On and Off-Grid Compressive Sensing for Channel Estimation in Millimeter-Wave Hybrid MIMO Systems

Kareem M. Attiah and Justin Kang
Department of Electrical and Computer Engineering
University of Toronto
kattiah@ece.utoronto.ca js.kang@mail.utoronto.ca

***Abstract*—We consider the problem of channel estimation in millimeter wave massive multiple-input multiple-output systems with a hybrid architecture. The existing strategies for channel estimation hinge on the use of a grid of finite resolution from which the angles of departure are taken. Unlike those strategies, we investigate a grid-less channel estimation approach does not assume that the angles come from a finite resolution grid, but rather are continuous variables. This underlying optimization problem can be reformulated as an exact semi-definite program for which efficient solvers are readily available. Numerical comparisons suggest that the performance of the off-grid scheme is better than that of the grid-base counterparts when the number of grid points is of the same order as the number of antennas. In this regime grid-based solutions suffer from the error floor associated with restricting the angle of departure to a discrete grid of finite resolution, while the off-grid scheme does not. In addition, computational complexity analysis reveal that these performance gains are realized only at the cost of a modest increase in complexity relative to the most widely-known grid-based algorithms.**

I. Introduction

Owing to its outstanding beamforming capabilities, Massive multiple-input multiple-output (MIMO) has established itself as one of the key enablers of millimeter wave (mmWave) systems in future cellular networks [1], [2]. While the adoption of massive MIMO technology allows for increased link performance, the practical deployment of the traditional fully-digital architecture is hindered by high energy consumption and stringent hardware requirements of radio-frequency (RF) components. In order to overcome such drawbacks, a novel framework, known as a hybrid architecture, is proposed in [3], [4] to replace the fully-digital system by an analog beamformer, implemented in the RF domain using low-cost analog phase shifters, and a low-dimensional digital beamformer.

In this work, we consider the problem of channel estimation in millimeter wave (mmWave) systems with hybrid beamforming using a small number of measurements. Due to the severe path loss associated with the mmWave environment, the channel matrix can effectively be represented as the sum of a small number of dominant channel paths, and hence possesses a sparse representation in the angular domain [5]. The conventional compressive sensing (CS)-based techniques for channel estimation in mmWave systems exploit this fact to perform on-grid recovery. In particular, these techniques make the key assumption that the angles of departure (AOD) are restricted to a discrete grid of finite resolution [6]–[8]. The theory of compressive sensing provides guarantees that these methods work well when the true AOD lies exactly on the grid. However, the lack of theoretical guarantees when the AOD falls outside the grid leads to a lack of clarity as to whether this is the best way of estimating the channel.

Distinct from the grid-based methods, in this work we utilize a gridless sparse-recovery algorithm to estimate the channel state information (CSI). In particular, this work is motivated by the theoretical findings in [9] for the off-grid sparse reconstruction of the sum of complex sinusoids of different frequencies. Such frequencies are not assumed to belong to a finite-resolution dictionary, but take on arbitrary values in the interval $[0, 1]$. Moreover, the exact recovery of such frequencies entails solving a semi-definte program corresponding to minimizing the so-called atomic norm.

We make use of this formulation to perform channel estimation without the additional requirement that the angles lie on a grid. To do this, we reformulate the channel estimation problem as an atomic norm minimization (ANM) problem. Because the ANM can be equivalently cast as a semi-definite program (SDP), the solution to the channel estimation problem is obtained using CVX [10]. Numerical comparisons reveal performance gains for off-grid channel estimation algorithm relative to the grid-based counterparts in the regime of grid spacing considered. This is intuitive since the ANM does not suffer from the error associated with grid restriction in the grid-based methods. Computational complexity analysis indicates that the cost of solving the original SDP formulation associated with the ANM is as high as $\mathcal{O}((N+1)^6)$ operations per iteration, which may suggest that the ANM comes with a heavy complexity toll. However, it was suggested in [11] that the atomic norm minimization can be efficiently solved using the Alternating Direction Method of Multipliers (ADMM) [12] in only $\mathcal{O}((N+1)^3)$ operations per iteration, thus entailing only a modest increase in computational cost compared to that of the most widely-known variation of the grid-based algorithms, known as the Orthogonal Matching Pursuit (OMP) [13].

