A GRIDLESS CS APPROACH FOR CHANNEL ESTIMATION IN HYBRID MASSIVE MIMO SYSTEMS

Khaled Ardah[1], *André L. F. de Almeida*[2], *and Martin Haardt*[1]

[1] Communications Research Laboratory (CRL), Ilmenau University of Technology, Ilmenau, Germany
[2] Wireless Telecom Research Group (GTEL), Federal University of Ceará, Fortaleza, Brazil
{khaled.ardah, martin.haardt}@tu-ilmenau.de, andre@gtel.ufc.br

ABSTRACT

Channel state information (CSI) estimation in hybrid analog-digital (HAD) millimeter-wave (mmWave) massive MIMO systems is a challenging problem due to the high channel dimension and reduced number of radio-frequency chains. However, exploiting the channel sparsity, several methods have been proposed leveraging the compressed sensing (CS) tools. Most of the prior works consider an approximate CS formulation by assuming that the channel parameters lie perfectly on a finite grid neglecting the grid mismatch effect. To resolve this issue, we propose a gridless CS approach that exploits the antenna array geometry. The proposed algorithm is based on an alternating optimization technique and is guaranteed to converge to a local minimum. Simulation results are provided to evaluate the effectiveness of the proposed algorithm.

Index Terms— Compressed sensing, hybrid analog-digital, CSI estimation, massive MIMO.

1. INTRODUCTION

The combination of mmWave and massive MIMO wireless technologies is seen as key enabler in future wireless networks [1]. However, due to high power and cost requirements, the use of fully-digital (FD) beamforming is very challenging. Recently, HAD beamforming architectures have been proposed to facilitate the practical implementation of massive MIMO systems by dividing the beamforming process between the analog and digital domains to reduce the number of the energy-hungry radio frequency (RF) chains [2]. To realize its advantages, CSI is required at transmitter, which is harder to estimate with HAD systems than with FD counterparts due to the reduced number of RF chains [3].

In massive MIMO systems, classical least square CSI estimation methods are impractical, since the required training overhead becomes overwhelming, due to the high channel dimension that consumes a large amount of communication resources [4]. Therefore, different approaches providing a higher spectral efficiency and low training overhead are required. In practice, it was observed by several measurement campaigns [5, 6] that the massive MIMO channel matrix in mmWave communication has a sparse structure in the angular domain due to the limited number of scatters comparing to the large number of antenna elements. Exploiting this sparse structure, CS tools [7] can be used to estimate the MIMO channel, where the problem can be turned into estimating the parameters of dominant channel paths, namely the angles-of-departures (AoDs), the angles-of-arrivals (AoAs), and the complex paths gains.

In the past few years, several CS-based massive MIMO channel estimation algorithms have been proposed [4, 8–11]. In [8], the authors consider a FD massive MIMO system and propose a distributed CS channel estimation approach while exploiting the channel correlation between adjacent frames and the spatially common sparsity within multiple subchannels. Differently, the algorithms proposed in [4, 9–11] consider HAD massive MIMO systems. However, the solutions proposed in [9–11] all consider an approximate CS formulation by assuming that the AoD and the AoA lie perfectly on a finite grid and an approximate solution is obtained using the orthogonal matching pursuit (OMP) technique [12]. In practice, however, the AoDs and AoAs follow a continuous distribution. Therefore, grid-based CS methods for massive MIMO channel estimation will suffer from grid-mismatch. To resolve this issue, one solution, as taken by [4], is to refine the estimated parameters by using, e.g., a Newton's method. Another solution is to exploit the known antenna array geometry to formulate a gridless CS method [13].

In this paper, differently from [4, 9–11], we take the second approach and propose a gridless CS algorithm for channel estimation considering a HAD massive MIMO system. The proposed algorithm extends the gridless SPARROW algorithm from [13] to the case of multiple antennas at both the transmitter and the receiver employing HAD beamforming architectures. In particular, we assume a uniform linear array (ULA), where the array steering vectors admit a Vandermonde structure [13]. Taking advantage of their Toeplitz structure, an iterative gridless CS algorithm is proposed based on semidefinite programming and alternating optimization techniques, which is proven to converge to a local minimum. Detailed simulation results are provided to evaluate the effectiveness of the proposed algorithm.

