Millimeter Wave Massive MIMO Channel Estimation Using Subspace Pursuit Greedy Algorithm

Yaseen A Mohammed 1 and Hatem H Abbas 2

1Electronic and Communications Engineering Department, Al-Nahrain University, Baghdad, Iraq
2Electronic and Communications Engineering Department, Al-Nahrain University, Baghdad, Iraq

Abstract. Achieving high data rate communications and increasing system reliability are the essentials for 5G networks. Massive MIMO and millimeter wave (mmWave) communications along with the ongoing developments in the underlying infrastructure promise to sharply increase the aggregate capacity. The communication system harnessing the odds of the unused spectrum in mmWave bands is vulnerable to the extreme path loss that severely weakens the strength of the information-bearing signal. Nevertheless, exploiting hybrid beamforming and efficient high-speed channel estimation algorithm can help address this problem and create the communication systems required for 5G. Our main contribution in this paper is to propose a very high speed accurate greedy algorithm called Subspace Pursuit (SP) and we also compare the performance of this algorithm with two other greedy algorithms. Simulation results show that SP algorithm achieves a noticeable improvement in comparison with Orthogonal Matching Pursuit (OMP) and Regularized Orthogonal Matching Pursuit (ROMP) algorithms in terms of both the spectral efficiency and runtime.

1. Introduction
A rapid growth in data traffic has been witnessed in the last years due to the significant increase of smartphone users and the massive growth in the population. Also, numerous IoT applications require a profoundly high data rate which cannot be supported by the congested radio spectrum below 3 GHz. All these reasons have compelled the technologists and researchers to invent novel technologies to meet these demands. Tremendous developments have to be made regarding the network infrastructure for the next generation (5G). Exploiting the unused wide bandwidth available at millimeter wave’s spectrum (20-300 GHz) are highly recommended to reach the aspirations [1] [2] [3] [4].

The primary concern that encounters the implementation of mmWave waves is the extremely high path loss which causes a severe attenuation to the transmitted signal [2] [5]. Also, the attenuation due to rain has a detrimental impact on the signal’s gain [6] [5]. However, exploiting the edge of a small wavelength, large scale number of antennas along with the hybrid beamforming architecture could be deployed and the signal transmitted over a high-rank communication channel can be steered with a high gain toward the intended user [7].

Using hybrid beamforming architecture, we can exploit the benefits of digital and analog beamforming [7], [8], [9]. Utilizing such a hybrid architecture will allow us to form and steer the beam by using the analog part while a digital one will be responsible for precoding the signal and for
implementing the signal processing techniques. Collectively, hybrid beamforming compromises between the required accuracy and the ease of implementation.

To reconstruct the precoding matrix for the sake of beamforming at the Base Station, it is paramount to have the channel coefficients that is regarded as a difficult mission due to the low signal to noise ratio (SNR). Nonetheless, considering the small number of non-zero elements of the channel matrix (the number of paths is in between (2-6) [5] [1] [6]), the channel matrix can be treated as a sparse matrix. Several greedy algorithms could be reasonably chosen to estimate the elements of the matrix. In [10], the authors propose a greedy algorithm deploying OMP, we found out that this algorithm in comparison with the Subspace Pursuit (SP) algorithm [11] [12] is slow and does not accurately calculate the channel coefficients as its attainable capacity is lower than the proposed algorithm. [13] uses Amplitude Message Passing (AMP) to approximate the channel. Nevertheless, the MMSE calculations are not convincing as the MMSE value plateaus and grows as the SNR increases. As far as our knowledge, the SP algorithm [11] [12] has not been used to predict the elements of the mmWave channel as it is presented in [14] for MIMO-OFDM radio channel and there are clear distinctions between the Radio and mmWave channels [2] [3] [4].

