Esfandiari, Majdoddin; Vorobyov, Sergiy A.; Heath, Robert W.

ADMM-Based Solution for mmWave UL Channel Estimation with One-Bit ADCs via Sparsity Enforcing and Toeplitz Matrix Reconstruction

Published in:
ICC 2023 - IEEE International Conference on Communications

DOI:
10.1109/ICC45041.2023.10279217

Published: 01/01/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

ADMM-based Solution for mmWave UL Channel Estimation with One-bit ADCs via Sparsity Enforcing and Toeplitz Matrix Reconstruction

Majdoddin Esfandiari[*], *Sergiy A. Vorobyov*[*], *and* *Robert W. Heath Jr.*[†]

[*]Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland
[†]Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, USA

ABSTRACT

Low-power millimeter wave (mmWave) multi-input multi-output communication systems can be enabled with the use of one-bit analog-to-digital converters. Owing to the extreme quantization, conventional signal processing tasks such as channel estimation are challenging, making uplink (UL) multiuser receivers difficult to implement. To address this issue, we first reformulate the UL channel estimation problem, and then combine the idea of ℓ_1 regularized logistic regression classification and Toeplitz matrix reconstruction in a properly designed optimization problem. Our new method is referred to as ℓ_1 regularized logistic regression with Toeplitz matrix reconstruction (L1-RLR-TMR). In addition, we develop a computationally efficient alternating direction method of multipliers (ADMM)-based implementation for the L1-RLR-TMR method. Numerical results demonstrate the performance of the L1-RLR-TMR method in comparison with other existing methods.

Index Terms— Multi-user MIMO, uplink channel estimation, one-bit ADC, angular domain, Toeplitz matrix reconstruction, ℓ_1 regularized logistic regression.

1. INTRODUCTION

Millimeter-wave (mmWave) multiple-input multiple-output (MIMO) communication is an approach to enhance the high data rate benefits provided by mmWave while enjoying the advantages of multi-user multiplexing offered by massive MIMO [1–4]. One approach to reduce transceiver power consumption is to use low-resolution ADCs/DACs (1-3 bits) [6, 8–10]. With one-bit data converters, the power consumption can be reduced to a minimum. Therefore, the methods developed for the case of high resolution converters for tasks like channel estimation or symbol detection have poor performance in the case of one-bit converters utilization which emphasize the need for designing methods consistent with one-bit converters setups.

Most one-bit receiver designs require channel estimation [7, 9, 11–13]. An near maximum-likelihood (nML) was proposed in [11], but its performance was not good at high signal-to-noise ratio (SNR). In [12], an algorithm known as Bussgang linear minimum mean squared error (BLMMSE) has been proposed that can handle the estimation of both flat and frequency-selective channels. BLMMSE approximates the nonlinear one-bit quantizer as a linear function by employing the Bussgang decomposition [16]. A generalized approximate message passing (GAMP) based algorithm has been presented in [9], in which the notion of compressive sensing (CS) is used to estimate the angular domain parameters of UL channels. Using ML direction-of-arrival estimation method together with restoring the lost amplitudes of one-bit measurements, [13] describes a method called the amplitude retrieval (AR). In [14], the problem of UL channel estimation is formulated as a binary classification task, in which the conventional support vector machine (SVM) and its modified version are adopted to handle the estimation of the uncorrelated and spatially correlated channels, respectively. Recently, a method known as the sparsity enforcing with Toeplitz matrix reconstruction (SE-TMR) has been presented in [15] where the combination of the angular domain sparsity and Toeplitz matrix reconstruction is leveraged to derive the SE-TMR estimator. The purpose of our work is to develop an UL channel estimation that not only has a good performance compared to other existing methods, but also requires a low computational complexity to implement.

Inspired by the use of the classification-based methods such as SVM, for instance in [14], as well as the promising underlying Toeplitz structure exploited in the SE-TMR method [15], we develop a computationally efficient UL channel estimation for one-bit ADCs. This new method for UL channel estimation for narrowband mmWave MIMO communications when the BS deploys one-bit ADCs is called ℓ_1 regularized logistic regression with Toeplitz matrix reconstruction (L1-RLR-TMR). A computationally efficient alternating direction method of multipliers (ADMM)-based implementation of the L1-RLR-TMR method is also developed. Numerical simulations are included to showcase the efficiency of the L1-RLR-TMR method compared to existing competitive methods.

Notations: Upper-case and lower-case bold-face letters denote matrices and vectors, respectively, while scalars are denoted by lower-case letters. The transpose, and Hermitian transpose are denoted by $\{\cdot\}^T$ and $\{\cdot\}^H$, respectively, while $\|\cdot\|_2$ and $\|\cdot\|_1$ are the ℓ_2 and ℓ_1 norms of a vector, and

$\|\cdot\|_F$ stands for the Frobenius norm of a matrix. The notation $\mathcal{T}(\cdot)$ stands for an operation of building a square Hermitian Toeplitz matrix with its first column being the bracketed vector. The $n\times n$ identity matrix is denoted by $\mathbf{I}_n$. The $n\times 1$ vector with all its entries equal to one is denoted by $\mathbf{1}_n$. The ith entry of the vector $\boldsymbol{\pi}$ is denoted by $[\boldsymbol{\pi}]_i$, while the entry in the intersection of the ith row and jth column of the matrix $\boldsymbol{\Pi}$ is denoted as $[\boldsymbol{\Pi}]_{i,j}$. The operator $\mathrm{vec}\{\cdot\}$ stacks the columns of a matrix into a long vector, while $\mathrm{unvec}\{\cdot\}$ forms a matrix by splitting the argument and putting them in the columns of that matrix. The operator $\mathrm{diag}\{\boldsymbol{\pi}\}$ generates a diagonal matrix by plugging the entries of the vector $\boldsymbol{\pi}$ into its main diagonal, while $\mathrm{blkdiag}\{\boldsymbol{\Pi}_1,\ldots,\boldsymbol{\Pi}_n\}$ generates a block-diagonal matrix using the bracketed matrices. The notation $\boldsymbol{\Pi}\succeq 0$ means that $\boldsymbol{\Pi}$ is Hermitian positive semidefinite. The least non-negative remainder in the division of a by b is denoted by $\mathrm{rem}(a,b)$. Finally, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ return the real and imaginary parts of the bracketed argument, respectively.

