Comparison of CS-Based Channel Estimation for Millimeter Wave Massive MIMO Systems

Xingbo Lu, Weiwei Yang *, Yueming Cai and Xinrong Guan

College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China; xingbo_lu@sina.com (X.L.); caiym@vip.sina.com (Y.C.); geniusg2017@gmail.com (X.G.)

* Correspondence: wwyang1981@163.com

Received: 25 August 2019; Accepted: 11 October 2019; Published: 15 October 2019

Abstract: Compressed sensing (CS) has great potential in channel estimation for millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems. To solve such a CS-based channel estimation problem, three categories of algorithms, namely convex relaxation algorithms, greedy iteration algorithms and Bayesian inference algorithms, are widely used. In this paper, with a unified massive MIMO framework, comprehensive comparisons among three categories of algorithms are presented in the perspective of the estimated accuracy, which is affected by the received signal-to-noise ratio (SNR), the number of resolvable paths, angular quantization error, the number of pilot symbols and hardware impairments. Specifically, it shows that convex relation algorithms achieve the best estimation accuracy at the high SNR range and it is mainly affected by the received SNR and transmitter’s hardware impairments. At the low SNR range, greedy iteration algorithms outperform others and the estimated accuracy is then limited by the angle quantization error. Furthermore, a tradeoff between the estimated error and complexity is achieved in Bayesian inference algorithms, although its estimated error is sensitive to the number of available pilot symbols.

Keywords: millimeter wave; massive MIMO; channel estimation; compressed sensing

1. Introduction

Millimeter wave (mmWave) communication has been widely considered as a key technology for 5G due to its potential to provide gigabits-per-second data rates by exploiting the large unused bandwidth [1]. Unfortunately, it is challenging for long range wireless communication because of the huge paths loss in mmWave band. To compensate the paths loss and extend the coverage, massive MIMO antennas are supposed to be deployed in mmWave communication systems. However, in fell-digital MIMO structure, the radio frequency (RF) chains will increase as the antennas become large scale, which results in unaffordable hardware cost and power consumption [2]. Thus, massive MIMO with phase shifter network based hybrid precoding and electromagnetic lens based beamspace has been considered [1–3]. Generally, it is difficult for such an efficient structure to obtain the full channel state information (CSI), since the limited number of RF chains only combines the measurement signal with smaller dimension compared to the entire channel matrix. Meanwhile, many pilot symbols need to be assigned during the channel estimation phase, leading to overwhelming pilot overhead. Fortunately, some inherent properties including low rank and structural sparsity of mmWave massive MIMO channels have been demonstrated in practical measurement environments [4,5]. By exploiting low rank and structural sparsity, some channel estimation algorithms have been proposed to efficiently recover the sparse channel matrix [6–14], including blind, semi-blind and pilot-aided estimation techniques.

For blind and semi-blind channel estimation techniques, the basic idea is to exploit some inherent features in a modulated signal or use a decision feedback method rather than many pilots to acquire...
the CSI [6–8]. Specifically, the blind 1-norm regularized channel estimation algorithm was proposed to solve the narrowband channel estimation problem in [6]. In [7], a novel blind signal detection scheme that simultaneously estimates the channel and data was proposed. In [8], based on certain available prior data obtained from the output of a soft-input soft-output channel decoder, a data-aided channel estimation scheme was proposed. In such a channel estimation scheme, the CSI can be coarsely estimated via a few pilot symbols, and then its precision can be further improved by data symbols. However, for such blind or semi-blind channel estimation schemes, only after a long time receiving the data symbols can the CSI be estimated accurately. In addition, the great computational complexity will lead to a long time delay, thus such algorithms cannot be applied to estimate the channel with relatively short coherence time or meet the demand of low time delay.

To efficiently acquire the CSI, pilot-aided channel estimation algorithms are widely studied [9–14], which can be divided into three categories, i.e., convex relaxation algorithms, greedy iterative algorithms and Bayesian inference algorithms. Specifically, based on the frequency-flat geometric channel model, the CSI of complex gains, angles of arrival (AoAs) and angles of departure (AoDs) were estimated through gradient descent algorithm in [9]. In [10], by exploiting sparse and low-rank structures of the mmWave massive MIMO channel with angle spread, the two-stage compressed sensing algorithm was proposed to reduce the sample complexity. In [11], the angle spread at receiver was estimated by greedy iterative algorithms and the normalized mean squared error (NMSE) was compared among different greedy iterative algorithms. In [12], the frequency selective channel was recovered with orthogonal matching pursuit (OMP) algorithm in the single carrier (SC) system and orthogonal frequency division multiplexing (OFDM) system, respectively. In [13], the off-grid sparse Bayesian learning algorithm was proposed to estimate CSI in an uplink multiuser mmWave massive MIMO system. In [14], the Bayesian compressed sensing (BCS) algorithm was adopted to recover the channel matrix based on a geometric channel model, where the transmitter’s hardware impairments were taken into consideration. Note that the prior works have investigated these channel estimation algorithms based on different assumptions and drawn some conclusions. However, these conclusions are only for specific parameter ranges under different scenarios. To the best of our knowledge, there is no comprehensive comparison among CS-based channel estimation algorithms with a unified framework for mmWave massive MIMO systems.

Inspired by the above analysis, in this paper, we formulate the sparse channel estimation as a normal CS problem and provide a general overview of the current mmWave massive MIMO channel estimation approaches. Different from the previous works, the purpose of our work is a comprehensive comparison among existing algorithms with a unified framework, including their basic assumptions, application conditions and key results. The main contributions and key results of this work are listed as follows.

