Abstract—We propose a MIMO channel estimation method for millimeter-wave (mmWave) and terahertz (THz) systems based on frequency-selective atomic norm minimization (FS-ANM). Due to the strong line-of-sight property of the channel in such high-frequency bands, prior knowledge on the ranges of angles of departure/arrival (AoD/AoA) can be obtained, which can be taken into account by the proposed channel estimator, to improve the estimation accuracy. Simulation results show that the proposed method can achieve considerable performance gain, as compared with existing approaches.

Index Terms—mmWave/THz channel estimation, frequency-selective atomic norm, MIMO.

I. INTRODUCTION

The mmWave/THz communication has been considered as a promising technique for future wireless communication systems [1], [2]. To compensate for the severe signal propagation loss at mmWave/THz band, the systems are expected to configure with massive antenna arrays at transceivers to achieve sufficient beamforming gains. For such MIMO systems, it is well known that the channel state information (CSI) is indispensable for reliable signal transmission and reception, and especially useful for designing efficient beamformers in mmWave/THz band [3]. However, channel estimation is challenging in mmWave/THz systems with a large number of antennas and low signal-to-noise ratio (SNR). Conventional channel estimation methods are based on the rich scattering assumption, so they are rather inefficient due to high training overhead and computational cost.

Based on the sparse representation of the MIMO channel [6], the channel estimation problem becomes equivalent to estimating only the AoD/AoAs of dominant paths and the corresponding path gains [4]. Leveraging on this, various channel estimators have been proposed in [5]–[13]. In particular, in [5], [6], closed-loop beam training based methods such as multistage beam search are proposed for channel estimation. The objective of beam scanning is to search for the best precoder-combiner pair by letting the transceiver scan the adaptive sounding beams chosen from pre-determined sounding beam codebooks. While such closed-loop methods have been adopted in practical systems, their performance tends to be limited by the training beam patterns selected from the pre-determined codebook since the beam are usually wider than the desired accuracy due to the limitation in hardware and RF circuits [5]. Different from the above closed-loop beam training techniques, the open-loop techniques perform explicit channel estimation using multiple signal classification (MUSIC) and compressive sensing (CS) methods by transmitting pilot symbols, which do not involve the time- and energy-consuming feedback process between the transceiver pair. In [7], a subspace-based channel estimation method that makes use of the MUSIC algorithm is proposed. A two-dimensional (2D) MUSIC algorithm for beamformed MIMO channel estimation is proposed in [8]. The MUSIC algorithm can result in worse performance than the CS-based algorithm in noisy environment. On the other hand, a number of CS-based channel estimators [2]–[15] have been proposed based on the virtual angular domain representation of MIMO channels [11], [12], which describes the channel with respect to some fixed basis functions of angles whose performance strongly depends on the designed dictionaries and gridding scheme in the angular domain. In order to make the grid fine enough and achieve high estimation accuracy, a refinement phase is proposed in [14] to seek the minimizer of the maximum likelihood cost function in the neighborhood of the current estimate. For reducing the basis mismatch, the grid update strategy is proposed in [15], which is implemented by exploiting the iterative surrogate function to push the grid points toward the true angles.

As an alternative, a gridless approach [16], which uses atomic norm minimization (ANM) to manifest the signal sparsity in the continuous parameter domain, has been proposed for several signal processing applications such as super-resolution frequency estimation [13], spectral estimation [17], angular estimation [19], [20] and uplink multiuser MIMO channel estimation [21]. Under certain conditions, ANM can achieve exact sparse signals reconstruction, avoiding the effects of basis mismatch which can plague grid-based CS techniques.

In mmWave/THz systems, the channels exhibit strong line-of-sight. Hence it is possible to obtain prior knowledge on the ranges of AoD/AoAs. In this paper, we propose a channel estimator that can incorporate such prior knowledge, based on frequency-selective atomic norm minimization.