2. UL CHANNEL ESTIMATION

2.1. System Model

Consider an UL multi-user mmWave MIMO system where the BS deploys a uniform linear array (ULA) with M antenna elements that serves K single antenna users.[1] Each antenna of the BS array is connected to two one-bit ADCs for converting the real and imaginary parts of the received signals, while all users are equipped with high-resolution DACs. The UL channel between user k and the BS is given by

$$\begin{aligned}\mathbf{h}_k &= \sum_{l=1}^{L_k}\sum_{m=1}^{M_{\text{path}}^{k,l}} \gamma_{k,l,m}\mathbf{a}(\theta_{k,l,m}) \\ &= [\mathbf{A}(\boldsymbol{\theta}_{k,1}),\cdots,\mathbf{A}(\boldsymbol{\theta}_{k,L_k})]\begin{bmatrix}\boldsymbol{\gamma}_{k,1}\\ \vdots \\ \boldsymbol{\gamma}_{k,L_k}\end{bmatrix} = \mathbf{A}(\boldsymbol{\theta}_k)\boldsymbol{\gamma}_k \quad (1)\end{aligned}$$

where L_k indicates the number of multipath clusters between the BS and the user k, the lth cluster consists of $M_{\text{path}}^{k,l}$ paths densely grouped around a mean direction with the corresponding angle spread [19], $\gamma_{k,l,m}$ and $\theta_{k,l,m}$ are respectively the gain and DOA of the mth path in the lth cluster, the steering vector is given by $\mathbf{a}(\theta_{k,l,m}) = [1, e^{-j\pi\sin(\theta_{k,l,m})},\cdots,e^{-j(M-1)\pi\sin(\theta_{k,l,m})}]^T \in \mathbb{C}^{M\times 1}$, $\boldsymbol{\theta}_{k,l} \triangleq [\theta_{k,l,1},\cdots,\theta_{k,l,M_{\text{path}}^{k,l}}]^T \in \mathbb{R}^{M_{\text{path}}^{k,l}\times 1}$ for $l=1,\cdots,L_k$, $\mathbf{A}(\boldsymbol{\theta}_{k,l}) \triangleq [\mathbf{a}(\theta_{k,l,1}),\cdots,\mathbf{a}(\theta_{k,l,M_{\text{path}}^{k,l}})] \in \mathbb{C}^{M\times M_{\text{path}}^{k,l}}$, $\boldsymbol{\gamma}_{k,l} \triangleq [\gamma_{k,l,1},\cdots,\gamma_{k,l,M_{\text{path}}^{k,l}}]^T \in \mathbb{C}^{M_{\text{path}}^{k,l}\times 1}$, $\boldsymbol{\theta}_k \triangleq [\boldsymbol{\theta}_{k,1}^T,\cdots,\boldsymbol{\theta}_{k,L_k}^T]^T$, $\mathbf{A}(\boldsymbol{\theta}_k) \triangleq [\mathbf{A}(\boldsymbol{\theta}_{k,1}),\cdots,\mathbf{A}(\boldsymbol{\theta}_{k,L_k})]$, and $\boldsymbol{\gamma}_k \triangleq [\boldsymbol{\gamma}_{k,1}^T,\cdots,\boldsymbol{\gamma}_{k,L_k}^T]^T$. Accordingly, the channel between the BS and K users is given as

$$\mathbf{H} = [\mathbf{h}_1,\cdots,\mathbf{h}_K] = [\mathbf{A}(\boldsymbol{\theta}_1)\boldsymbol{\gamma}_1,\cdots,\mathbf{A}(\boldsymbol{\theta}_K)\boldsymbol{\gamma}_K]\,. \quad (2)$$

The users transmit a pilot sequence of length N_s ($N_s \geq K$) during the training stage, and the received signal at the BS is

$$\mathbf{Y} = \mathcal{Q}(\mathbf{HS}+\mathbf{N}) \quad (3)$$

where $\mathcal{Q}(\cdot) \triangleq \mathrm{sign}(\Re\{\cdot\})+j\mathrm{sign}(\Im\{\cdot\})$ denotes the element-wise one-bit quantizer which only preserves the sign of arguments and its output, i.e., takes values from the set $\mathcal{S} = \{1+j, 1-j, -1+j, -1-j\}$, $\mathbf{S}\in\mathbb{C}^{K\times N_s}$ represents the orthogonal pilot matrix, and $\mathbf{N}\in\mathbb{C}^{M\times N_s}$ is the additive circularly symmetric complex Gaussian noise with zero mean and variance σ^2. The task is to restore (scaled) $\mathbf{H}\in\mathbb{C}^{M\times K}$ from the received signal $\mathbf{Y}\in\mathbb{C}^{M\times N_s}$.

