

## Article

# Precise Channel Estimation Approach for a mmWave MIMO System

Prateek Saurabh Srivastav <sup>1,2</sup> , Lan Chen <sup>1,\*</sup>  and Arfan Haider Wahla <sup>1,2</sup> 

<sup>1</sup> Institute of Microelectronics of Chinese Academy of Sciences, Beijing 100029, China; prateek@ime.ac.cn (P.S.S.); irfan.hayd@gmail.com (A.H.W.)

<sup>2</sup> School of Electronic, Electrical and Communication, University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: chenlan@ime.ac.cn; Tel.: +86-010-82995754

Received: 7 May 2020; Accepted: 24 June 2020; Published: 26 June 2020



**Abstract:** Channel estimation is a formidable challenge in mmWave Multiple Input Multiple Output (MIMO) systems due to the large number of antennas. Therefore, compressed sensing (CS) techniques are used to exploit channel sparsity at mmWave frequencies to calculate fewer dominant paths in mmWave channels. However, conventional CS techniques require a higher training overhead for efficient recovery. In this paper, an efficient extended alternation direction method of multipliers (Ex-ADMM) is proposed for mmWave channel estimation. In the proposed scheme, a joint optimization problem is formulated to exploit low rank and channel sparsity individually in the antenna domain. Moreover, a relaxation factor is introduced which improves the proposed algorithm's convergence. Simulation experiments illustrate that the proposed algorithm converges at lower Normalized Mean Squared Error (NMSE) with improved spectral efficiency. The proposed algorithm also ameliorates NMSE performance at low, mid and high Signal to Noise (SNR) ranges.

**Keywords:** millimeter wave; Ex-ADMM; channel estimation; beamforming; sparsity; low rank approximations; matrix calculations

## 1. Introduction

In accordance with recent research trends, millimeter-wave (mmWave) communication has been found to be a potential candidate for next-generation Wireless Local Area Network (WLAN) and 5G cellular systems [1]. mmWave has a large amount of unused bandwidth, which can provide higher frequencies, higher spectral efficiency, and a relevant rise in channel capacity and support to connect millions of devices with high reliability and low latency [2–7]. However, the poor scattering nature of the mmWave channel causes path loss problems; as a result, fewer dominant paths remain in the mmWave channel [8–10]. Therefore, the channel matrix can be reconstructed by exploiting information from the remaining paths. In order to reconstruct the mmWave channel, the few dominant paths remaining in the channel matrix have to be estimated at the receiver's end. Conventional channel estimators, such as least-square (LS) estimators and maximum likelihood (ML) estimators, require a higher training overhead, which is not possible for any hybrid mmWave MIMO system. Considering the hardware architecture complexities of mmWave MIMO systems, channel estimation becomes a difficult task. In the literature, two types of approaches are used for mmWave channel estimation. In the first approach, the mmWave channel is estimated by exploiting the sparsity of the channel matrix in the virtual beamspace domain, whereas, in the second approach, the estimation is performed by exploiting the low rank properties of the channel matrix in the antenna domain. In [11–14], a compressive sensing (CS)-based channel estimation approach was used for a mmWave MIMO system. The basic idea behind this approach is based on the technique in which estimators have

to search for a pair angle in a predefined dictionary matrix, which further depends upon the training information. Such estimators exploit sparsity in a predefined dictionary matrix. For high resolution, another asymmetric approach was proposed in [15]. Under this method, atomic norm is used to formulate the angles of arrival (AoAs)/angles of departure (AoDs), finding problems. A beam codebook design-based approach was proposed in [16], in which, based on the information in static dictionaries, a codebook is designed. All the aforementioned techniques are CS-based approaches. For the case of high-dimensional training information, CS-based estimators suffer from high computational load, which can further reduce the spectral efficiency. A matrix completion (MC)-based approach was proposed in [17,18], in which the estimator exploits sparsity and low-rank properties of mmWave channels by two independent procedures.

In this paper, an extended alternating direction method of multipliers (Ex-ADMM) mmWave massive MIMO channel estimation technique is proposed for more accurate estimation rates. mmWave exhibits some important and peculiar effects over the wireless channel, i.e., the channel experiences an unspecified amount of spread over the angular domain due to its easy scattering nature [19]. As a result, a jointly sparse and low-rank channel matrix in the angular domain can be obtained. This the main motivation behind the formulation of a joint optimization problem for efficient recovery of the channel matrix proposed in this paper. ADMM was recently proposed in [20], has received attention for its easy implementation. Fundamentally, ADMM can exploit both the sparsity and low rank properties of any data matrix. Researchers have found that, in many real-world applications, additional information in a data matrix (known as side information) helps complete the matrix with few entries to obtain more accurate estimations. Therefore, the use of side information for matrix completion and factorization has been introduced in various research areas, like statistical signal processing, image processing, statistical learning, computer vision and so on [21–24]. Side information is very useful to describe the row and column entries of any matrix, and it has been further shown to reduce the complexity of completing a matrix [25,26]. In this work, side information theory is used to obtain the optimal solution for a joint optimization problem. The main contributions of this work are as follows:

1. A joint optimization estimation problem for a mmWave massive MIMO system based on the Ex-ADMM algorithm is formulated. With the help of side information in matrix completion theory, a training procedure compatible with the hybrid beamforming (HBF) structure, leveraging the low rank and sparsity in angular domains, is designed. In Ex-ADMM, the joint optimization problem is further divided into several subproblems, which are then solved individually.
2. An Ex-ADMM is proposed for better estimation of the channel matrix. This algorithm is originally derived from [18,26–28], and it is more efficient than the other iterative algorithms described in the results section. In the proposed Ex-ADMM algorithm, a nuclear norm is used for the low-rank approximations of the channel matrix, and  $l_1$  – norm is used to enforce the sparsity, consecutively. In addition, a relaxation factor is introduced to enhance the system performance.

Simulation results demonstrate the performance of the proposed Ex-ADMM algorithm, along with orthogonal matching pursuit (OMP) [16], two-stage sparse representation (TSSR) [29], vector approximate message passing (VAMP) [30] and the alternating direction method of multipliers (ADMM) [18] in terms of normalized mean squared error (NMSE), achievable spectral efficiency (ASE) and convergence.

This paper is organized as follows. The mmWave channel model is presented in Section 2. In Section 3, a mmWave channel estimation problem based on Ex-ADMM is formulated. Section 4 provides the training procedure details for the proposed Ex-ADMM. A detailed description of the proposed algorithm is provided in Section 5. Section 6 illustrates the computational complexities of proposed algorithm with respect to OMP, TSSR, VAMP and ADMM. Simulation results and discussion, describing the superiority of proposed algorithm over the benchmark algorithms in terms of NMSE, ASE and convergence, are provided in Section 7. Finally, the conclusion is presented in Section 8.

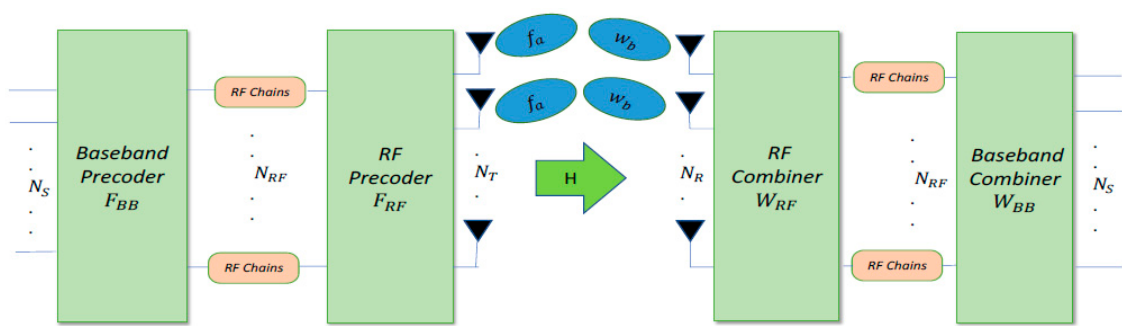
**Notation:** We use the following notation throughout the paper.  $\alpha$ ,  $a$ , and  $A$  denote a scalar, a vector and a matrix, respectively.  $A^T$ ,  $A^H$  and  $A^*$  indicate  $A$ 's transpose, conjugate transpose and

conjugate, respectively.  $\|(\cdot)\|_F$ ,  $\|(\cdot)\|_*$  and  $\|(\cdot)\|_1$  indicate the Frobenius norm, nuclear norm (sum of all singular values), and  $l_1$  – norm, respectively. The operands  $\circ$  and  $\otimes$  represent matrix Hadamard and Kronecker products.  $\text{vec}(\cdot)$  and  $\text{unvec}(\cdot)$  represent the concatenation of columns of the matrix into vectors and the reverse operation, respectively. Expected value is represented by  $E\{\cdot\}$ .  $A \in \{0,1\}^{M \times N}$  indicates that  $A$ 's elements are taken independently with equal probability from the binary set  $\{0,1\}$ .