II. SYSTEM MODEL

We consider a downlink MIMO communication system working at mmWave/THz band, where a BS equipped with $N_t$ antennas transmits data to a UE equipped with $N_r$ antennas. The channel $\mathbf{H}$ can be expressed as

$$\mathbf{H} = \sum_{l=1}^{L} \alpha_l \mathbf{a}(N_r, \phi_l) \mathbf{a}^H(N_t, \theta_l),$$

(1)

where $\alpha_l \sim \mathcal{CN}(0, \bar{P}/\rho)$ is the complex gain of the $l$th path, $l = 1, \ldots, L$, with $\bar{P}$ and $\rho$ denoting the average power gain and the average path loss between the BS and the UE respectively. $\mathbf{a}(N_t, \theta_l)$ and $\mathbf{a}(N_r, \phi_l)$ denote the antenna array.

Yicheng Xu, Hongyun Chu, Xiaodong Wang
response vectors of the BS and the UE respectively. In this paper, we consider the uniform linear arrays (ULA), where array response is in the form of
\[
a(N, \phi) = [1, e^{j2\pi \phi/N}, ..., e^{j2\pi (N-1)\phi/N}]^T.
\]
(2)

In (1), \(\phi_l = (d/\lambda) \sin(\phi_l)\) and \(\theta_l = (d/\lambda) \sin(\theta_l)\), with \(\lambda\) denoting the signal wavelength, \(d\) denoting the interval between adjacent antenna elements, and \(\phi_l, \theta_l\) being the UE’s azimuth AoA and the BS’ azimuth AoD of the \(l\)th path respectively.

In this paper, we assume that the ranges of the AoD/AoAs are known a priori, i.e., \(\forall l, \theta_l \in \Omega_l, \phi_l \in \Omega_2\) with \(\Omega_l, \Omega_2 \subset [0, 2\pi]\). Thus we have \(\theta_l \in \mathcal{I}_1\) and \(\phi_l \in \mathcal{I}_2\), with \(\mathcal{I}_1, \mathcal{I}_2 \subset [-d/\lambda, d/\lambda], l = 1, ..., L\). Without loss of generality, we set \(d/\lambda = 1/2\) in this paper.

The channel \(\mathbf{H}\) in (1) can be written in the matrix form as
\[
\mathbf{H} = \mathbf{A}_t \mathbf{A}_r^H, \quad \theta_l \in \mathcal{I}_1, \phi_l \in \mathcal{I}_2,
\]
(3)
where \(\mathbf{A} = \text{diag}(\alpha_1, ..., \alpha_L)\), and the matrices \(\mathbf{A}_t = [\alpha(N_t, \theta_1), ..., \alpha(N_t, \theta_L)]\) and \(\mathbf{A}_r = [\alpha(N_r, \phi_1), ..., \alpha(N_r, \phi_L)]\) contain the array response of the BS and the UE respectively.

To estimate the channel matrix, the transmitter transmits \(S\) distinct beams during \(N\) successive time slots, i.e., in the \(s\)-th time slot, the beamforming vector \(\mathbf{f}_s \in \mathbb{C}^{N_t}\) is selected from a DFT codebook of dimension \(N_t\). Thus the received signal of the \(s\)-th time slot can be expressed as
\[
\mathbf{y}_s = \mathbf{H} \mathbf{f}_s \mathbf{x}_s + \mathbf{n}_s,
\]
(4)
where \(\mathbf{n}_s \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{N_t})\) is the additive white Gaussian noise with \(\mathbf{I}_{N_t}\) denoting the \(N_t \times N_t\) identity matrix, and \(\mathbf{x}_s\) denotes the pilot symbol in the \(s\)-th time slot. The receiver collects \(\mathbf{y}_s \in \mathbb{C}^{N_t \times 1}\) for \(s = 1, ..., S\), and concatenates them to obtain the signal matrix
\[
\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_S] = \mathbf{H} \mathbf{F} \mathbf{X} + \mathbf{N},
\]
(5)
where \(\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_S] \in \mathbb{C}^{N_t \times S}\) consists of the beamforming vectors of the \(S\) time slots, \(\mathbf{X} = \text{diag}(\mathbf{x}_1, ..., \mathbf{x}_S) \in \mathbb{C}^{S \times S}\), and \(\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_S] \in \mathbb{C}^{N_t \times S}\).

