

Massive MIMO Channel Estimation for Millimeter Wave Systems via Matrix Completion

Evangelos Vlachos , George C. Alexandropoulos , Senior Member, IEEE, and John Thompson, Fellow, IEEE

Abstract—Millimeter wave (mmWave) massive multiple input multiple output (MIMO) systems realizing directive beamforming require reliable estimation of the wireless propagation channel. However, mmWave channels are characterized by high variability that severely challenges their recovery over short training periods. Current channel estimation techniques exploit either the channel sparsity in the beamspace domain or its low-rank property in the antenna domain, nevertheless, they still require large numbers of training symbols for the satisfactory performance. In this letter, we present a novel channel estimation algorithm that jointly exploits the latter two properties of mmWave channels to provide more accurate recovery, especially for shorter training intervals. The proposed iterative algorithm is based on the alternating direction method of multipliers and provides the global optimum solution to the considered convex mmWave channel estimation problem with fast convergence properties.

Index Terms—Alternating direction method of multiplier (ADMM), beamforming, channel estimation, massive multiple input multiple output (MIMO), matrix completion, millimeter wave.

I. INTRODUCTION

NEAR-OPTIMAL beamforming (BF) performance in millimeter wave (mmWave) massive multiple input multiple output (MIMO) systems employing Hybrid analog/digital BF (HBF) architectures necessitates reliable channel state information (CSI) knowledge. This knowledge is, however, very challenging to acquire in practice due to the very large numbers of transceiver antenna elements and the high channel variability [1]. Several approaches requiring receiver feedback have been lately proposed for designing BF vectors adequate for CSI estimation [2], [3]. On the other side, static dictionaries or beam training techniques without receiver feedback have also been adopted for beam codebook designs [4]–[6]. In these studies, CSI estimation has been treated as a compressive sensing problem [7], where the orthogonal matching pursuit (OMP) algorithm [8] has usually been adopted to recover the sparse channel

Manuscript received July 12, 2018; revised September 1, 2018; accepted September 2, 2018. Date of publication September 17, 2018; date of current version September 28, 2018. The work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/P000703/1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Luca Sanguinetti. (*Corresponding author: Evangelos Vlachos*)

E. Vlachos and J. Thompson are with the Institute for Digital Communications, University of Edinburgh, Edinburgh EH9 3JL, U.K. (e-mail: e.vlachos@ed.ac.uk; j.s.thompson@ed.ac.uk).

G. C. Alexandropoulos is with the Mathematical and Algorithmic Sciences Lab, Paris Research Center, Huawei Technologies France SASU, Boulogne-Billancourt 92100, France (e-mail: george.alexandropoulos@huawei.com).

Color versions of one or more of the figures in this letter are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2018.2870533

gain vector. However, the performance of the aforementioned channel estimation techniques is usually limited by the codebook design, since beam dictionaries suffer from power leakage due to the discretization of the angles of arrival (AoA) and departure. Very recently in [9], mmWave CSI estimation that exploits both the sparsity and low-rank properties of mmWave MIMO channels via a two independent stages procedure (one stage per each property) was proposed.

In this letter, we present a novel *joint optimization formulation* for mmWave massive MIMO channel estimation incorporating both the sparsity and low-rank properties, which possesses a global optimum solution due to its convexity property. To achieve the optimum solution, we capitalize on a recently developed theory of matrix completion with side information [10], which we deploy together with the channel's beamspace representation. We develop an algorithm based on the alternating direction method of multipliers (ADMM) [11] for efficient recovery of massive MIMO channel matrices. It is shown through representative simulation results that the proposed algorithm exhibits faster convergence and improved performance in terms of mean squared error (MSE) for channel estimation with short training length, when compared with other state-of-the-art techniques [4], [9], [12], [13].

Notation: Fonts α , \mathbf{a} , and \mathbf{A} denote a scalar, a vector, and a matrix, respectively. \mathbf{A}^T , \mathbf{A}^* , \mathbf{A}^H , and $\|\mathbf{A}\|_F$ represent \mathbf{A} 's transpose, conjugate transpose, Hermitian transpose, and Frobenius norm. Operands \circ and \otimes denote the matrix Hadamard and Kronecker products, respectively. $\text{vec}(\cdot)$ concatenates the columns of a matrix into a vector, and $\text{unvec}(\cdot)$ is the inverse operation; $\|\mathbf{A}\|_* \triangleq \sum_{i=1}^r \sigma_i$ is the nuclear norm with σ_i 's being the r singular values of \mathbf{A} ; $\|\mathbf{A}\|_1 \triangleq \max_{1 \leq j \leq N} \sum_{i=1}^M |[\mathbf{A}]_{i,j}|$ ($\mathbf{A} \in \mathbb{C}^{M \times N}$) with $[\mathbf{A}]_{i,j}$ denoting \mathbf{A} 's (i, j) th element; $\mathcal{E}\{\cdot\}$ is the expected value. $\mathbf{A} \in \{0, 1\}^{M \times N}$ implies that \mathbf{A} 's elements are taken independently and with equal probability from the binary set $\{0, 1\}$.