2. SYSTEM MODEL

We consider a HAD massive MIMO system between a single transmitter-receiver pair as shown in Fig. 1. The transmitter has M_t antennas and single RF chain, while the receiver has M_r antennas and N_r RF chains. We assume a training-based parameters estimation approach, where the training period is divided into K training/transmission times, indexed by $k = 1, \ldots, K$. At the time instance k, the transmitter transmits the precoded signal $\mathbf{x}^{[k]} = \mathbf{f}^{[k]} s^{[k]}$, where $\mathbf{f}^{[k]} \in \mathbb{C}^{M_t}$ is the precoding vector and $s^{[k]}$ is the training symbol. At the other end, the receiver uses the *same* combining matrix $\mathbf{W} \in \mathbb{C}^{M_r \times N_r}$ to combine the received signal at each time instance. The received signal at time instance k is

$$\mathbf{y}^{[k]} = \mathbf{W}^{\mathrm{H}} \mathbf{H} \mathbf{x}^{[k]} + \mathbf{z}^{[k]} \in \mathbb{C}^{N_r}, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the MIMO channel matrix, $\mathbf{z}^{[k]} = \mathbf{W}^{\mathrm{H}} \mathbf{n}^{[k]} \in \mathbb{C}^{N_r}$ is the noise term after combining, and $\mathbf{n}^{[k]} \in \mathbb{C}^{M_r}$ is the additive white Gaussian noise with variance σ^2. At the end of K transmission times, the snapshots at receiver are concatenated in a measurement matrix $\mathbf{Y}$ as: $\mathbf{Y} = [\mathbf{y}^{[1]} \ldots \mathbf{y}^{[K]}] \in \mathbb{C}^{N_r \times K}$.

We assume a geometric channel model composed of L paths given as [3]

$$\mathbf{H} = \mathbf{A}_r \mathbf{D} \mathbf{A}_t^{\mathrm{T}} \in \mathbb{C}^{M_r \times M_t}. \tag{2}$$

In (2), $\mathbf{A}_r = [\mathbf{a}_{M_r}(\theta_1) \ldots \mathbf{a}_{M_r}(\theta_L)] \in \mathbb{C}^{M_r \times L}$ and $\mathbf{A}_t = [\mathbf{a}_{M_t}(\phi_1) \ldots \mathbf{a}_{M_t}(\phi_L)] \in \mathbb{C}^{M_t \times L}$ contain the L receive and transmit array steering vectors $\mathbf{a}_{M_r}(\theta_\ell)$ and $\mathbf{a}_{M_t}(\phi_\ell)$, respectively, where θ_ℓ and ϕ_ℓ denote the ℓ-th AoA and the AoD. Further, $\mathbf{D} = \mathrm{diag}\{\alpha_1, \ldots, \alpha_\ell\} \in \mathbb{C}^{L \times L}$ is a diagonal matrix containing the complex path gains $\alpha_\ell \in \mathbb{C}, \forall \ell$, on its diagonal.

Given the measurement matrix $\mathbf{Y}$, our problem is to estimate the channel parameters, the AoDs, the AoAs, and the complex path gains of dominant L paths.

3. GRID-BASED FORMULATION

Let us assume that the AoD and AoA, i.e., $\theta_\ell, \phi_\ell, \forall \ell$, fall on a uniform finite grid of $[I \times J]$ points quantizing the angular range of interest. Based on this assumption,

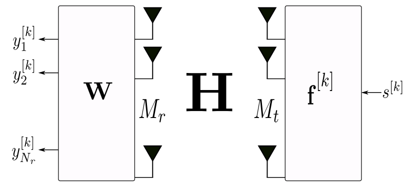


Fig. 1. A hybrid analog-digital system architecture.

we have $\theta_\ell \in \{\theta_1, \ldots, \theta_I\}, \forall l$, and $\phi_\ell \in \{\phi_1, \ldots, \phi_J\}, \forall l$. Define the following dictionary matrices

$$\tilde{\mathbf{A}}_r = [\mathbf{a}_{M_r}(\theta_1) \ldots \mathbf{a}_{M_r}(\theta_I)] \in \mathbb{C}^{M_r \times I}, \tag{3}$$

$$\tilde{\mathbf{A}}_t = [\mathbf{a}_{M_t}(\phi_1) \ldots \mathbf{a}_{M_t}(\phi_J)] \in \mathbb{C}^{M_t \times J}. \tag{4}$$

Then, using $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_t$, the channel matrix $\mathbf{H}$ can be represented by an L-sparse matrix $\tilde{\mathbf{D}}$ as $\mathbf{H} = \tilde{\mathbf{A}}_r \tilde{\mathbf{D}} \tilde{\mathbf{A}}_t^{\mathrm{T}}$, where $\tilde{\mathbf{D}} \in \mathbb{R}^{I \times J}$ contains L nonzero entries: their positions indicates the active AoD and AoA, where their values represent the complex path gains.