## 2. mmWave Channel Model

Let us consider a  $N_R \times N_T$  mmWave massive MIMO system based on the recent models described in [5–7]. This system is equipped with  $N_T$  transmitters at the base station (BS) and  $N_R$  receivers at the mobile station (MS),  $N_S$  parallel data streams, and radio frequency (RF) chains, such that  $N_{RF} \leq \min(N_T, N_R)$  at the transmitter and receiver. At the transmitter,  $N_{RF}$  chains are present, such that  $N_S < N_{RF} < N_T$ . The mmWave MIMO system shown in Figure 1 is a combination of two consecutive joint segments: a digital MIMO baseband  $F_{BB} \in \mathbb{C}^{N_{RF} \times N_S}$  and analog RF precoder  $F_{RF} \in \mathbb{C}^{N_T \times N_{RF}}$ . Similarly, at the receiver, the signal is processed by the consecutive joint segments of the RF combiner  $W_{RF} \in \mathbb{C}^{N_R \times N_{RF}}$  and baseband combiner  $W_{BB} \in \mathbb{C}^{N_{RF} \times N_S}$ , respectively. In order to get the closest estimation, the transmitter employs  $N_T^{\text{Beam}} \leq N_T$  pilot beam patterns, denoted as  $\{f_a \in \mathbb{C}^{N_T \times 1} : \|f_a\|_2^2 = 1\}$ . Meanwhile, at the receiver end, the receiver employs  $N_R^{\text{Beam}} \leq N_R$  pilot beam patterns, denoted as  $\{w_b \in \mathbb{C}^{N_R \times 1} : \|w_b\|_2^2 = 1\}$  [16], where  $a$  and  $b$  are the transmitter's training precoding vector and receiver's training combining vector, respectively. For enabling communication, an analog HBF [11] is used, which opens a path for the transmitter to apply a baseband precoder by steering the beam to the receiver. The  $b$ -th received vector for the  $a$ -th transmitted beam is determined by

$$y_{b,a} = w_b^H H f_a x_a + w_b^H \tilde{v} \quad (1)$$



**Figure 1.** Block structure of a typical mmWave Multiple Input Multiple Output (MIMO) system.

Here,  $x_a$  is the pilot vector symbol emitted from the transmitter,  $\tilde{v}$  is the observation noise, with zero mean and  $\sigma_v^2 K_{N_R}$  variance, i.e.,  $CN(0, \sigma_v^2 K_{N_R})$ , and  $H \in \mathbb{C}^{N_R \times N_T}$  is the channel matrix of the mmWave MIMO system. The received vector  $y$  in Equation (1), which is expressed as  $y \triangleq [y_{1,b} \dots y_{N_T^{\text{Beam}},b}]^T \in \mathbb{C}^{N_R^{\text{Beam}} \times 1}$ , can be written in generalized form as follows:

$$y = w^H H f x + q \quad (2)$$

Here, the combiner noise vector  $q = w_b^H \tilde{v} \in \mathbb{C}^{N_R^{\text{Beam}} \times 1}$ .

Therefore, the received signal matrix  $Y$  can be expressed as

$$Y = W^H H F X + Q \quad (3)$$

Here, the received signal matrix  $Y \triangleq [y_1, \dots, y_{N_T^{\text{Beam}}}] \in \mathbb{C}^{N_R^{\text{Beam}} \times N_T^{\text{Beam}}}$ , complex combining matrix  $W \triangleq [w_1, \dots, w_{N_R^{\text{Beam}}}] \in \mathbb{C}^{N_R \times N_R^{\text{Beam}}}$ , and precoding matrix  $F \triangleq [f_1, \dots, f_{N_T^{\text{Beam}}}] \in \mathbb{C}^{N_T \times N_T^{\text{Beam}}}$ .  $X \in$

$\mathbb{C}^{N_T^{\text{Beam}} \times N_T^{\text{Beam}}}$  and  $Q \in \mathbb{C}^{N_R^{\text{Beam}} \times N_T^{\text{Beam}}}$  are independent and identically distributed (I.I.D) complex additive white gaussian noise (AWGN), with zero mean and  $\sigma_q^2$  variance  $\mathcal{CN}(0, \sigma_q^2)$ . In this paper, we assume pilot symbols are identical, so  $X = \sqrt{P_t} K_{N_T^{\text{Beam}}}$ , where  $P_t$  is the average transmitted pilot power [12,16].

According to the hybrid mmWave MIMO structure, the transmitted and received matrices are decomposed at the transmitter and receiver ends, such that  $F = F_{\text{BB}} F_{\text{RF}}$ , and  $W = W_{\text{RF}} W_{\text{BB}}$ , respectively. Hence, Equation (3) of the received signal matrices  $Y$  can be re-written as

$$\begin{aligned} Y &\triangleq \sqrt{P_t} W_{\text{BB}}^H W_{\text{RF}}^H H F_{\text{BB}} F_{\text{RF}} + Q \\ Y &\triangleq \sqrt{P_t} W^H H F + Q \end{aligned} \quad (4)$$

Here,  $F_{\text{RF}} \in \mathbb{C}^{N_T \times N_T}$  and  $W_{\text{RF}} \in \mathbb{C}^{N_R \times N_R}$  are the transmitted and received beamforming matrices, respectively.  $F_{\text{BB}} \in \mathbb{C}^{N_T \times N_T^{\text{Beam}}}$  and  $W_{\text{BB}} \in \mathbb{C}^{N_R \times N_R^{\text{Beam}}}$  are the transmitted and received baseband processing matrices, respectively.  $W$  is the combiner, such that  $W \in \{0, 1\}^{N_R}$ , and  $F$  is the precoder, such that  $F \in \{0, 1\}^{N_T}$ .

We adopted the geometric virtual model (GV) of mmWave MIMO system described in [12,17]. According to the GV model, channel  $H$  is described as

$$H \triangleq \sum_{l=1}^{L_p} \alpha_l a_R(\Phi_R^{(l)}, \theta_R^{(l)}) a_T^H(\Phi_T^{(l)}, \theta_T^{(l)}) \quad (5)$$

Here,  $L_p$  is the total number of propagation paths,  $\alpha_l$  is the complex channel gain of the  $l$ -th path, which can be obtained from the random complex gaussian distributions, and  $\mathcal{CN}(\frac{0,1}{2})$ .  $a_T^H(\Phi_T^{(l)}, \theta_T^{(l)}) \in \mathbb{C}^{N_T}$  and  $a_R(\Phi_R^{(l)}, \theta_R^{(l)}) \in \mathbb{C}^{N_R}$  are the array response vectors (ARV) at the transmitters and receivers, respectively.  $\Phi_T^{(l)}, \theta_T^{(l)}$  and  $\Phi_R^{(l)}, \theta_R^{(l)}$  are the elevation and azimuth AoA and AoD angles at the transmitters and receivers, respectively. These angles are generated by Laplacian distributions, whose means are uniformly distributed over  $(0, 2\pi)$ . The ARV of a uniform linear array (ULA) [31,32] is described by  $a(\theta) = \frac{1}{\sqrt{N}} [1, e^{-j\frac{2\pi}{\lambda} d \cos(\theta)}, \dots, e^{-j\frac{2\pi}{\lambda} (N-1) d \cos(\theta)}]^T$ . Here,  $\lambda$  is the wavelength,  $d$  is the antenna spacing and  $\theta$  is the even function of the ARV. The channel model described in Equation (5) can be further written in matrix form (described in a virtual beamspace representation model [33,34]) as follows:

$$H = A_R Z A_T^H \quad (6)$$

Here,  $A_R \in \mathbb{C}^{N_R \times N_R}$  and  $A_T \in \mathbb{C}^{N_T \times N_T}$  are the unitary matrices with the ARV of the receivers and transmitters [33], respectively, where  $A_R \triangleq [a_R(\Phi_1, \theta_1), a_R(\Phi_2, \theta_2) \dots a_R(\Phi_{L_p}, \theta_{L_p})]$  and  $A_T \triangleq [a_T(\Phi_1, \theta_1), a_T(\Phi_2, \theta_2) \dots a_T(\Phi_{L_p}, \theta_{L_p})]$ . By using matrix properties, we can say that  $A_R^H A_R = K_{N_R}$  and  $A_T^H A_T = K_{N_T}$ , with  $I_N$  being the  $N \times N$  identity matrix. In Equation (6),  $Z$  is the sparse matrix (which contains few virtual channel gains of greater amplitude) with dimensions  $Z \in \mathbb{C}^{N_R \times N_T}$ .

### 3. mmWave Channel Estimation Problem Formulation

In this section, a channel estimation problem for a mmWave MIMO system is formulated. Without loss of generality, matrix decomposition methods are used for the completion of a low-rank matrix with the help of partially observed data [35,36]. By following the given conditions, Equation (6) can be written in decomposed form as  $H = A_R D A_T^H$ , where  $D$  is the submatrix of  $Z$ . The optimization problem

for the joint recovery of channel state information (CSI) matrix  $H$  and the decomposed beamspace for the unknown sparse matrix  $D$  thereof can be described as

$$\begin{aligned} & \underset{H,D}{\text{minimize}} \quad \Gamma_H \|H\|_* + \Gamma_D \|D\|_1 \\ & \text{Subject to} \quad \Psi \circ H = H_\Psi \text{ and } H = A_R D A_T^H \end{aligned} \quad (7)$$

Here, the side information of  $A_T$  and  $A_R$  is used to find the missing values of matrix  $H$ . Originally,  $H = A_R D A_T^H$  is the decomposed form of Equation (6). Here, the sum of the singular values of a matrix is obtained by imposing a nuclear norm, which also represents the tightest convex lower bound for any matrix. In this way, the low rank property of matrix  $H$  is exhibited by the nuclear norm, however,  $l_1$  - norm bounds  $D$  to enforce the sparsity. In contrast, the weighting factors  $\Gamma_H$  and  $\Gamma_D$ , which depend upon the number of propagation paths, are always considered as positive integers [35].

The positions of non-zero-unit values, which are distributed uniformly, such that  $\Psi = \{1, 2, 3, \dots, N_R N_T\}$  [28,37] in matrix  $\Psi$  of Equation (7), are chosen in a random fashion. In this way, one can say that the matrix  $\Psi$  has  $M$  ones and  $(N_R N_T - M)$  zeros. The position of non-zero values in  $H_\Psi$  is the same as the position of non-zero values in matrix  $\Psi$ , where matrix  $H_\Psi$  is the subsampled estimated matrix, followed by matrix  $\Psi$ . The estimation error of  $H$  depends upon the estimation accuracy of  $H_\Psi$ 's elements and the  $M$  non-zero values of  $H_\Psi$ .

#### 4. Proposed Extended ADMM (Ex-ADMM)-Based Channel Estimation Scheme

In the following section, the proposed Ex-ADMM is described in detail. From Section 3, Equation (7) is clearly a convex function with two objectives. Therefore, there are many possible ways to get the global optimal solution. However, the best ways to solve the convex problem are first-order methods, which only require the first-order information of the optimization problem. Since these methods are computationally expensive, alternating optimization techniques (AOTs) [26] are used to obtain the optimal solution to a convex optimization problem, due to their less complex structure and easy handling. ADMM [20] is one very popular and efficient AOT. To solve the convex optimization problem described in Equation (7), an extension of ADMM [28] known as extended ADMM (Ex-ADMM) is proposed for the channel estimation of a mmWave MIMO system.