- In this paper, with a unified massive MIMO framework, we compare the NMSE performance among three categories of algorithms, which is affected by the received SNR, the number of resolvable paths and pilot symbols, angular quantization error, hardware impairments and computational complexity. Through comprehensive comparison, the characteristics and application conditions of each algorithm are revealed and the factors that affect the estimated error and computational complexity are also presented.
- Through theoretical analysis and simulation results, we show that convex relation algorithms achieve the best estimation accuracy at the high SNR range and it is mainly affected by the received SNR and transmitter’s hardware impairments. At the low SNR range, greedy iteration algorithms outperform others and the estimated accuracy is then limited by the angle quantization error. Furthermore, a tradeoff between the estimated error and complexity is achieved in Bayesian inference algorithms, although its estimated error is sensitive to the number of available pilot symbols.
- We also analyze the overall computational complexities of three categories of algorithms and visually represent them by the running time. Through illustrating the runtime of different
algorithms versus the sparseness, we show that the computational complexity in the convex relaxation algorithm is the highest, and it even squarely increases with the sparseness in the gradient descent-based convex algorithm, while that in greedy iteration algorithms is minimum and grows linearly with the sparseness. In contrast to them, the computational complexities of Bayesian inference algorithm decreases as the sparseness increases.

The rest of the this paper is organized as follows. After describing the system model and channel model in Section 2, we formulate the channel estimation as a compressed sensing (CS)-based sparse signal recovery problem in Section 3. A comparison of sparse signal recovery algorithms, namely convex relaxation algorithms, greedy iteration algorithms and Bayesian inference algorithms, are presented in Section 4. Simulation results are provided in Section 5. Finally, we draw conclusions in Section 6.

Throughout our discussions, bold uppercase \( \mathbf{A} \) denotes a matrix and lowercase \( \mathbf{a} \) denotes a vector. Superscripts \( \mathbf{A}^\ast \), \( \mathbf{A}^T \), \( \mathbf{A}^H \), and \( \mathbf{A}^{-1} \) denote the conjugate, the transpose, the conjugate transpose, and the inverse of a matrix \( \mathbf{A} \), respectively. \( \text{vec} (\mathbf{A}) \) denotes a vector obtained through the vectorization of a matrix \( \mathbf{A} \). For \( M \times N \) matrices \( \mathbf{A} \) and \( \mathbf{B} \), \( \mathbf{A} \otimes \mathbf{B} \) denotes the \( M^2 \times N^2 \) matrix of Kronecker product between \( \mathbf{A} \) and \( \mathbf{B} \).

\[ \| \mathbf{a} \|_0, \| \mathbf{a} \|_1, \text{ and } \| \mathbf{a} \|_2 \] are the \( l_0 \), \( l_1 \), and \( l_2 \) norms, respectively.

2. System Model

2.1. System Model

Without loss of generality, we consider a typical mmWave point-to-point hybrid massive MIMO system, as shown as Figure 1. Let \( N_T \) and \( N_R \) denote the number of antennas at the transmitter and receiver, respectively. For simplicity, we assume the number of RF chains deployed at both sides is equal, i.e., \( N^T_{RF} = N^R_{RF} = N_{RF} \). Since the RF chains are costly and due to power consumption, the number of RF chains is much smaller than that of antennas, i.e., \( \min \{ N_T, N_R \} \gg N_{RF} \).

For such an mmWave massive MIMO system with hybrid structure, the transmitter and receiver communicate via \( N_s \) streams where \( N_s \leq N_{RF} \). Specifically, the transmitter employs a baseband precoder \( \mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_s} \) followed by an analog phase shifter precoder \( \mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_{RF}} \). After the combined transmitter precoding matrix \( \mathbf{F} = \mathbf{F}_{RF} \mathbf{F}_{BB} \in \mathbb{C}^{N_T \times N_s} \), the transmitted signals can be written as \( \mathbf{x} = \mathbf{Fs} \in \mathbb{C}^{N_T \times 1} \), where the \( i \)th element of \( \mathbf{x} \) is the transmitted signal at the \( i \)th antenna, \( \mathbf{s} \) is the transmitted symbol vector which satisfies \( \mathbb{E} \{ \mathbf{s} \cdot (\mathbf{s})^H \} = \frac{P}{N_s} \mathbf{1} \), and \( P \) is the total power of transmitted symbols [12]. During the channel estimation phase, it is worth noting that the pilot power and pilot pattern also affect the performance of sparse channel estimation, which need to be carefully designed. However, to simplify expression, we assume the pilot symbols are evenly placed and transmitted with equal power.
At the receiver, the received signals on all antennas are combined with a combiner \( \mathbf{W} \in \mathbb{C}^{N_R \times N_t} \) composed of the RF combiner \( \mathbf{W}_{RF} \in \mathbb{C}^{N_R \times N_{RF}} \) and baseband combiner \( \mathbf{W}_{BB} \in \mathbb{C}^{N_{RF} \times N_t} \), which can be expressed as

\[
y = \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{s} + \mathbf{W}^H \mathbf{n}
\]

where \( \mathbf{H} \in \mathbb{C}^{N_T \times N_R} \) denotes the channel matrix discussed in the next subsection and \( \mathbf{n} \sim \mathcal{CN} ( \mathbf{0}, \sigma^2 \mathbf{I}_{N_t} ) \) denotes the symmetric complex Gaussian distributed additive noise vector at the receiver.