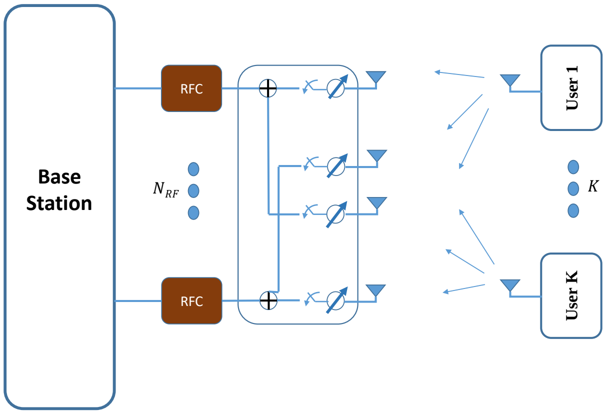

Fig. 1. Block Diagram of the system architecture. The BS is equipped with analog phase shifters with switches that map the observations of the antennas into the RF chains.

II. System and Channel Model

A. System Model

We consider a TDD massive MIMO system operating in a frequency flat mmWave environment shown in Fig. 1. The base-station (BS) is equipped with M antennas and communicates with K single antenna users. Specifically, we focus on the problem of channel estimation between the BS and the users. Channel estimation is of importance to the design of such system since it serves as an intermediate step for the design of other components (e.g., precoding matrix design [6]). To perform estimation, the channel is sensed by the BS through uplink pilot training. Since TDD operation is assumed, the uplink and downlink channels are identical [14]. As a result, the BS can acquire the CSI once it has observed the uplink pilot transmissions. Moreover, in a multi-user scenario this channel sensing process is performed in a TDMA fashion, where each user is assigned a block of L interference-free transmission slots during which a pilot sequence is communicated to the BS. Let $x_k^{(\ell)}$ denote the ℓ-th pilot transmission by user k, where $\ell = 1, \ldots, L$, the received baseband vector at the BS is given by:

$$\mathbf{z}_k^{(\ell)} = \mathbf{W}^{(\ell)}(\mathbf{h}_k x_k^{(l)} + \mathbf{n}_k^{(l)}), \tag{1}$$

where $\mathbf{W}^{(\ell)}$ is the combining matrix for time slot ℓ at the BS and $\mathbf{n}_k^{(\ell)} \in\sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$ is the corresponding circularly-symmetric, additive white Gaussian noise vector. We assume M is large and the BS employs a hybrid architecture with switches adopted in [15]. In particular, to reduce the implementation complexity of the system, the BS antennas are internally connected to N_{RF} RF chains, with $K \leq N_{\text{RF}} \leq M$, where the connections are implemented in the analog domain using phase shifters and can be enabled or disabled using switches. This entails a slight modification to the traditional hybrid architecture in [3] which only consists of analog phase shifters. However, as argued in [15] the cost of adding the switches to internal connections is marginal. This concept is illustrated in Fig. 1. With this assumption, the combining matrix in the ℓ-th time slot is:

$$\mathbf{W}^{(\ell)} = \mathbf{A}^{(\ell)} \odot \mathbf{W}_{\text{RF}}^{(\ell)}, \tag{2}$$

where $\odot$ is the Hadamard (i.e., element-wise) product, $\mathbf{W}_{\text{RF}}^{(\ell)} \in \mathbb{C}^{N_{\text{RF}} \times M}$ is the analog precoding matrix whose entries are constrained to unity, i.e., $|[\mathbf{W}_{\text{RF}}]_{ij}| = 1$, and the matrix $A^{(\ell)} \in \{0,1\}^{N_{\text{RF}} \times M}$ is binary valued with entries correspond to the state of the switches. Finally, by vertically stacking all BS observations for user k, the overall received pilot vector is given by:

$$\mathbf{z}_k = \sqrt{P} \underbrace{\begin{bmatrix} \mathbf{A}^{(1)} \odot \mathbf{W}_{\text{RF}}^{(1)} \\ \vdots \\ \mathbf{A}^{(L)} \odot \mathbf{W}_{\text{RF}}^{(L)} \end{bmatrix}}_{\mathbf{\Phi}} \mathbf{h}_k + \begin{bmatrix} \tilde{\mathbf{n}}_k^{(1)} \\ \vdots \\ \tilde{\mathbf{n}}_k^{(L)} \end{bmatrix}, \tag{3}$$