In this paper, a greedy algorithm is proposed and a comparison is made with the conventional OMP [13] [10] and ROMP [15] [16] algorithms. We use three evaluation metrics (Minimum Mean Square Error (MMSE), Capacity, and Runtime) to assess these algorithms. Eventually, we come to prove that the proposed algorithm makes sure to rapidly predict the elements of the matrix and to significantly increase the attainable capacity approaching the optimal one.

The rest of the paper is organized as follows. Section II describes the system model. Section III illustrates the proposed channel estimation algorithm. Section IV provides the resultant figures evaluating the proposed algorithm. Finally, section V gives a brief conclusion about the problem and its effective solution.

2. System Model

We consider a mmWave Massive MIMO communication system provided with a hybrid beamforming architecture and quantizers similar to the systems in [13], [9]. The BS and MS utilize $N_{\text{RT}}$ and $N_{\text{MR}}$ number of antennas, respectively. The number of RF chains used at both ends are denoted by $L$. The number of bits deployed at ADCs is $n_{\text{ADC}}$. $N_s$ represents the data streams’ number and can be taken from $[1 , \min(L_{\text{RT}},L_{\text{MR}})]$. For simplicity, an equal number of RF chains is presumed at the BS and MS. The BS is equipped with $L \times N_s$ baseband precoder $F_{\text{BB}}$ which is followed by $N_{\text{RT}} \times L$ RF precoder $F_{\text{RF}}$. The discretized signal after the precoder can be written as:

$$\textbf{x} = F_{\text{T}} \textbf{s}$$  (1)

Where $F_{\text{T}} = F_{\text{RF}} F_{\text{BB}}$, and $\textbf{s}$ denotes the transmitted symbols. For the purpose of normalization, the vector $\textbf{s}$ should be designed such that $E[\textbf{s}\textbf{s}^H] = (\frac{P}{N_s})I_{N_s}$, where $P$ is the average total transmit power. In addition, the baseband beamforming matrix $F_{\text{BB}}$ is normalized by forcing each element of the RF beamforming matrix $F_{\text{RF}}$ to have a power of $N_{\text{RT}}^{-1}$ and by assuring that $||F_{\text{T}}||_F^2 = N_s$.

After the transmission is taken place across a wide band Massive MIMO channel, the received signal can be written in the form:

$$\textbf{r} = H F_{\text{T}} \textbf{s} + \textbf{n}$$  (2)

Where $H$ is the $N_{\text{MR}} \times N_{\text{RT}}$ mmWave channel matrix between the BS and MS, and $\textbf{n} \sim \mathcal{N}(0, \sigma^2 I_{N_{\text{MR}}})$ represents the additive white Gaussian noise.

The combiner $W_{\text{T}}$ at the MS is formulated by the multiplication of the RF and BS combiners $W_{\text{RF}}$ and $W_{\text{BB}}$ that is employed to have the following received signal:

$$\textbf{y} = W_{\text{T}}^H \textbf{r} = W_{\text{T}}^H \textbf{r}$$  (3)

We will illustrate the proposed algorithm for the downlink system. Nevertheless, applying the same algorithm to the uplink model is also applicable with some modifications such that the uplink channel model is utilized and the roles of combiners and precoders flipped.
A geometric mmWave channel model with a limited number of $K$ scatterers is adopted where each scatterer is suggested to take part in a single propagation path between the BS and the MS. Thereby, the $u^{th}$ tap channel $H_u$ can be formulated as $[5], [4], [3]$

$$H_u = \frac{N_{Mr}N_{Br}}{\rho} \sum_{k=1}^{K} \alpha_k a_{MS}(\theta_k) a_{BS}(\phi_k)$$

(4)