To exploit the above sparse representation, we first write the measurement matrix $\mathbf{Y} \in \mathbb{C}^{N_r \times K}$ in its vectorized form, i.e., $\mathbf{y} = \mathsf{vec}(\mathbf{Y}) \in \mathbb{C}^{KN_r}$, by utilizing the property $\mathsf{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A}) \mathsf{vec}(\mathbf{B})$, where $\otimes$ denotes the Kronecker product. Thus, $\mathbf{y}$ can be written as

$$\mathbf{y} = \underbrace{(\mathbf{X}^{\mathrm{T}} \otimes \mathbf{W}^{\mathrm{H}})}_{\mathbf{\Phi} \in \mathbb{C}^{KN_r \times M_r M_t}} \underbrace{(\tilde{\mathbf{A}}_t \otimes \tilde{\mathbf{A}}_r)}_{\mathbf{\Psi} \in \mathbb{C}^{M_r M_t \times IJ}} \tilde{\mathbf{d}} + \mathbf{z} = \mathbf{M} \tilde{\mathbf{d}} + \mathbf{z}, \tag{5}$$

where $\mathbf{X} = [\mathbf{x}^{[1]} \ldots \mathbf{x}^{[K]}] \in \mathbb{C}^{M_t \times K}$, $\mathbf{M} = \mathbf{\Phi\Psi} \in \mathbb{C}^{KN_r \times IJ}$ is the sensing matrix, $\tilde{\mathbf{d}} = \mathsf{vec}(\tilde{\mathbf{D}}) \in \mathbb{C}^{IJ}$ is the vectorized sparse vector, and $\mathbf{z} = \mathsf{vec}(\mathbf{Z}) \in \mathbb{C}^{KN_r}$ is the vectorized noise vector after combining, where $\mathbf{Z} = [\mathbf{z}^{[1]} \ldots \mathbf{z}^{[K]}] \in \mathbb{C}^{N_r \times K}$. Given the matrix $\mathbf{M}$, the MIMO channel parameters estimation is equivalent to estimating the nonzero entries in vector $\tilde{\mathbf{d}}$ as

$$\min_{\tilde{\mathbf{d}}} \|\tilde{\mathbf{d}}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{M}\tilde{\mathbf{d}}\|_{\mathrm{F}}^2 \leq \sigma^2, \tag{6}$$

where $\|\tilde{\mathbf{d}}\|_0$ is the pseudo-norm, which counts the nonzero entries in $\tilde{\mathbf{d}}$. Problem (6) is nonconvex and NP-hard due to the pseudo-norm formulation of its objective function. However, one can achieve an approximate solution by relaxing the objective function by the $\|\tilde{\mathbf{d}}\|_1$ or using a greedy search solutions like the OMP technique [9, 10]. However, it was shown recently in [13] that the *relaxed* version of problem (6) can be written equivalently in a convex form as

$$\min_{\mathbf{s} \in \mathbb{R}_+^{IJ}} \mathrm{Tr}((\mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{H}} + \lambda \mathbf{I}_{KN_r})^{-1} \mathbf{y}\mathbf{y}^{\mathrm{H}}) + \mathrm{Tr}(\mathbf{S}), \tag{7}$$

where $\mathbf{s}$ is the design variable vector, $\mathbf{S} = \mathsf{diag}(\mathbf{s}) \in \mathbb{R}_+^{IJ \times IJ}$, and $\lambda > 0$ is a predetermined regularization parameter, which is generally chosen in accordance to the noise power[1] [13]. Problem (7) can be equivalently written in a semidefinite programing form as [13]

$$\begin{aligned} \min_{\beta>0,\mathbf{s}\in\mathbb{R}_+^{IJ}} \quad & \mathrm{Tr}(\mathbf{S})+\beta \\ \text{s.t.} \quad & \begin{bmatrix} \beta & \mathbf{y}^{\mathrm{H}} \\ \mathbf{y} & \mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{H}}+\lambda\mathbf{I}_{KN_r} \end{bmatrix} \succeq 0, \end{aligned} \tag{8}$$