In a practical communication system, the non-ideal characteristics of components, including nonlinear power amplifier [14], inphase/quadrature-phase imbalance [15], phase noise [16] and carrier frequency offset [17], will greatly affect the estimation performance. Especially in mmWave band, the use of cheap equipment is encouraged, which leads to unignorable hardware impairments. To analyze the effect of hardware impairments on the performance of channel estimation algorithms, we take the residual transmitter hardware impairments into consideration and model them as the additive distortion noise [18]. Specifically, the additive distortion noise denoted by \( \mathbf{e}_t \in \mathbb{C}^{N_T \times 1} \) can be modeled as a Gaussian distributed random variable and its power is proportional to the pilot symbols power. That is, \( \mathbf{e}_t \sim \mathcal{CN} ( \mathbf{0}, \Lambda ) \) and \( \Lambda \overset{\Delta}{=} k_e (q_{11}, \ldots, q_{N_{RF} N_T})^T \), where \( q_{ii} \) denotes the \( i \)th element of covariance matrix \( \mathbf{Q} \overset{\Delta}{=} E (\mathbf{x} \mathbf{x}^H) \) and \( k_e \) denotes the proportionality coefficient [19]. In addition, we assume that the hardware impairments at the transmitter are uncorrelated with the transmitted signals \( \mathbf{x} \) and ignore the hardware impairments at the receiver as they can be simply treated as additive noise during the channel estimation phase. Then, the combined signals at the receiver can be modeled as

\[
y = \mathbf{W}^H \mathbf{H} (\mathbf{F}_t + \mathbf{e}_t) + \mathbf{W}^H \mathbf{n}
\]

2.2. Channel Model

In a practical massive MIMO system, the transmitters, such as base stations, are typically located at high elevation with fewer surrounding scatterers, while the receivers, such as user equipment, are surrounded by rich scatterers at low elevation. Thus, each path departing from the transmitter may arrive at several reception directions [11]. Moreover, the channel is time-varying as the multipath delay and frequency offset are common and it is often useful to represent the channel response in frequency domain, which can be expressed as [20,21]

\[
\mathbf{H} (t, f) = \sum_{l=1}^{L_{AoA}} \sum_{q=1}^{L_{AoA}} a_{l,q} e^{j2\pi (v_l q - \tau_l d)} \mathbf{a}_R (\theta_{l,q}^R) \mathbf{a}_T^H (\theta_{l,q}^T)
\]

where \( L \) denotes the number of paths that depart from the transmitter with different AoDs; \( L_{AoA} \) denotes that each of departure path will contribute \( L_{AoA} \) reception directions; \( a_{l,q}, v_l, \tau_l \) and \( \theta_{l,q}^R \), respectively, denote the complex gain, Doppler shift, delay and the AoA of the \( l \)th departure path; \( \theta_{l,q}^T \in [0, 2\pi) \) denotes the AoD of the \( l \)th path; and \( \mathbf{a}_T (\theta_{l,q}^T) \) and \( \mathbf{a}_R (\theta_{l,q}^R) \) denote the antenna array response vectors at the transmitter and receiver, respectively. For a uniform linear array (ULA) antenna, the array response vectors at transmitter and receiver can be presented as

\[
\mathbf{a} (\theta_{l,q}^T) = \begin{bmatrix}
e^{j2\pi d \sin \theta_{l,q}^T / \lambda}, \ldots, ne^{j2\pi (N-1)d \sin \theta_{l,q}^T / \lambda}
\end{bmatrix}^T
\]

\[
\mathbf{a} (\theta_{l,q}^R) = \begin{bmatrix}
e^{j2\pi d \sin \theta_{l,q}^R / \lambda}, \ldots, ne^{j2\pi (N-1)d \sin \theta_{l,q}^R / \lambda}
\end{bmatrix}^T
\]

where \( \lambda \) is the wavelength of signal and \( d \) is the antenna spacing, which is typically assumed as \( d = 0.5 \lambda \).

It is difficult to estimate such a channel matrix as five parameters need to be estimated in each path: the complex gain \( a_l \), the angle of arrival \( \theta_{l,q}^R \), the angle of departure \( \theta_{l,q}^T \), the Doppler shift \( v_l \) and the delay \( \tau_l \). Besides, although there are only 2–8 paths on average in the mmWave band [22],
the actual number of paths also needs to be estimated, which makes it more difficult to estimate the CSI. Therefore, it is necessary to simplify the channel model according to the practical scenarios.

Suppose that the channel is slowly varying over the signal duration of interest $T$, which means the channel is quasi-static and the Doppler frequency shift can be ignored [21,23]. Based on such assumption, we can rewrite Equation (3) as

$$H = \sum_{l=1}^{L} \sum_{q=1}^{L_{\text{OA}}} a_{l,q} e^{-j2\pi\tau_{l,q}\theta_{R}^{l,q}} a_{R}^{H}(\theta_{T}^{l})$$  

(5)

In addition, if we further assume that the bandwidth $B$ of channel is sufficiently small (i.e., $B\tau_{l} \ll 1 \forall l, l = 1, \cdots, L$), the channels of mmWave massive MIMO system can be simplified as a narrowband frequency-flat physical channel model [11]. Then, Equation (5) can be simplified as

$$H = \sum_{l=1}^{L} a_{l} a_{R}^{H}(\theta_{R}^{l}) a_{T}^{H}(\theta_{T}^{l})$$  

(6)