where we set $x_k^{(\ell)} = \sqrt{P}, \forall \ell$ for simplicity, and take $\tilde{\mathbf{n}}_k^{(\ell)} \triangleq \mathbf{W}^{(\ell)} \mathbf{n}_k^{(\ell)}$ to be the corresponding noise vector for user k, and P as the transmitted power. Finally, we also denote the channel sensing matrix by $\mathbf{\Phi} \in \mathbb{C}^{LN_{\text{RF}} \times M}$. The design of such matrix follows from the design of the analog phase shifter matrices $\mathbf{W}_{\text{RF}}^{(\ell)}$, and the design of the switching matrices $\mathbf{A}^{(\ell)}$. Following [15], [16], we assume that the BS has Q-bits quantized phase shifters and thus the coefficients of analog matrices are of the form $e^{j2\pi i/2^Q}$, where $i \in \{0, \ldots, 2^Q - 1\}$ and $j = \sqrt{-1}$. In addition, for reasons that will become clear later on, we choose $\mathbf{A}^{(\ell)}, \forall \ell$ in such a manner that only N_{RF} of M antennas are active. Thus, over the course of L pilot transmissions of user k, the BS is able to observe the noisy channels of distinct LN_{RF} antenna elements. If we define $\mathbf{A} \triangleq \left[\left(\mathbf{A}^{(1)}\right)^T \ldots \left(\mathbf{A}^{(L)}\right)^T \right]^T$, this requirement is equivalent to setting one entry in the each row of the matrix $\mathbf{A}$ to one and the rest of the entries to zero, where no two rows are identical.

By horizontally stacking the observations from all users, we finally obtain

$$\mathbf{Z} = \mathbf{\Phi} \mathrm{H} + \mathbf{N}, \tag{4}$$

where $\mathbf{Z} \triangleq [\mathbf{z}_1 \ldots \mathbf{z}_K] \mathbb{C}^{LN_{\text{RF}} \times K}$ is the observation matrix containing the overall received pilot vectors during the entire training phase for all users, $\mathbf{H} \triangleq [\mathbf{h}_1 \ldots \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ is the channel matrix for all users, and $\mathbf{N} \triangleq [\tilde{\mathbf{n}}_1 \ldots \tilde{\mathbf{n}}_K] \in \mathbb{C}^{M_{\text{RF}} \times K}$ is the corresponding noise matrix after combining.

B. Channel Model

The expression in (4) describes the relationship between the known observation (i.e., $\mathbf{Z}$) and the unknown parameter (i.e., $\mathbf{H}$) in the channel estimation problem. In the traditional multi-path rich-scattering environment, the entries of the channel matrix $\mathbf{H}$ are uncorrelated and thus conventional estimation techniques such as the MMSE must be employed to obtain a channel estimate. In this case, the number of pilot transmissions needed for effective channel recovery is at least $\lceil M/N_{\text{RF}} \rceil$. The propagation environment we consider here is that of a mmWave channel [1]. In this model the user channels can be written as

$$\mathbf{h}_k = \sum_{i=0}^{L_p} \alpha_i \mathbf{a}\left(\phi_i\right), \quad \alpha_i \sim \mathcal{CN}(0,1) \quad \phi_i \sim \text{Unif}([0,2\pi]), \tag{5}$$

where L_p is the number of paths, α_i and ϕ_i are the path gains and AOD for path i, and Unif$([a,b])$ is the uniform distribution with parameters a and b. Finally, $\mathbf{a}(\phi)$ is the antenna array response. For a uniform planner array, this is given by:

$$\mathbf{a}\left(\phi\right) = \left[1, e^{j\frac{2\pi d}{\lambda}\sin\left(\phi\right)}, \ldots, e^{j\frac{2\pi d}{\lambda}\sin\left(\phi\right)(M-1)}\right]^T, \tag{6}$$

where λ is the operating wavelength and d is the antenna elements separation. Due to the sparsity associated with such a model, one can employ compressive sensing schemes to perform channel recovery using a modest number of pilot transmissions.