where β is a nonnegative real number. Problem (8) can be solved using any convex solver, like CVX [14]. The equivalence between the relaxed problem (6) and problem (8) is achieved by noting that for a given minimizer $\hat{\mathbf{S}}$ to problems (8), the minimizer $\hat{\tilde{\mathbf{d}}}$ is given as

$$\hat{\tilde{\mathbf{d}}} = \hat{\mathbf{S}}\mathbf{M}^{\mathrm{H}}(\mathbf{M}\hat{\mathbf{S}}\mathbf{M}^{\mathrm{H}}+\lambda\mathbf{I}_{TL_r})^{-1}\mathbf{y}. \tag{9}$$

4. GRIDLESS-BASED FORMULATION

In practice, the AoDs and AoAs follow a continuous distribution. Therefore, grid-based parameter estimation methods often suffer from grid-mismatch. Moreover, their accuracy and complexity highly depends on the grid quantization. The very fine quantization not only increases the estimation accuracy, but also its complexity. To resolve the grid mismatch and reduce the estimation complexity, in the following we propose a gridless implementation of problem (8) for the special case of ULA of m isotropic antennas, the steering vector for a given angle x can be written as $\mathbf{a}_m(x) = [1, e^{-j\pi\cos x}, \ldots, e^{-j\pi(m-1)\cos x}]^{\mathrm{T}}$, assuming half-wavelength inter-element spacing [13].

We start-off by expanding the term $\mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{H}}$ in problem (8) as (recalling that $\mathbf{M}=\mathbf{\Phi}\mathbf{\Psi}=\mathbf{\Phi}(\tilde{\mathbf{A}}_t\otimes\tilde{\mathbf{A}}_r)$)

$$\begin{aligned} \mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{H}} &= \mathbf{\Phi}(\tilde{\mathbf{A}}_t\otimes\tilde{\mathbf{A}}_r)\mathbf{S}(\tilde{\mathbf{A}}_t^{\mathrm{H}}\otimes\tilde{\mathbf{A}}_r^{\mathrm{H}})\mathbf{\Phi}^{\mathrm{H}} \\ &\overset{(a)}{=} \textstyle\sum_{\ell=1}^{L}\mathbf{\Phi}(\tilde{\mathbf{A}}_t\otimes\tilde{\mathbf{A}}_r)\mathbf{S}_\ell(\tilde{\mathbf{A}}_t^{\mathrm{H}}\otimes\tilde{\mathbf{A}}_r^{\mathrm{H}})\mathbf{\Phi}^{\mathrm{H}} \\ &\overset{(b)}{=} \textstyle\sum_{\ell=1}^{L}\mathbf{\Phi}(\tilde{\mathbf{A}}_t\mathbf{S}_{t,\ell}\tilde{\mathbf{A}}_t^{\mathrm{H}}\otimes\tilde{\mathbf{A}}_r\mathbf{S}_{r,\ell}\tilde{\mathbf{A}}_r^{\mathrm{H}})\mathbf{\Phi}^{\mathrm{H}}, \end{aligned} \tag{10}$$

where (a) is obtained by noting that $\mathbf{S}$ can be written as $\mathbf{S}=\sum_{\ell=1}^{L}\mathbf{S}_l$, where $\mathbf{S}_\ell\in\mathbb{C}^{IJ\times IJ}$ is the diagonal matrix containing the ℓ-th nonzero entry of $\mathbf{S}$, i.e., $\mathbf{s}_\ell$. Likewise, (b) is obtained by noting that $\mathbf{S}_\ell$ can be written as $\mathbf{S}_\ell=(\mathbf{S}_{t,\ell}\otimes\mathbf{S}_{r,\ell})$, where $\mathbf{S}_{r,\ell}\in\mathbb{C}^{I\times I}$ and $\mathbf{S}_{t,\ell}\in\mathbb{C}^{J\times J}$ are diagonal matrices containing one nonzero at the (i,i)-th and (j,j)-th entries, respectively, equal to $\sqrt{\mathbf{s}_\ell}$ and then by utilizing the Kronecker product property of $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC}\otimes\mathbf{BD})$.