II. SYSTEM AND CHANNEL MODELS

We consider an $N_R \times N_T$ massive MIMO system operating over quasi-static mmWave channels, and adopting analog BF with switches [4] for the purpose of channel estimation. This cost and energy efficient BF scheme, which is sufficient for the channel estimation presented in this letter, can be realized with any available HBF architectures [14]. Assuming that the channel $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ remains static during the transmission of T unit power training symbols $s[t] \in \mathbb{C}$, $\forall t = 1, 2, \dots, T$, the postprocessed received signal at the N_R -element receiver (RX) is expressed as $r[t] \triangleq \sqrt{P_t} \mathbf{w}^T \mathbf{H} \mathbf{f} s[t] + n[t]$, where P_t is the transmitter (TX) power, $\mathbf{w} \in \{0, 1\}^{N_R}$ and $\mathbf{f} \in \{0, 1\}^{N_T}$ denote the RX combining and TX precoding vectors, respectively, and

$n[t]$ represents the zero-mean complex additive white Gaussian noise with variance σ_n^2 .

We adopt the geometric representation [3], [9] for the mmWave MIMO channel, according to which \mathbf{H} is given by

$$\mathbf{H} \triangleq \sum_{k=1}^{N_p} \alpha_k \mathbf{a}_R(\phi_R^{(k)}, \theta_R^{(k)}) \mathbf{a}_T^H(\phi_T^{(k)}, \theta_T^{(k)}) \quad (1)$$

where N_p denotes the number of propagation paths and α_k is the gain of the k th path drawn from the complex Gaussian distribution $\mathcal{CN}(0, 1/2)$. $\mathbf{a}_T^H(\phi_T^{(k)}, \theta_T^{(k)}) \in \mathbb{C}^{N_T}$ and $\mathbf{a}_R(\phi_R^{(k)}, \theta_R^{(k)}) \in \mathbb{C}^{N_R}$ represent the TX and RX array response vectors, respectively, which are expressed as described in [1, Sec. II.C] for uniform arrays. $\phi_T^{(k)}, \theta_T^{(k)}$ and $\phi_R^{(k)}, \theta_R^{(k)}$ are the physical elevation and azimuth angles of departure and arrival, respectively, which are generated according to the Laplace distribution [15]. An alternative representation for \mathbf{H} is based on the beamspace model [16], [17] that is defined as

$$\mathbf{H} = \mathbf{D}_R \mathbf{Z} \mathbf{D}_T^H \quad (2)$$

where $\mathbf{D}_R \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{D}_T \in \mathbb{C}^{N_T \times N_T}$ are unitary matrices based on the discrete Fourier transform (DFT). For uniform linear arrays (ULAs), \mathbf{D}_R and \mathbf{D}_T are the normalized DFT matrices, whereas for the planar case they are the normalized Khatri-Rao products of DFT matrices [16]. It holds for both cases that $\mathbf{D}_R^H \mathbf{D}_R = \mathbf{I}_{N_R}$ and $\mathbf{D}_T^H \mathbf{D}_T = \mathbf{I}_{N_T}$ with \mathbf{I}_N being the $N \times N$ identity matrix. Also in (2), $\mathbf{Z} \in \mathbb{C}^{N_R \times N_T}$ contains only a few virtual channel gains with high amplitude, i.e., it is a sparse (or compressible) matrix.

III. PROPOSED MMWAVE MIMO CHANNEL ESTIMATION

A. Problem Formulation

Matrix completion [13] for the recovery of the unknown elements of a matrix \mathbf{H} has been recently extended to incorporate side knowledge about the structure or properties of \mathbf{H} [18]. Motivated by this idea, we consider the beamspace representation of \mathbf{H} given by (2) as its side information; particularly, we assume that the unknown \mathbf{H} is decomposed as $\mathbf{D}_R \mathbf{S} \mathbf{D}_T^H$ with \mathbf{S} being the unknown matrix. On this premise, we formulate the following constrained optimization problem (OP) for the joint recovery of the unknown CSI matrix \mathbf{H} and its beamspace representation via the unknown sparse channel gain matrix \mathbf{S}

$$\begin{aligned} & \min_{\mathbf{H}, \mathbf{S}} \tau_H \|\mathbf{H}\|_* + \tau_S \|\mathbf{S}\|_1 \\ & \text{subject to } \boldsymbol{\Omega} \circ \mathbf{H} = \mathbf{H}_\Omega \text{ and } \mathbf{H} = \mathbf{D}_R \mathbf{S} \mathbf{D}_T^H \end{aligned} \quad (3)$$

where \mathbf{H} 's nuclear norm in the objective imposes its low-rank property, whereas the ℓ_1 -norm of \mathbf{S} enforces its sparse structure. Also, constraint $\mathbf{H} = \mathbf{D}_R \mathbf{S} \mathbf{D}_T^H$ refers to \mathbf{H} 's representation given by (2), and the weighting factors $\tau_H, \tau_S > 0$ depend in general on the number of the mmWave channel propagation paths.