III. Grid Based Channel Estimation

Compressive sensing is a tool for finding sparse solutions to undetermined linear systems. From (5) each user channel is formed from a linear combination of relatively few "atoms" $\mathbf{a}(\phi)$, and can thus can be represented as a sparse vector in some dictionary. A traditional approach to exploiting this sparsity is to quantize the continuous parameters ϕ_i. This is done by forming a grid with G_b points $\varphi_1, \ldots, \varphi_{G_b}$. If there exists points on the grid $\phi_i^{'} \approx \phi_i$, then from (5) we can write:

$$\mathbf{h}_k \approx \sum_{i=0}^{L_p} \alpha_i \mathbf{a}\left(\phi_i^{'}\right), \quad \phi_i^{'} \in \left\{\varphi_1, \ldots, \varphi_{G_b}\right\}. \tag{7}$$

It is clear that the accuracy of the approximation in (7) depends on how close each $\phi_i^{'}$ on the grid is to the continuous parameter ϕ_i. In this grid based approach we introduce a dictionary matrix:

$$\mathbf{\Psi} = \left[\mathbf{a}(\varphi_1), \ldots, \mathbf{a}(\varphi_{G_b})\right]. \tag{8}$$

Using this dictionary matrix each user channel can be represented as a transformation of a sparse vector:

$$\mathbf{h}_k \approx \mathbf{\Psi}\mathbf{h}_k^S, \quad \left\|\mathbf{h}_k^S\right\|_0 = L_p. \tag{9}$$

This sparse representation enables us to write the problem of channel estimation as a compressive sensing problem:

$$\begin{aligned} \underset{\mathbf{h}_k^S}{\arg\min} \quad & \left\|\mathbf{h}_k^S\right\|_0 \\ \text{subject to} \quad & \left\|\mathbf{\Phi\Psi}\mathbf{h}_k^S - \mathbf{z}_k\right\|_2 \leq \eta, \end{aligned} \tag{10}$$

where η should be chosen as a function of noise power. This problem is not convex, but under certain circumstances it is approximable [17]. Let $\mathbf{\Phi\Psi} = \mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_{G_b}]$. The solution to (10) can be approximated if the "mutual coherence" μ of $\mathbf{B}$ satisfies:

$$\mu(\mathbf{B}) = \max_{i \neq j} \frac{\left|\mathbf{b}_i^{\mathrm{H}}\mathbf{b}_j\right|}{\left\|\mathbf{b}_i\right\|_2 \cdot \left\|\mathbf{b}_j\right\|_2} < \frac{1}{2L_p - 1} \tag{11}$$

Note that (11) is related to the so-called RIP condition, which has similar implications. Our choice of $\mathbf{A}$ in (2) and the introduction of switches, is important for reducing the mutual coherence of $\mathbf{B}$. In the following sections, we consider two different techniques which can be used to approximate the solution to this optimization problem.

A. Orthogonal Matching Pursuit

OMP [13] is a greedy approach for approximating the solution of the combinatorial problem (10). We consider two variants of this algorithm. In the first variant given in Algorithm 1, the solver is not aware of the number of paths, L_p which comprise the channel. Instead, a threshold is provided to the algorithm, and the termination condition is based on a comparison between the residual error and the threshold. The threshold is chosen empirically to minimize the mean square error. Another variant of OMP is considered when L_p is known. This version differs from Algorithm 1, only in the fact that the "while" condition is $\left\|\mathbf{h}^S\right\|_0 \leq L_p$.

Algorithm 1: Threshold Orthogonal Matching Pursuit

Input
Sensing matrix and dictionary $\mathbf{\Phi}$ and $\mathbf{\Psi}$, measurement vector: $\mathbf{z}_k$, threshold: δ;
Initialize: $t \leftarrow 0$, $\mathbf{r} \leftarrow \mathbf{z}$, $\mathbf{h}^S \leftarrow \mathbf{0} \in \mathbb{C}^{G_b}$, $\mathcal{I}_t \leftarrow \emptyset$;
while $\left\|\mathbf{r}\right\|_2 \geq \delta$ **do**
  $g^* \leftarrow \text{argmax} \left\|\mathbf{a}_g^H\right\|_2$;
  $t \leftarrow t+1$;
  $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{g\}$;
  $x^* \leftarrow (\mathbf{\Phi\Psi})_{\mathcal{I}_t}^{\dagger}\,\mathbf{z}$;
  $\mathbf{r} \leftarrow \mathbf{z} - (\mathbf{\Phi\Psi})_{\mathcal{I}_t}\mathbf{x}^*$;
  $\mathbf{h}^S \leftarrow \mathbf{x}^*$
end
Output: Channel Estimate $\hat{\mathbf{h}^S}$

