta^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_{:,k} = 0 \quad (\text{B.4})$$

Using lemma $\mathbf{A}^\dagger = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ and knowing that $\mathbf{P}_{\Omega(\mathbf{Q})} \mathbf{P}_{\Omega(\mathbf{Q})}^H = \mathbf{P}_{\Omega(\mathbf{Q})}$, (B.4) is equivalent to

$$\Psi_{:,k}^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_{:,k} - \Psi_{:,k}^H \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta (\mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta)^{\dagger} \Psi_{:,k}^H [\mathbf{P}_{\Omega(\mathbf{Q})} - \mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta (\mathbf{P}_{\Omega(\mathbf{Q})} \Psi_\delta)^{\dagger}] \Psi_{:,k} = 0 \quad (\text{B.5})$$

But in noisy measurement condition, $\mathbf{Y} = \Psi \mathbf{H}_v + \mathbf{N}$, (B.5) is not satisfied and for any $j \notin \chi$ and $k \in \chi$ we have,

$$\Psi_k^H [\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)}] \Psi_k < \Psi_j^H [\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)}] \Psi_j \quad (\text{B.6})$$

where $\hat{\mathbf{Q}}$ is estimated noise subspace from the noisy model of \mathbf{Y} . (B.6) means that we should find columns such that minimize (B.5). Therefore we have,

$$\min_{k \in \{1, \dots, N_t N_r\} / \delta} \Psi_{:,k}^H [\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)}] \Psi_{:,k} \quad (\text{B.7})$$

On the other hand, we know that $\Omega(\mathbf{U}_\delta) = \Omega(\hat{\mathbf{Q}})^\perp$, then $\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} = \mathbf{P}_{\Omega(\mathbf{U}_\delta)}^\perp$.

Lemma B.2. The term of $\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)}$ in (B.7) is equal to the orthogonal complement projection onto the $\Omega([\mathbf{U}_\delta, \Psi_\delta])$.

Proof. We know that.

$$\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)} = \hat{\mathbf{Q}} \hat{\mathbf{Q}}^H - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)} = \mathbf{P}_{\mathbf{U}_\delta}^\perp - \mathbf{P}_{\Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)} \quad (\text{B.8})$$

It means that (B.8) is orthogonal complement projection onto $\Omega(\hat{\mathbf{Q}}) \cap \Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)^\perp$. In the other side we have,

$$\Omega([\mathbf{U}_\delta, \Psi_\delta]) = \Omega(\mathbf{U}_\delta) + \Omega(\Psi_\delta) = \Omega(\mathbf{U}_\delta) + \Omega(\mathbf{P}_{\hat{\mathbf{Q}}} \Psi_\delta) \quad (\text{B.9})$$

Noise subspace for $\Omega([\mathbf{U}_\delta, \Psi_\delta])$ is equivalent to $\Omega([\mathbf{U}_\delta, \Psi_\delta])^\perp$. By the projection update rule into the noise subspace we have,

$$\Omega([\mathbf{U}_\delta, \Psi_\delta])^\perp = (\Omega(\mathbf{U}_\delta) + \Omega(\mathbf{P}_{\hat{\mathbf{Q}}} \Psi_\delta))^\perp = \Omega(\mathbf{U}_\delta)^\perp \cap \Omega(\mathbf{P}_{\hat{\mathbf{Q}}} \Psi_\delta)^\perp \quad (\text{B.10})$$

where (B.10) is equivalent to $\Omega(\hat{\mathbf{Q}}) \cap \Omega(\mathbf{P}_{\Omega(\hat{\mathbf{Q}})} \Psi_\delta)^\perp$ because of $\Omega(\mathbf{U}_\delta)^\perp = \Omega(\hat{\mathbf{Q}})$.

Subsequently, we can conclude that (B.7) is equal to

$$\max_{k \in \{1, \dots, N_t N_r\} / \delta} \Psi_{:,k}^H \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])} \Psi_{:,k} \quad (\text{B.11})$$

By applying property of a projection matrix $\mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])} = \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])}^H \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])} = \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])}^H \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])}$ we can rewrite (B.11) as

$$\max_{k \in \{1, \dots, N_t N_r\} / \delta} \Psi_{:,k}^H \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])} \mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])} \Psi_{:,k} = \max_{k \in \{1, \dots, N_t N_r\} / \delta} \|\mathbf{P}_{\Omega([\mathbf{U}_\delta, \Psi_\delta])} \Psi_{:,k}\|_2 \quad (\text{B.12})$$

References

- [1] T.S. Rappaport, R.W. Heath Jr, R.C. Daniels, J.N. Murdock, Millimeter Wave Wireless Communications, Pearson Education, 2014.
- [2] P. Wang, Y. Li, L. Song, B. Vucetic, Multi-gigabit millimeter wave wireless communications for 5G: from fixed access to cellular networks, Commun. Mag., IEEE 53 (2015) 168–178.
- [3] A.L. Swindlehurst, E. Ayanoglu, P. Heydari, F. Capolino, Millimeter-wave massive MIMO: the next wireless revolution?, IEEE Commun. Mag. 52 (2014) 56–62.
- [4] M.R. Akdeniz, L. Yuanpeng, M.K. Samimi, S. Shu, S. Rangan, T.S. Rappaport, et al., Millimeter wave channel modeling and cellular capacity evaluation, Sel. Areas Commun., IEEE J. 32 (2014) 1164–1179.
- [5] T.S. Rappaport, S. Shu, R. Mayzus, Z. Hang, Y. Azar, K. Wang, et al., Millimeter