Matrix $\boldsymbol{\Omega} \in \{0, 1\}^{N_R \times N_T}$ is composed of M ones and $N_R N_T - M$ zeros; hence, $\|\boldsymbol{\Omega}\|_0 = M$. The positions of its unity elements are randomly chosen in a uniform fashion over the set $\Omega \triangleq \{1, 2, \dots, N_R N_T\}$ [13], [18]. The matrix \mathbf{H}_Ω represents the subsampled estimated channel matrix and contains M nonzero entries following the same pattern with $\boldsymbol{\Omega}$. These entries are derived prior to the solution of (3), based on the training procedure, which is described in Section III-B. Clearly, \mathbf{H} 's estimation error from (3) depends on the value of M ($0 \leq M \leq N_R N_T$) and the

estimation accuracy of \mathbf{H}_Ω 's elements. Note also that (2) may introduce additional errors due to the angle discretization effect [1], [3].

B. Proposed ADMM-Based Solution

The OP of (3) is a two-objective convex problem, and thus, it possesses a global optimum that can be efficiently found via alternating optimization techniques [11]. We first introduce the two auxiliary matrix variables $\mathbf{Y} \in \mathbb{C}^{N_R \times N_T}$ and $\mathbf{C} \triangleq \mathbf{Y} - \mathbf{D}_R \mathbf{S} \mathbf{D}_T^H$ to reformulate the targeted OP in the following equivalent form:

$$\begin{aligned} & \min_{\mathbf{H}, \mathbf{Y}, \mathbf{S}, \mathbf{C}} \tau_H \|\mathbf{H}\|_* + \tau_S \|\mathbf{S}\|_1 + \frac{1}{2} \|\mathbf{C}\|_F^2 + \frac{1}{2} \|\boldsymbol{\Omega} \circ \mathbf{Y} - \mathbf{H}_\Omega\|_F^2 \\ & \text{subject to } \mathbf{H} = \mathbf{Y} \text{ and } \mathbf{C} = \mathbf{Y} - \mathbf{D}_R \mathbf{S} \mathbf{D}_T^H. \end{aligned} \quad (4)$$

Note that now the third and fourth terms in the objective take into account possible noise on \mathbf{H}_Ω in the problem formulation. Although OP in (4) seems more complex than that in (3), it has separate blocks of variables (i.e., a separable cost function). This property enables ADMM utilization, and consequently, the augmented Lagrangian function of (4) is given by

$$\begin{aligned} \mathcal{L}_1(\mathbf{H}, \mathbf{Y}, \mathbf{S}, \mathbf{C}, \mathbf{Z}_1, \mathbf{Z}_2) & \triangleq \tau_H \|\mathbf{H}\|_* + \tau_S \|\mathbf{S}\|_1 + \frac{1}{2} \|\mathbf{C}\|_F^2 \\ & + \frac{1}{2} \|\boldsymbol{\Omega} \circ \mathbf{Y} - \mathbf{H}_\Omega\|_F^2 + \text{tr}(\mathbf{Z}_1^H (\mathbf{H} - \mathbf{Y})) + \frac{\rho}{2} \|\mathbf{H} - \mathbf{Y}\|_F^2 \\ & + \text{tr}(\mathbf{Z}_2^H (\mathbf{Y} - \mathbf{D}_R \mathbf{S} \mathbf{D}_T^H - \mathbf{C})) + \frac{\rho}{2} \|\mathbf{Y} - \mathbf{D}_R \mathbf{S} \mathbf{D}_T^H - \mathbf{C}\|_F^2 \end{aligned}$$

where $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathbb{C}^{N_R \times N_T}$, are dual variables (the Lagrange multipliers) adding the constraints of (4) to the cost function, and ρ denotes ADMM's stepsize. According to standard ADMM, at the ℓ th algorithmic iteration with $\ell = 0, 1, \dots$ the following separate subproblems need to be solved

$$\mathbf{H}^{(\ell+1)} = \arg \min_{\mathbf{H}} \mathcal{L}_1(\mathbf{H}, \mathbf{Y}^{(\ell)}, \mathbf{S}^{(\ell)}, \mathbf{C}^{(\ell)}, \mathbf{Z}_1^{(\ell)}, \mathbf{Z}_2^{(\ell)}) \quad (5)$$