In order to solve the optimization problem described in Equation (7) by Ex-ADMM, we reformulate the optimization problem, and two auxiliary matrices,  $I \in \mathbb{C}^{N_R \times N_T}$  and  $J \triangleq I - A_R D A_T^H$ , are introduced. As a result, the reformulated optimization problem can be re-written as,

$$\begin{aligned} & \underset{H,I,D,J}{\text{minimize}} \quad \Gamma_H \|H\|_* + \Gamma_D \|D\|_1 + \frac{1}{2} \|J\|_F^2 + \frac{1}{2} \|\Psi \circ I - H_\Psi\|_F^2 \\ & \text{Subject to} \quad H = I \text{ and } J = I - A_R D A_T^H \end{aligned} \quad (8)$$

The first term in the optimization problem described in Equation (8) is the side information of the matrix, i.e.,  $A_R$  and  $A_T$ , along with the virtual channel gain in matrix  $Z$ . The second term is the information on subsampled virtual channel gain. The third term in the optimization problem is described by Equation (8), and considers the discretization error, and the fourth term is the AWGN noise. Afterwards, Equation (8) can be re-written with consideration of the augmented Lagrangian function (ALF) as follows:

$$\begin{aligned} \mathcal{L}_A(H, I, D, J, P_1, P_2) & \triangleq \Gamma_H \|H\|_* + \Gamma_D \|D\|_1 + \frac{1}{2} \|J\|_F^2 + \frac{1}{2} \|\Psi \circ I - H_\Psi\|_F^2 + \text{tr}(P_1^H (H - I)) \\ & + \frac{\beta}{2} \|H - I\|_F^2 + \text{tr}(P_2^H (I - A_R D A_T^H - J)) + \frac{\beta}{2} \|I - A_R D A_T^H - J\|_F^2 \end{aligned} \quad (9)$$

Here,  $P_1$  and  $P_2 \in \mathbb{C}^{N_R \times N_T}$  are Lagrange multipliers.  $\beta$  is the step size of Ex-ADMM and  $\beta > 0$ . For the next iteration, the Ex-ADMM can split Equation (9) into six sub-equations and solve it alternatively, i.e.,

$$H^{(l+1)} = \underset{H}{\operatorname{argmin}} \mathcal{L}_A(H, P_1^{(l)}, P_2^{(l)}, I^{(l)}, D^{(l)}, J^{(l)}) \quad (10)$$

$$P_1^{(l+1)} = P_1^{(l)} + \beta(H^{(l+1)} - I^{(l)}) \quad (11)$$

$$P_2^{(l+1)} = P_2^{(l)} + \beta(I^{(l)} - A_R D^{(l)} A_T^H - J^{(l)}) \quad (12)$$

$$I^{(l+1)} = \underset{I}{\operatorname{argmin}} \mathcal{L}_A(H^{(l+1)}, P_1^{(l+1)}, P_2^{(l+1)}, I, D^{(l)}, J^{(l)}) \quad (13)$$

$$D^{(l+1)} = \underset{D}{\operatorname{argmin}} \mathcal{L}_A(H^{(l+1)}, P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D, J^{(l)}) \quad (14)$$

$$J^{(l+1)} = \underset{J}{\operatorname{argmin}} \mathcal{L}_A(H^{(l+1)}, P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D^{(l+1)}, J) \quad (15)$$

In Equations (10)–(15),  $H^{(l+1)}$  is involved in every step; henceforth,  $H^{(l+1)}$  is known as an intermediate variable. In contrast, variables  $I$ ,  $D$  and  $J$  are known as essential variables, and Lagrange multipliers  $P_1$  and  $P_2$  are recognized as dual variables. Another ADMM algorithm mentioned in [18] is also able to solve the optimization problem in (9) but the proposed Ex-ADMM algorithm converges at a lower NMSE as compared to the ADMM and provides better ASE performance as well.

Basically, the Ex-ADMM algorithm is a combination of cyclical ADMM and a relaxation factor [28,38–40]. In general, ADMM updates its order in an arbitrary way, i.e.,  $H \rightarrow I \rightarrow D \rightarrow J \rightarrow P_1 \rightarrow P_2$ . In contrast, the Ex-ADMM updates its order in a cyclical way, i.e.,  $H \rightarrow P_1 \rightarrow P_2 \rightarrow I \rightarrow D \rightarrow J$ . In the proposed Ex-ADMM, reordering of the ADMM is only done to ensure the channel matrix  $H$  satisfies the first-order optimality conditions [39]; therefore, right after the updating of  $H^{(l+1)}$ , dual variables  $P_1$  and  $P_2$  are updated and then, at last, the essential variables  $I$ ,  $D$  and  $J$  are updated. A relaxation technique, along with a relaxation parameter, is used to relax the essential variables [41,42]. Consequently, for the relaxation of the essential variables, we can assume that the cyclical order of dual and essential variables in Equations (11)–(15) is a block variable, i.e.,  $\tilde{z}^{(l)} = (\tilde{P}_1^{(l)}, \tilde{P}_2^{(l)}, \tilde{I}^{(l)}, \tilde{D}^{(l)}, \tilde{J}^{(l)})$ . Thus, the final relaxed variable,  $z^{(l+1)} = (P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D^{(l+1)}, J^{(l+1)})$ , can be generated as

$$z^{(l+1)} = z^l - \gamma(z^l - \tilde{z}^{(l)})$$

Here, roughly speaking, the tilde ( $\sim$ ) variables (i.e.,  $\tilde{P}_1^{(l+1)}, \tilde{P}_2^{(l+1)}, \tilde{I}^{(l+1)}, \tilde{D}^{(l+1)}$  and  $\tilde{J}^{(l+1)}$ ) are auxiliary variables, and they can be updated using the main variables (i.e.,  $P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D^{(l+1)}$  and  $J^{(l+1)}$ ), and these variables can be obtained using the Equations (11)–(15) as follows:

$$\tilde{P}_1^{(l+1)} = P_1^{(l)} + \beta(H^{(l+1)} - I^{(l)}) \quad (16)$$

$$\tilde{P}_2^{(l+1)} = P_2^{(l)} + \beta(I^{(l)} - A_R D^{(l)} A_T^H - J^{(l)}) \quad (17)$$

$$\tilde{I}^{(l+1)} = \underset{I}{\operatorname{argmin}} \mathcal{L}_A(H^{(l+1)}, P_1^{(l+1)}, P_2^{(l+1)}, I, D^{(l)}, J^{(l)}) \quad (18)$$

$$\tilde{D}^{(l+1)} = \underset{D}{\operatorname{argmin}} \mathcal{L}_A(H^{(l+1)}, P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D, J^{(l)}) \quad (19)$$

$$\tilde{J}^{(l+1)} = \underset{J}{\operatorname{argmin}} \mathcal{L}_A(H^{(l+1)}, P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D^{(l+1)}, J) \quad (20)$$

Therefore, the final relaxed variable can be expressed as

$$\begin{pmatrix} P_1^{(l+1)} \\ P_2^{(l+1)} \\ I^{(l+1)} \\ D^{(l+1)} \\ J^{(l+1)} \end{pmatrix} = \begin{pmatrix} P_1^{(l)} \\ P_2^{(l)} \\ I^{(l)} \\ D^{(l)} \\ J^{(l)} \end{pmatrix} - \gamma \begin{pmatrix} P_1^{(l)} \\ P_2^{(l)} \\ I^{(l)} \\ D^{(l)} \\ J^{(l)} \end{pmatrix} - \begin{pmatrix} \tilde{P}_1^{(l)} \\ \tilde{P}_2^{(l)} \\ \tilde{I}^{(l)} \\ \tilde{D}^{(l)} \\ \tilde{J}^{(l)} \end{pmatrix} \quad (21)$$

From the above discussion, it is clear that the Equations (10) and (21) have several advantages: i.e., the optimization problem described in Equation (9) turns into six individual subproblems, and these can be solved further without any strict conditions. Hence, the global optimal solution can be derived effortlessly. This helps reduce the computational complexity and storage requirements. For the next iteration, the values of the subproblems described in Equations (10) and (21) have to be updated. In order to do so, first of all, the closed-form solutions of Equations (10)–(15) need to be obtained. The advantages of Ex-ADMM over ADMM are described further in Section 5.

#### 4.1. Procedure for Updating $H^{(l+1)}$

In order to obtain the closed-form solution for  $H$ , reformulate  $\mathcal{L}_A$  to  $\mathcal{L}_1$  and consider all terms with respect to  $H$  in Equation (9). Keeping only the terms that are related to it,

$$\begin{aligned} \mathcal{L}_1 &\triangleq \argmin_H \Gamma_H \|H\|_* + \text{tr}(P_1^H (H - I)) + \frac{\beta}{2} \|H - I\|_F^2 \\ &= \Gamma_H \|H\|_* + \frac{\beta}{2} \|H - (I^{(l)} - \frac{1}{\beta} P_1^{(l)})\|_F^2 \end{aligned} \quad (22)$$

Here, Equation (22) is the solution of Equation (10). To get the closed-form solution, a singular value thresholding (SVT) operator [43] is implemented on Equation (22):

$$H^{(l+1)} = U \text{diag}(\{\text{sign}(h_i) \max(h_i, 0)\}_{1 \leq i \leq r}) V^H \quad (23)$$

Here,  $U \in \mathbb{C}^{N_r \times r}$  and  $V \in \mathbb{C}^{N_r \times r}$  are the side singular vector matrices of the matrices  $(I^{(l)} - \frac{1}{\beta} P_1^{(l)})$  and  $h_i \triangleq \mu_i - \frac{\Gamma}{\beta}$ , respectively, where  $\Gamma$  is the SVT threshold operator and  $\mu_i$  denotes the  $r$  singular values.