III. CHANNEL ESTIMATION USING FS-ANM

A. Single Rx Antenna

When the UE has only one antenna, i.e., \(N_t > 1, N_r = 1\), then \(y_s\) is a scalar. We denote
\[
\hat{y} = [y_1, ..., y_S]^H = \mathbf{X}^H \mathbf{F}^H \hat{\mathbf{h}} + \hat{\mathbf{n}},
\]
(6)
where \(\hat{\mathbf{h}} = \mathbf{H}^H \mathbf{F}^H \hat{\mathbf{h}} + \hat{\mathbf{n}} = [n_1, ..., n_s]^H\).

To solve the off-grid problem, we employ the FS atomic norm to enforce the sparsity of \(\hat{\mathbf{h}}\). First, we briefly introduce the concept of FS Vandermonde decomposition and FS atomic norm [25].

Define \(\mathcal{I} = (f_l, f_H) \subset [-\frac{1}{2}, \frac{1}{2}]\) as a frequency interval, and trigonometric polynomial
\[
\beta(f) = r_1 z^{-1} + r_0 + r_{-1} z,
\]
(7)
where \(z = e^{j2\pi f}, r_1 = e^{j(\pi f + f_H)}, r_0 = -2 \cos(\pi(f - f_L)) + f_L, r_{-1} = r_1^*\). Then \(\beta(f)\) is always positive for \(f \in \mathcal{I}\), and negative for \(f \in [-\frac{1}{2}, \frac{1}{2}] \setminus \mathcal{I}\).

Given \(\mathcal{I} \subset [-\frac{1}{2}, \frac{1}{2}]\), a Toeplitz matrix \(\mathbf{T} \in \mathbb{C}^{N \times N}\) with \(r = \text{rank}(\mathbf{T}) \leq N - 1\) admits a unique FS Vandermonde decomposition as \(\mathbf{T} = \sum_{k_k} c_k \mathbf{a}(N, f_k) \mathbf{a}^H(N, f_k)\) with \(f_k \in \mathcal{I}\), if and only if
\[
\begin{align*}
\mathbf{T} & \succeq 0, \\
\mathbf{T}_{\beta} & \succeq 0,
\end{align*}
\]
(8)
where \(\mathbf{T} = \text{Toeplitz}(t)\) is generated by a complex sequence \(t = [t_{-N+1}, t_{-N+2}, ..., t_{N-1}]^T\), where Toeplitz denotes the Toeplitz matrix whose first column is the last \(N\) elements of the input vector, \(c_k > 0\), \(\mathbf{T}_{\beta}\) is a Toeplitz matrix defined as \(\mathbf{T}_{\beta} = \sum_{k=1} c_k \mathbf{a}(N, f_k) \mathbf{a}(N, f_k)^H\). The FS atomic set is defined as \(\mathcal{A}_T = \{\mathbf{a}(N, f_k) : f_k \in \mathcal{I}\}\). The FS atomic norm is then
\[
\|\mathbf{h}\|_{\mathcal{A}_T} = \inf_{\mathbf{a}(N, f_k) \in \mathcal{A}_T} \left\{ \sum_{k=1} c_k \|\mathbf{a}(N, f_k)\| : \mathbf{h} = \sum_{k=1} c_k \mathbf{a}(N, f_k) \right\}.
\]
(9)

Note that (9) is equivalent to the following semi-definite program (SDP) [25]
\[
\begin{align*}
\|\mathbf{h}\|_{\mathcal{A}_T} = \inf_{\mathbf{t} \in \mathbb{C}^{(N_t-1) \times 1}} \frac{1}{2N_t} \text{Tr}(\text{Toeplitz}(t)) + \frac{t}{2} \\
\text{s.t.} \quad \text{Toeplitz}(t) \|\mathbf{h}\| \geq 0, \quad \mathbf{T}_{\beta} \succeq 0,
\end{align*}
\]
(10)
where \(\text{Tr}()\) denotes the trace, \(\geq 0\) indicates a semidefinite matrix, \(t = \sum_{i=1}^L |\alpha_i|\), and \(\mathbf{T}_{\beta}\) is defined in (8).