$\rho$ is the average path-loss between the BS and MS, and $\alpha_k$ is the complex gain of the $k^{th}$ path. The amplitude of these gains belong to Rayleigh distribution, in other words, $\alpha_k \sim N(0, \bar{P}_k)$, $k = 1, 2, \ldots, K$ with the average power gain $\bar{P}_k$. The azimuth angles of arrival and departure ($\theta_k, \phi_k$), respectively, are uniformly distributed between $0$ and $2\pi$. Taking the azimuth into consideration and leaving the elevation results in 2D beamforming since all the scatterers are treated to change the direction of the signal in the azimuth only. Lastly, $a_{MS}(\theta_k)$ and $a_{BS}(\phi_k)$ are the array steering vectors at the BS and MS, respectively. The array response vector at the MS can be expressed as $[5][3][4]$

$$a_{MS}(\theta_k) = \left[ \frac{1}{\sqrt{N_{Mr}}} \right]_1^N e^{j\left(\frac{2\pi}{\lambda}d\sin(\theta_k)\right)}, \ldots, e^{j\left((N_{Mr}-1)\frac{2\pi}{\lambda}d\sin(\theta_k)\right)}$$

(5)

Where $\lambda$ denotes the wavelength, and $d$ is the distance between the antenna elements. The array steering vector at the BS can be found in the same manner. In this paper, it is assumed that any prior knowledge of the channel does not exist. The channel can be simplified as$[17], [2]$

$$H = A_R \text{diag}(\alpha) A_T^H$$

(6)

Where

$$\alpha = \left[ \frac{N_{Mr}N_{Br}}{\rho} \right] \left[ \alpha_1, \alpha_2, \ldots, \alpha_K \right]$$

$$A_T = [a_T(\phi_1), a_T(\phi_2), \ldots, a_T(\phi_K)]$$

$$A_R = [a_R(\theta_1), a_R(\theta_2), \ldots, a_R(\theta_K)]$$

Using the narrowband channel model and combining the channels of all clusters, the wideband channel model can be expressed as

$$H = \sum_{u=1}^{N_c} H_u$$

(7)

Where $N_c$ denotes the number of clusters.

The received matrix ($Y$) can be vectorized and reformulated as follow$[13], [18]$

$$y_v = \text{vec}(W_T^H H F_r s + W_T^H n)$$

$$y_v = \sqrt{P} (F_r \otimes W_T^H) A_T^* \circ A_R \alpha + n_o$$

(8)

Utilizing the values of Angle Of Departures/Angle Of Arrivals, AODs/AOAs, after quantization, the received vector can be rewritten as$[18], [13]$

$$y = \sqrt{P} \Psi z + n$$

(9)

Where $y$ are the obtainable observations and $\Psi$ is considered to be the sensing matrix which can be found by $\Psi = (F_r \otimes W_T^H) A_D$. $A_D$ is a dictionary matrix whose elements can be calculated by exploiting all the AOA/AOD’s possible quantized values. The $i^{th}$ column of the dictionary matrix can be determined by $a_T(i) \otimes a_R(i)$. The vector $z$ represents the paths’ gains of the quantized directions. These gains rely on the number of paths. Considering the sparsity of the mmWave Massive MIMO channel, this channel has a limited number of paths due to the high attenuation at these bands which greatly diminishes the strength of the transmitted signals. Sometimes, the number of paths gets reduced to one, in this case, the communication channel can support one link which is considered as a SISO channel outperforming the ordinary SISO channel by leveraging the array gains only. The sensing matrix is exploited to recover the transmitted sequence. Then, by harnessing the existence of $A_D$, we can calculate the channel matrix.
3. Compressed Sensing Channel Estimation

Due to the limited number of paths, the mmWave Massive MIMO channel matrix is considered to be sparse (Referring to section I). We propose a greedy algorithm to calculate the non-zero elements of the matrix $\mathbf{H}$.