$$\mathbf{Y}^{(\ell+1)} = \arg \min_{\mathbf{Y}} \mathcal{L}_1(\mathbf{H}^{(\ell+1)}, \mathbf{Y}, \mathbf{S}^{(\ell)}, \mathbf{C}^{(\ell)}, \mathbf{Z}_1^{(\ell)}, \mathbf{Z}_2^{(\ell)}) \quad (6)$$

$$\mathbf{S}^{(\ell+1)} = \arg \min_{\mathbf{S}} \mathcal{L}_1(\mathbf{H}^{(\ell+1)}, \mathbf{Y}^{(\ell+1)}, \mathbf{S}, \mathbf{C}^{(\ell)}, \mathbf{Z}_1^{(\ell)}, \mathbf{Z}_2^{(\ell)}) \quad (7)$$

$$\mathbf{C}^{(\ell+1)} = \arg \min_{\mathbf{C}} \mathcal{L}_1(\mathbf{H}^{(\ell+1)}, \mathbf{Y}^{(\ell+1)}, \mathbf{S}^{(\ell+1)}, \mathbf{C}, \mathbf{Z}_1^{(\ell)}, \mathbf{Z}_2^{(\ell)}) \quad (8)$$

$$\mathbf{Z}_1^{(\ell+1)} = \mathbf{Z}_1^{(\ell)} + \rho(\boldsymbol{\Omega} \circ \mathbf{Y}^{(\ell+1)} - \mathbf{H}_\Omega) \quad (9)$$

$$\mathbf{Z}_2^{(\ell+1)} = \mathbf{Z}_2^{(\ell)} + \rho(\mathbf{Y}^{(\ell+1)} - \mathbf{D}_R \mathbf{S}^{(\ell+1)} \mathbf{D}_T^H - \mathbf{C}^{(\ell+1)}). \quad (10)$$

Note that for the initialization $\ell = 0$: $\mathbf{H}^{(0)} = \mathbf{Z}_1^{(0)} = \mathbf{Z}_2^{(0)} = \mathbf{0}$.

To proceed with the formulation of the proposed algorithm, we derive closed form solutions for the problems (5)–(8). First, to solve (5), we reformulate \mathcal{L}_1 to \mathcal{L}_2 as follows, where the terms not affecting the minimization over \mathbf{H} were removed and the term $\|\frac{1}{\rho} \mathbf{Z}_1^{(\ell-1)}\|_F^2$ was added

$$\mathcal{L}_2(\mathbf{H}) \triangleq \tau_H \|\mathbf{H}\|_* + \frac{\rho}{2} \left\| \mathbf{H} - \left(\mathbf{Y}^{(\ell)} - \frac{1}{\rho} \mathbf{Z}_1^{(\ell)} \right) \right\|_F^2. \quad (11)$$

Given the Lagrangian in (11), the solution of (5) is obtained from the singular value thresholding (SVT) operator [13]

$$\mathbf{H}^{(\ell+1)} = \text{Udiag}(\{\text{sign}(\xi_i) \max(\xi_i, 0)\}_{1 \leq i \leq r}) \mathbf{V}^H \quad (12)$$

where $\mathbf{U} \in \mathbb{C}^{N_R \times r}$ and $\mathbf{V} \in \mathbb{C}^{N_R \times r}$ are the left and right singular vector matrices of the matrix $(\mathbf{Y}^{(\ell)} - \frac{1}{\rho} \mathbf{Z}_1^{(\ell)})$, respectively, and $\xi_i \triangleq \sigma_i - \tau/\rho$ with σ_i s denote its r singular values. Similarly, to derive the solution of (6), we reformulate \mathcal{L}_1 to the following Lagrangian function \mathcal{L}_3 of \mathbf{Y} :

$$\begin{aligned} \mathcal{L}_3(\mathbf{Y}) &\triangleq \|\mathbf{\Omega} \circ \mathbf{Y} - \mathbf{H}_\Omega\|_F^2 + \frac{\rho}{2} \left\| \frac{1}{\rho} \mathbf{Z}_1^{(\ell)} + \mathbf{H}^{(\ell+1)} - \mathbf{Y} \right\|_F^2 \\ &+ \frac{\rho}{2} \left\| \frac{1}{\rho} \mathbf{Z}_2^{(\ell)} + \mathbf{C}^{(\ell)} - \mathbf{Y} + \mathbf{D}_R \mathbf{S}^{(\ell)} \mathbf{D}_T^H \right\|_F^2 \end{aligned} \quad (13)$$

which can be equivalently expressed based on the Krockecker vectorization and the Hadamard element-wise property as

$$\begin{aligned} \mathcal{L}_3(\mathbf{y}) &= \frac{1}{2} \|\mathbf{A}\mathbf{y} - \mathbf{h}_\Omega\|_2^2 + \frac{\rho}{2} \left\| \frac{1}{\rho} \mathbf{z}_1^{(\ell)} + \mathbf{h}^{(\ell+1)} - \mathbf{y} \right\|_2^2 \\ &+ \frac{\rho}{2} \left\| \frac{1}{\rho} \mathbf{z}_2^{(\ell)} + \mathbf{c}^{(\ell)} - \mathbf{y} + \mathbf{B}\mathbf{s}^{(\ell)} \right\|_2^2. \end{aligned} \quad (14)$$