In case of ULAs, we know that the matrices $\tilde{\mathbf{A}}_t$ and $\tilde{\mathbf{A}}_r$ admit a Vandermonde structure such that the matrices $\tilde{\mathbf{A}}_t\mathbf{S}_{t,\ell}\tilde{\mathbf{A}}_t^{\mathrm{H}}$ and $\tilde{\mathbf{A}}_r\mathbf{S}_{r,\ell}\tilde{\mathbf{A}}_r^{\mathrm{H}}$ exhibit a Toeplitz structure [13], i.e., $\tilde{\mathbf{A}}_t\mathbf{S}_{t,\ell}\tilde{\mathbf{A}}_t^{\mathrm{H}}=\mathcal{T}(\mathbf{u}_{t,\ell})$ and $\tilde{\mathbf{A}}_r\mathbf{S}_{r,\ell}\tilde{\mathbf{A}}_r^{\mathrm{H}}=\mathcal{T}(\mathbf{u}_{r,\ell})$, where $\mathbf{u}_{t,\ell}\in\mathbb{C}^{M_t}$, $\mathbf{u}_{r,\ell}\in\mathbb{C}^{M_r}$, and $\mathcal{T}(\mathbf{v})$ denotes a Hermitian Toeplitz matrix with $\mathbf{v}$ as its first column. Therefore, we can write

$$\mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{H}} = \textstyle\sum_{\ell=1}^{L}\mathbf{\Phi}(\mathcal{T}(\mathbf{u}_{t,\ell})\otimes\mathcal{T}(\mathbf{u}_{r,\ell}))\mathbf{\Phi}^{\mathrm{H}}. \tag{11}$$

Note that the later function is nonconvex due to the multilinear relation between $\mathcal{T}(\mathbf{u}_{t,\ell})$ and $\mathcal{T}(\mathbf{u}_{r,\ell})$. To resolve this issue, we propose to use an alternating optimization technique by solving for one variable at a time, while keeping the other fixed. Let $\mathcal{T}(\mathbf{u}_r)\overset{\text{def}}{=}\sum_{\ell=1}^{L}\mathcal{T}(\mathbf{u}_{r,\ell})$ and $\mathcal{T}(\mathbf{u}_t)\overset{\text{def}}{=}\sum_{\ell=1}^{L}\mathcal{T}(\mathbf{u}_{t,\ell})$. Then, we propose to approximate the function (11) as

$$\mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{H}} \approx \mathbf{\Phi}(\mathcal{T}(\mathbf{u}_t)\otimes\mathcal{T}(\mathbf{u}_r))\mathbf{\Phi}^{\mathrm{H}}. \tag{12}$$

The rationale behind this is by noting that for any given and fixed matrix $\mathbf{A}$, the function $\sum_{\ell=1}^{L}\mathbf{\Phi}(\mathcal{T}(\mathbf{u}_{t,\ell})\otimes\mathbf{A})\mathbf{\Phi}^{\mathrm{H}}$ is equivalent to $\mathbf{\Phi}(\mathcal{T}(\mathbf{u}_t)\otimes\mathbf{A})\mathbf{\Phi}^{\mathrm{H}}$, where $\mathcal{T}(\mathbf{u}_t)$ is defined above. Thus, letting $\mathbf{A}=\sum_{\ell=1}^{L}\mathcal{T}(\mathbf{u}_{r,\ell})$, function (12) follows directly.

From (12), a solution to problem (8) can be obtained in a gridless form by using an alternating optimization technique. Algorithm 1 summarizes the proposed solution steps, where (i) indicates the iteration index. In Step 3 we update $\mathcal{T}(\mathbf{u}_t)$, while fixing $\mathcal{T}(\mathbf{u}_r)$ and in Step 4 we update $\mathcal{T}(\mathbf{u}_r)$, while fixing $\mathcal{T}(\mathbf{u}_t)$.

Algorithm 1 Gridless CS for HAD MIMO Channel Estimation.