The main incentives of using the SP algorithm to predict the channel coefficients are: this algorithm is distinct for the low complexity in comparison with other greedy algorithms as the upper bound complexity can be approximated to $O(mN \log K)$ [11] (where $K$ is the channel sparsity, $m$ and $N$ are the number of sparse measurement matrix rows and columns, respectively) when signal sparsity rises slowly. Whereas the other greedy algorithms which are widely used in the state of the art literature such as OMP [19] [10] have a complexity of roughly $O(KmN)$ which indicates that SP is much more efficient in term of the complexity. Also, it is worth mentioning that this algorithm is profoundly suitable for noiseless and noisy communication channels.

In order to reconstruct the channel matrix, the algorithm first identifies the subspace, using no more than $s$ columns of the sensing matrix $\Psi$, where the observed vector $y$ lies. Next, after finding the right subspace, the pseudoinversion process is utilized to determine the non-zero elements. The method which is deployed to find the $s$ columns distinguishes the SP algorithm is: subsets of $s$ columns are examined in a group in order to refine at each stage an initially selected prediction for the subspace. More precisely, the list of $s$ columns of $\Psi$ is calculated and the algorithm also applies a basic test in the spanned space and then, the list is refined. If the current prediction for the right spanning space does not consist of $y$, the correct candidates are retained and the wrong ones are dismissed while same number of novel candidates are incorporated.

The authors of this paper are aware of a similar algorithm that is used in the literature that is called Compressive Sampling Matching Pursuit (CoSaMP) [15] [20] [21] as this algorithm appears to have the same characteristics. However, the way of how the candidates are added is considerably different. In fact, the SP algorithm exploits $K$ new candidates to be added in each iteration in contrast to the CoSaMP algorithm that it uses $2K$ new ones. Hence, the SP algorithm is computationally more efficient.

In order to perform the estimation process, the transmitted signal is sent by the BS to the receiver across the communication channel. The received vector ($y$) is used to calculate the paths’ gains of the quantized directions ($z$) using the SP algorithm shown below. Then, $z$ is utilized to estimate $\mathbf{H}$.

In order to reconstruct the channel matrix, the algorithm first identifies the subspace, using no more than $s$ columns of the sensing matrix $\Psi$, where the observed vector $y$ lies. Next, after finding the right subspace, the pseudoinversion process is utilized to determine the non-zero elements. The method which is deployed to find the $s$ columns distinguishes the SP algorithm is: subsets of $s$ columns are examined in a group in order to refine at each stage an initially selected prediction for the subspace. More precisely, the list of $s$ columns of $\Psi$ is calculated and the algorithm also applies a basic test in the spanned space and then, the list is refined. If the current prediction for the right spanning space does not consist of $y$, the correct candidates are retained and the wrong ones are dismissed while same number of novel candidates are incorporated.

Algorithm 1 is illustrated in brief as follows. First, the initialization process takes place in steps 1 and 2. We want to point out that maxi is chosen according to the performance. We executed the algorithm using different values of this parameter and we have found out that no need to increase the value to more than 10. The identification stage takes place in steps 3 and 4 in which the subspace is selected. Step 5 makes sure the non-zero elements by using the LS method. The pruning operation is carried out in step 6 which ensures that the unwanted components are removed. The algorithm never stops unless a stopping criterion is done or the maximum number of iteration is reached. Finally, in the remaining steps, $z$ is utilized to find the wideband channel matrix $\mathbf{H}$.

\textbf{Algorithm 1 Subspace Pursuit}

Input: $\Psi$ sensing matrix, observation vector $y$, the length of the approximated vector $z$, and the number of iteration $\text{maxi}$

1. Initialize: $r^0 = y$, $z^0 = 0$, $\Gamma^0 = \emptyset$
2. for $n = 1$; $n := n + 1$ till $\text{maxi}$
3. $\Omega = \{ s \text{ indices corresponding to the s largest magnitude entries in the vector } \Psi^\top(y - \Psi z^{n-1}) \}$
4. $\bar{\Gamma}^n = \Gamma^{n-1} \cup \Omega$
5. $z^n = \arg\min_{z \in \mathbb{R}^N} \{ \| y - \Psi z \|_2, \text{supp}(z) \subseteq \bar{\Gamma}^n \}$
6. $\Gamma^n = \{ s \text{ indices corresponding to the s largest magnitude elements of } z^n \}$
7. \( z^n = \arg\min_{z \in \mathbb{R}^N} \{ \| y - \Psi z \|_2, \text{supp}(z) \subseteq \Gamma^n \} \)