2.2. Logistic Regression for Binary Classification

Consider a binary classification task where the goal is to separate a training data set $\mathcal{D} = \{(\mathbf{x}_p, y_p)\}_{p=1}^P$ with $\mathbf{x}_p$ and $y_p \in \{\pm 1\}$ being a feature and a corresponding binary label, respectively. The logistic regression classifier separates the data space into two regions via finding vectors $\mathbf{w}$ and $\mathbf{b}$ that minimize

$$\min_{\mathbf{w},\mathbf{b}} \sum_{p=1}^{P} \log\left(1+e^{-y_p(\mathbf{w}^T\mathbf{x}_p)}\right) \quad (4)$$

where $\mathbf{w}$ and $\mathbf{b}$ are referred to as the weight vector and the bias, respectively [17, 18].

2.3. Proposed UL Channel Estimation

Because of the clustered-based representation of $\mathbf{h}_k$ in (1), the channel between the BS and the kth user can be still considered to be sparse with respect to an angular-based dictionary in spite of being composed of many closely located paths. In other words, $\mathbf{h}_k$ can be approximated by a linear combination of a few atoms of a proper angular-based dictionary (e.g., the normalized discrete Fourier transform (DFT)). We approximate each $\mathbf{h}_k$ as a linear combination of the corresponding L_k basis paths with L_k DOAs and path gains.[2] Hence, (2) can be reformulated as

$$\mathbf{H} = [\mathbf{h}_1,\cdots,\mathbf{h}_K] = [\mathbf{A}(\bar{\boldsymbol{\theta}}_1)\bar{\boldsymbol{\gamma}}_1,\cdots,\mathbf{A}(\bar{\boldsymbol{\theta}}_K)\bar{\boldsymbol{\gamma}}_K] \quad (5)$$

where $\bar{\boldsymbol{\theta}}_k \triangleq [\bar{\theta}_{k,1},\cdots,\bar{\theta}_{k,L_k}]^T \in \mathbb{R}^{L_k\times 1}$ and $\bar{\boldsymbol{\gamma}}_k \triangleq [\bar{\gamma}_{k,1},\cdots,\bar{\gamma}_{k,L_k}]^T \in \mathbb{C}^{L_k\times 1}$ denote respectively the DOAs and path gains associating with L_k basis paths between the BS and user k. Manipulating (5), we obtain

$$\mathbf{H} = \mathbf{A}\boldsymbol{\Gamma}\bar{\mathbf{G}} = \bar{\mathbf{H}}\bar{\mathbf{G}} \quad (6)$$

[1] The assumption of single antenna users is made for simplicity, while a more practical configuration is to consider users with multiple antennas.

[2] Note that increasing the number of atoms for approximating each $\mathbf{h}_k$ leads to a more accurate modeling indeed, however, it also increases the computational complexity.

where $\mathbf{A} \triangleq [\mathbf{A}(\bar{\boldsymbol{\theta}}_1), \cdots, \mathbf{A}(\bar{\boldsymbol{\theta}}_K)] \in \mathbb{C}^{M\times L}$, $\boldsymbol{\Gamma} \triangleq \text{diag}\{[\bar{\boldsymbol{\gamma}}_1^T, \bar{\boldsymbol{\gamma}}_2^T, \ldots, \bar{\boldsymbol{\gamma}}_K^T]^T\} \in \mathbb{C}^{L\times L}$, $\bar{\mathbf{G}} \triangleq \text{blkdiag}\{\mathbf{1}_{L_1}, \mathbf{1}_{L_2}, \ldots, \mathbf{1}_{L_K}\} \in \mathbb{R}^{L\times K}$, $\bar{\mathbf{H}} \triangleq \mathbf{A}\boldsymbol{\Gamma} \in \mathbb{C}^{M\times L}$, and $L \triangleq \sum_{k=1}^{K} L_k$. Thanks to $\mathbf{A}$ and $\boldsymbol{\Gamma}$ being respectively Vandermonde and diagonal matrices, we have

$$\bar{\mathbf{H}}\bar{\mathbf{H}}^H = \mathcal{T}(\mathbf{u}) \tag{7}$$

where $\mathbf{u} \in \mathbb{C}^M$ and $[\mathbf{u}]_1$ is a real number according to (7). On the other hand, with the DOA-based structure of $\mathbf{H}$ through $\mathbf{A}$, $\mathbf{H}$ can be multiplied by a proper angular-based dictionary matrix like the normalized DFT matrix to make each column of the resultant sparse, i.e.,

$$\mathbf{X}(\bar{\boldsymbol{\theta}}) = \mathbf{F}\mathbf{H} = \mathbf{F}\bar{\mathbf{H}}\bar{\mathbf{G}} \tag{8}$$

where $\mathbf{F} \in \mathbb{C}^{M\times M}$ denotes the normalized DFT matrix.

Applying the vectorization operator to (3) together with the use of (6), we have

$$\mathbf{y} \triangleq \text{vec}\{\mathbf{Y}\} = \mathcal{Q}\Big(\left((\bar{\mathbf{G}}\mathbf{S})^T \otimes \mathbf{I}_M\right) \bar{\mathbf{h}} + \mathbf{n}\Big) \tag{9}$$

where $\bar{\mathbf{h}} \triangleq \text{vec}\{\bar{\mathbf{H}}\}$ and $\mathbf{n} \triangleq \text{vec}\{\mathbf{N}\}$. For convenience in later derivations, the notation in (9) is converted to the real domain as

$$\mathbf{y}_R \triangleq [\Re\{\mathbf{y}\}^T, \Im\{\mathbf{y}\}^T]^T = \bar{\mathbf{S}}\bar{\mathbf{h}}_R \tag{10}$$

where $\bar{\mathbf{H}} \triangleq \bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I = [\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2, \ldots, \bar{\mathbf{h}}_M]^T$, $\bar{\mathbf{h}}_R \triangleq [\text{vec}\{\bar{\mathbf{H}}_R\}^T, \text{vec}\{\bar{\mathbf{H}}_I\}^T]^T$, and

$$\begin{aligned}\bar{\mathbf{S}} &\triangleq \begin{bmatrix} \Re\{(\bar{\mathbf{G}}\mathbf{S})^T \otimes \mathbf{I}_M\} & -\Im\{(\bar{\mathbf{G}}\mathbf{S})^T \otimes \mathbf{I}_M\} \\ \Im\{(\bar{\mathbf{G}}\mathbf{S})^T \otimes \mathbf{I}_M\} & \Re\{(\bar{\mathbf{G}}\mathbf{S})^T \otimes \mathbf{I}_M\} \end{bmatrix} \\ &= [\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \ldots, \bar{\mathbf{s}}_{2MN_s}]^T. \end{aligned} \tag{11}$$