- wave mobile communications for 5G cellular: it will work!, *Access*, IEEE 1 (2013) 335–349.
- [6] T.S. Rappaport, F. Gutierrez, E. Ben-Dor, J.N. Murdock, Y. Qiao, J. Tamir, Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications, *Antennas Propag., IEEE Trans.* 61 (2013) 1850–1859.
- [7] Y.C. Eldar, G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.
- [8] O. El Ayach, S. Rajagopal, S. Abu-Surra, P. Zhouyue, R.W. Heath, Spatially sparse precoding in millimeter wave MIMO systems, *Wirel. Commun., IEEE Trans.* 13 (2014) 1499–1513.
- [9] A. Alkhateeb, O. El Ayach, G. Leus, R.W. Heath, Channel estimation and hybrid precoding for millimeter wave cellular systems, *Sel. Top. Signal Process., IEEE J.* 8 (2014) 831–846.
- [10] H.Ghauch, M.Bengtsson, T.Kim, M.Skoglund, Subspace estimation and decomposition for hybrid analog-digital millimetre-wave MIMO systems, 2014.
- [11] P.Schniter, A.Sayeed, Channel estimation and precoder design for millimeter-wave communications: The sparse way, in: *Signals, Systems and Computers the 48th ASILOMAR Conference onpp.* 273–277, 2014.
- [12] W.U. Bajwa, J. Haupt, A.M. Sayeed, R. Nowak, Compressed channel sensing: a new approach to estimating sparse multipath channels, *Proc. IEEE* 98 (2010) 1058–1076.
- [13] J.D. Blanchard, M. Cermak, D. Hanle, J. Yirong, Greedy algorithms for joint sparse recovery, *Signal Process., IEEE Trans.* 62 (2014) 1694–1704.
- [14] J.M. Kim, O.K. Lee, J.C. Ye, Improving noise robustness in subspace-based joint sparse recovery, *IEEE Trans. Signal Process.* 60 (2012) 5799–5809.
- [15] R.C. Hansen, *Phased Array Antennas* 213, John Wiley & Sons, 2009.
- [16] H. Sooyoung, K. Taejoon, D.J. Love, J.V. Krogmeier, T.A. Thomas, A. Ghosh, Millimeter wave beamforming for wireless backhaul and access in small cell networks, *Commun., IEEE Trans.* 61 (2013) 4391–4403.
- [17] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 2012.
- [18] A.M. Sayeed, Deconstructing multiantenna fading channels, *Signal Process., IEEE Trans.* 50 (2002) 2563–2579.
- [19] P.Schniter, A.Sayeed, "A sparseness-preserving virtual MIMO channel model, in: *Proceedings 38th Annu. Conference Inf. Sci. Syst. (CISS'04)*pp. 36–41, 2004.
- [20] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci.* 100 (2003) 2197–2202.
- [21] J.D.Blanchard, J.Tanner, K.Wei, "CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion," *completion*, 2014.
- [22] S.Foucort, "Recovering jointly sparse vectors via hard thresholding pursuit, *Proceedings Sampling Theory and Applications (SampTA)*], (May 2–6 2011), 2011.
- [23] J.A. Tropp, Algorithms for simultaneous sparse approximation. Part II: Convex relaxation, *Signal Process.* 86 (2006) 589–602.
- [24] J.D. Blanchard, M.E. Davies, Recovery guarantees for rank aware pursuits, *Signal Process. Lett., IEEE* 19 (2012) 427–430.
- [25] M.E. Davies, Y.C. Eldar, Rank awareness in joint sparse recovery, *Inf. Theory, IEEE Trans.* on 58 (2012) 1135–1146.
- [26] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, *Acoust., Speech Signal Process., IEEE Trans.* on 33 (1985) 387–392.
- [27] K. Lee, Y. Bresler, M. Junge, Subspace methods for joint sparse recovery, *Inf. Theory, IEEE Trans.* on 58 (2012) 3613–3641.
- [28] K. Jong Min, L. Ok Kyun, Y. Jong Chul, "compressive music: revisiting the link between compressive sensing and array signal processing, *Inf. Theory, IEEE Trans.* 58 (2012) 278–301.
- [29] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.
- [30] M.A. Davenport, M.B. Wakin, Analysis of orthogonal matching pursuit using the restricted Isometry property, *Inf. Theory, IEEE Trans.* 56 (2010) 4395–4401.
- [31] S.F. Cotter, B.D. Rao, K. Engan, K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, *Signal Process., IEEE Trans.* 53 (2005) 2477–2488.
- [32] W. Xiong, J. Cao, S. Li, Sparse signal recovery with unknown signal sparsity, *EURASIP J. Adv. Signal Process.* 2014 (2014) 1–8.
- [33] P.Á.Wedin, On angles between subspaces of a finite dimensional inner product space, in *Matrix Pencils*, ed: Springer, pp. 263–285, 1983.
- [34] J.M.Kim, O.K.Lee, J.C.Ye"Compressive MUSIC with optimized partial support for joint sparse recovery, in: *Information Theory Proceedings (ISIT) IEEE International Symposium* on pp. 658–662, 2011.
- [35] D.Wipf, J.-M.Yun, Q.Ling, "Augmented Bayesian Compressive Sensing," in *2015 Data Compression Conference*pp. 123–132, 2015.
- [36] A.Alkhateeb, G.Leusz, R.W.Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*pp. 2909–2913, 2015.
- [37] I. Ziskind, M. Wax, Maximum likelihood localization of multiple sources by alternating projection, *IEEE Trans. Acoust., Speech, Signal Process.* 36 (1988) 1553–1560.
- [38] L.N. Trefethen, D. Bau III, *Numerical Linear Algebra* 50, SIAM, 1997.