Although there are abundant scatterers surrounding the ends, the penetration and reflection loss are also significant at mmWave band. Thus, a geometric channel model is widely used in the prior work [9,10,24], which can be expressed as

$$H = \sum_{l=1}^{L} a_{l} a_{R}^{H}(\theta_{R}^{l}) a_{T}^{H}(\theta_{T}^{l})$$  

(7)

To simplify the expression, we adopt the multipath narrowband frequency-flat channel model in Equation (7) to compare three categories channel estimation algorithms. Note that Equation (7) can be further expressed by a more compact way

$$H = A_{R} \Delta_{d} A_{T}^{H}$$  

(8)

where $\Delta_{d} \in \mathbb{C}^{L \times L}$ is a diagonal matrix that contains non-zero elements of channel gains, and $A_{R} \in \mathbb{C}^{N_{R} \times L}$ and $A_{T} \in \mathbb{C}^{N_{T} \times L}$ contain the columns $a_{R}(\theta_{R}^{l})$ and $a_{T}(\theta_{T}^{l})$, respectively.

With the multipath sparsity of physical channel, a discrete virtual angular domain (VAD) channel model is widely used [9–13]. Assume that the AoAs and AoDs are taken from a uniform grid of size $G$, i.e., $\theta_{R}^{l}, \theta_{T}^{l} \in \left\{0, \frac{2\pi}{G}, \cdots, \frac{2\pi(G-1)}{G}\right\}$ with $G \gg L$. Then, Equation (8) can be rewritten as

$$H = \tilde{A}_{R} \Delta_{d}^{\circ} \tilde{A}_{T}^{H} + E_{Q}$$  

(9)

where $\Delta_{d}^{\circ} \in \mathbb{C}^{G \times G}$ is a $L$-sparse VAD channel matrix with $L$ non-zero elements in the positions corresponding to the AoAs and AoDs, $\tilde{A}_{T} \in \mathbb{C}^{N_{T} \times G}$ and $\tilde{A}_{R} \in \mathbb{C}^{N_{R} \times G}$ are the transformation dictionaries that conation the array response vectors corresponding to the angles in the grid, and $E_{Q}$ denotes the quantization error that is bounded, as analyzed in [25]. In addition, a suitable choice of $G$ is crucial as a larger $G$ can reduce the quantization error but increases the computational complexity, vice versa. Typically, it is assumed that the number of grids $G$ is larger than the number of antennas [25].

3. Formulation of the Channel Estimation Problem via Compressed Sensing

In this section, we formulate a technique via CS theory to estimate the complex gain $a_{l}$, the angle of arrival $\theta_{R}^{l}$ and the angle of departure $\theta_{T}^{l}$ with a unified framework, leveraging the sparsity of mmWave channel.
To formulate a sparse signal recovery problem of channel estimation, we substitute Equation (9) into Equation (2) and use the result of Kronecker product (i.e., $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$) to transform it as

$$\text{vec}(\mathbf{y}) = \left( (\mathbf{F}_s + \mathbf{e}_l)^T \otimes \mathbf{W}^H \right) (\hat{\mathbf{A}}^*_T \otimes \hat{\mathbf{A}}_R) \text{vec}(\Delta_s^p) + (\mathbf{F}_s + \mathbf{e}_l)^T \otimes (\mathbf{W})^H \text{vec}(\mathbf{E}_Q) + \text{vec}(\mathbf{W}^H \mathbf{n}) \tag{10}$$

According to CS theory, sufficient measurement vectors are needed to guarantee the sparse vector recovered with high probability. Assume that $M$ training frames are occupied and both sides use pseudorandomly precoders and combiners during the channel estimation phase. After collecting the combined signals at the receiver, Equation (10) can be extended to a standard expression based on CS theory.

$$\begin{bmatrix} \text{vec}(\mathbf{y}^{(1)}) \\ \vdots \\ \text{vec}(\mathbf{y}^{(M)}) \end{bmatrix}_{\mathbf{y}} = \begin{bmatrix} (\mathbf{F}_s^{(1)} \mathbf{s}^{(1)} + \mathbf{e}_l^{(1)})^T \otimes (\mathbf{W}^{(1)})^H \\ \vdots \\ (\mathbf{F}_s^{(M)} \mathbf{s}^{(M)} + \mathbf{e}_l^{(M)})^T \otimes (\mathbf{W}^{(M)})^H \end{bmatrix}_{\mathbf{f}} \begin{bmatrix} (\hat{\mathbf{A}}^*_T)^{(1)} \otimes (\hat{\mathbf{A}}_R)^{(1)} \\ \vdots \\ (\hat{\mathbf{A}}^*_T)^{(M)} \otimes (\hat{\mathbf{A}}_R)^{(M)} \end{bmatrix}_{\mathbf{h}} + \begin{bmatrix} \text{vec}(\Delta_s^p)^{(1)} \\ \vdots \\ \text{vec}(\Delta_s^p)^{(M)} \end{bmatrix}_{\mathbf{e}_N} + \begin{bmatrix} \text{vec}(\mathbf{e}_N^{(1)}) \\ \vdots \\ \text{vec}(\mathbf{e}_N^{(M)}) \end{bmatrix}_{\mathbf{e}_N} \tag{11}$$