1: Input: $\mathbf{y}, \mathbf{\Phi}, \lambda$ and initial $\mathbf{u}_r^{(0)}$. Set $i=1$
2: **while** not converged **do**
3: Update $\mathcal{T}(\mathbf{u}_t^{(i)})$ for given $\mathcal{T}(\mathbf{u}_r^{(i-1)})$ by solving

$$\begin{aligned} \min_{\beta_t>0,\mathcal{T}(\mathbf{u}_t)\succeq 0} \quad & \mathrm{Tr}(\mathcal{T}(\mathbf{u}_t^{(i)}))+\beta_t \\ \text{s.t.} \quad & \begin{bmatrix} \beta_t & \mathbf{y}^{\mathrm{H}} \\ \mathbf{y} & \mathbf{\Phi}(\mathcal{T}(\mathbf{u}_t^{(i)})\otimes\mathcal{T}(\mathbf{u}_r^{(i-1)}))\mathbf{\Phi}^{\mathrm{H}}+\lambda\mathbf{I}_{KN_r} \end{bmatrix} \succeq 0. \end{aligned} \tag{13}$$

4: Update $\mathcal{T}(\mathbf{u}_r^{(i)})$ for given $\mathcal{T}(\mathbf{u}_t^{(i)})$ by solving

$$\begin{aligned} \min_{\beta_r>0,\mathcal{T}(\mathbf{u}_r^{(i)})\succeq 0} \quad & \mathrm{Tr}(\mathcal{T}(\mathbf{u}_r))+\beta_r \\ \text{s.t.} \quad & \begin{bmatrix} \beta_r & \mathbf{y}^{\mathrm{H}} \\ \mathbf{y} & \mathbf{\Phi}(\mathcal{T}(\mathbf{u}_t^{(i)})\otimes\mathcal{T}(\mathbf{u}_r^{(i)}))\mathbf{\Phi}^{\mathrm{H}}+\lambda\mathbf{I}_{KN_r} \end{bmatrix} \succeq 0. \end{aligned} \tag{14}$$

5: **end while**

The alternating optimization proposed in Algorithm 1 has guaranteed monotonic convergence of the objective function to a stationary (locally optimal) point if each step has a unique optimum [15, Proposition 2.7.1]. To have unique Vandermonde decompositions, the conditions of $\mathsf{rank}(\mathcal{T}(\mathbf{u}_r))<M_r$ and $\mathsf{rank}(\mathcal{T}(\mathbf{u}_t))<M_t$ must be satisfied, which can be guaranteed by appropriately choosing the regularization parameter λ [13].

At the convergence of Algorithm 1, we can estimate the associated AoAs and AoDs parameters from $\mathcal{T}(\mathbf{u}_r)$

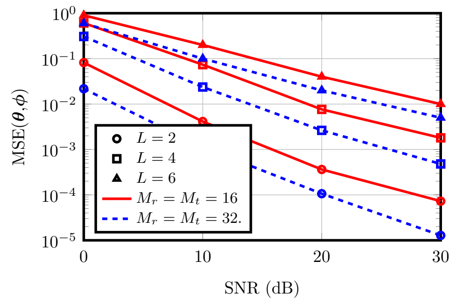
[1] Here we note that for small λ, $\mathbf{s}$ tends to have large nonzero entries, and vice-versa otherwise.