According to [5], the 1D channel estimation can be formulated as the following optimization problem:
\[
\hat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \mathbb{C}^{N_t \times 1}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}^H \mathbf{F}^H \mathbf{F} \hat{\mathbf{h}}\|_2^2 + \mu \|\mathbf{h}\|_{\mathcal{A}_T},
\]
(11)
where \(\mu > 0\) is the weight factor. In practice, we set \(\mu \simeq \sigma_n \sqrt{\ln(N_t)}\).

The problem in (11) has \(n = \mathcal{O}(N_t)\) free variables and \(m = 2\) linear matrix inequations (LMI), and the \(i\)-th LMI has size of \(k_{i-x} \times k_i\) with \(k_i = \mathcal{O}(N_t)\). It follows from [28] that a primal-dual algorithm for (11) has a computational complexity on the order of
\[
\left(1 + \sum_{i=1}^m k_i^3 \right) \frac{1}{2} \left( n^2 + n \sum_{i=1}^m k_i^2 + \sum_{i=1}^m k_i^3 \right) = \mathcal{O}(N_t^{4.5}).
\]
(12)

By arguments similar to those above, the original atomic norm method [26] in the absence of prior knowledge has the same computational complexity of \(\mathcal{O}(N_t^{4.5})\).

B. Multiple Rx Antennas

For the case of \(N_t, N_r > 1\), \(\mathbf{Y}\) in (5) is vectorized as
\[
\mathbf{y} = \text{vec} (\mathbf{Y}) = (\mathbf{X}^T \mathbf{F} \otimes \mathbf{I}) \hat{\mathbf{h}} + \hat{\mathbf{n}},
\]
(13)
where $I$ is the identity matrix of size $N_r$, $\hat{h} = \text{vec}(H) = \sum_{l=1}^{L} \alpha_l a^*((N_l, \theta_l) \otimes \alpha(N_r, \phi_l)$ and $\hat{v} = \text{vec}(N)$, with $\otimes$ being the Kronecker product.

Before solving the problem, we first extend the FS atomic norm in (23) to the 2D case as follows.

For the 2D case, we define $I_1 = (f_{l_1}, f_{h_1}) \subset [-\frac{1}{2}, \frac{1}{2}]$, $I_2 = (f_{l_2}, f_{h_2}) \subset [-\frac{1}{4}, \frac{1}{4}]$, and a 2-level Toeplitz matrix $T(V) \in \mathbb{C}^{N_l N_r \times N_l N_r}$ formed by the elements of $V$, which is defined as $V = \{v_{N_r-1}, v_{N_r-2}, ..., v_{-N_r-1}\}$, with $v_j = [v_j(-N_r + 1), v_j(-N_r + 2), ..., v_j(N_r - 1)]^T \in \mathbb{C}^{2N_r-1 \times 1}$, $j = -N_r + 1, -N_r + 2, ..., N_r - 1$. More specifically, $T(V)$ is in the form of

$$[T(V)]_{pq} = \text{Toep}(v_{p-q}),$$
$$[\text{Toep}(v_j)]_{mn} = v_j(m-n),$$

where $\text{Toep}(\cdot)$ denotes the Toeplitz matrix whose first column is the last $N_l$ elements of the input vector, $j = -N_r + 1, -N_r + 2, ..., N_r - 1$, with $1 \leq p, q \leq N_l$, denoting the block indices and $1 \leq m, n \leq N_r$ denoting the element indices. Similar to (7), we can write $\beta_1(f)$ and $\beta_2(f)$ according to $I_1$ and $I_2$, whose parameters are denoted as $\alpha_{l_1 j}$ and $\alpha_{l_2 j}$, respectively. Then the corresponding 2-level Toeplitz matrices, i.e., $T_{\beta_1} \in \mathbb{C}^{(N_l-1)N_r \times (N_l-1)N_r}$ and $T_{\beta_2} \in \mathbb{C}^{N_l(N_l-1)N_r \times (N_l-1)N_r}$, are given by

$$[|T_{\beta_1}|]_{pq} = \sum_{j=-N_r+1}^{N_r-1} r_{1 j} v_{p-q}(m-n-j),$$
$$[T_{\beta_2}]_{pq} = \sum_{j=-N_r+1}^{N_r-1} r_{2 j} \text{Toep}(v_{p-q-j}),$$

where $1 \leq p, q \leq N_r - 1$.