Until the algorithm reaches the maximum iteration or the stopping criteria is achieved
8. end for
9. \( \mathbf{H} = \mathbf{A}_D z \)
10. Output: \( \mathbf{H} \)

4. Numerical Results
In this section, we compare the performance of the proposed algorithm with the OMP, ROMP, and CoSaMP algorithms presented in [10] [13] [15] [16] [15] [20] [21]. We use three evaluation metrics which are the capacity, MMSE, and runtime metrics to find out the best algorithm in terms of speed and accuracy. The utilized parameters are: \( L_{\text{BI}} = N_s = L_{\text{MR}} = 4, \ N_{\text{Mr}} = 16, \ N_r = 4, \ N_{\text{BT}} = 32, \ K = 2 \) (number of paths in each cluster), \( n_{\text{ADC}} = 4, \ N = 16 \) symbols per frame at the receiver before the quantizer, \( M = 80 \) number of observations or frames, \( G_r = 32, \ G_t = 64 \) receiver and transmitter sections grids, respectively.

Starting with the first evaluation metric which is the MMSE calculations. As you can see in Fig.1 that the SP algorithm shows a consistent behavior along with variant values of SNR as its MMSE plummets rapidly with almost a constant high slope. The OMP algorithm shows a similar behavior except when the SNR enhances to be around -15 dB, it diverts and the gap expands as the SNR increases. On the other hand, ROMP illustrates an outstanding performance until the SNR reaches about -10 dB, the absolute slope decreases and then at around -5 dB, it plateaus. Lastly, the CoSaMP algorithm shows a sharp consistent decrease as the SNR grows, and with some improvements shown in some segments.

The proposed algorithm is further evaluated using the spectral efficiency metric represented in Fig.2 by comparing it with the optimal case which is called Perfect Channel State Information CSI. It is shown that this algorithm can nearly achieve the performance of the optimal case as the difference is as small as about 2 bps/Hz when the SNR is 20 dB. All the algorithms’ behaviors assemble each other until the SNR becomes about 0 dB, SP, OMP, and CoSaMP algorithms begin to show their remarkable performance transcending ROMP algorithm. Also, it is worth mentioning that SP, OMP, and CoSaMP algorithms achieve the same capacity performance along the range of SNR. Therefore, it is important to consider the last metric in order to induce the most efficient algorithm.

![Figure 1. The MMSE of the proposed algorithm as compared with the OMP algorithm.](image)
To ensure that the SP algorithm is computationally effective. Table 1 explains a comparison among algorithms in terms of the runtime. It also turns out that the SP algorithm is the fastest which affirms the superiority of this algorithm.

![Figure 2](image)

**Figure 2.** The achievable capacity of the proposed algorithm in comparison with the OMP algorithm taking the optimal case (Perfect CSI) as a reference.

| Algorithm’s Name | Runtime(s) |
|------------------|------------|
| SP               | 0.0100     |
| CoSaMP           | 0.0163     |
| OMP              | 0.0185     |
| ROMP             | 0.0246     |

**Table 1.** A comparison between the runtimes of OMP and SP algorithms (CPU INTEL ® CORE i5, 8th Generation, 8 GB RAM).

5. **Conclusion**
In this paper, we proposed a greedy algorithm to estimate the mmWave channel coefficients. Three evaluation metrics were used to figure out the most effective algorithm. After the simulation was carried out, we have discovered that SP, CoSaMP, and OMP algorithms approximately accomplish the same capacity with some minor differences along the SNR range. Nevertheless, the last metric which is the runtime proved that SP is the superior algorithm as it spent less time than others to precisely calculate the channel coefficients. For future work, it is expected that the performance of the SP algorithm can be elevated in terms of the capacity and runtime using techniques that substitute the LS equations that are used two times in the algorithm.