where $\Phi^{(m)} = (\mathbf{F}_s^{(m)} \mathbf{s}^{(m)} + \mathbf{e}_l^{(m)})^T \otimes (\mathbf{W}^{(m)})^H$, $m \in [1, M]$ is defined as the measurement matrix, which is assumed to be known at the receiver for standard CS-based model. However, the transmitter hardware impairments will prevent the receiver from fully knowing the real measurement matrix. $\Psi^{(m)} = (\hat{\mathbf{A}}^*_T)^{(m)} \otimes (\hat{\mathbf{A}}_R)^{(m)}$ is defined as the orthogonal dictionary matrix. $\mathbf{h}^{(m)} = \text{vec}(\Delta_s^p)^{(m)} \in \mathbb{C}^{2 \times 1}$ is a sparse vector that contains all complex channel gains. $\mathbf{e}_N^{(m)} = (\mathbf{F}_s^{(m)} \mathbf{s}^{(m)} + \mathbf{e}_l^{(m)})^T \otimes (\mathbf{W}^{(m)})^H \text{vec}(\mathbf{E}_Q^{(m)}) + \text{vec}(\mathbf{W}^{(m)} \mathbf{n}^{(m)})$ is defined as the additive noise vector that contains the quantization error, hardware impairments and additive white Gaussian noise (AWGN) at the receiver.

For such a sparse signal recovery problem, it has been widely accepted [9–14] that the parameters, including the complex gain $a_i$, the angle of arrival $\theta_i^R$ and the angle of departure $\theta_i^T$, can be estimated by the following optimization problem

$$\min_{\mathbf{h}} \| \mathbf{h} \|_0 \quad \text{s.t.} \quad \| \mathbf{y} - \Phi \Psi \mathbf{h} \|_2^2 < \varepsilon \tag{12}$$

where $\varepsilon$ is a parameter that denotes the tolerance of estimation error and the variance of the noise is a suitable choice for it.

4. A Comparison of Sparse Signal Recovery Algorithms

In this section, the properties of different algorithms are revealed and the factors that affect the estimated accuracy and computational complexity are also presented. Further, the estimated error and complexity are compared for three kinds of algorithms.

4.1. Convex Relaxation Algorithms

(1) CVX-based Algorithm: It is hard to solve an optimization problem when the objective function to minimize the $l_0$ norm is nonconvex. Fortunately, $l_1$ norm has been proved to approximate it well and widely used in prior works [9–11]. By transforming the objective function from $l_0$ norm to $l_1$ norm, the optimization problem in Equation (12) can be formulated as a convex optimization problem, which can be expressed as

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \| \mathbf{h} \|_1 + \lambda \| \mathbf{y} - \Phi \Psi \mathbf{h} \|_2 \tag{13}$$
where $\lambda$ controls the tradeoff between the sparseness and estimation error and it is a key parameter that affects the rate of convergence, which needs to be updated in each iteration.

To work out such a convex optimization problem, there are numerous toolboxes such as CVX and SeDumi [26,27]. For the existing toolboxes, we can directly use them since the algorithms have been integrated into software such as MATLAB. However, the toolboxes are restricted by the hardware because the size of channel matrix may exceed the hardware memory when the transmitter and receiver deploy massive MIMO antennas. In addition, transforming the sparse channel recovery as a convex optimization problem will cause some performance loss and the quantization error of angles will limit the estimated performance. Besides, some extra procedures are included in the common toolboxes, which will increase the computational complexity of algorithm.

(2) IR-based Algorithm: Except the existing toolboxes, authors in [9] adopted the gradient descent and iterative reweighting (IR) method to solve the minimum $l_0$ norm problem. By replacing the objective function with a log-sum function, the optimization problem in Equation (12) can be rewritten as

$$
\min_{\mathbf{h}, \mathbf{\theta}_T, \mathbf{\theta}_R} S^{(i)}(\mathbf{h}, \mathbf{\theta}_T, \mathbf{\theta}_R) = \lambda^{-1} \mathbf{h}^H \mathbf{D}^{(i)} \mathbf{h} + \| \mathbf{y} - \Phi \Psi \mathbf{h} \|^2_2 
$$

where $\mathbf{D}^{(i)}$ is a diagonal matrix and its elements are the estimation of path gains at the $i$th iteration. Then, by defining the maximal tolerance of channel gain error and the corresponding optimal value of objective function $S$ at the $i$th iteration, we can update the $\hat{\mathbf{\theta}}_R^{(i+1)}$ and $\hat{\mathbf{\theta}}_T^{(i+1)}$ and compute the residue in each iteration via gradient descent method until the error is less than the threshold or iteration reaches the maximum.

By using the gradient descent and IR method, the quantization error of angles and limited hardware memory can be effectively avoided. However, the complexity increases as complex matrix operations such as derivation are involved, which cannot meet the requirement of low delay. To reduce the computational complexity, some preconditioning procedures can be adopted such as the suitable choice of initial candidates and the iteration step size [9].