In (14), $\mathbf{B} \triangleq \mathbf{D}_T^* \otimes \mathbf{D}_R$ and $\mathbf{A} \triangleq \sum_{i=1}^{N_R} \text{diag}([\mathbf{\Omega}]_i)^T \otimes \mathbf{E}_{ii}$ where $[\mathbf{\Omega}]_i$ denoting $\mathbf{\Omega}$ s i th row and \mathbf{E}_{ii} obtained from the $N_R \times N_R$ all-zero matrix after inserting a unity value at its (i, i) th position. Also, small boldfaced letters are the $\text{vec}(\cdot)$ results of their capital equivalents. Then, (14) for (6) is minimized with

$$\begin{aligned} \mathbf{y}^{(\ell+1)} &= (\mathbf{A}^H \mathbf{A} + 2\rho \mathbf{I})^{-1} (\mathbf{z}_1^{(\ell)} + \rho \mathbf{h}^{(\ell+1)} \\ &+ \mathbf{A}^H \mathbf{h}_\Omega + \mathbf{z}_2^{(\ell)} + \rho \mathbf{c}^{(\ell)} + \rho \mathbf{B}\mathbf{s}^{(\ell)}) \end{aligned} \quad (15)$$

which is finally used to obtain $\mathbf{Y}^{(\ell+1)} = \text{unvec}(\mathbf{y}^{(\ell+1)})$.

To find \mathbf{S} solving (7), we reformulate \mathcal{L}_1 to \mathcal{L}_4 as follows:

$$\mathcal{L}_4(\mathbf{S}) \triangleq \tau_S \|\mathbf{S}\|_1 + \frac{\rho}{2} \left\| \mathbf{D}_R^H \left(\frac{1}{\rho} \mathbf{Z}_2^{(\ell)} - \mathbf{C}^{(\ell)} + \mathbf{Y}^{(\ell+1)} \right) \mathbf{D}_T + \mathbf{S} \right\|_F^2 \quad (16)$$

where we have used the property that \mathbf{D}_T and \mathbf{D}_R are unitary matrices. By performing vectorization, \mathcal{L}_4 is equivalent to a standard LASSO problem [19], namely

$$\arg \min_{\mathbf{s}} \tau_S \|\mathbf{s}\|_1 + \frac{\rho}{2} \|\mathbf{s} - \mathbf{v}^{(\ell+1)}\|_2^2 \quad (17)$$

where $\mathbf{s}^{(\ell)} \triangleq \text{vec}(\mathbf{S}^{(\ell)})$ and $\mathbf{v}^{(\ell+1)} \triangleq \text{vec}(\mathbf{V}^{(\ell+1)})$ with

$$\mathbf{v}^{(\ell+1)} \triangleq \mathbf{D}_R^H \left(\frac{1}{\rho} \mathbf{Z}_2^{(\ell)} - \mathbf{C}^{(\ell)} + \mathbf{Y}^{(\ell+1)} \right) \mathbf{D}_T. \quad (18)$$

The solution of (17) is, thus, given by

$$\begin{aligned} \mathbf{s}^{(\ell+1)} &= \text{sign}(\text{Re}(\mathbf{v}^{(\ell+1)})) \circ \max(|\text{Re}(\mathbf{v}^{(\ell+1)})| - \tau'_S, 0) \\ &+ \text{sign}(\text{Im}(\mathbf{v}^{(\ell+1)})) \circ \max(|\text{Im}(\mathbf{v}^{(\ell+1)})| - \tau'_S, 0) \end{aligned} \quad (19)$$

where $\tau'_S \triangleq \tau_S/\rho$, and the $\max(\cdot)$ and the sign operator $\text{sign}(\cdot)$ are applied component wise. The resulting vector in (19) is then transformed into matrix form as $\mathbf{S}^{(\ell+1)} = \text{unvec}(\mathbf{s}^{(\ell+1)})$.