Define the 2D FS atomic norm as $b(\theta, \phi) = a(N_l, \theta) \otimes a(N_r, \phi) \in \mathbb{C}^{N_l N_r}$, and the set of 2D FS atoms as $A_{\beta_1, \beta_2} = \{b(\theta, \phi) | \theta \in I_1, \phi \in I_2\}$. Then the 2D FS atomic norm of any signal $p$ with respect to $A_{\beta_1, \beta_2}$ is defined as

$$\|p\|_{A_{\beta_1, \beta_2}} = \inf_{b(\theta, \phi) \in A_{\beta_1, \beta_2}} \frac{1}{2} \|T(V)\| \geq \frac{1}{2},$$

where the $T(V)$, $T_{\beta_1}$ and $T_{\beta_2}$ are defined in (14), (15) and (16) respectively.

The proof of Lemma 2 can be found in the appendix.

For the 2D FS atomic norm, we have the corresponding SDP formulation as follows:

**Lemma 2** Let $\lambda_1$ and $\lambda_2$ be the eigenvalues of $T(V)$, then

$$\|p\|_{A_{\beta_1, \beta_2}} = \inf_{\lambda_1 \geq 0, \lambda_2 \geq 0} \left\{ \frac{1}{2} \text{Tr}(T(V)) + \frac{1}{2} \right\},$$

where $T(V)$, $T_{\beta_1}$ and $T_{\beta_2}$ are defined in (14), (15) and (16) respectively.

The proof of Lemma 1 can be found in the appendix.
decomposition \[23\]. Therefore, it suffices to show \( \theta_k \in I_1 \) and \( \phi_k \in I_2 \) under the condition \( T_{\beta_1} \geq 0 \) and \( T_{\beta_2} \geq 0 \). According to \(14\), the element of \( \mathcal{T}(\mathbf{V}) \) is given by

\[
[\mathcal{T}(\mathbf{V})]_{mn} = v_{p-q}(m-n) = \sum_{k=1}^{r} c_k e^{i 2\pi (m-n) \theta_k e^{i 2\pi (p-q) \phi_k}}.
\]

Then we have

\[
[\mathcal{T}_{\beta_1}]_{mn} = \sum_{j=1}^{1} r_{1,j} v_{p-q}(m-n-j)
\]

\[
= \sum_{j=1}^{1} r_{1,j} \sum_{k=1}^{r} c_k e^{i 2\pi (m-n-j) \theta_k e^{i 2\pi (p-q) \phi_k}}
\]

\[
= \sum_{k=1}^{r} c_k e^{i 2\pi (m-n) \theta_k e^{i 2\pi (p-q) \phi_k}} \sum_{j=1}^{1} r_{1,j} e^{-i 2\pi j \theta_k} e^{-i 2\pi j \phi_k}
\]

\[
= \sum_{k=1}^{r} c_k \beta_1(\theta_k) e^{i 2\pi (m-n) \theta_k e^{i 2\pi (p-q) \phi_k}},
\]

and hence

\[
T_{\beta_1} = \sum_{k=1}^{r} c_k \beta_1(\theta_k) \mathbf{b}_1(\theta_k, \phi_k) \mathbf{b}_1^H(\theta_k, \phi_k)
\]

\[
= \mathbf{B}_1 \mathbf{A}_1 \mathbf{B}_1^H,
\]

where \( \mathbf{b}_1(\theta_k, \phi_k) = a(N_t-1, \theta_k) \otimes a(N_r, \phi_k) \), \( \mathbf{B}_1 = [\mathbf{b}_1(\theta_1, \phi_1), \ldots, \mathbf{b}_1(\theta_r, \phi_r)] \).