4.2. Greedy Iterative Algorithms

The typical representation of the greedy iterative is OMP algorithm. Its main idea is to find the most relevant column contained in the measurement matrix to the residue in the $i$th iteration [25].

$$
j^{(i)} = \arg \max_{j \notin \hat{\Omega}^{(i-1)}} \left\| \left( \mathbf{r}^{(i)} \right)^H \left( \Phi \Psi \right)_j \right\|_2 
$$

where $\mathbf{r}^{(i)}$ is the residual vector, $(\Phi \Psi)_j$ denotes the $j$th column of the measurement matrix and $\hat{\Omega}$ denotes the estimation of the common support. $\hat{\Omega}$ can be updated by

$$
\hat{\Omega}^{(i)} = \hat{\Omega}^{(i-1)} \cup j^{(i)} 
$$

After the sparse common support $\Omega$ is estimated, the least square estimation algorithm can be used to estimate the full channel, which is given as follows:

$$
\hat{\mathbf{h}} = \left( \Omega^H \Omega \right)^{-1} \Omega \mathbf{y} 
$$

Except OMP algorithm, there are some modified greedy iteration algorithms such as subspace pursuit (SP) algorithm to solve such a problem [11]. For SP algorithm, the relevant $L$ columns are selected in each iteration to reduce iterations. For greedy iterative algorithms, a suitable criterion to end the procedure is crucial. If the sparseness of channel is known in advance, the algorithms will be stopped when the iterations reach the sparseness; otherwise, the maximum iteration will be set
larger than the real channel sparseness to ensure the recovery accuracy. However, with regard to SP algorithm, there is an extra threshold to exceed since numerous columns are selected in each iteration.

The defect of greedy iteration algorithms is that they cannot avoid the quantification error of angle even when $G$ is large enough. Especially at the high SNR range, the quantization error is a key factor that plays a decisive role. This indicates that the performance based on greedy iterative algorithms is similar because the quantification error of angle is a bottleneck for such a sparsity channel estimation problem.

Although the performance loses a lot by replacing the real angle of arrival and departure by discrete grid angle, the computational complexity is relatively lower than convex relaxation algorithms. This is due to the fact that the greedy iterative algorithms do not involve the complicated matrix operations such as derivation. Hence, the greedy iterative algorithms are candidates to solve such a sparse channel estimation problem as the low delay requirement for next wireless communication can be satisfied well.

4.3. Bayesian Inference Algorithms

The Bayesian inference algorithms are based on the Bayesian theorem and their main idea can be summarized as follows [28]. Firstly, assume the sparsity prior distribution of parameters that need to be estimated. Secondly, transform the prior to posterior distribution based on Bayesian theorem. Thirdly, find the optimal value of parameters to maximize the posterior probability. The prior assumption is crucial as it controls the computational complexity and recovery accuracy of algorithms. To enlarge the sparseness and formulate a closed-form expression after Bayesian inference, the hierarchical sparsity prior is widely used in the sparsity Bayesian learning, particularly in the relevance vector machine (RVM) [28]. Specifically, define a zero-mean Gaussian prior of each element in $h$, and then the probability density function (PDF) can be expressed as

$$ p(h|\alpha) = \prod_{i=1}^{L} N(h_i|0,\alpha_i^{-1}) $$

where $\alpha_i$ is the precision of a Gaussian density function. A Gamma prior is considered over $\alpha$

$$ p(\alpha|a,b) = \prod_{i=1}^{L} \Gamma(\alpha_i|a,b) $$

Based on Bayesian inference, we can express the posterior for $h$ as a Gaussian distribution with the mean and variance given as follows

$$ \begin{align*}
\mu &= \alpha_0 \Sigma \Phi^T h \\
\Sigma &= (\alpha_0 \Phi^T \Phi + A)^{-1}
\end{align*} $$

where $A = \text{diag}(\alpha_1,\alpha_2,\ldots,\alpha_L)$. By marginalizing over the $h$, the logarithm marginal likelihood for $a$ and $a_0$ can be expressed as

$$ L(\alpha,a_0) = \log \int p(y|h,a_0) p(h|\alpha) \, dh $$

$$ = -\frac{1}{2} \left[ K \log (2\pi) + \log |C| + y^T C^{-1} y \right] $$

where $C = \sigma^2 I + \Phi A^{-1} \Phi^T$. By using the type-II ML [29] to maximize Equation (21), the parameters $\alpha$ and $a_0$ can be updated as follows

$$ \begin{align*}
\alpha_i^{\text{new}} &= \frac{\gamma_i}{\mu_i^2} \\
p_0^{-1} &= \frac{\|y - \Phi \mu\|^2}{K \sum_i \gamma_i}
\end{align*} $$

where $\mu_i$ is the $i$th element of $\mu$ and $\gamma_i = 1 - \alpha_i \Sigma_{ii}$. 
It is worth noting that the channel estimation in mmWave massive MIMO involves complex signal recovery so that the traditional Bayesian inference algorithm designed to recover the real signal in [30,31] needs to be improved to fit the complex signal. A feasible method is to slip the real and imaginary components as follows

$$\begin{bmatrix}
R(y) \\
I(y)
\end{bmatrix} = \begin{bmatrix}
R(\Phi \Psi) & -I(\Phi \Psi) \\
I(\Phi \Psi) & R(\Phi \Psi)
\end{bmatrix} \begin{bmatrix}
R(h) \\
I(h)
\end{bmatrix} + \begin{bmatrix}
R(e_N) \\
I(e_N)
\end{bmatrix}$$

(23)

where \(R(\cdot)\) denotes the real component and \(I(\cdot)\) denotes the imaginary component. As a result, the dimensions will increase so a simplified model for Bayesian inference is essential and the main motivation of adopting the RVM is because of its highly efficient computation.

Based on above analysis, the Bayesian compressed sensing (BCS) algorithm can be used to recover the sparse channel through proper conversion. In addition, as the BCS algorithm highly depends on the statistical distribution of parameters and more available data at receiver can fit the prior distribution better, the sparseness and the number of pilot symbols are key factors that affect the recovery accuracy. Moreover, the BCS algorithm may break out early when the recovery accuracy reaches the threshold.