B. Basis Pursuit Denoising

While the OMP is considered a cheap algorithm for estimating the channel, a more elaborate approach, known as Basis Pursuit Denoising (BPD) [18] is often used to approximate the solution to the compressive sensing problem (10). The formulation of Basis Pursuit Denoising is given by:

$$\underset{\tilde{\mathbf{h}}_k}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{\Phi\Psi}\tilde{\mathbf{h}}_k - \mathbf{z}\right\|_2^2 + \nu\left\|\tilde{\mathbf{h}}_k\right\|_1. \tag{12}$$

It is an instance of a convex quadratic program, and can be solved efficiently by many standard methods, for example, homotopy continuation [19]. The form of the objective in (12) is a weighted sum of the square-residual error of the reconstructed channel and the l_1-norm of the estimated channel in the angular domain, where it is known to be sparse. Here the l_1-norm serves as a convex proxy for the "l_0-norm" in (10) (which is not actually a norm). The scaling hyperparameter ν controls the relative importance of sparsity and residual error in the optimal solution $\hat{\mathbf{h}^S}$. As ν becomes large, (12) favors sparser solutions, while as ν becomes smaller, solutions which more faithfully reconstruct the observed measurement are favoured. Practically, ν can be chosen to minimize the normalized mean square error of the channel estimation via sweeping the parameter over a range of values.

IV. Off-Grid Channel Estimation

The common element in all of the channel estimation techniques which were previously described is the introduction of a finite dictionary matrix $\mathbf{\Psi}$. Off-Grid Compressive Sensing techniques use a more advanced mathematical framework to avoid the use of a finite dictionary—instead using an infinite one. We start by defining such an infinite dictionary:

$$\mathcal{A} = \left\{ \mathbf{a}(\phi)\alpha : \phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right], \alpha \in \mathbb{C}, |\alpha| = 1 \right\}. \tag{13}$$

If we consider (5) in the context of the dictionary $\mathcal{A}$, we can write $\mathbf{h}_k$ as:

$$\mathbf{h}_k = \sum_{i=1}^{L_p} \mathbf{v}_i \quad \mathbf{v_i} \in \mathcal{A}. \tag{14}$$

We now seek a convex function which can be used to indicate sparsity over this dictionary, similar to the l_1-norm which was used in the basis pursuit denoising formalism. This problem has been considered previously in [9]. In that work, the "atomic" norm formalizes this idea. This norm is given by:

$$\begin{aligned} \|\mathbf{h}_k\|_{\mathcal{A}} &= \inf\{g > 0 : \mathbf{h}_k \in g \cdot \operatorname{conv}(\mathcal{A})\} \\ &= \inf\left\{ \sum_i b_i : \mathbf{h}_k = b_i \sum_i \mathbf{a}_i, b_i > 0, \mathbf{a}_i \in \mathcal{A} \right\}. \end{aligned} \tag{15}$$

The atomic norm $\|\cdot\|_{\mathcal{A}}$, is not as simple as the l_1-norm, however, it has been shown in [9] that this norm can be calculated by the following SDP:

$$\begin{aligned} &\underset{t, \mathbf{u}}{\text{minimize}} && \frac{1}{2}(t + u_1) \\ &\text{subject to} && \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{h}_k \\ \mathbf{h}_k^{\mathrm{H}} & t \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{16}$$

Thus, despite the relatively complicated formulation of the atomic norm, computationally, it remains tractable. In addition, the authors of [9] further show that if $\mathbf{h}_k$ is observed exactly on a subset of sufficient length $N = \mathcal{O}(\log L_p \log M)$, then the ANM guarantees exact reconstruction of the entire vector in the noiseless case. Thus, we can see that our choice of $\mathbf{A}$ in (2) serves a second purpose of allowing us to observe the channel in LN_{RF} positions, up to a unit modulus scaling.