Note that $\mathbf{y}_R \in \{\pm 1\}^{2MN_s\times 1}$ and $\bar{\mathbf{h}}_R \in \mathbb{R}^{2ML\times 1}$. In addition, $\bar{\mathbf{h}}_m^T \in \mathbb{C}^{1\times L}$ with $m \in \{1, \ldots, M\}$ and $\bar{\mathbf{s}}_t^T \in \mathbb{R}^{1\times 2ML}$ with $t \in \{1, \ldots, 2MN_s\}$ denote the mth and tth rows of $\bar{\mathbf{H}}$ and $\bar{\mathbf{S}}$, respectively. Using (4), (7)-(8), and (10)-(11), the minimization problem associated with the L1-RLR-TMR method can be written as

$$\begin{aligned} \min_{\bar{\mathbf{h}}_R, \mathbf{u}} & \|\bar{\mathbf{F}}\bar{\mathbf{h}}_R\|_1 + \lambda \sum_{t=1}^{2MN_s} \log\left(1 + e^{-\kappa[\mathbf{y}_R]_t(\bar{\mathbf{s}}_t^T\bar{\mathbf{h}}_R)}\right) \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{I}_L & (\bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I)^H \\ \bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I & \mathcal{T}(\mathbf{u}) \end{bmatrix} \succeq 0 \\ & \|\bar{\mathbf{h}}_m\|_2^2 = c\,, \;\; m = 1, \ldots, M \end{aligned} \tag{12}$$

where $\bar{\mathbf{F}} \triangleq \begin{bmatrix} \Re\{\bar{\mathbf{G}}^T \otimes \mathbf{F}\} & -\Im\{\bar{\mathbf{G}}^T \otimes \mathbf{F}\} \\ \Im\{\bar{\mathbf{G}}^T \otimes \mathbf{F}\} & \Re\{\bar{\mathbf{G}}^T \otimes \mathbf{F}\} \end{bmatrix}$, $\lambda > 0$ is a regularization parameter, and $\kappa > 1$ is added for accelerating the convergence of the ADMM-based implementation of the L1-RLR-TMR method in the next Subsection. Note that the two constrains of (12) are added to impose (7), where c is a tunable parameter. We suggest to set $c = 1$. Moreover, we set $\kappa = 10$. The minimization problem presented in (12) is non-convex because of the equality constraints.

2.4. The ADMM Implementation

To develop the ADMM-based method, an auxiliary variable $\mathbf{Z}$ is first introduced in order to modify (12) as

$$\begin{aligned} \min_{\bar{\mathbf{h}}_R, \mathbf{u}} & \|\bar{\mathbf{F}}\bar{\mathbf{h}}_R\|_1 + \lambda \sum_{t=1}^{2MN_s} \log\left(1 + e^{-\kappa[\mathbf{y}_R]_t(\bar{\mathbf{s}}_t^T\bar{\mathbf{h}}_R)}\right) \\ \text{s.t.} \quad & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_L & (\bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I)^H \\ \bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I & \mathcal{T}(\mathbf{u}) \end{bmatrix} \\ & \mathbf{Z} \succeq 0 \\ & \|\bar{\mathbf{h}}_m\|_2^2 = c\,, \;\; m = 1, \ldots, M \end{aligned} \tag{13}$$

Consequently, the scaled augmented Lagrangian of (13) is expressed as

$$\begin{aligned} &\mathcal{L}_\rho(\bar{\mathbf{h}}_R, \mathbf{u}, \mathbf{Z}, \boldsymbol{\Lambda}) = \|\bar{\mathbf{F}}\bar{\mathbf{h}}_R\|_1 \\ &+\lambda \sum_{t=1}^{2MN_s} \log\left(1 + e^{-\kappa[\mathbf{y}_R]_t(\bar{\mathbf{s}}_t^T\bar{\mathbf{h}}_R)}\right) \\ &+\frac{\rho}{2}\left\|\mathbf{Z} - \begin{bmatrix} \mathbf{I}_L & (\bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I)^H \\ \bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I & \mathcal{T}(\mathbf{u}) \end{bmatrix} + \boldsymbol{\Lambda}\right\|_F^2 \end{aligned} \tag{14}$$

where $\rho > 0$ is a penalty parameter, and $\boldsymbol{\Lambda}$ is the dual variable. For convenience, $\mathbf{Z}$ and $\boldsymbol{\Lambda}$ are partitioned as $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_0 & (\mathbf{Z}_R + j\mathbf{Z}_I)^H \\ \mathbf{Z}_R + j\mathbf{Z}_I & \mathbf{Z}_1 \end{bmatrix}$, and $\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_0 & (\boldsymbol{\Lambda}_R + j\boldsymbol{\Lambda}_I)^H \\ \boldsymbol{\Lambda}_R + j\boldsymbol{\Lambda}_I & \boldsymbol{\Lambda}_1 \end{bmatrix}$. Therefore, the updating rules of the ADMM for solving (12) are