**Remark 1:**
The overall computational complexities of three categories of algorithms are compared in Table 1. For convex relaxation such as IR-based algorithm, its computational complexity will grow squarely with the sparseness increase, since there are gradients of \(\theta_T\) and \(\theta_R\) that need to be calculated in each iteration [9]. For greedy iteration algorithms, its computational complexity will grow linearly with the sparseness due to the number of iterations depends on the sparseness [25]. With respect to Bayesian inference algorithms, its computational complexity will mainly depend on the amount of available pilot symbols at receiver where more available pilot symbols will let the algorithm break out early [30]. In Table 1, we can see that the computational complexity of all algorithms is quite high when deployed on massive MIMO antennas. The CS-based algorithms proposed in the previous works mainly aim at improving the estimation accuracy rather than reducing the computational complexity. However, reducing the computational complexity without significantly decreasing the accuracy is a direction that is worth studying.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convex Relaxation Algorithms [9]</td>
<td>(O\left(N_{RF}^2 N_{RF}^T (N_T + N_R) L^2\right))</td>
</tr>
<tr>
<td>Greedy Iterative Algorithms [25]</td>
<td>(O\left(N_T N_R L\right))</td>
</tr>
<tr>
<td>Bayesian Inference Algorithms [30]</td>
<td>(O\left(N_R \log^2 N_R\right))</td>
</tr>
</tbody>
</table>

5. Simulation Results

In this section, we evaluate the complexity and NMSE performance of three categories of algorithms through Monte Carlo simulation with the following parameters: (1) Both transmitter and receiver are equipped with ULAs where \(N_T = N_R = 16\) and \(N_{RF}^T = N_{RF}^R = N_{RF} = 4\). (2) The \(\left\{\sin \theta\right\}\) appearing in the array response vector is uniformly distributed in \([-1,1]\) [25]. (3) The received SNR is defined as \(\text{SNR}_n = 10 \log_{10} \left(\frac{P}{\sigma^2}\right)\) dB. In addition, the hardware impairment at transmitter is defined as \(\text{SNR}_e = 10 \log_{10} \left(\frac{1}{k_e}\right)\) dB, where the ratio coefficient \(k_e\) reflects the degree of hardware impairment. Without loss of generality, we set \(k_e = 0.05\), which is larger than that in the long-term evolution (LTE) standard as the signals in mmWave band are more vulnerable to hardware impairments [14]. In addition, we take the NMSE as the performance metric to evaluate the estimated quality, which is defined as

$$\text{NMSE} = E \left\{ \frac{\|\hat{H} - H\|_2^2}{\|H\|_2^2} \right\}$$

(24)

where \(\hat{H}\) denotes the estimated channel.
5.1. Comprehensive Comparison of Estimation Quality

In Figure 2, we compare the NMSE performance of different methods with respect to received SNR. We can observe that the performance of greedy iterative algorithms are best at the low SNR range. This is because the noise will be treated as the non-zero elements, which badly destroy the sparsity of the channel and reduce the amount of available pilot symbols. However, the gradient descent and iterative reweighting method are best at the high SNR range. It can also be observed that the performance of greedy iterative algorithms will remain unchanged with the SNR improving at the high SNR range because the estimated error caused by noise becomes smaller and smaller, but the quantization error decided by the resolution of angle quantization remains unchanged. Meanwhile, we can see the NMSE of Bayesian inference algorithm linearly decreases with the SNR improving, approximatively.

In Figure 3, we illustrate the NMSE performance versus the sparseness $L$. We can see that the NMSE of greedy iteration algorithms remains unchanged with the sparseness increasing, approximately, since the most relevant columns are selected in each iteration. The performance of IR algorithm also remains unchanged. Moreover, it is better than others at the low sparseness range. In addition, the performance of BCS and CVX algorithms improve as the sparseness increases. This is because the assumptions of prior distribution and the condition that let $l_1$ norm replace $l_0$ norm are fitted better when the sparseness increases. Furthermore, we can observe that the performance of BCS algorithm is better than greedy iterative algorithms when the sparseness is larger than 3 and the gap even reaches 8 dB when the sparseness is 7.

Figure 4 depicts pilot symbol frames versus the NMSE performance where each frame is assumed to be length of 16 pilot symbols. It is obvious that the NMSE becomes smaller with the increasing of pilot symbols for the three categories of algorithms. However, for the OMP and BCS algorithms, the NMSE curves become flat as the number of pilot symbol frames becomes large enough. The NMSE of channel estimated by convex relaxation algorithm such as basis pursuit (BP) algorithm improves with the pilot symbol frames increasing. The figure shows that the performance of BP and BCS algorithms even make the channel estimation results unusable at the low SNR range, but they are obviously better than OMP algorithm at the high SNR range. The figure also shows that the OMP algorithm is optimal in the range of low SNR with a few pilot symbols. After all, the estimation quality of OMP algorithm is least affected by the pilot symbols, while the BP algorithm is most affected.

![Figure 2](image_url)  
**Figure 2.** NMSE of the estimated CSI versus received SNR with sparseness $L = 5$, pilot symbol frames $M = 16$, angle quantization $G = 32$ and without hardware impairment.
Figure 3. NMSE of the estimated CSI versus sparseness with received SNR = 10 dB, pilot symbol frames $M = 16$, angle quantization $G = 32$ and without hardware impairment.

Figure 4. NMSE of the estimated CSI versus received pilot symbols with sparseness $L = 5$, received SNR = 10 dB, angle quantization $G = 32$ and without hardware impairment.