Atomic norm denoising [20] is a simple augmentation of the basis pursuit denoising described in Section III-B to handle the noisy case. The overall form is similar to that of (12), but with the l_1-norm replaced with the atomic norm:

$$\underset{\mathbf{h}_k}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{\Phi}\mathbf{h}_k - \mathbf{z}\|_2^2 + \nu \|\mathbf{h}_k\|_{\mathcal{A}}. \tag{17}$$

Equation (17) is also an SDP. This can be seen by including the definition of the atomic norm in (16) with the optimization

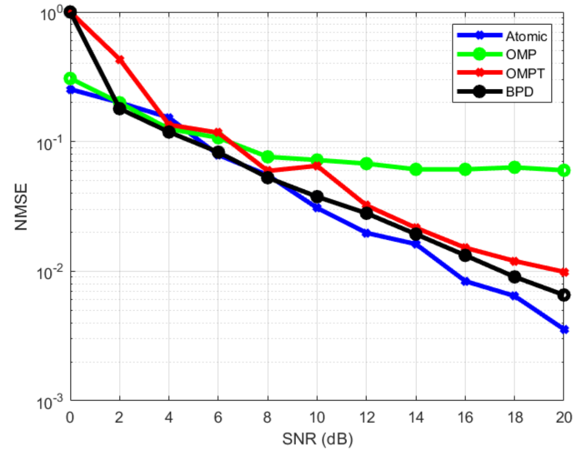

Fig. 2. Performance comparison of the ANM, BPD, and the two variants of the OMP in terms of the Normalized Mean Square Error.

problem in (17). This leads to the following formulation for the sparse channel estimation problem:

$$\begin{aligned} &\underset{t, \mathbf{u}, \mathbf{h}_k}{\text{minimize}} && \frac{1}{2}\|\mathbf{\Phi}\mathbf{h}_k - \mathbf{z}\|_2^2 + \frac{\nu}{2}(t + u_1) \\ &\text{subject to} && \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{h}_k \\ \mathbf{h}_k^{\mathrm{H}} & t \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{18}$$

V. Numerical Analysis

In this section, we assess the performance of off-grid channel estimation against the grid-based approaches discussed in Section IV. To do this, we consider a BS with $M = 64$ antennas and $N_{\text{RF}} = 16$ RF chains. The BS serves $K = 10$ users. The sparse channel has $L_p = 2$ paths and we allocate $L = 2$ pilot transmissions per user. Thus, we can observe only $LN_{\text{RF}} = 32$ noisy observations. The matrices $\mathbf{W}_{\text{RF}}$ and $\mathbf{A}$ are chosen according to the construction in Section II in a random fashion. Finally, we take $d/\lambda = 1/2$.

A. Performance Comparison

In Fig. 2, we plot the Normalized Mean squared Error (NMSE) $\mathbb{E}\{(\|\hat{\mathbf{H}} - \mathbf{H}\|_{\mathrm{F}}^2)/\|\mathbf{H}\|_{\mathrm{F}}^2\}$ against the Signal-to-Noise Ratio SNR $= 10\log\frac{P}{\sigma^2}$ for the ANM, BPD, and the two variations of the OMP considered in Section IV, denoting the variation with ℓ_2 termination condition by OMPT. For the grid-based approaches, we choose the grid size $G_b = 2M$. Focusing on medium and high SNR ranges, we see that the ANM outperforms the other three approaches. This is expected since it does not suffer from the error floor associated with the grid restrction in the grid-based schemes. This error floor is particularly visible in the case of OMP since it detects only the highest two directions in the dictionary rather than than relying on an ℓ_2 termination condition.

B. Performance and Complexity Trade-off

In this section we study the performance and complexity trade-off of grid based channel estimation schemes. Fig.3,

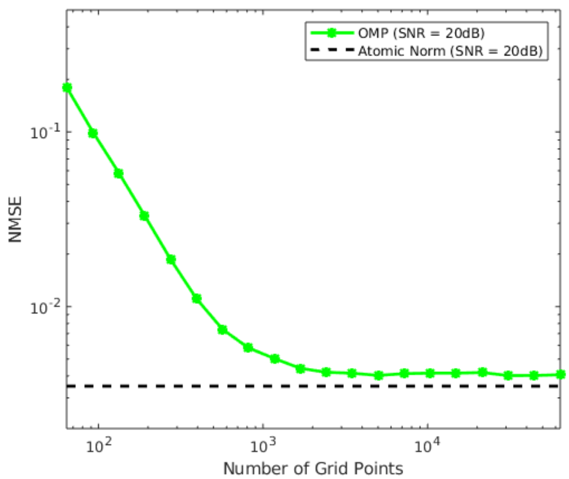

Fig. 3. On-Grid Channel Estimation is compared to Off-Grid Channel Estimation in NMSE, as a function of the grid size G_b. $L = 2$, $K = 10$, $M = 64$, $N_{\text{RF}} = 16$, $Q_b = 7$, $N_b = 50$.