According to \(16\) and \(14\), we have

\[
[\mathcal{T}_{\beta_2}]_{mn} = \sum_{j=1}^{3} r_{2,j} v_{p-q-j}(m-n).
\]

Similarly, we can get

\[
T_{\beta_2} = \mathbf{B}_2 \mathbf{A}_2 \mathbf{B}_2^H,
\]

where \( \mathbf{B}_2 = [\mathbf{b}_2(\theta_1, \phi_1), \ldots, \mathbf{b}_2(\theta_r, \phi_r)] \), with \( \mathbf{b}_2(\theta_k, \phi_k) = a(N_t, \theta_k) \otimes a(N_r-1, \phi_k) \).

Since \( r \leq \min(N_t(N_t-1), N_r-1, N_r) \), \( \mathbf{B}_1 \) and \( \mathbf{B}_2 \) have full column ranks. Using \(18\), \(25\) and \(27\), we have

\[
\{ \text{diag}(c_1 \beta_1(\theta_1), \ldots, c_r \beta_1(\theta_r)) \} = \mathbf{B}_1 \mathbf{A}_1 \mathbf{B}_1^H \geq 0,
\]

\[
\{ \text{diag}(c_1 \beta_2(\theta_1), \ldots, c_r \beta_2(\theta_r)) \} = \mathbf{B}_2 \mathbf{A}_2 \mathbf{B}_2^H \geq 0,
\]

where \( \mathbf{A}_1 \) denotes the matrix pseudo-inverse operator. Thus \( \forall k \), \( c_k \beta_1(\theta_k) \geq 0 \), \( c_k \beta_2(\theta_k) \geq 0 \). Since \( c_k > 0 \), we have \( \beta_1(\theta_k) \geq 0 \) and \( \beta_2(\phi_k) \geq 0 \). By the property of \( \beta(f) \), we finally have \( \theta_k \in I_1 \), \( \phi_k \in I_2 \), \( k = 1, \ldots, r \).

The “only if” part can be shown by similar arguments. Given \( \mathcal{T}(\mathbf{V}) = \sum_{k=1}^{r} c_k \mathbf{b}(\theta_k, \phi_k) \mathbf{b}^H(\theta_k, \phi_k) \), it is evident that \( \mathcal{T}(\mathbf{V}) \geq 0 \). Then on the basis of \(25\), \(27\) and the property of \( \beta(f) \), we have \( T_{\beta_1} \geq 0 \) and \( T_{\beta_2} \geq 0 \).

**Proof of Lemma 2.** Let \( F^* \) be the optimal objective value of \(19\). We need to show that \( \| \mathbf{p} \|_{\mathcal{A}_{I_1,I_2}} = F^* \).

To begin with, we first show that \( F^* \leq \| \mathbf{p} \|_{\mathcal{A}_{I_1,I_2}} \). Let \( \mathbf{p} = \sum_{k=1}^{r} c_k \mathbf{b}(\theta_k, \phi_k) \mathbf{b}^H(\theta_k, \phi_k) \) be an 2D FS Vandermonde decomposition of \( \mathbf{p} \) on \( I_1 \) and \( I_2 \), with \( \| \psi_k \|^2 = 1 \). Then let \( \mathbf{V} \) conform to \( \mathcal{T}(\mathbf{V}) = \sum_{k=1}^{r} c_k \mathbf{b}(\theta_k, \phi_k) \mathbf{b}^H(\theta_k, \phi_k) \) and \( t = \sum_{k=1}^{r} c_k \). By
Thus the constructed $t$ and $V$ are a feasible solution to the problem \((19)\), with the objective value calculated as

$$ \frac{1}{2N_tN_r} \text{Tr}(T(V)) + \frac{1}{2} \sum c_k. \quad (30) $$

Therefore, it holds that $F^* \leq \sum c_k$. Since the inequality holds for any FS atomic decomposition of $p$ on $I_1$ and $I_2$, we have that $F^* \leq \|p\|_{A_{x_1,x_2}}$ based on the definition of $\|p\|_{A_{x_1,x_2}}$.