6. **References**
[1] T S Rappaport et al., “Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!,” IEEE Access, vol. 1, no. 6, pp. 335–349, 2013.
[2] M K Samimi and T S Rappaport, “Ultra-wideband statistical channel model for non line of sight millimeter-wave urban channels,” IEEE GLOBECOM, pp. 3483–3489, 2014.
[3] Y Azar et al., “28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York city,” IEEE International Conference on Communications, 2013, pp. 5143–5147.
G R Maccartney and T S Rappaport, “73 GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City,” ICC, 2014, pp. 4862–4867.

S Ju, T S Rappaport, and N Y U Tandon, “Millimeter-wave Extended NYUSIM Channel Model for Spatial Consistency,” GLOBECOM, Abu Dhabi, UAE, 2018, pp. 1–6.

S Sun, T S Rappaport, M Shafi, P Tang, J Zhang, and P J Smith, “Propagation Models and Performance Evaluation for 5G Millimeter-Wave Bands,” IEEE Trans. Veh. Technol., vol. 67, no. 9, pp. 8422–8439, 2018.

S Sun, T S Rappaport, and M Shaft, “Hybrid beamforming for 5G millimeter-wave multi-cell networks,” INFOCOM, 2018, pp. 589–596.

S Sun, T S Rappaport, R W Heath, A Nix, and S Rangan, “MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?,” IEEE Commun. Mag., vol. 52, no. 12, pp. 110–121, 2014.

A Alkhateeb, O El Ayach, G Leus, and R W Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE Journal on Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831–846, 2014.

J Lee, G T Gil, and Y H Lee, “Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems Millimeter Wave Communications,” IEEE Transactions on Communications, 2016, vol. 64, no. 6, pp. 2370–2386.

W Dai and O Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2230–2249, 2009.

C B Song, S T Xia, and X J Liu, “Improved analysis for subspace pursuit algorithm in terms of restricted isometry constant,” IEEE Signal Process. Lett., vol. 21, no. 11, pp. 1365–1369, 2014.

J Sung, J Choi, and B L Evans, “Narrowband Channel Estimation for Hybrid Beamforming Millimeter Wave Communication Systems with One-Bit Quantization,” ICASSP, 2018, pp. 3914–3918.

W Wu, R Chiueh, and F Tseng, “Channel Estimation for OFDM Systems with Subspace Pursuit Algorithm,” ICGCS, 2010, pp. 269–272.

E Hernández and D Vera, “Greedy type algorithms for RIP matrices. A study of two selection rules,” Madrid University, 2018.

H Pan, Y Xue, L Mei, and F Gao, “An improved CoSaMP sparse channel estimation algorithm in OFDM system,” ICSPTC, 2015, pp. 1–4.

S A Busari, K M S Huq, S Mumtaz, L Dai, and J Rodriguez, “Millimeter-Wave Massive MIMO Communication for Future Wireless Systems: A Survey,” IEEE Commun. Surv. Tutorials, vol. 20, no. 2, pp. 836–869, 2018.

I Chafaa and M Djeddou, “Improved channel estimation in mmWave communication system,” DAT, 2017.

A Manoj and A P Kannu, “Channel estimation strategies for multi-user mm wave systems,” IEEE Trans. Commun., vol. 66, no. 11, pp. 5678–5690, 2018.

D Needell and J A Tropp, “CoSaMP,” Commun. ACM, vol. 53, no. 12, p. 93, 2010.

H Pan, Y Xue, L Mei, and F Gao, “An improved CoSaMP sparse channel estimation algorithm in OFDM system,” ICSPTC, 2015, no. 2, pp. 2–5.