#### 4.2. Procedure for Updating $I^{(l+1)}$

To update  $I^{(l+1)}$ , reformulate  $\mathcal{L}_A$  to  $\mathcal{L}_2$ , and, to get the closed-form solutions for  $I$ , consider all terms of  $I$  in Equation (9) and differentiate with respect to  $I$ :

$$\begin{aligned} \mathcal{L}_2 &\triangleq \arg \min_I \frac{1}{2} \|\Psi \circ I - H_\Psi\|_F^2 + \text{tr}(P_1^H (H - I)) + \frac{\beta}{2} \|H - I\|_F^2 \\ &\quad + \text{tr}(P_2^H (I - A_R D A_T^H - J)) + \frac{\beta}{2} \|I - A_R D A_T^H - J\|_F^2 \\ \mathcal{L}_2 &= \frac{\partial \mathcal{L}_A}{\partial I} = \frac{\partial}{\partial I} (\frac{1}{2} \|\Psi \circ I - H_\Psi\|_F^2 + \text{tr}(P_1^H (H - I)) + \frac{\beta}{2} \|H - I\|_F^2 \\ &\quad + \text{tr}(P_2^H (I - A_R D A_T^H - J)) + \frac{\beta}{2} \|I - A_R D A_T^H - J\|_F^2) \\ \mathcal{L}_2 &= \Psi \circ I - H_\Psi - P_1 - \beta (H - I) - P_2 - \beta (J - I + A_R D A_T^H) \end{aligned} \quad (24)$$

Setting  $\mathcal{L}_2(I)$  to zero, i.e.,  $\mathcal{L}_2 = 0$ ,

$$\begin{aligned} \Psi \circ I - H_\Psi - P_1 - \beta (H - I) - P_2 - \beta (J - I + A_R D A_T^H) &= 0 \\ \Psi \circ I + 2\beta I &= H_\Psi + P_1 + \beta (H) + P_2 + \beta J + A_R D A_T^H \end{aligned}$$

$$\mathbf{I} = (\mathbf{A} + 2\beta\mathbf{K})^{-1}(\mathbf{P}_1 + \beta(\mathbf{H}) + \mathbf{P}_2 + \beta\mathbf{J} + \beta\mathbf{B}\mathbf{D}) \quad (25)$$

where  $\mathbf{K}$  is the identity matrix,  $\mathbf{A} \triangleq \sum_{i=1}^{N_R} \text{diag}([\Psi]_k)^T \otimes \mathbf{E}_{kk} [\Psi]_i$  exhibits the  $k$ -th row, and  $\mathbf{E}_{kk}$  is derived by inserting unit values in the  $N_R \times N_R$  zero matrix at its  $(k,k)$ -th position and  $\mathbf{B} \triangleq \mathbf{A}_T^* \otimes \mathbf{A}_R$ .

Therefore, the closed-form solution of  $\mathbf{i}$  for the  $(l+1)$  iteration is

$$\mathbf{i}^{(l+1)} = (\mathbf{A}^H \mathbf{A} + 2\beta\mathbf{K})^{-1} \left( \mathbf{p}_1^{(l)} + \beta \mathbf{h}^{(l+1)} + \mathbf{A}^H \mathbf{h}_\Psi + \mathbf{p}_2^{(l)} + \beta \mathbf{j}^{(l)} + \beta \mathbf{B} \mathbf{d}^{(l)} \right) \quad (26)$$

To get the final solution, unvectorized  $\mathbf{i}^{(l+1)}$ , i.e.,

$$\mathbf{I}^{(l+1)} = \text{unvec}(\mathbf{i}^{(l+1)}) \quad (27)$$

#### 4.3. Procedure for Updating $\mathbf{D}^{(l+1)}$

To get the solution for  $\mathbf{D}$ , reformulate  $\mathcal{L}_A$  to  $\mathcal{L}_3$  and consider all terms corresponding to  $\mathbf{D}$  in Equation (9). Keeping only the terms that are related to it,

$$\begin{aligned} \mathcal{L}_3 &\triangleq \underset{\mathbf{D}}{\text{argmin}} \Gamma_D \|\mathbf{D}\|_1 + \text{tr} \left( \mathbf{P}_2^H (\mathbf{I} - \mathbf{A}_R \mathbf{D} \mathbf{A}_T^H - \mathbf{J}) \right) + \frac{\beta}{2} \|\mathbf{I} - \mathbf{A}_R \mathbf{D} \mathbf{A}_T^H - \mathbf{J}\|_F^2 \\ &= \Gamma_D \|\mathbf{D}\|_1 + \frac{\beta}{2} \|\mathbf{A}_R^H \left( \frac{1}{\beta} \mathbf{P}_2^{(l+1)} - \mathbf{J}^{(l)} + \mathbf{I}^{(l+1)} \right) \mathbf{A}_T\|_F^2 \end{aligned} \quad (28)$$

Here,  $\mathbf{A}_R$  and  $\mathbf{A}_T$  are the unitary matrices. To get the closed-form solution of  $\mathcal{L}_3$ , Equation (28) is transformed into the standard least absolute shrinkage and selection operator (LASSO) problem [44]. In order to do this, vectorization is performed on Equation (28):

$$\underset{\mathbf{D}}{\text{argmin}} \Gamma_D \|\mathbf{D}\|_1 + \frac{\beta}{2} \|\mathbf{A}_R^H \left( \frac{1}{\beta} \mathbf{P}_2^{(l+1)} - \mathbf{J}^{(l)} + \mathbf{I}^{(l+1)} \right) \mathbf{A}_T\|_F^2 \quad (29)$$

Let us assume  $\mathbf{V}^{(l+1)} = \mathbf{A}_R^H \mathbf{A}_T \left( \frac{1}{\beta} (\mathbf{P}_2^{(l+1)} - \mathbf{J}^{(l+1)} + \mathbf{I}^{(l+1)}) \right)$  and  $\mathbf{v}^{(l+1)} = \text{vec}(\mathbf{V}^{(l+1)})$ . Therefore, Equation (29) can be written as

$$\underset{\mathbf{d}}{\text{argmin}} \Gamma_d \|\mathbf{d}\|_1 + \frac{\beta}{2} \|\mathbf{d} - \mathbf{v}^{(l+1)}\|_F^2 \quad (30)$$

Thus, to obtain the estimate of  $\mathbf{d}$  in Equation (30), a soft thresholding operator is imposed for  $(l+1)$  iterations:

$$\begin{aligned} \mathbf{d}^{(l+1)} &= \text{sign}(\text{Re}(\mathbf{v}^{(l+1)})) \circ \max(|\text{Re}(\mathbf{v}^{(l+1)})| - \Gamma'_d, 0) \\ &\quad + \text{sign}(\text{Im}(\mathbf{v}^{(l+1)})) \circ \max(|\text{Im}(\mathbf{v}^{(l+1)})| - \Gamma'_d, 0) \end{aligned} \quad (31)$$

$\Gamma'_d$  is the scaled version of  $\Gamma_d$  and  $\Gamma'_d \triangleq \frac{\Gamma_d}{\beta}$ . Hence, the value of  $\mathbf{D}^{(l+1)}$  is obtained by

$$\mathbf{D}^{(l+1)} = \text{unvec}(\mathbf{d}^{(l+1)}) \quad (32)$$

#### 4.4. Procedure for Updating $\mathbf{J}^{(l+1)}$

To update  $\mathbf{J}^{(l+1)}$ , reformulate  $\mathcal{L}_A$  to  $\mathcal{L}_4$ , and, to obtain the closed-form solution for  $\mathbf{J}$ , take all terms corresponding to  $\mathbf{J}$  in Equation (9) and differentiate with respect to  $\mathbf{J}$ :

$$\begin{aligned} \mathcal{L}_4 &\triangleq \underset{\mathbf{J}}{\text{argmin}} \frac{1}{2} \|\mathbf{J}\|_F^2 + \left( \mathbf{P}_2^H (\mathbf{I} - \mathbf{A}_R \mathbf{D} \mathbf{A}_T^H - \mathbf{J}) \right) + \frac{\beta}{2} \|\mathbf{I} - \mathbf{A}_R \mathbf{D} \mathbf{A}_T^H - \mathbf{J}\|_F^2 \\ \mathcal{L}_4 &= \frac{\partial \mathcal{L}_A}{\partial \mathbf{J}} = \frac{\partial}{\partial \mathbf{J}} \left( \frac{1}{2} \|\mathbf{J}\|_F^2 + \left( \mathbf{P}_2^H (\mathbf{I} - \mathbf{A}_R \mathbf{D} \mathbf{A}_T^H - \mathbf{J}) \right) + \frac{\beta}{2} \|\mathbf{I} - \mathbf{A}_R \mathbf{D} \mathbf{A}_T^H - \mathbf{J}\|_F^2 \right) \end{aligned}$$

$$\mathcal{L}_4 = (1 + \beta)J - \beta(I - A_R D A_T^H + \frac{P_2^H}{\beta})$$

By setting  $\mathcal{L}_4$  to zero, i.e.,  $\mathcal{L}_4 = 0$ ,

$$J = \frac{\beta}{1 + \beta} (I - A_R D A_T^H + \frac{P_2^H}{\beta}) \quad (33)$$

Hence, for iteration  $(l + 1)$ ,

$$J^{(l+1)} = \frac{\beta}{\beta + 1} \left( I^{(l+1)} - A_R D^{(l+1)} A_T^H + \frac{1}{\beta} P_2^{(l)} \right) \quad (34)$$

The dual variables  $P_1$  and  $P_2$  can be directly updated using the  $H$ ,  $I$ ,  $D$  and  $J$  variables.

## 5. Algorithm Description

Algorithm 1 demonstrates the proposed Ex-ADMM-based mmWave channel estimation scheme, which is originally derived from the ADMM described in [18,28]. In the proposed Ex-ADMM algorithm, first of all, the parameters  $H^{(0)} = P_1^{(0)} = P_2^{(0)} = I^{(0)} = D^{(0)} = J^{(0)} = \tilde{P}_1^{(0)} = \tilde{P}_2^{(0)} = \tilde{I}^{(0)} = \tilde{D}^{(0)} = \tilde{J}^{(0)} = 0$  are initialized [18].