Next we will show that $F^* \geq \|p\|_{A_{x_1,x_2}}$. We suppose that $(t^*, V^*)$ is the optimal solution to \((19)\). By the fact that $T(V^*) \succeq 0$, $T_{\beta_1} \succeq 0$ and $T_{\beta_2} \succeq 0$, according to lemma 1, $\hat{T}(V^*)$ has an FS Vandermonde decomposition on $I_1$ and $I_2$ given by

$$ \hat{T}(V^*) = \sum_{k=1}^{r^*} c^*_k b(\theta^*_k, \phi^*_k) b^H(\theta^*_k, \phi^*_k). \quad (31) $$

Since $T(V^*) \succeq 0$, $p$ lies in the range space of $\hat{T}(V^*)$ and thus has an FS atomic decomposition given by

$$ p = \sum_{k=1}^{r^*} c^*_k b(\theta^*_k, \phi^*_k) \psi^*_k, \text{ } \|\psi^*_k\|^2 = 1, \theta^*_k \in I_1, \phi^*_k \in I_2, \quad (32) $$

which achieves the FS atomic norm. Furthermore, it holds that

$$ t^* \succeq p^H [T(V^*)]^\dagger p = \sum_{k=1}^{r^*} c^*_k, $$

$$ \frac{1}{N_tN_r} \text{Tr}(T(V^*)) = \sum_{k=1}^{r^*} c^*_k. \quad (33) $$

Thus we have

$$ F^* = \frac{1}{2N_tN_r} \text{Tr}(T(V^*)) + \frac{t^*}{2} \sum c^*_k \geq \|p\|_{A_{x_1,x_2}}. \quad (34) $$

Since $F^* \leq \|p\|_{A_{x_1,x_2}}$ and $F^* \geq \|p\|_{A_{x_1,x_2}}$ have both been shown, we conclude that $F^* \geq \|p\|_{A_{x_1,x_2}}$. Thus lemma 2 is proved.

REFERENCES

[1] P. Wang, Y. Li, L. Song, B. Vucetic, “Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks”, IEEE Commun. Mag., vol. 53, no. 1, pp. 168-178, Jan. 2015.

[2] V. W. Wong, R. Schober, D. W. K. Ng, L.-C. Wang, Key Technologies for 5G Wireless Systems, Cambridge, U.K./Cambridge Univ. Press, 2017.

[3] E. A. Omar, S. Rajagopal, S. Abu-Surra, Z. Pi, R. W. Heath, Jr., “Spatially sparse precoding in millimeter wave MIMO systems”, IEEE Trans. Wirel. Commun., vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[4] L. Zhao, G. Geraci, T. Yang, D. W. K. Ng, J. Yuan, “A tone-based AoA estimation and multiuser precoding for millimeter wave massive MIMO”, IEEE Trans. Commun., vol. 65, no. 12, pp. 5209-5225, Dec. 2017.

[5] J. Wang, Z. Lan, C. Pyo, T. Baykas, C. S. Sum, M. A. Rahman, R. F. F. Kojima, I. L. H. Harada, S. Kato, “Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems”, IEEE J. Sel. Areas Commun., vol. 27, no. 8, pp. 1390-1399, Oct. 2009.

[6] A. Alkhateeb, O. El Ayach, G. Leus, R. W. Heath, Jr., “Channel estimation and hybrid precoding for millimeter wave cellular systems”, IEEE J. Sel. Top. Signal Process., vol. 8, no. 5, pp. 831-846, Oct. 2014.

[7] T. S. Rappaport, R. W. Heath, R. C. Daniels, J. N. Murdock, Millimeter Wave Wireless Communications, Englewood Cliffs, NJ, USA:Prentice-Hall, 2015.

[8] Z. Guo, X. Wang, W. Heng, “Millimeter-wave channel estimation based on two-dimensional beamspace MUSIC method”, IEEE Trans. Wirel. Commun., vol. 16, no. 8, pp. 5384-5394, Aug. 2017.

[9] X. Rao, W. K. N. Lau, “Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems”, IEEE Trans. Signal Process., vol. 62, no. 12, pp. 3261-3271, Jun. 2014.