and $\mathcal{T}(\mathbf{u}_t)$ using, e.g., the standard ESPRIT [16] or unitary ESPRIT [17] techniques to obtain $\hat{\mathbf{A}}_r$ and $\hat{\mathbf{A}}_t$ matrices. Specifically, let the singular value decomposition of $\mathcal{T}(\mathbf{u}_r)$ be given as $\mathcal{T}(\mathbf{u}_r) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}} = \mathbf{U}_s\boldsymbol{\Sigma}_s\mathbf{V}_s^{\mathrm{H}} + \mathbf{U}_n\boldsymbol{\Sigma}_n\mathbf{V}_n^{\mathrm{H}}$, where $\boldsymbol{\Sigma}_s$ is a sub-matrix of $\boldsymbol{\Sigma}$ containing the largest L singular values, while $\mathbf{U}_s$ and $\mathbf{V}_s$ are sub-matrices containing the corresponding L left and right singular vectors, respectively. Let $\overline{\mathbf{U}}_s$ and $\underline{\mathbf{U}}_s$ denote the sub-matrices after removing the first and last rows of $\mathbf{U}_s$, respectively. Let $\boldsymbol{\Upsilon} = \overline{\mathbf{U}}_s^{+}\underline{\mathbf{U}}_s$, where $^{+}$ denotes pseudo-inverse. Then, the ℓ-th AoA $\hat{\theta}_\ell$ is recovered as $\hat{\theta}_\ell = \arccos\left(\frac{\arg(e_\ell)}{\pi}\right)$ [16], where e_ℓ is the ℓ-th eigenvalue of $\boldsymbol{\Upsilon}$. A similar method can be used to recover $\hat{\phi}_\ell, \forall \ell$, from $\mathcal{T}(\mathbf{u}_t)$. Note that, since the angles recovery is performed separately on $\mathcal{T}(\mathbf{u}_r)$ and $\mathcal{T}(\mathbf{u}_t)$, pairing cannot be guaranteed. To resolve this issue, a low-complexity grid-based approach can be used in a similar way to problem (6) by forming the dictionary matrices $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_t$ from the estimated angles $\hat{\theta}_1, \ldots, \hat{\theta}_L$ and $\hat{\phi}_1, \ldots, \hat{\phi}_L$, respectively. The resulting optimization problem can be solved using a greedy search solution such as the OMP technique [9, 10]. Alternatively, high resolution parameter estimation techniques that provide automatic pairing can be used [18, 19]. After that, the path gains $\hat{\mathbf{s}} \in \mathbb{R}^L$ can be recovered by solving the following simple linear system[2]: $\hat{\mathbf{A}}_r\hat{\mathbf{s}} = \mathbf{u}_r$. Then, the complex path gains can be recovered in accordance to equation (9) as $\hat{\tilde{\mathbf{d}}} = \hat{\tilde{\mathbf{S}}}\tilde{\mathbf{M}}^{\mathrm{H}}(\tilde{\mathbf{M}}\hat{\tilde{\mathbf{S}}}\tilde{\mathbf{M}}^{\mathrm{H}} + \lambda\mathbf{I}_{KN_r})^{-1}\mathbf{y}$, where $\tilde{\mathbf{M}} = \boldsymbol{\Phi}(\hat{\mathbf{A}}_t \diamond \hat{\mathbf{A}}_r) \in \mathbb{R}^{KN_r \times L}$ and $\hat{\tilde{\mathbf{S}}} = \mathsf{diag}(\hat{\mathbf{s}}) \in \mathbb{R}^{L\times L}$, where $\diamond$ denotes the Khatri-Rao product.

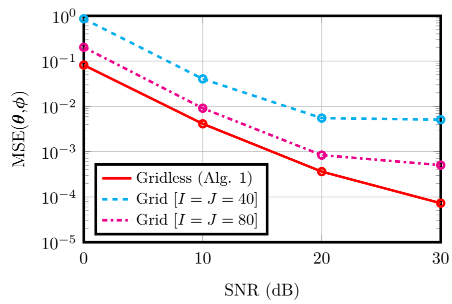
5. NUMERICAL RESULTS

In this section, we show some simulation results to evaluate the performance of the proposed Algorithm 1. We assume that the channel matrix $\mathbf{H}$ has at most $L = 6$ paths, where the AoAs are $\boldsymbol{\theta} = \{7^\circ, 13^\circ, 23^\circ, 52^\circ, 67^\circ, 81^\circ\}$ and the AoDs are $\boldsymbol{\phi} = \{9^\circ, 17^\circ, 32^\circ, 57^\circ, 72^\circ, 83^\circ\}$. We update the analog precoder $\mathbf{F} \in \mathbb{C}^{M_t\times K}$ by the steering vectors $\mathbf{f}^{[k]}(\bar{\phi}_k), k = 1, \ldots, K$, where $\{\bar{\phi}_k\}$ uniformly divide the range $[0^\circ, 90^\circ]$ and $\mathbf{f}^{[k]}(\bar{\phi}_k) = 1/\sqrt{M_t}\mathbf{a}_{M_t}(\bar{\phi}_k)$, so that $\|\mathbf{f}^{[k]}(\bar{\phi}_k)\| = 1$. Likewise, we update the analog decoder $\mathbf{W} \in \mathbb{C}^{M_r\times N_r}$ by the steering vectors $\mathbf{w}(\bar{\theta}_i), i = 1, \ldots, N_r$, where $\{\bar{\theta}_i\}$ uniformly divide the range $[0^\circ, 90^\circ]$ and $\mathbf{w}(\bar{\theta}_i) = 1/\sqrt{M_r}\mathbf{a}_{M_r}(\bar{\theta}_i)$. Moreover, the training signals $s^{[k]} = 1, \forall k$, $\alpha_\ell \sim \mathcal{CN}(0, 1), \forall l$, and $\mathbf{n}^{[k]} \sim \mathcal{CN}(0, \sigma^2), \forall k$. We define the signal-to-noise ratio (SNR) as $\mathrm{SNR} = \frac{1}{\sigma^2}$ and set $\lambda = \sigma\sqrt{M_r M_t \log(M_r M_t)}$ [13]. The mean-square-error (MSE) is defined as $\mathrm{MSE}(\boldsymbol{\theta},\boldsymbol{\phi}) = 1/\sqrt{L}\big(\sum_{\ell=1}^{L}|\theta_\ell - \hat{\theta}_\ell|^2 + \sum_{\ell=1}^{L}|\phi_\ell - \hat{\phi}_\ell|^2\big)$.