The main objective of the proposed algorithm is to update  $H^{(l+1)}$  (in step 2). This step is the most crucial task, which is done by deriving the Lagrangian  $\mathcal{L}_A$  to  $\mathcal{L}_1$ , and then implementing the SVT operator on Equation (22). In order to efficiently update  $H^{(l+1)}$ , the SVT operator computes the  $\mu_i$  singular values of  $h_i$ . At every instant, the proposed algorithm is required to run the subsampled version of  $H$  as an input, which is known as  $H_\Psi$ . For the training procedure, the mmWave MIMO model detailed in Section 2, inspired from [12,45], has been adopted. For the general case, let us assume that only a single set of transmitter and receiver antennas are operational at each illustration of  $t$ . According to the adopted hardware structure described in Section 2, the transmitter's training precoding vectors and the receiver's training combining vectors have non-zero (unit) values at their respective  $ij$ -th position in matrix  $\Psi$ . At any  $t$ -th training illustration, followed by the matrix  $\Psi$ , the subsampled matrix  $H_\Psi$  also has the estimated non-zero values in the  $ij$ -th positions. In contrast, the length of the training symbols is equal to the position of the non-zero elements, i.e.,  $T = M$  and  $M \ll N_R N_T$ . This factor is the stopping criterion of the proposed Ex-ADMM algorithm. In step (3) of the proposed Ex-ADMM algorithm, to maintain the cyclical order, dual variables  $\tilde{P}_1^{(l+1)}$  and  $\tilde{P}_2^{(l+1)}$  are updated first, with the help of Equations (16) and (17). The values of the dual variables  $P_1$  and  $P_2$  can be directly calculated using Equations (23), (26), (31) and (34). In step (4), the first subproblem,  $\tilde{I}^{(l+1)}$ , has been updated using Equation (18). Steps (5) and (6) are used to update the second and third subproblems,  $\tilde{D}^{(l+1)}$  and  $\tilde{J}^{(l+1)}$ , using Equations (19) and (20), respectively.  $(P_1^{(l+1)}, P_2^{(l+1)}, I^{(l+1)}, D^{(l+1)}$  and  $J^{(l+1)})$  have been updated in step (7) using Equation (21)). The relaxation factor  $\gamma$  provides better NMSE performance and a slightly improved convergence rate for all training lengths. Although Ex-ADMM is better than the other benchmark algorithms described in this paper, there is a major drawback regarding the knowledge produced during its implementation. When the proposed Ex-ADMM algorithm is implemented to solve sparse optimization problems on  $I^{(l+1)}, D^{(l+1)}, J^{(l+1)}, P_1^{(l+1)}$  and  $P_2^{(l+1)}$ , these parameters start losing their sparse nature. The reason behind this problem is that  $I^{(l+1)}, D^{(l+1)}, J^{(l+1)}, P_1^{(l+1)}$  and  $P_2^{(l+1)}$  are the algebraic combinations of  $(I^{(l)} \text{ and } \tilde{I}^{(l)}), (D^{(l)} \text{ and } \tilde{D}^{(l)}), (J^{(l)} \text{ and } \tilde{J}^{(l)}), (P_1^{(l)} \text{ and } \tilde{P}_1^{(l)})$  and  $(P_2^{(l)} \text{ and } \tilde{P}_2^{(l)})$ , respectively. The most expedient way to overcome this problem is, when the proposed algorithm reaches stopping criterion, to run only the cyclical part of the algorithm for one final iteration to ensure the sparsity (i.e., step 8).

**Algorithm 1. Ex-ADMM based mmWave MIMO Channel Estimation Scheme**


---

**Require** : Subsampled matrix  $H_\Psi$ , side information matrices  $A_R$  and  $A_T$ , and the set of indices of observed entries in  $\Psi$

**Input** :  $H_\Psi$ ,  $\Psi$ ,  $A_R$ ,  $A_T$ ,  $\rho$ ,  $\gamma$ ,  $\Gamma_H$ ,  $\Gamma_Z$  and  $I_{\max}$

**Output** : Estimated output channel matrix  $\hat{\mathbf{H}} = \mathbf{H}^{(I_{\max})}$

**Initialization:**  
 $\mathbf{H}^{(0)} = \mathbf{P}_1^{(0)} = \mathbf{P}_2^{(0)} = \mathbf{I}^{(0)} = \mathbf{D}^{(0)} = \mathbf{J}^{(0)} = \tilde{\mathbf{P}}_1^{(0)} = \tilde{\mathbf{P}}_2^{(0)} = \tilde{\mathbf{I}}^{(0)} = \tilde{\mathbf{D}}^{(0)} = \tilde{\mathbf{J}}^{(0)} = 0$

**Step 1** : **for**  $l = 0, 1, 2 \dots I_{\max} - 1$  **do**

**Step 2** : Update  $\mathbf{H}^{(l+1)}$  by using Equation (23).

**Step 3** : Update  $\tilde{\mathbf{P}}_1^{(l+1)}$  and  $\tilde{\mathbf{P}}_2^{(l+1)}$  by using Equations (16) and (17).

**Step 4** : Update  $\tilde{\mathbf{I}}^{(l+1)}$  by using the Equation (18).

**Step 5** : Update  $\tilde{\mathbf{D}}^{(l+1)}$  by using the Equation (19).

**Step 6** : Update  $\tilde{\mathbf{J}}^{(l+1)}$  by using the Equation (20).

**Step 7** : Update  $(\mathbf{P}_1^{(l+1)}, \mathbf{P}_2^{(l+1)}, \mathbf{I}^{(l+1)}, \mathbf{D}^{(l+1)} \text{ and } \mathbf{J}^{(l+1)})$  by using Equation (21), where  $\gamma$  is the relaxation factor.

**Step 8** : meets stopping criterion **do**  
 $i = l + 1$   
 repeat step 2 to 6.

**Step 9** : **end do**

**Step 10** : **end for**

---

**6. Computational Complexity**

In this section, four benchmark algorithms, namely, TSSR [46], VAMP [47], OMP [16] and ADMM [18], are compared with the proposed Ex-ADMM algorithm in terms of complexity.

For the general case, TSSR is much faster than any one-stage method. The computational complexity of TSSR is dependent upon the calculation of its maximum diagonal elements and the smaller number of off-diagonal elements in its sparse matrix. The complexity order of the TSSR algorithm is  $\mathcal{O}(n \log(n))$ , [46], where  $n$  is the full column rank of the targeted sparse matrix.

In VAMP, the complexity is dominated by the matrix vector multiplication. The computational complexity of VAMP is the order of  $\mathcal{O}(N_R N_T L \log(N_R N_T L))$  [47], where  $N_R$  and  $N_T$  is the number of receivers and transmitters, respectively, and  $L$  is number of channel paths or the sparsity level of the channel.

In OMP, the computational complexities depend upon the sparsity of the dictionary matrix and the number of grids. Mathematically, this can be expressed as  $\mathcal{O}(L \ln G^2)$ . Here,  $L$  is the number of paths or the sparsity level of the channel and  $G$  is the grid of the dictionary matrix in which the AoAs and AoDs are distributed uniformly [16]. The computational complexity of ADMM reflects the number of iterations,  $I_{\max}$ , and the number of transmitting and receiving antennas, i.e.,  $N_T$  and  $N_R$ , respectively. In ADMM, appropriately, matrix factorization can be performed offline, and only the matrix vector product has to be calculated online. This reduces the computational complexities of the ADMM in a significant way. The complexity of ADMM can be expressed as  $\mathcal{O}(N_R^2 N_T)$  [18], where  $N_R$  and  $N_T$  are the numbers of receivers and transmitters, respectively. The proposed Ex-ADMM algorithm is originally derived from ADMM. Therefore, the complexity of the proposed Ex-ADMM algorithm remains the same as that of ADMM, i.e.,  $\mathcal{O}(N_R^2 N_T)$ .

**7. Simulation Results and Discussion**

In this section, the numerical results are explained to showcase the accomplishment and performance of the proposed algorithm (Ex-ADMM) with different standard algorithms.

### 7.1. System Model

The HBF system architecture was adopted as described in [45]. A total of 64 transmitting and receiving antennas, i.e.,  $N_T = N_R = 64$ , were considered at the BS and MS. The antennas were assumed to be in a ULA configuration. Laplacian distributions with  $55^\circ$  standard deviation were used to produce the Azimuth angle of the AoAs and AoDs. Phase shifters had the quantized phase, whereas spacing between antennas was determined by  $d = \frac{\lambda}{2}$ .

### 7.2. Channel Model

A mmWave channel with two paths and two clusters is considered here. Over the range of  $[0, 2\pi]$ , AoAs and AoDs were uniformly distributed. The noise is complex additive white gaussian noise (AWGN), with zero mean and  $\sigma_q^2$  variance. A signal-to-noise Ratio (SNR) of 30 dB, mathematically defined as  $\text{SNR} \triangleq \sigma_q^{-2}$ , was used for simulation. The frequency of the mmWave channel model used for simulation was 90GHz.

### 7.3. Simulation Environment

In this section, we interpret the results of a comparison between the proposed Ex-ADMM algorithm and the TSSR, OMP, VAMP and ADMM algorithms, in terms of ASE, NMSE, the convergence of Ex-ADMM with respect to ADMM and SVT, and the effect of NMSE on the proposed Ex-ADMM, TSSR, OMP, VAMP and ADMM algorithms for multiple paths. An average of 100 independent iterations and 100 Monte Carlo realizations were considered for the simulations [48]. Three different training symbol lengths were considered for training purposes, i.e.,  $T = 400, 800$  and  $1200$ .

### 7.4. Results and Discussions

OMP determines the sparsity of the channel or dictionary matrix; VAMP also uses this concept with a few differences. On the other hand, TSSR individually exploits the low rank property of the channel matrix along with the sparsity. At first, an SVT operator was implemented to recapture the channel matrix  $H$ . The SVT threshold operator  $\Gamma = \beta \|H_\psi\|$ , while  $\beta = \frac{3M}{N_R N_T}$  was fixed to exploit the low rank property. Channel sparsity depends upon the number of paths in the mmWave channel, i.e., the sparsity of the channel is fixed with the number of propagation paths. For the case of the Ex-ADMM algorithm, the weighting factor of the channel matrix is  $\Gamma_H = \beta \|H_\psi\|$ , where  $\beta$  is set to be 0.005, and the weighting factor with respect to the sparse matrix  $D$  is obtained by  $\Gamma_D = \frac{0.1}{(1 - 10 \log(\sigma_q^2))}$ . Under the aforementioned constraints, an analysis was conducted, and results are explained as follows.