$$\left(\bar{\mathbf{h}}_R^{l+1}, \mathbf{u}^{l+1}\right) = \arg \min_{\bar{\mathbf{h}}_R, \mathbf{u}} \mathcal{L}_\rho(\bar{\mathbf{h}}_R, \mathbf{u}, \mathbf{Z}^l, \boldsymbol{\Lambda}^l)$$

$$\text{s.t.} \quad \|\bar{\mathbf{h}}_m\|_2^2 = c\,, \;\; m = 1, \ldots, M \tag{15}$$

$$\mathbf{Z}^{l+1} = \left[\mathbf{D}^{l+1} - \boldsymbol{\Lambda}^l\right]_+ \tag{16}$$

$$\boldsymbol{\Lambda}^{l+1} = \boldsymbol{\Lambda}^l + \mathbf{Z}^{l+1} - \mathbf{D}^{l+1} \tag{17}$$

where $\mathbf{D}^{l+1} \triangleq \begin{bmatrix} \mathbf{I}_L & (\bar{\mathbf{H}}_R^{l+1} + j\bar{\mathbf{H}}_I^{l+1})^H \\ \bar{\mathbf{H}}_R^{l+1} + j\bar{\mathbf{H}}_I^{l+1} & \mathcal{T}(\mathbf{u}^{l+1}) \end{bmatrix}$, and the notation $(\cdot)^l$ represents the estimates at lth iteration. In addition, $[\cdot]_+$ is a projection function used in (16) to project the argument onto the positive semidefinite cone via carrying out the eigenvalue decomposition and setting all the negative eigenvalues to zero. As it is presented in (15), a minimization problem has to be solved to estimates $\bar{\mathbf{h}}_R^{l+1}$ and $\mathbf{u}^{l+1}$. (15) is convex with respect to $\mathbf{u}$ with the closed-form solution as

$$\mathbf{u}^{l+1} = \mathbf{W}\left(\mathcal{T}^*(\mathbf{Z}_1^l + \boldsymbol{\Lambda}_1^l)\right) \tag{18}$$

where $\mathbf{W} \triangleq \text{diag}\{[\frac{1}{M}, \frac{1}{2(M-1)}, \ldots, \frac{1}{2}]^T\}$, and $\mathcal{T}^*(\cdot)$ denotes the Toeplitz adjoint operator. However, estimating $\bar{\mathbf{h}}_R^{l+1}$ can be handled using an inner ADMM formulation for solving

(15) with respect to $\bar{\mathbf{h}}_R$. In doing so, an auxiliary variable $\mathbf{w}$ is introduced to modify (15) as

$$\begin{aligned}\bar{\mathbf{h}}_R^{l+1} = \arg\min_{\bar{\mathbf{h}}_R} \|\mathbf{w}\|_1 + \lambda \sum_{t=1}^{2MN_s} \log\Big(1+e^{-\kappa[\mathbf{y}_R]_t(\bar{\mathbf{s}}_t^T\bar{\mathbf{h}}_R)}\Big) \\ + \frac{\rho}{2}\Big\|\mathbf{Z}^l - \begin{bmatrix}\mathbf{I}_L & (\bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I)^H \\ \bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I & \mathcal{T}(\mathbf{u}^l)\end{bmatrix} + \mathbf{\Lambda}^l\Big\|_F^2 \\ \text{s.t.} \quad \|\bar{\mathbf{h}}_m\|_2^2 = c\,, \;\; m=1,\ldots,M \\ \bar{\mathbf{F}}\bar{\mathbf{h}}_R - \mathbf{w} = \mathbf{0} \end{aligned} \tag{19}$$

Hence, its scaled augmented Lagrangian of (19) is given as

$$\begin{aligned}\mathcal{L}_{\bar{\rho}}(\bar{\mathbf{h}}_R, \mathbf{w}, \mathbf{v}) = \|\mathbf{w}\|_1 + \lambda \sum_{t=1}^{2MN_s} \log\Big(1+e^{-\kappa[\mathbf{y}_R]_t(\bar{\mathbf{s}}_t^T\bar{\mathbf{h}}_R)}\Big) \\ + \frac{\rho}{2}\Big\|\mathbf{Z}^l - \begin{bmatrix}\mathbf{I}_L & (\bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I)^H \\ \bar{\mathbf{H}}_R + j\bar{\mathbf{H}}_I & \mathcal{T}(\mathbf{u}^l)\end{bmatrix} + \mathbf{\Lambda}^l\Big\|_F^2 \\ + \frac{\bar{\rho}}{2}\|\bar{\mathbf{F}}\bar{\mathbf{h}}_R - \mathbf{w} + \mathbf{v}\|_2^2 \end{aligned} \tag{20}$$

where $\bar{\rho} > 0$ is a penalty parameter associated with the inner ADMM, and $\mathbf{v}$ is the dual variable. In each outer ADMM iteration l, the inner ADMM update rules are

$$\begin{aligned}\bar{\mathbf{h}}_R^+ &= \arg\min_{\bar{\mathbf{h}}_R} \mathcal{L}_{\bar{\rho}}(\bar{\mathbf{h}}_R, \mathbf{w}^-, \mathbf{v}^-) \\ &\quad \text{s.t.} \quad \|\bar{\mathbf{h}}_m\|_2^2 = c\,, \;\; m=1,\ldots,M \qquad (21) \\ \mathbf{w}^+ &= \mathcal{S}_{1/\bar{\rho}}\Big(\bar{\mathbf{F}}\bar{\mathbf{h}}_R^+ + \mathbf{v}^-\Big) \qquad (22) \\ \mathbf{v}^+ &= \mathbf{v}^- + \bar{\mathbf{F}}\bar{\mathbf{h}}_R^+ - \mathbf{w}^+ \qquad (23)\end{aligned}$$