Fig. 2(a) shows $\mathrm{MSE}(\boldsymbol{\theta},\boldsymbol{\phi})$ versus SNR for Alg. 1 with



(a) Evaluating the Algorithm 1 performance.



(b) Grid vs. Gridless ($M_r = M_t = 16$, $L = 2$)

Fig. 2. MSE vs. SNR ($N_r = 8$, $K = 8$).

six different simulation scenarios. From Fig. 2(a), we can see that Alg. 1 is able to recover the channel parameters with satisfactory performance. Note that increasing the number of channel paths L degrades the $\mathrm{MSE}(\boldsymbol{\theta},\boldsymbol{\phi})$ performance. However, it is also shown that the $\mathrm{MSE}(\boldsymbol{\theta},\boldsymbol{\phi})$ performance improves by increasing the number of antenna elements.

Fig. 2(b) shows $\mathrm{MSE}(\boldsymbol{\theta},\boldsymbol{\phi})$ versus SNR comparing between the proposed gridless approach, Alg. 1, and the grid-based approach, problem (6), where a solution to which is obtained using a OMP technique [12]. From Fig. 2(b), we can see that the $\mathrm{MSE}(\boldsymbol{\theta},\boldsymbol{\phi})$ of the grid-based approach improves when increasing the quantization level. However, it seems to have a saturate performance when increasing the SNR level above 20dB. This is due to the grid-mismatch effect, since the channel AoAs, $\boldsymbol{\theta}$, and AoDs, $\boldsymbol{\phi}$, defined above fall off the quantization grids. Obviously, increasing the quantization level decreases the grid-mismatch effect, in expense of increasing the computational complexity.

6. CONCLUSIONS

We have proposed a gridless CS aided channel estimation algorithm for HAD MIMO systems based on an alternating optimization technique. The proposed formulation is achieved by capitalizing on the antenna array geometry, which resolves the grid-mismatch issue that appears with existing grid-based CS methods. Our simulation results showed the effectiveness of the proposed algorithm.

[2]Note that the path gains $\hat{\mathbf{s}}$ can be equivalently recovered from linear system $\hat{\mathbf{A}}_t\tilde{\mathbf{s}} = \mathbf{u}_t$ according to (10).

7. REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun 2011.

[2] K. Ardah, G. Fodor, Y. C. B. Silva, W. C. Freitas, and F. R. P. Cavalcanti, "A unifying design of hybrid beamforming architectures employing phase shifters or switches," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 243–11 247, Nov 2018.

[3] R. W. Heath, 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.

[4] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Proc. Information Theory and Applications Workshop*, Feb 2012, pp. 234–239.

[5] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun 2017.

[6] X. Gao, L. Dai, and A. M. Sayeed, "Low rf-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr 2018.

[7] M. Rani, S. B. Dhok, and R. B. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.

[8] R. Zhang, H. Zhao, and J. Zhang, "Distributed compressed sensing aided sparse channel estimation in FDD massive MIMO system," *IEEE Access*, vol. 6, pp. 18 383–18 397, 2018.

[9] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[10] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "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.

[11] J. Lee, G. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE Global Communications Conference*, Dec 2014, pp. 3326–3331.

[12] S. K. Sahoo and A. Makur, "Signal recovery from random measurements via extended orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2572–2581, May 2015.

[13] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1483–1497, Mar 2018.

[14] M. Grant, S. Boyd, and Y. Ye, *CVX: MATLAB software for disciplined convex programming*, 2008.

[15] D. Bertsekas, *Nonlinear Programming.* Athena Scientific, 1999.

[16] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul 1989.

[17] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J. Nossek, "2D unitary esprit for efficient 2D parameter estimation," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 1995, pp. 2096–2099.

[18] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, Feb 1996.

[19] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul 2008.