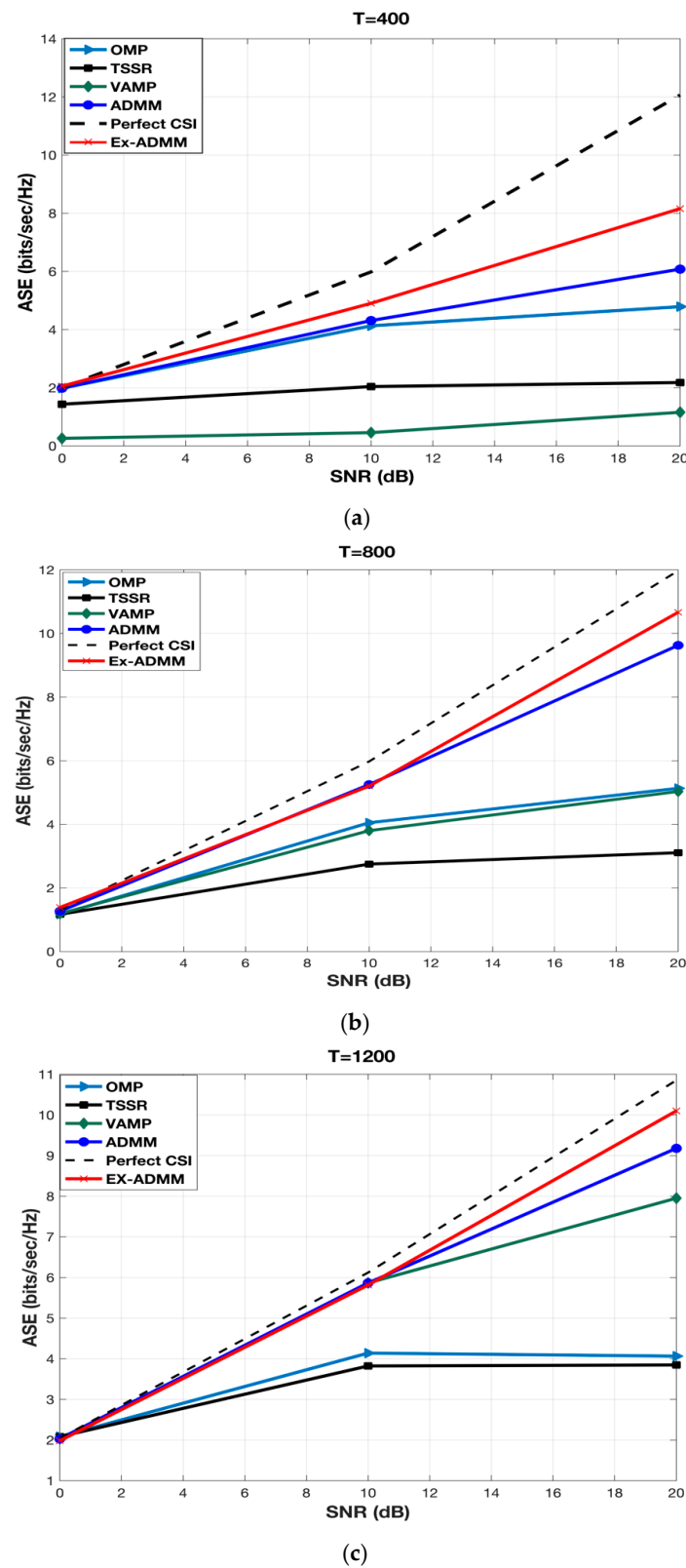
#### 7.4.1. Comparison of ASE

In Figure 2a–c, the ASE (in bits/sec/Hz) at the SNR points for OMP, TSSR, VAMP, ADMM and the proposed Ex-ADMM algorithm with perfect CSI is shown. The following expression is used to calculate the ASE [49,50],

$$\text{ASE} = E \left\{ \log_2 \det(K_{N_R} + (N_R N_T (\sigma_q^2 + \text{NMSE})^{-1} H H^H) \right\}$$

The performance of the OMP in terms of ASE at low-to-mid SNR points is moderate, and it gets worse at mid-to-high SNR points as the training symbols' length is increased, i.e.,  $T < 400$ . As is clear from Figure 2a, for all SNR points, VAMP performs very poorly for smaller training symbol lengths; i.e.,  $T < 400$ . When the training symbol length is increased, i.e.,  $T \geq 800$ , the performance of VAMP increases significantly, as depicted in Figure 2b,c. It was found that, for different  $T$  values, the performance of TSSR is very bad at almost every SNR point relative to OMP, TSSR, VAMP, ADMM and Ex-ADMM. ADMM, at all SNR points, performed better than OMP, TSSR and VAMP for all training symbol lengths. In the case of Ex-ADMM, for  $T = 400$ , Ex-ADMM outperformed the OMP, TSSR, VAMP and ADMM algorithms at all SNR points. For  $T = 800$  and  $1200$ , at low-to-mid SNR

points, ADMM and Ex-ADMM performed similarly, but, as the SNR range increased from mid to high, Ex-ADMM outperformed ADMM. Moreover, for high training symbol lengths, Ex-ADMM was very close to achieving the perfect CSI.



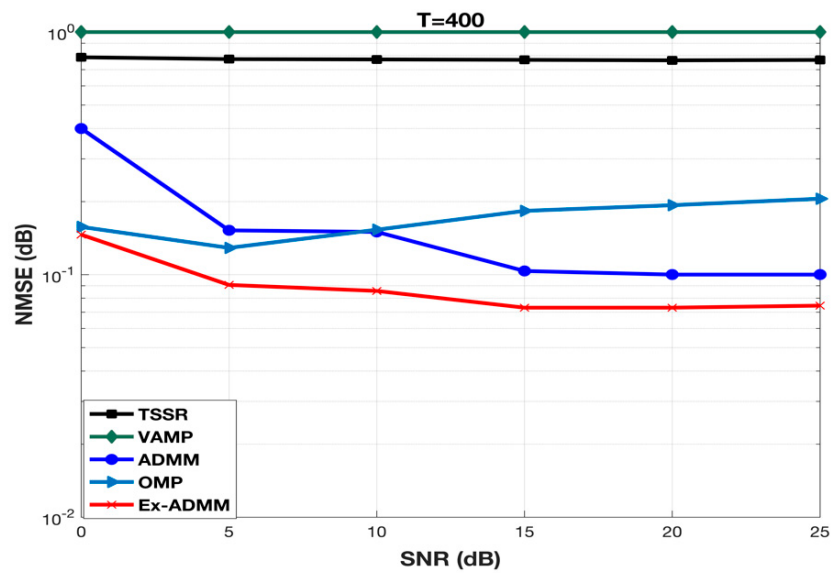
**Figure 2.** (a–c) ASE for different transmit SNRs for a  $64 \times 64$  mmWave MIMO channel at  $T = 400$ ,  $T = 800$  and  $T = 1200$ .

#### 7.4.2. Comparison of NMSE

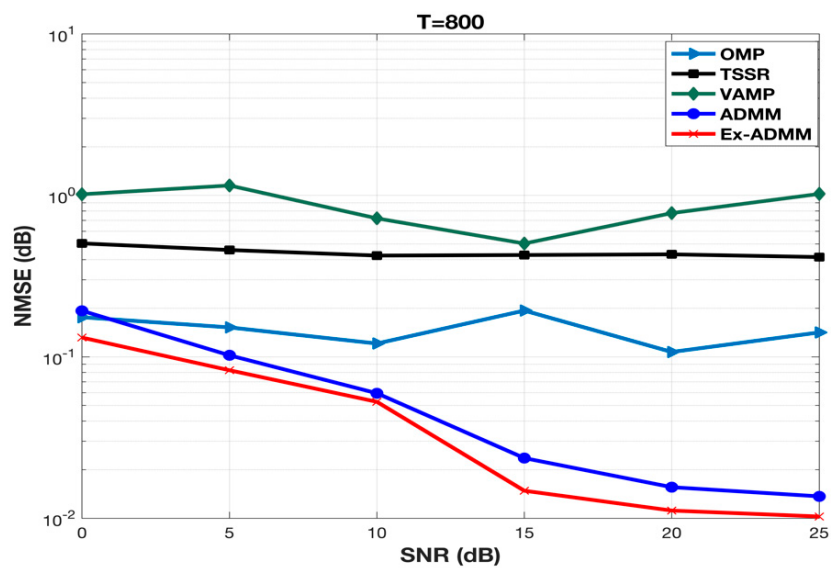
The performance of OMP, TSSR, VAMP, ADMM and the proposed Ex-ADMM algorithm in terms of NMSE, at different SNR points with respect to different training symbol lengths  $T$ , is depicted in Figure 3a–c. The following relation is used to calculate NMSE:

$$\text{NMSE} \triangleq \mathbb{E} \left( 10 \log_{10} \frac{\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2} \right)$$

where  $\hat{\mathbf{H}}$  is the estimated channel matrix and  $\mathbf{H}$  is the true channel matrix.  $\hat{\mathbf{H}}$  is determined by  $\mathbf{H}^{(I_{\max})}$ . The performance of OMP at different SNR points is almost constant for all training symbol lengths, and, due to its huge dictionary matrix, does not change for any increment of  $T$ . Thus, OMP may suffer from a discretization problem. One can therefore say that the OMP is not capable enough to recover small training symbols.

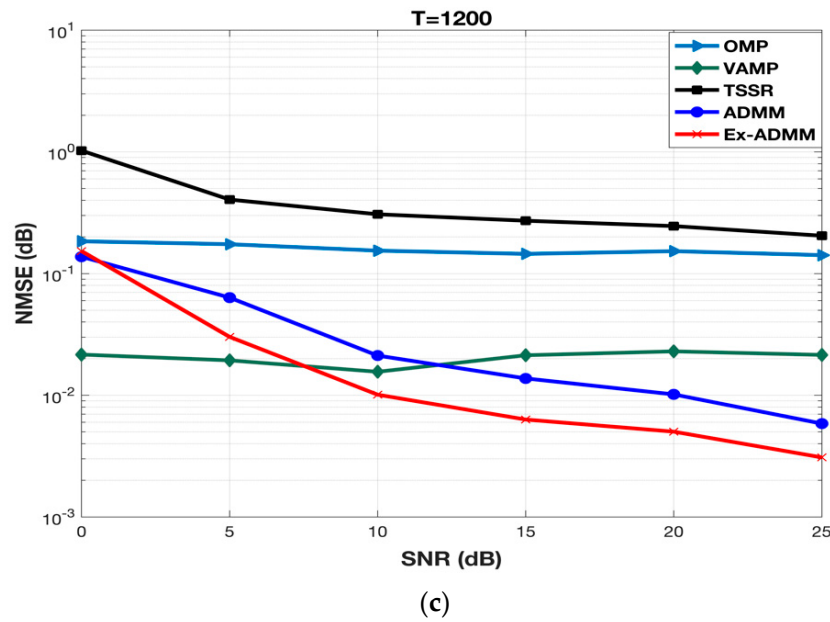


(a)



(b)

Figure 3. Cont.