verifies the intuitive expectation that increasing the grid size results in a decrease in the estimation error for the grid based scheme. As the number of grid points increases, the estimation error decreases at a decreasing rate. In the region where G_b is small, Fig.3 shows that in a log-log plot the error decreases linearly. In this region, the linear fit results in the power law: $NMSE \approx 639{G_b}^{-1.534}$. As G_b becomes larger, however, the on-grid estimation error asymptotically approaches the atomic norm estimation error. Fig.3 shows that the advantage

OMP	ADMM	Interior Point
$LN_{rf}(G_b + L_p^2)$	$(LN_{rf})^3$	$(LN_{rf})^6$

TABLE I
COMPLEXITY ANALYSIS OF CHANNEL ESTIMATION SCHEMES.

of using the atomic norm for the purpose of channel estimation is limited, in the sense that nearly identical results can be achieved with a fine grid. Thus, the choice of algorithm is dominated primarily by the difference in computational costs, as well as estimation error requirements. For example, from Fig.3 if we wish to achieve results close to that of atomic norm denoising, we might choose $G_b \geq 1500$. If we solve (17) by ADMM the complexity of the algorithm grows with $(LN_{rf})^3$, while the number of computations in OMP is $LN_{rf}(G_b+L_p^2)$. Since the required $G_b > (LN_{rf})^2$, Atomic Norm Denoising may be faster, although the complexity measures stated here are approximate, and only accurate in order.

VI. Conclusion

In this work, we examined an off-grid channel estimation strategy for massive MIMO systems in mmWave propagation environments and compared it with the traditional strategy where AODs are restricted to lie on a finite grid. We cast the off-grid estimation problem into an exact convex optimization problem for which computationally efficient algorithm exists. Furthermore, we studied the computational complexity of both the grid-based and off-grid algorithms. Our contribution is a numerical example which shows that the increase in complexity of off-grid schemes is modest when compared against the complexity of OMP with the required number of grid points to achieve similar performance to atomic norm denoising. Finally, we remark that we have not included convergence analysis due to space limitations.

References

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] X. Zhang, A. F. Molisch, and S.-Y. Kung, “Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection,” *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Oct. 2005.

[4] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

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

[6] A. Alkhateeb, O. El Ayach, G. Leus, and R. 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.

[7] T. J. Shaw and G. C. Valley, “Angle of arrival detection using compressive sensing,” in *2010 18th European Signal Processing Conference*, Aug 2010, pp. 1424–1428.

[8] J. Lee, G.-T. Gil, and Y. H. Lee, “Channel estimation via orthogonal matching pursuit for hybrid mimo systems in millimeter wave communications,” *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.

[9] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.

[10] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” Mar. 2014.

[11] Y. Li and Y. Chi, “Off-the-grid line spectrum denoising and estimation with multiple measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2016.

[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[13] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[14] G. S. Smith, “A direct derivation of a single-antenna reciprocity relation for the time domain,” *IEEE Trans. Antennas Propag.*, vol. 52, no. 6, pp. 1568–1577, June 2004.

[15] J. Deng, O. Tirkkonen, and C. Studer, “MmWave channel estimation via atomic norm minimization for multi-user hybrid precoding,” *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 2018-April, pp. 1–6, 2018.

[16] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Jan. 2016.

[17] J. A. Tropp, “Just relax: convex programming methods for identifying sparse signals in noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[18] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001. [Online]. Available: https://doi.org/10.1137/S003614450037906X

[19] D. M. Malioutov, M. Cetin, and A. S. Willsky, “Homotopy continuation for sparse signal representation,” in *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5, 2005, pp. v/733–v/736 Vol. 5.

[20] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.