To finally solve (8) for \mathbf{C} , we reformulate \mathcal{L}_1 as follows:

$$\mathcal{L}_5(\mathbf{C}) = \frac{1}{2} \|\mathbf{C}\|_F^2 + \frac{\rho}{2} \left\| \frac{1}{\rho} \mathbf{Z}_2^{(\ell)} + \mathbf{Y}^{(\ell+1)} - \mathbf{D}_R \mathbf{S}^{(\ell+1)} \mathbf{D}_T^H - \mathbf{C} \right\|_F^2$$

which is strictly convex with respect to \mathbf{C} . Taking the derivative and equating it to zero yields the closed form solution

$$\mathbf{C}^{(\ell+1)} = \frac{\rho}{\rho + 1} \left(\mathbf{Y}^{(\ell+1)} - \mathbf{D}_R \mathbf{S}^{(\ell+1)} \mathbf{D}_T^H + \frac{1}{\rho} \mathbf{Z}_2^{(\ell)} \right). \quad (20)$$

Algorithm 1: ADMM-based MIMO Channel Estimation.

Input: \mathbf{H}_Ω , $\mathbf{\Omega}$, \mathbf{D}_R , \mathbf{D}_T , ρ , τ_H , τ_S , and I_{\max}
Output: $\mathbf{H}^{(I_{\max})}$

Initialization: $\mathbf{H}^{(0)} = \mathbf{S}^{(0)} = \mathbf{C}^{(0)} = \mathbf{Z}_2^{(0)} = \mathbf{Z}_1^{(0)} = \mathbf{0}$
1: **for** $\ell = 0, 1, \dots, I_{\max} - 1$ **do**
2: Update $\mathbf{H}^{(\ell+1)}$ using (12).
3: Update $\mathbf{Y}^{(\ell+1)} = \text{unvec}(\mathbf{y}^{(\ell+1)})$ using (15).
4: Update $\mathbf{S}^{(\ell+1)} = \text{unvec}(\mathbf{s}^{(\ell+1)})$ using (19).
5: Update $\mathbf{C}^{(\ell+1)}$ using (20).
6: Update $\mathbf{Z}_1^{(\ell+1)}$ and $\mathbf{Z}_2^{(\ell+1)}$ using (9) and (10).
7: **end for**

Expressions (9) and (10) including the dual variable updates can be straightforwardly computed using (15), (19), and (20).

The previously described ADMM steps constituting the proposed mmWave massive MIMO CSI estimation technique are summarized in Algorithm 1. After a predefined number of algorithmic iterations I_{\max} , the output of this algorithm is the estimated MIMO channel matrix $\mathbf{H}^{(I_{\max})}$.

1) *Computational Complexity*: The computational complexity of Algorithm 1 depends on the numbers of TX and RX antennas N_T and N_R , as well as the number of iterations I_{\max} . The most demanding step is in line 2 with the update $\mathbf{H}^{(\ell+1)}$ that requires the computation of σ_i s for the SVT operator. In general, the complexity of this computation for a $N_R \times N_T$ matrix is $\mathcal{O}(N_R^2 N_T)$ [20, Ch. 8.6]. However, the dominant singular values and vectors can be efficiently computed via incomplete singular value decomposition methods (e.g., Lanczos bidiagonalization algorithm [21]) or via subspace tracking, thus, the complexity can be further reduced to $\mathcal{O}(N_R N_T)$.

2) *Channel Subsampling \mathbf{H}_Ω* : Algorithm 1 will run at every channel coherence interval requiring as input the estimation of a subsampled version of \mathbf{H} . We adopt the training procedure described in Section II to estimate $M \ll N_R N_T$ nonzero elements of \mathbf{H}_Ω at respective training instances (i.e., $T = M$). Specifically, to estimate the (i, j) th nonzero element of \mathbf{H}_Ω at the t th training instance, we use the training symbol $s[t]$ and set $\mathbf{w} = \mathbf{e}_i$ and $\mathbf{f} = \mathbf{e}_j$ as the RX combining and TX precoding vectors, respectively, having zeros except for their i th and j th positions, respectively, which contain ones. In fact, only one pair of TX and RX antennas is activated at each instance t . These training BF vectors can be efficiently realized with any available HBF architecture [14] by including switches to the analog phase shifters. We note that in conventional estimation of sparse mmWave channels with analog BF training vectors implemented via antenna switches [4], $\mathbf{w} \in \{0, 1\}^{N_R}$ and $\mathbf{f} \in \{0, 1\}^{N_T}$ are used at each training instance t . Then, all T postprocessed received signals are used for designing \mathbf{H}_Ω s estimations. This training procedure is inherently different from the aforementioned proposed one for estimating \mathbf{H}_Ω .

IV. SIMULATION RESULTS AND DISCUSSION

We consider the examples of 32×32 and 64×64 MIMO systems equipped with ULAs at both TX and RX sides and operating over a 90 GHz mmWave channel. The azimuth AoA and departure have been generated from the Laplace distribution with standard deviation 50° . As benchmark CSI estimation techniques we have considered:

1) OMP [4];

- 2) vector approximate message passing (VAMP) [12];
- 3) SVT [13]; and
- 4) Two-Stage estimation exploiting both Sparsity and low Rankness (TSSR) [9].