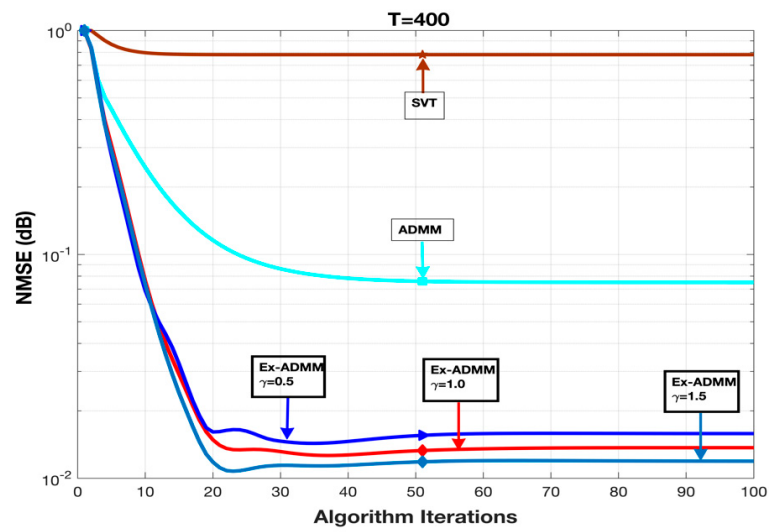
**Figure 3.** (a–c) NMSE for different transmit SNRs for a  $64 \times 64$  mmWave MIMO channel at  $T = 400$ ,  $T = 800$  and  $T = 1200$ .

On the other side, it was found that VAMP exhibited poor performance at  $T < 800$ , but, as depicted in Figure 3b,c, the performance of VAMP increases as the training symbol length increases. VAMP is used to calculate statical information on sparse signals, which is impossible in case of small training symbol lengths, but, at low-to-mid SNR points, as the training symbol length increases to  $T > 800$ , it exhibits a significant improvement in NMSE performance compared to other algorithms. TSSR is used for two-stage estimation. In the simulation results, TSSR is not capable of recovering the estimated values for low and high training lengths. TSSR exploits the low rank and sparsity of any channel matrix individually. In contrast, ADMM has the capacity to exploits low rank and sparsity together, with the channel matrix for any training symbol length.

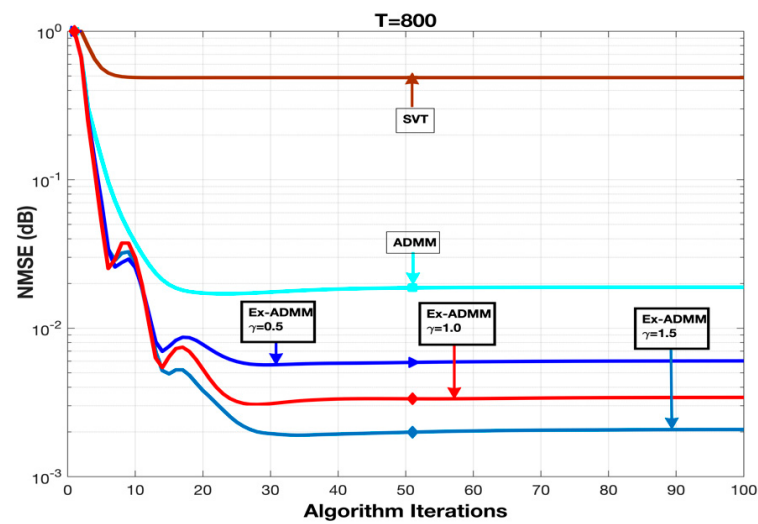
Estimation, in the proposed Ex-ADMM algorithm, is done by SVT implemented on  $H$ , which exploits the low rank property of  $H$  and  $l_1$  – norm, enforced on the submatrix to ensure the sparsity. Estimation of  $H_{\psi}$  becomes noisier, as the proposed Ex-ADMM algorithm and ADMM lack array gain, but the noise is not severe enough to cause a major effect on the estimated values. The proposed Ex-ADMM algorithm performed better than OMP, TSSR, VAMP and ADMM in terms of NMSE at different SNR points with respect to different training symbol lengths.

#### 7.4.3. Comparison of Convergence

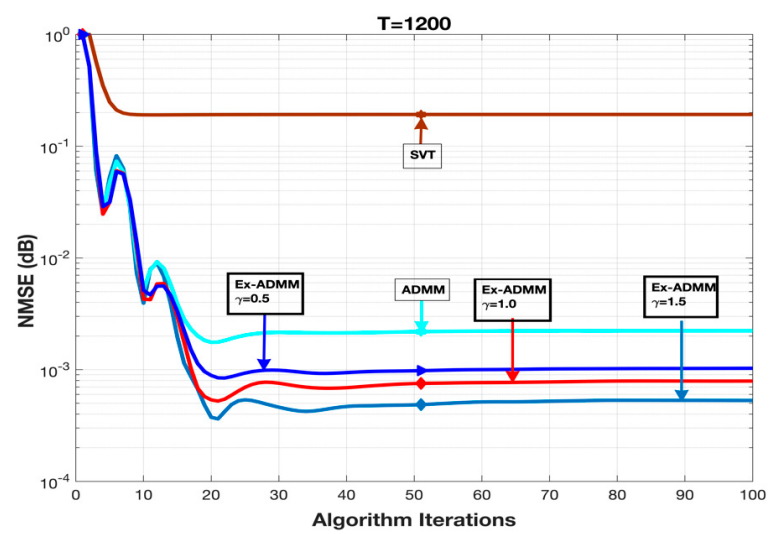
In Figure 4a–c, three values of  $\gamma$ , i.e.,  $\gamma = 0.5$ , 1 and 1.5 are used for simulation to demonstrate the effect of the relaxation factor on convergence with respect to the NMSE of the proposed Ex-ADMM algorithm. As the value of  $\gamma$  increases from 0.5 to 1 and then to 1.5, the convergence of proposed algorithm gets slightly better and converges to smaller NMSE values for all training symbol lengths (i.e.,  $T = 400$ , 800 and 1200) as compared to ADMM and SVT. Effects on NMSE for different algorithms of multiple paths are shown in Figure 4d. Due to the poor scattering nature of mmWave, the performance of all algorithms gets worse with escalation of the number of paths,  $L_p$ . However, the performance of the proposed algorithm is still better than that of the others.



(a)

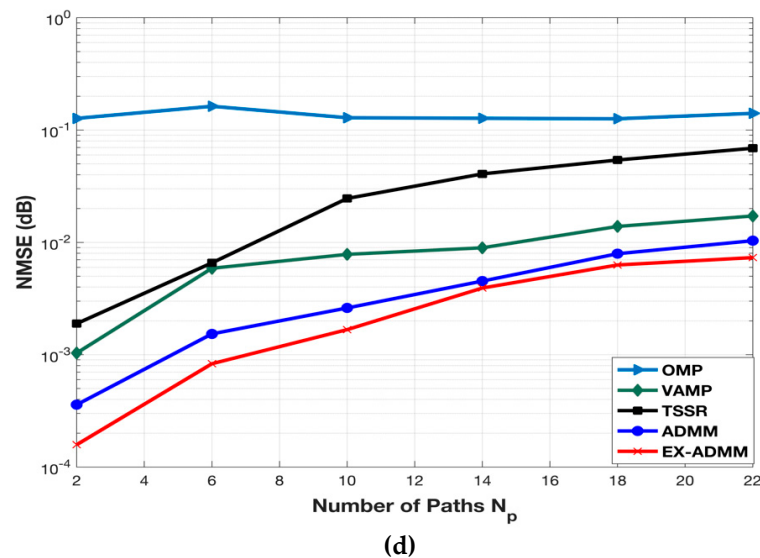


(b)



(c)

Figure 4. Cont.



**Figure 4.** (a–c) NMSE at a 30-dB transmit SNR for a  $64 \times 64$  mmWave MIMO channel at  $T = 400$ ,  $T = 800$  and  $T = 1200$ , with respect to algorithmic iteration for different  $\gamma$  values. (d) The effect of NMSE for multiple paths  $N_p$  at  $T = 2000$ .

## 8. Conclusions

In this paper, an extended version of ADMM (Ex-ADMM) was proposed for mmWave channel estimation. In the proposed scheme, a joint optimization problem was formulated, exploiting the sparsity and low rank properties of the channel matrix. The proposed Ex-ADMM algorithm exploits both the properties of targeted optimization problems by breaking them into several subproblems. These subproblems are then solved effortlessly, by acquiring their closed-form solutions independently. A relaxation factor was introduced to converge the proposed algorithm to smaller NMSE values for all training symbol lengths. Comprehensive simulation experiments were performed to validate the performance of Ex-ADMM. The proposed Ex-ADMM algorithm outperformed other benchmark algorithms in terms of ASE, NMSE and convergence rate.

**Author Contributions:** Conceptualization, P.S.S.; methodology, P.S.S. and A.H.W.; validation, P.S.S. and A.H.W.; writing, review and editing, L.C., A.H.W. and P.S.S.; funding acquisition, L.C.; supervision, L.C. and A.H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by National Major Special Funding (NMSF) under grant 2018ZX03001006-002.

**Acknowledgments:** The first author, P.S.S., hereby acknowledges the University of the Chinese Academy of Sciences (CAS) and the World Academy of Sciences (TWAS) for financial support for Ph.D. studies under the CAS-TWAS Fellowship. The first author would also like to acknowledge Pankaj Chaturvedi, Post-Doctoral Fellow, Y.M.C., Tsinghua University, Beijing, China for his help in mathematical expressions.