In Figure 5, we illustrate the NMSE performance versus the angle quantization $G$. We can see that the performance of BP algorithm is hardly affected by angle quantization, which means it can avoid the quantization error as the AoAs and AoDs are accurately estimated by gradient descent method in each iteration. However, the estimation quality of BP algorithm is worse than OMP and BCS algorithms when the angle quantization is larger than 20. For the OMP algorithm, the MNSE decreases as angle quantization increases, which means the angle quantization is a key factor that affects the estimation performance of OMP algorithm. As to the BCS algorithm, the NMSE curve goes down rapidly and then stays flat, which indicates there is an optimal angle quantization $G$ for it.

In Figure 6, we show the effect of transmitter’s hardware impairments on channel estimation quality, where the noise contained in the SNR is the AWGN without considering the angle quantization error and the hardware impairments. We can see that the estimation quality of all algorithms is affected by the transmitter’s hardware impairments, especially at the high SNR range. This is because the
transmitter’s hardware impairment can be regarded as the extra noise and it is the main component of noise at the high SNR range. It is worth mentioning that the hardware impairments will increase with the increase of transmission power, which means simply increasing the transmission power cannot avoid the hardware impairments. Therefore, estimation and compensation are necessary for a practical mmWave massive MIMO system to combat them.

Figure 5. NMSE of the estimated CSI versus received angle quantization (G) with received SNR = 10 dB, sparseness L = 5, pilot symbol frames M = 40 and without hardware impairment.

Figure 6. NMSE of the estimated CSI versus received SNR with hardware impairment = 13 dB, sparseness L = 5, angle quantization G = 32 and pilot symbol frames M = 40.

In Figure 7, we illustrate the success probability of channel estimation versus the number of antennas at the transmitter. Without loss of generality, we set $2N_{RF}^T = N_T$, $G_T = 2N_T$ and $\gamma_{th} = -10$ dB, which indicates the number of RF chains $N_{RF}$ and the uniform grid of angle quantization $G$ linearly
increase when the antennas as transmitter increase, and the channel estimation is successful when the estimated error is less than $-10$ dB. We can observe that the success probability increases and then remains unchanged as the number of antennas increases for OMP algorithm because the angle quantization error at transmitter limits the success probability when the number of antennas is small, but the angle quantization error at receiver is the decisive factor when the antennas of transmitter is large enough. The same phenomenon can be observed in the BCS algorithm. However, different from the OMP algorithm, the number of available pilot symbols is the main factor that limits the success probability of BCS algorithm. As for BP algorithm, the number of antennas has little effect on the success probability, since the NMSE of BP algorithm is mainly affected by SNR at the receiver.

![Figure 7](image)

**Figure 7.** Success probability of channel estimation versus the number of antennas ($N_T$) with received $SNR = 10$ dB, $N_R = 4$, $N_R^{RF} = 4$, $\gamma_{th} = -10$ dB, sparseness $L = 5$ and pilot symbol frames $M = 16$ and without hardware impairment.

Based on above simulation results and analysis, a comprehensive comparison of different algorithms can be formed. Specifically, the convex relaxation algorithms are mainly affected by received SNR and transmitter’s hardware impairments, while the angle quantization error is the key factor that limits the performance of greedy iteration algorithms and all referred factors have an impact on Bayesian inference algorithms.

### 5.2. Computation Complexity versus Sparseness

Regarding to the computation complexity, a visual method is to compare the runtime of each estimator [11]. We performed simulations in MATLAB R2015b using a 3.0 GHz Inter Core i5-7400 CPU with 8 GB of memory and used 5000 Monte Carlo simulation trials to average the results. To ensure the fairness, we only treated the direct computational procedure as the runtime [11].

In Figure 8, we can observe that the runtime of IR algorithm increases squarely with the sparseness, approximately. At the low sparseness range, calculating the gradient has no significant contribution to runtime but it is the main source when the sparseness increases. The runtime of CVX algorithm remains unchanged but much longer than others as the CVX is a common toolbox and its runtime is not affected by the sparseness. It is worth mentioning that the runtime of BCS algorithm decreases with the sparseness increasing because the algorithm will break out ahead when the accuracy is satisfied. On the other hand, it can been observed that the runtime of greedy iterative algorithms is shortest and...
grows linearly with the sparseness increasing, since there are no complicated operations of the matrix. Among the greedy iterative algorithms, the runtime of OMP algorithm is longest as only one column is selected in each iteration.

![Figure 8. Runtime of the estimation algorithms versus the sparseness with received SNR = 10 dB, angle quantization $G = 32$, pilot symbol frames $M = 16$ and without hardware impairment.](image)

6. Conclusions

In this paper, with a unified massive MIMO framework, we overview the CS-based mmWave channel estimation algorithms, namely convex relaxation algorithms, greedy iteration algorithms and Bayesian inference algorithms. The properties of three categories of algorithms are presented and how the estimation quality is affected by the sparseness, pilot symbols, angle quantization and transmitter’s hardware impairment is also compared. Acquiring a tradeoff between the performance and computation complexity, the greedy iterative algorithm SP outperforms others at the low SNR region but the BCS algorithm is the best at the high SNR range. Furthermore, it is possible to extend our work to other sparse signal recovery problems as the three categories of algorithms are common methods.

Author Contributions: X.L. and W.Y. conceived the model; X.L. performed the performance analysis and simulation results; W.Y. and Y.C. analyzed the numerical and simulation; X.L. wrote the paper; and W.Y. and X.G. provided some suggestions and revised the paper.

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 61771487 and 61471393).

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).