[10] W. Ding, F. Yang, W. Dai, J. Song, “Time-frequency joint sparse channel estimation for MIMO-OFDM systems”, IEEE Commun. Lett., vol. 19, no. 1, pp. 58-61, Jan. 2015.

[11] W. U. Bajwa, J. Haupt, A. M. Sayeed, R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels”, Proc. IEEE, vol. 98, no. 6, pp. 1058-1076, Jun. 2010.

[12] A. M. Sayeed, “Deconstructing multiantenna fading channels”, IEEE Trans. Signal Process., vol. 50, no. 10, pp. 2563-2579, Oct. 2002.

[13] J. Lee, G. T. Gil, Y. H. Lee, “Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications”, IEEE Trans. Commun., vol. 64, no. 6, pp. 2370-2386, Jun. 2016.

[14] Z. Marzi, D. Ramsamy, U. Madhow, “Compressive channel estimation and tracking for large arrays in mm-wave picocells”, IEEE J. Sel. Top. Signal Process., vol. 10, no. 3, pp. 514-527, Apr. 2016.

[15] C. Hu, L. Dai, T. Mir, Z. Gao, J. Fang, “Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding”, IEEE Trans. Veh. Technol., vol. 67, no. 9, pp. 8954-8958, Sep. 2018.

[16] S. Pejoski, V. Kafezidis, “Estimation of sparse time dispersive channels in pilot aided OFDM using atomic norm”, IEEE Commun. Lett., vol. 4, no. 4, pp. 397-400, Aug. 2015.

[17] B. N. Bhaskar, G. Tang, B. Recht, “Atomic norm denoising with applications to line spectral estimation”, IEEE Trans. Signal Process., vol. 61, no. 23, pp. 5987-5999, Dec. 2013.

[18] J. Steinwandt, F. Roemer, C. Steffens, M. Haardt, M. Pesavento, “Gridless super-resolution direction finding for strictly noncircular sources based on atomic norm minimization”, Proc. 50th Asilomar Conf. Signals Systems and Computers, pp. 1518-1522, Nov. 2016.

[19] Y. Zhang, G. Zhang, X. Wang, “Array covariance matrix-based atomic norm minimization for off-grid coherent direction-of-arrival estimation”, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 3196-3200, Mar. 2017.

[20] Z. Tian, Z. Zhang, Y. Wang, “Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization”, in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 3071-3075, Mar. 2017.

[21] P. Zhang, L. Gan, S. Sun, C. Ling, “Atomic norm denoising-based channel estimation for massive multiuser mimo systems”, in Proc. IEEE Int. Conf. Commun. (ICC), pp. 4564-4569, Jun. 2015.

[22] A. Alkhateeb, O. E. Ayach, G. Leus, R. W. Heath, Jr., “Channel estimation and hybrid precoding for millimeter wave cellular systems”, IEEE J. Sel. Top. Signal Process., vol. 8, no. 5, pp. 831-846, Jul. 2014.

[23] Z. Yang, L. Xie, “Frequency-selective Vandermonde decomposition of Toeplitz matrices with applications”, Signal Process., vol. 142, pp. 157-167, Jan. 2018.

[24] D. Yang, L. -L. Yang, L. Hanzo, “DFT-based beamforming weight-vector codebook design for spatially correlated channels in three-antenna precoding aided multiuser downlink”, Proc. IEEE Int. Conf. Commun. (ICC), pp. 1-5, May 2010.

[25] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, “Compressed sensing off the grid”, IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7465-7490, Nov. 2013.

[26] Z. Yang, L. Xie, P. Stoica, “Vandermonde decomposition of multilevel Toeplitz matrices with applications to multidimensional super-resolution”, IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3685-3701, Jun. 2016.

[27] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, Z. Wang, “Fast channel tracking for terahertz beamspace massive MIMO systems”, IEEE Trans. Veh. Technol., vol. 66, no. 7, pp. 5689-5696, Jul. 2017.

[28] A. Ben-Tal, A. K. Nemirovski, Lectures on modern convex optimization: Analysis algorithms and engineering applications, PA, Philadelphia:SIAM, 2001.