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

## References

1. Cheng, X.; Tang, C.; Zhang, Z. Accurate Channel Estimation for Millimeter-Wave MIMO Systems. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5159–5163. [\[CrossRef\]](#)
2. Kolawole, O.Y.; Biswas, S.; Singh, K.; Ratnarajah, T. Transceiver Design for Energy-Efficiency Maximization in mmWave MIMO IoT Networks. *IEEE Trans. Green Commun. Netw.* **2020**, *4*, 109–123. [\[CrossRef\]](#)
3. Prasad, R.; Vandendorpe, L. An overview of millimeter wave indoor wireless communication systems. In Proceedings of the Proceedings of 2nd IEEE International Conference on Universal Personal Communications, Ottawa, ON, Canada, 12–15 October 1993; Volume 2, pp. 885–889.

4. Luo, F.-L.; Zhang, C. 5G Millimeter-wave Communication Channel and Technology Overview. In *Signal Processing for 5G: Algorithms and Implementations*; IEEE: Piscataway, NJ, USA, 12–15 October 2016; pp. 354–371. ISBN 9781119116479.
5. Torkildson, E.; Madhow, U.; Rodwell, M. Indoor Millimeter Wave MIMO: Feasibility and Performance. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 4150–4160. [[CrossRef](#)]
6. Akoum, S.; El Ayach, O.; Heath, R.W. Coverage and capacity in mmWave cellular systems. In Proceedings of the 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 4–7 November 2012; pp. 688–692.
7. Smolders, A.B.; Reniers, A.C.F.; Johannsen, U.; Herben, M.H.A.J. Measurement and calibration challenges of microwave and millimeter-wave phased-arrays. In Proceedings of the 2013 International Workshop on Antenna Technology (iWAT), Karlsruhe, Germany, 4–6 March 2013; pp. 358–361.
8. Alkhateeb, A.; Mo, J.; Gonzalez-Prelcic, N.; Heath, R.W. MIMO Precoding and Combining Solutions for Millimeter-Wave Systems. *IEEE Commun. Mag.* **2014**, *52*, 122–131. [[CrossRef](#)]
9. Wei, L.; Hu, R.Q.; Qian, Y.; Wu, G. Key elements to enable millimeter wave communications for 5G wireless systems. *IEEE Wirel. Commun.* **2014**, *21*, 136–143.
10. Akdeniz, M.R.; Liu, Y.; Samimi, M.K.; Sun, S.; Rangan, S.; Rappaport, T.S.; Erkip, E. Millimeter Wave Channel Modeling and Cellular Capacity Evaluation. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1164–1179. [[CrossRef](#)]
11. Méndez-Rial, R.; Rusu, C.; González-Prelcic, N.; Alkhateeb, A.; Heath, R.W. Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches? *IEEE Access* **2016**, *4*, 247–267. [[CrossRef](#)]
12. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W. Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 831–846. [[CrossRef](#)]
13. Cao, Z.; Geng, H.; Chen, Z.; Chen, P. Sparse-Based Millimeter Wave Channel Estimation With Mutual Coupling Effect. *Electronics* **2019**, *8*, 358. [[CrossRef](#)]
14. Lu, X.; Yang, W.; Cai, Y.; Guan, X. Comparison of CS-Based Channel Estimation for Millimeter Wave Massive MIMO Systems. *Appl. Sci.* **2019**, *9*, 4346. [[CrossRef](#)]
15. Han, Y.; Lee, J. Asymmetric channel estimation for multi-user millimeter wave communications. In Proceedings of the 2016 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 19–21 October 2016; pp. 4–6.
16. Lee, J.; Gil, G.; Lee, Y.H. Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications. *IEEE Trans. Commun.* **2016**, *64*, 2370–2386. [[CrossRef](#)]
17. Li, X.; Fang, J.; Li, H.; Wang, P. Millimeter Wave Channel Estimation via Exploiting Joint Sparse and Low-Rank Structures. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 1123–1133. [[CrossRef](#)]
18. Vlachos, E.; Alexandropoulos, G.C.; Thompson, J. Massive MIMO Channel Estimation for Millimeter Wave Systems via Matrix Completion. *IEEE Signal Process. Lett.* **2018**, *25*, 1675–1679. [[CrossRef](#)]
19. Schniter, P.; Sayeed, A. Channel estimation and precoder design for millimeter-wave communications: The sparse way. In Proceedings of the 2014 48th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2–5 November 2014; pp. 273–277.
20. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]
21. Abernethy, J.; Bach, F.; Evgeniou, T.; Vert, J.-P. A new approach to collaborative filtering: Operator estimation with spectral regularization. *J. Mach. Learn. Res.* **2009**, *10*, 803–826.
22. Natarajan, N.; Dhillon, I.S. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics* **2014**, *30*, i60–i68. [[CrossRef](#)] [[PubMed](#)]
23. Chen, T.; Zhang, W.; Lu, Q.; Chen, K.; Zheng, Z.; Yu, Y. SVDFeature: A toolkit for feature-based collaborative filtering. *J. Mach. Learn. Res.* **2012**, *13*, 3619–3622.
24. Menon, A.K.; Chitrapura, K.-P.; Garg, S.; Agarwal, D.; Kota, N. Response prediction using collaborative filtering with hierarchies and side-information. In Proceedings of the 17th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, Manchester Grand Hyatt, San Diego, CA, USA, 21–24 August 2011; pp. 141–149.
25. Chiang, K.-Y.; Hsieh, C.-J.; Dhillon, I. Robust principal component analysis with side information. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016; pp. 2291–2299.

26. Lu, J.; Liang, G.; Sun, J.; Bi, J. A sparse interactive model for matrix completion with side information. In Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 4071–4079.
27. Du, J.; Li, J.; He, J.; Guan, Y.; Lin, H. Low-Complexity Joint Channel Estimation for Multi-User mmWave Massive MIMO Systems. *Electronics* **2020**, *9*, 301. [\[CrossRef\]](#)
28. Ma, F.; Ni, M.; Zhang, X.; Yu, Z. Solving Lasso: Extended ADMM is more efficient than ADMM. In Proceedings of the 2015 Chinese Automation Congress (CAC), Wuhan, China, 27–29 November 2015; pp. 55–58.
29. Peng, C.; Cheng, H.; Ko, M. An efficient two-stage sparse representation method. *Int. J. Pattern Recognit. Artif. Intell.* **2016**, *30*, 1651001. [\[CrossRef\]](#)
30. Schniter, P.; Rangan, S.; Fletcher, A.K. Vector approximate message passing for the generalized linear model. In Proceedings of the 2016 50th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 6–9 November 2016; pp. 1525–1529.
31. Ayach, O.E.; Rajagopal, S.; Abu-Surra, S.; Pi, Z.; Heath, R.W. Spatially Sparse Precoding in Millimeter Wave MIMO Systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1499–1513. [\[CrossRef\]](#)
32. Forenza, A.; Love, D.J.; Heath, R.W. Simplified Spatial Correlation Models for Clustered MIMO Channels With Different Array Configurations. *IEEE Trans. Veh. Technol.* **2007**, *56*, 1924–1934. [\[CrossRef\]](#)
33. Sayeed, A.M. Deconstructing multiantenna fading channels. *IEEE Trans. Signal Process.* **2002**, *50*, 2563–2579. [\[CrossRef\]](#)
34. Brady, J.; Behdad, N.; Sayeed, A.M. BeamSpace MIMO for Millimeter-Wave Communications: System Architecture, Modeling, Analysis, and Measurements. *IEEE Trans. Antennas Propag.* **2013**, *61*, 3814–3827. [\[CrossRef\]](#)
35. Keshavan, R.H.; Montanari, A.; Oh, S. Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **2010**, *56*, 2980–2998. [\[CrossRef\]](#)
36. Lin, Z.; Chen, M.; Ma, Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv* **2010**, arXiv:1009.5055.
37. Chiang, K.-Y.; Hsieh, C.-J.; Dhillon, I.S. Matrix completion with noisy side information. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, December 2015; pp. 3447–3455.
38. Nishihara, R.; Lessard, L.; Recht, B.; Packard, A.; Jordan, M.I. A general analysis of the convergence of ADMM. *arXiv* **2015**, arXiv:1502.02009.
39. Cai, X.; Gu, G.; He, B.; Yuan, X. A proximal point algorithm revisit on the alternating direction method of multipliers. *Sci. China Math.* **2013**, *56*, 2179–2186. [\[CrossRef\]](#)
40. Fortin, M.; Glowinski, R. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*; Elsevier: Amsterdam, The Netherlands, 2000; ISBN 008087536X.
41. Xu, Z.; Figueiredo, M.A.T.; Yuan, X.; Studer, C.; Goldstein, T. Adaptive Relaxed ADMM: Convergence Theory and Practical Implementation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7234–7243.
42. He, B.-S.; Xu, M.-H.; Yuan, X.-M. Block-Wise ADMM with a Relaxation Factor for Multiple-Block Convex Programming. *J. Oper. Res. Soc. China* **2018**, *6*, 485–505. [\[CrossRef\]](#)
43. Cai, J.-F.; Candes, E.; Shen, Z. A Singular Value Thresholding Algorithm For Matrix Completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [\[CrossRef\]](#)
44. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [\[CrossRef\]](#)
45. Molisch, A.F.; Ratnam, V.V.; Han, S.; Li, Z.; Nguyen, S.L.H.; Li, L.; Haneda, K. Hybrid Beamforming for Massive MIMO: A Survey. *IEEE Commun. Mag.* **2017**, *55*, 134–141. [\[CrossRef\]](#)
46. He, R.; Hu, B.; Zheng, W.-S.; Guo, Y. Two-stage sparse representation for robust recognition on large-scale database. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, GA, USA, 11–15 July 2010.
47. Rangan, S.; Schniter, P.; Fletcher, A.K. Vector approximate message passing. *IEEE Trans. Inf. Theory* **2019**, *65*, 6664–6684. [\[CrossRef\]](#)
48. Bucklew, J.A.; Radeke, R. On the Monte Carlo simulation of digital communication systems in Gaussian noise. *IEEE Trans. Commun.* **2003**, *51*, 267–274. [\[CrossRef\]](#)

49. Berriche, L.; Abed-Meraim, K.; Belfiore, J. Investigation of the channel estimation error on MIMO system performance. In Proceedings of the 2005 13th European Signal Processing Conference, Antalya, Turkey, 4–8 September 2005; pp. 1–4.
50. Yoo, T.; Goldsmith, A. Capacity and power allocation for fading MIMO channels with channel estimation error. *IEEE Trans. Inf. Theory* **2006**, *52*, 2203–2214.



© 2020 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/>).