where $\mathcal{S}_\alpha(\cdot)$ is the soft thresholding operator. Note that the notation relations to l is omitted for simplicity in (21)–(23), and also the notation $(\cdot)^+$ is adopted to show the updates related to the inner ADMM iterations. Finally, we exploit the projected gradient descent (PGD) to handle the minimization problem (21) where the gradient of its objective with respect to $\bar{\mathbf{h}}_R$ is given as

$$\begin{aligned}\bigtriangledown\mathcal{L}_{\bar{\rho}}(\bar{\mathbf{h}}_R) = -\lambda \sum_{t=1}^{2MN_s} \frac{\kappa[\mathbf{y}_R]_t}{1+e^{\kappa[\mathbf{y}_R]_t(\bar{\mathbf{s}}_t^T\bar{\mathbf{h}}_R)}}\bar{\mathbf{s}}_t + 2\rho\Big(\bar{\mathbf{h}}_R - \mathbf{q}^l\Big) \\ + \bar{\rho}\Big(\bar{\mathbf{F}}^T\bar{\mathbf{F}}\bar{\mathbf{h}}_R - \bar{\mathbf{F}}^T\left(\mathbf{w}^- - \mathbf{v}^-\right)\Big)\end{aligned} \tag{24}$$

where $\mathbf{q}^l \triangleq [\text{vec}\{\mathbf{Z}_R^l + \mathbf{\Lambda}_R^l\}^T, \text{vec}\{\mathbf{Z}_I^l + \mathbf{\Lambda}_I^l\}^T]^T$. It is worth mentioning that we initialize the aforementioned PGD by the maximum ratio estimate [11] as

$$\bar{\mathbf{h}}_R^{(0)} = \frac{\tilde{\mathbf{X}}_R^T\mathbf{1}_{2MN}}{\|\tilde{\mathbf{X}}_R^T\mathbf{1}_{2MN}\|_2}\sqrt{Mc} \tag{25}$$

where $\tilde{\mathbf{X}}_R \triangleq \text{diag}\{\mathbf{y}_R\}\bar{\mathbf{S}}$. The steps of the proposed UL channel estimator are outlined in Algorithm 1. The following definitions of the primal and dual variables are used for the ADMM implementation [22] in Algorithm 1:

$$\begin{aligned}\mathbf{x}^{l+1} &\triangleq \Big[\left(\bar{\mathbf{h}}_R^{l+1}\right)^T, \text{vec}\{\Re\{\mathcal{T}(\mathbf{u}^{l+1})\}\}^T, \\ &\quad \text{vec}\{\Im\{\mathcal{T}(\mathbf{u}^{l+1})\}\}^T\Big]^T, \;\; \mathbf{z}^{l+1} \triangleq \Big[\text{vec}\{\mathbf{Z}_R^{l+1}\}^T, \\ &\quad \text{vec}\{\mathbf{Z}_I^{l+1}\}^T, \text{vec}\{\Re\{\mathbf{Z}_1^{l+1}\}\}^T, \text{vec}\{\Im\{\mathbf{Z}_1^{l+1}\}\}^T\Big]^T \\ \mathbf{y}^{l+1} &\triangleq \rho\Big[\text{vec}\{\mathbf{\Lambda}_R^{l+1}\}^T, \text{vec}\{\mathbf{\Lambda}_I^{l+1}\}^T, \text{vec}\{\Re\{\mathbf{\Lambda}_1^{l+1}\}\}^T, \\ &\quad \text{vec}\{\Im\{\mathbf{\Lambda}_1^{l+1}\}\}^T\Big]^T, \\ \epsilon_{\text{pri}}^{l+1} &\triangleq \sqrt{2M(L+M)}\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\max\{\|\mathbf{x}^{l+1}\|_2, \|\mathbf{z}^{l+1}\|_2\} \\ \epsilon_{\text{dual}}^{l+1} &\triangleq \sqrt{2M(L+M)}\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\|\mathbf{y}^{l+1}\|_2\end{aligned}$$

Note that in the steps related to the PGD iterations in Algorithm 1, the notation $\tilde{h}$ is adopted for updating $\bar{h}_R$ to avoid confusion. Moreover, the number of iterations is limited to 1000 for the outer ADMM.