Note that OMP and VAMP exploit only the sparsity of the channel matrix, while SVT capitalizes only on its low-rank property. TSSR exploits both properties by first employing the SVT operator to recover the channel matrix, and then uses it as input to VAMP. The maximum number of iterations for SVT, VAMP, and the proposed algorithm was set to $I_{\max} = 100$, which from our experiments was verified as adequate for their convergence. For SVT, we used $\tau = \rho \|\mathbf{H}_\Omega\|$ with $\rho = \frac{3M}{(N_R N_T)}$, for OMP and VAMP algorithms the parameter for channel sparsity was set to N_p , and we have considered $\tau_H = \rho \|\mathbf{H}_\Omega\|$ with $\rho = 0.005$ and $\tau_S = \frac{0.1}{(1-10 \log(\sigma_n^2))}$ for the proposed Algorithm 1.

We compare the considered CSI estimation techniques both in terms of the normalized MSE (NMSE) performance and achievable spectral efficiency (ASE) in bits/second/hertz. For the SVT and proposed techniques that are based on matrix completion, the speed of convergence was also investigated for different values of the training instances T . Denoting by $\hat{\mathbf{H}}$ the estimation for the true channel \mathbf{H} with any of the considered techniques, NMSE was numerically evaluated as follows:

$$\text{NMSE} \triangleq \mathcal{E}\{10 \log_{10} \|\hat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2\}. \quad (21)$$

Note that for I_{\max} algorithmic iterations for the proposed technique, $\hat{\mathbf{H}}$ is given by $\mathbf{H}^{(I_{\max})}$. In addition, we have computed the following lower bound for ASE [22], [23]:

$$\text{ASE} \triangleq \mathcal{E}\{\log_2 \det(\mathbf{I}_{N_R} + (N_T N_R (\sigma_n^2 + \text{NMSE}))^{-1} \mathbf{H} \mathbf{H}^H)\}. \quad (22)$$

For both latter expressions, the expectations were obtained from averaging 100 independent Monte Carlo realizations.

It is demonstrated that the proposed technique outperforms OMP, VAMP, and TSSR in terms of the NMSE performance for low training lengths T . As shown, the OMP performance is not improved over the training length T or signal-to-noise ratio (SNR) due to the discretization error of the AoA. Also, VAMP is incapable of recovering the 64×64 MIMO channel matrix for small numbers (< 800) of training symbols. This happens because VAMP is based on the calculation of the statistical information of the sparse signal, which cannot be captured for small T . However, for $T \geq 800$ and low-to-medium transmit SNR ($10 \log_{10}(P_t/\sigma_n^2)$ in decibel) values (< 16 dB), VAMP provides improved NMSE compared to the proposed algorithm. This behavior is due to the different training procedures between these two techniques. Specifically, the proposed training lacks of array gain, and hence, \mathbf{H}_Ω estimation will be in general more noisy than channel estimation with VAMP. Nevertheless, this noisy estimation becomes less impactful and less severe as transmit SNR increases. It is also evident in Fig. 1 that TSSR, which is based on successive application of SVT and VAMP algorithms, cannot recover the channel for small T values. This indicates that the independent treatment of each stage does not permit the joint exploitation of the channel sparsity and low-rank properties.

The fast convergence of the proposed algorithm, even for very small training lengths T , is illustrated in Fig. 2(a), and compared with that of the standard SVT technique. As shown, the proposed algorithm converges to smaller NMSE performances for all T values. In Fig. 2(b), we depict the impact of the N_p

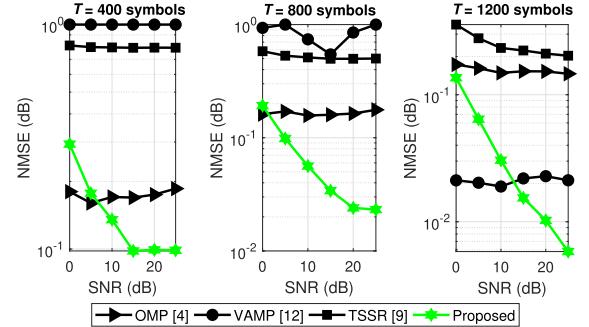


Fig. 1. NMSE w.r.t. transmit SNR for a 64×64 MIMO channel with $N_p = 2$ and different T values.

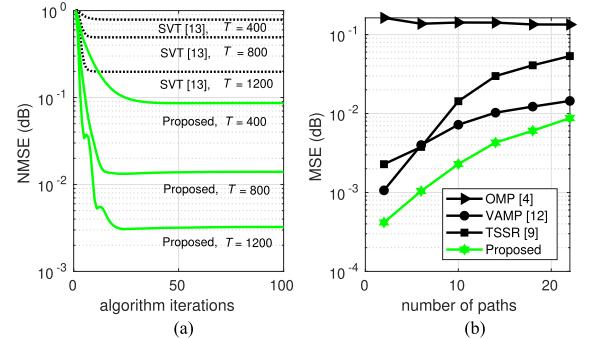


Fig. 2. NMSE for a 64×64 MIMO channel and 30 dB transmit SNR w.r.t. (a) algorithmic iterations and different T . (b) N_p for $T = 2000$.