3. SIMULATION RESULTS

The proposed L1-RLR-TMR method is evaluated and compared to other competitive algorithms in this section. The nML [11], AR [13], and SE-TMR [15] methods are used for comparison. The pilot sequence is constructed as a circularly shifted replica of a Zadoff-Chu (ZC) sequence of length N_s where each row is orthogonal to the others, i.e., $\mathbf{S}\mathbf{S}^H = N_s\mathbf{I}_\text{K}$. The SNR and normalized mean square error (NMSE) are respectively defined as $\text{SNR} \triangleq 10\log_{10}\left(\frac{\|\mathbf{H}\mathbf{S}\|_F^2}{MN_s\sigma^2}\right)$ and $\text{NMSE} \triangleq \frac{1}{KN}\sum_{k=1}^K\sum_{n=1}^N\left\|\frac{\hat{\mathbf{h}}_k^{(n)}}{\|\hat{\mathbf{h}}_k^{(n)}\|_2} - \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|_2}\right\|_2^2$, where $\hat{\mathbf{h}}_k^{(n)}$ stands for the kth column of $\hat{\mathbf{H}}$ estimated in the nth Monte Carlo run with $\mathbf{h}_k$ being the actual kth column of $\mathbf{H}$, and N is the total number of Monte Carlo trials considered as $N = 200$ here. We consider $\lambda = 1$ for the SE-TMR and L1-RLR-TMR methods, and $K = 8$. We set the number of channel clusters and within cluster multipaths for all users to be the same, i.e., $L_1 = \cdots = L_K$ and $M_{\text{path}}^{1,1} = \cdots = M_{\text{path}}^{1,L_1} = \cdots = M_{\text{path}}^{K,1} = \cdots = M_{\text{path}}^{1,L_K} = 100$. UL DOAs are generated randomly once and remain the same throughout all Monte Carlo trials, and the channel path gains are distributed as $\mathcal{CN}(0,1)$. Fig. 1 depicts the NMSE of the methods tested versus SNR for the scenario that $M = 16$, $N_s = 128$, $L_k = 1$ for all users, and the angle spread of 8 degrees within each cluster. m,,It shows that the performance of L1-RLR-TMR is comparable to that of the SE-TMR method at high-SNR regime, although the SE-TMR method is implemented using CVX and has high complexity [21]. In Fig. 2, the performance of the methods tested is presented for the setup of $M = 16$, $N_s = 128$, $L_k = 2$, and the within cluster angle spreads are 8 and 10 degrees for all users. The efficiency of L1-RLR-TMR is confirmed at high-SNR regime compared to other methods tested. Particularly, Fig. 2 shows that the

Algorithm 1: ADMM implementation for L1-RLR-TMR

Initialization

1: Set the initial $\bar{\mathbf{h}}_R^{(0)}$ as (25). In addition, set $\lambda = 1$, $\rho = 10$, $\bar{\rho} = 1$, $c = 1$, and $\kappa = 10$.

2: Set $i_{\max} = 10$, the step size $\beta = 0.01$ and the termination threshold $\epsilon = 10^{-3}$ for the PGD. In addition, set the termination thresholds of the outer ADMM as $\epsilon^{\text{abs}} = \epsilon^{\text{rel}} = 10^{-4}$.

3: Set

$\mathbf{Z}_0^0 = \mathbf{I}_M$, $\mathbf{Z}_R^0 = \bar{\mathbf{H}}_R^0 = \text{unvec}\left\{\left[\bar{\mathbf{h}}_R^{(0)}\right]_{1:ML}\right\}$,

$\mathbf{Z}_I^0 = \bar{\mathbf{H}}_I^0 = \text{unvec}\left\{\left[\bar{\mathbf{h}}_R^{(0)}\right]_{ML+1:2ML}\right\}$, $\mathbf{w}^- = \bar{\mathbf{F}}\bar{\mathbf{h}}_R^{(0)}$,

$\mathbf{u}^0 = \mathbf{W}\left(\mathcal{T}^*\left(\bar{\mathbf{H}}_R^0 + j\bar{\mathbf{H}}_I^0\right)\right)$, $\mathbf{Z}_1^0 = \mathcal{T}(\mathbf{u}^0)$, $\bar{\mathbf{h}}_R^- = \bar{\mathbf{h}}_R^{(0)}$.

4: Initialize $\mathbf{\Lambda}^0$ and $\mathbf{v}^-$ as all zero matrix and vector, respectively. In addition, set $l = 0$.

Outer ADMM

5: **while** $e_{\text{pri}}^{l+1} \geq \epsilon_{\text{pri}}^{l+1}$ & $e_{\text{dual}}^{l+1} \geq \epsilon_{\text{dual}}^{l+1}$

Inner ADMM

6: **for** $i = 1 : i_{max}$

7: Reset $k = 1$.

8: **while** $\|\tilde{\mathbf{h}}^{(k)} - \tilde{\mathbf{h}}^{(k-1)}\|_2 \geq \epsilon\|\tilde{\mathbf{h}}^{(k-1)}\|_2$

9: $\tilde{\mathbf{h}}^{(k)} = \tilde{\mathbf{h}}^{(k-1)} - \beta \bigtriangledown \mathcal{L}_{\bar{\rho}}(\tilde{\mathbf{h}}^{(k-1)})$ via (24).

10: $\tilde{\mathbf{H}}_1 \triangleq \text{unvec}\left\{\left[\tilde{\mathbf{h}}^{(k)}\right]_{1:ML}\right\}$.

11: $\tilde{\mathbf{H}}_2 \triangleq \text{unvec}\left\{\left[\tilde{\mathbf{h}}^{(k)}\right]_{ML+1:2ML}\right\}$

12: $\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_1 + j\tilde{\mathbf{H}}_2$.

13: **for** $m = 1 : M$

14: $[\tilde{\mathbf{H}}]_{m,:} = \frac{[\tilde{\mathbf{H}}]_{m,:}}{\|[\tilde{\mathbf{H}}]_{m,:}\|_2}\sqrt{c}$

15: **end for**

16: $\tilde{\mathbf{h}}^{(k)} = [\text{vec}\{\Re\{\tilde{\mathbf{H}}\}\}^T, \text{vec}\{\Im\{\tilde{\mathbf{H}}\}\}^T]^T$

17: **if** rem(k,10) = 0

18: $\beta = \frac{\beta}{2}$

19: **end if**

20: **end while**

21: Set $\bar{\mathbf{h}}_R^+ = \tilde{\mathbf{h}}^{(k)}$. Update $\mathbf{w}^+$ and $\mathbf{v}^+$ via (22) and (23), respectively.

22: **end for**

23: Set $\bar{\mathbf{h}}_R^{l+1} = \bar{\mathbf{h}}_R^+$. Update $\mathbf{u}^{l+1}$, $\mathbf{Z}^{l+1}$, and $\mathbf{\Lambda}^{l+1}$ using (18), (16), and (17), respectively.

24: Construct $\mathbf{x}^{l+1}$, $\mathbf{z}^{l+1}$, $\mathbf{y}^{l+1}$, $\epsilon_{\text{pri}}^{l+1}$, and $\epsilon_{\text{dual}}^{l+1}$. Then, calculate $e_{\text{pri}}^{l+1} = \|\mathbf{x}^{l+1} - \mathbf{z}^{l+1}\|_2$, $e_{\text{dual}}^{l+1} = \|\rho(\mathbf{z}^{l+1} - \mathbf{z}^l)\|_2$

25: **if** $e_{\text{pri}}^{l+1} > 2\, e_{\text{dual}}^{l+1}$

26: $\rho = 2\rho$ and $\mathbf{\Lambda}^{l+1} = \frac{1}{2}\mathbf{\Lambda}^{l+1}$

27: **else if** $e_{\text{dual}}^{l+1} > 2\, e_{\text{pri}}^{l+1}$

28: $\rho = \frac{\rho}{2}$ and $\mathbf{\Lambda}^{l+1} = 2\mathbf{\Lambda}^{l+1}$

29: **end if**

30: **end while**

31: Reshape $\bar{\mathbf{H}}$, and recover $\mathbf{H}$ using (6).

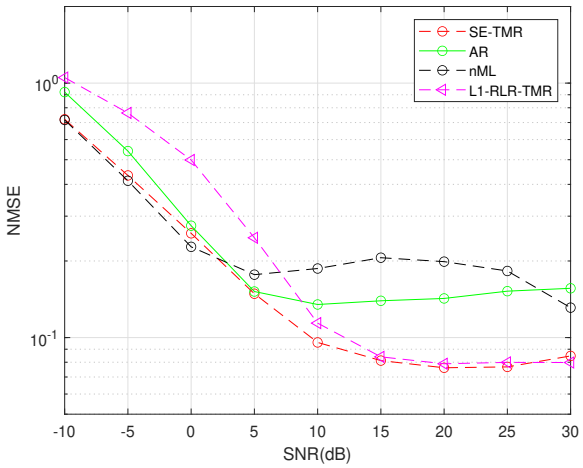

Fig. 1. NMSE vs. SNR for $M = 16$, $N_s = 128$, and $L_k = 1$.

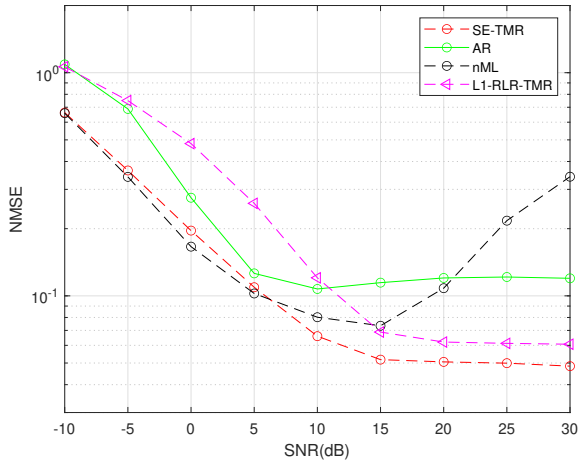

Fig. 2. NMSE vs. SNR for $M = 16$, $N_s = 128$, and $L_k = 2$.

performance of L1-RLR-TMR implemented by the ADMM is comparable with that of the SE-TMR implemented using CVX.

4. CONCLUSION

A novel method called L1-RLR-TMR is proposed for estimating UL channels in mmWave multi-user MIMO communications when the BS uses one-bit ADCs. The main idea of the L1-RLR-TMR method is to recover the UL channel as the solution of an optimization problem designed by considering the ℓ_1 regularized logistic regression classification as well as the notion of Toeplitz matrix reconstruction. To make the use of L1-RLR-TMR method practical, we developed an ADMM-based implementation for it with low computational cost. Numerical results validate the efficiency of the L1-RLR-TMR method compared to other competitive methods.

5. REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Millimeter-wave massive MIMO: The next wireless revolution?," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolio, "Five disruptive technology directions for 5G," *IEEE Comm. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[4] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roth, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[5] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[6] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. Int. ITG Workshop Smart Antennas*, Bremen, Germany, Feb. 2010, pp. 143–150.

[7] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath Jr., "Uplink Performance of Wideband Massive MIMO With One-Bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[8] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[9] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[10] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[11] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[12] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[13] C. Qian, X. Fu, and N. D. Sidiropoulos, "Amplitude retrieval for channel estimation of MIMO systems with one-bit ADCs," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1698–1702, Nov. 2019.

[14] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "SVM-based channel estimation and data detection for one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2086–2099, Mar. 2021.

[15] M. Esfandiari, S. A. Vorobyov, and R. W. Heath, "Sparsity enforcing with Toeplitz matrix reconstruction method for mmWave UL channel estimation with one-bit ADCs," in *Proc. IEEE 12th Sensor Array multichan. Signal Process. Workshop*, Trondheim, Norway, Jun. 2022, pp. 141–145.

[16] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, 1952.

[17] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*. Springer, 2009.

[18] A. Jung, *Machine learning: The basics*. Springer Nature, 2022.

[19] "5G channel model for bands up to 100 GHz (v2.3)," Tech. Rep., 2016. [Online]. Available: http://www.5gworkshops.com/5gcm.html

[20] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial-and-frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[21] M. Grant, S. Boyd, and Y. Ye, CVX: Matlab software for disciplined convex programming. 2008 [Online]. Available: http://stanford.edu/∼boyd/cvx

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via alternating direction method of multipliers," *Found. Trends. Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.