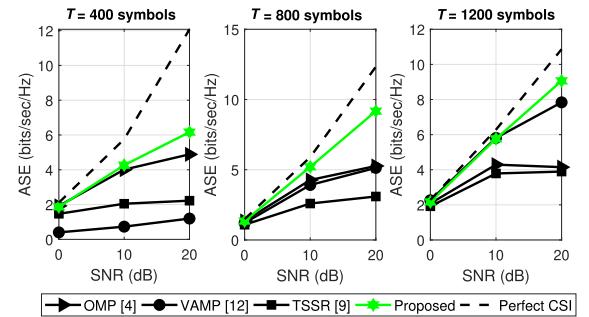


Fig. 3. ASE w.r.t. transmit SNR for a 32×32 MIMO channel with $N_p = 2$ and different T values.

value (i.e., the number of mmWave channel propagation paths) to NMSE of all considered estimation techniques. As expected, CSI estimation performance gets worse as N_p increases; however, the proposed technique outperforms all others providing good NMSE for quite large N_p values. We finally present the ASE performance of all techniques for $N_p = 2$ in Fig. 3. In this figure, ASE with perfect CSI is also sketched. Clearly, the proposed technique outperforms all others for all considered T values. Interestingly, the higher NMSE of the proposed technique for low-to-mid SNR values and $T = 1200$ shown in Fig. 1 does not affect its superiority in terms of ASE.

Putting all the above together, the proposed matrix completion estimation of mmWave massive MIMO channels leveraging jointly their sparsity and low-rank properties outperforms the state-of-the-art techniques requiring short training periods. Our performance results showed that the convergence of the proposed ADMM-based iterative approach is relatively fast.

REFERENCES

- [1] R. W. Heath, Jr., N. González-Prelcic, 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, pp. 436–453, Apr. 2016.
- [2] K. Venugopal, A. Alkhateeb, R. W. Heath, Jr., and N. González-Prelcic, “Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 6493–6497.
- [3] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [4] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [5] J. Lee, G. T. Gil, and 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.
- [6] G. C. Alexandropoulos and S. Chouvardas, “Low complexity channel estimation for millimeter wave systems with hybrid A/D antenna processing,” in *Proc. IEEE GLOBECOM*, Washington D.C., USA, Dec. 2016, pp. 1–6.
- [7] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [8] T. T. Cai and L. Wang, “Orthogonal matching pursuit for sparse signal recovery with noise,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
- [9] X. Li, J. Fang, H. Li, and P. Wang, “Millimeter wave channel estimation via exploiting joint sparse and low-rank structures,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.
- [10] J. Lu, G. Liang, J. Sun, and J. Bi, “A sparse interactive model for matrix completion with side information,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 4071–4079, 2016.
- [11] S. Boyd, N. Parikh, E. C. B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [12] P. Schniter, S. Rangan, and A. K. Fletcher, “Vector approximate message passing for the generalized linear model,” in *Proc. Asilomar CSSC*, Pacific Grove, CA, USA, Nov. 2016, pp. 1525–1529.
- [13] J. F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. Opt.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [14] A. F. Molisch *et al.*, “Hybrid beamforming for massive MIMO: A survey,” *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.
- [15] A. Forenza, D. J. Love, and R. W. Heath, Jr., “Simplified spatial correlation models for clustered MIMO channels with different array configurations,” *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, Jul. 2007.
- [16] A. M. Sayeed, “Deconstructing multiantenna fading channels,” *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [17] J. Brady, N. Behdad, and A. M. Sayeed, “Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Oct. 2013.
- [18] K.-Y. Chiang, C.-J. Hsieh, and I. S. Dhillon, “Matrix completion with noisy side information,” *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 3447–3455, 2015.
- [19] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *J. Roy. Statist. Soc., B*, vol. 58, no. 1, pp. 267–288, 1996.
- [20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.
- [21] R. M. Larsen, “Lanczos bidiagonalization with partial reorthogonalization,” *DAIMI Rep. Series*, vol. 27, no. 537, pp. 1–61, 1998.
- [22] T. Yoo and A. Goldsmith, “Capacity and power allocation for fading MIMO channels with channel estimation error,” *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [23] L. Berriche, K. Abed-Meraim, and J. C. Belfiore, “Investigation of the channel estimation error on MIMO system performance,” in *Proc. Eur. Sig. Proces. Conf.*, Antalya, Turkey, Sep. 2005, pp. 1–4.