Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6698424, 11 pages
https://doi.org/10.1155/2021/6698424

Research Article

Deep Learning-Based Symbol-Level Precoding for Large-Scale Antenna System

Changxu Xie, Huiqin Du, and Xialing Liu

College of Information Science and Technology, Jinan University, Guangzhou 510632, China

Correspondence should be addressed to Huiqin Du; thuiqin.du@jnu.edu.cn

Received 24 December 2020; Revised 20 February 2021; Accepted 17 March 2021; Published 30 March 2021

Academic Editor: K. Shankar

In this work, we consider a multiple input multiple-output system with large-scale antenna array which creates unintended multiuser interference and increases the power consumption due to the large number of radio frequency (RF) chains. The antenna selective symbol level precoding design is developed by minimizing the symbol error rate (SER) with limits of available RF chains. The ℓ_0-norm constrained nonconvex problem can be approximated as ℓ_1-minimization, which is further solved by alternating direction method of multipliers (ADMM) approach. The basic ADMM scheme is mapped into iterative construction process where the optimum solution is obtained by taking deep learning network as building block. Moreover, because that the standard ADMM algorithm is sensitive to the selection of hyperparameters, we further introduce the back propagation process to train the parameters. Simulation results show that the proposed deep learning ADMM scheme can achieve significantly low SER performance with small activated subset of transmit antennas.

1. Introduction

With the severe spectrum shortage in conventional cellular network, the large-scale antenna system in the millimeter-wave (mmWave) bands has been considered as a potential solution to meet the constantly growing demand by the users for higher data rate. The mmWave systems are capable of accommodating a large number of half-wavelength-based antennas while maintaining a compact form factor. High-gain multiple-input-and-multiple-output (MIMO) beamformers are advocated to compensate for their substantial path loss. There are many challenges regarding the implementation of conventional fully digital beamforming in large-scale antenna systems, such as complexity, energy consumption, and cost [1, 2]. Therefore, hybrid beamforming structures and symbol-level precoding (SLP) with limits of radio frequency (RF) chains are developed as alternative choices.

Hybrid beamforming that is comprised of a linear network of variable phase shifters in the RF domain, has typically been combined with baseband digital beamforming. For the multi-user downlink mmWave communication system, the inter-user interference is suppressed by zero-forcing- (ZF-) based hybrid beamforming algorithms [3, 4] and regularized ZF hybrid beamforming [5, 6]. The sum-rate maximization is addressed in [7], where the analog beamformer is optimized via alternating optimization approach. The maximization of the minimum total transmit rate with optimal power allocation is further developed via low-complexity path-following algorithms [8]. However, the limitation on the number of RF chains makes the designs of hybrid beamforming considerable challenging due to the unit-modulus constraints imposed on the analog beamformers. Unless the number of users is less than half the number of RF chains, a highly nonconvex design problems could result in computational intractability [9].

On the other hand, the SLP has been developed by tracking the interference symbol-by-symbol and turning the induced interference into constructive interference [10]. With the knowledge of data and channel state information (CSI), the early work on SLP focuses on the adaptation on linear precoding method, such as ZF-SLP beamformer [11, 12], signal-to-interference-and-noise ratio (SINR)-SLP [13–15], and minimum mean square error (MMSE)-SLP [16]. Recently, the SLP schemes have been widely combined with optimizations to achieve further performance improvements [17–19].

There are additional studies on constructive precoding including per-antenna power constraint [20], noise-robust SLP [21], symbol error rate (SER) minimization [22], and nonlinear channels [23]. Note that the implementation of existing techniques is usually based on standard optimization tools with high computational complexity. Considering the limited computational capabilities of communication systems, their implementation is infeasible, especially in the large-scale antenna array systems. In order to reduce the power consumption of digital RF chains, the SLP technique assisted with antenna selection (AS-SLP) is designed via the orthogonal matching pursuit (OMP) [24] and coordinated descent (CD) algorithms [25].

In this work, we consider mmWave communication system with the large-scale antenna array in which only a subset of transmit antennas is activated. In order to reduce the power consumption of RF chains, the proposed AS-SLP design is developed to minimize the achievable SER by using a subset of activated transmit antennas. The underlying combinatorial optimization problem is further transferred into a regularized ℓ_1-norm form. The optimum solution is achieved via alternating direction method of multipliers (ADMM) algorithm. However, the improper regularization parameters of the ℓ_1-norm optimization may degrade the SER performance, and the specification of the hyperparameters is a challenge. To overcome this difficulty, a deep architecture dubbed as ADMM-Net is introduced to link the iterative algorithm to the deep learning architecture. The hyperparameters become learnable and are optimized via back propagation by minimizing the corresponding loss function using the gradient algorithm. The main contributions of this work can be summarized as follows.

(i) The SLP design is developed to jointly select the optimum subset of RF chains and minimize the SER between the desired and received symbols over large-scale MIMO system. Instead of using exhaustive research, the optimization problem is approximated as a regularized ℓ_1-norm constrained problem and further solved by ADMM algorithm

(ii) Different from the traditional iteration-based or optimization-based algorithms, the ADMM process is mapped into a deep ADMM-Net with the data flow graph, in which the nth iteration corresponds to the n-stage data flow and three types of operations are mapped into three types of nodes. Under this structure, the proposed ADMM-Net is unfold without any extreme learning machine. The optimum solution can be iteratively recovered by constructive blocks

(iii) The back propagation process with the standard gradient-based optimizer is introduced to overcome the difficulty in determining the hyperparameters. The corresponding hyperparameters are learnt from back propagation by minimizing the gradient of the loss function. Assisted with back propagation, the proposed ADMM-Net can further improve the SER performance

(iv) The mean square error (MSE) and SER performance of the proposed ADMM-Net algorithm are compared with the state-of-the-art AS-SLP algorithms. Simulation results demonstrate that the proposed ADMM-Net scheme can reduce the achievable MSE and SER considerably and consume the low transmit power with optimum subset of transmit antennas

The remainder of the paper is organized as follows. Section 2 presents the AS-SLP model and formulates the optimization problem. Section 3 presents the detail of proposed ADMM-Net. The training procedure based on back propagation is established in Section 4. Simulation results are discussed in Section 5. Finally, Section 6 concludes the paper.

Notations: Matrices and vectors are typefaced using slanted bold uppercase and lowercase letters, respectively. $\mathbf{I}$ is an identity matrix. $\mathbb{C}^{m\times n}$ is used to describe the complex space of $m \times n$ matrices, and $\mathscr{CN}(\mathbf{m}, \mathbf{\Sigma})$ denotes a complex Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$. For a vector like $\mathbf{a}$, $\|\mathbf{a}\|_{i=1,2}$ are the Euclidean distance.

2. System Model and Problem Formulation

In this work, we consider a downlink multiuser MIMO system in which the base station (BS) equipped with N_t transmit antennas communicates K user terminals with single antenna. Considering high cost and power consumption of RF components, each transmission only activates a subset of M_t transmit antennas ($N_t \geq M_t$). The active transmit antennas are selected by switches that connect with the available RF chains. A mmWave channel $\tilde{\mathbf{H}} \in \mathbb{C}^{K\times N_t}$ is considered [26], such as

$$\tilde{\mathbf{H}} = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_r(\phi_l^r, \theta_l^r)\mathbf{a}_t^H(\phi_l^t, \theta_l^t), \tag{1}$$

where L is the total number of transmission paths and α_l is the complex gain of the lth path of i.i.d. $\mathscr{CN}(0,1)$. The variables $\phi_l^{r(t)} \in [0, 2\pi]$ and $\theta_l^{r(t)} \in [0, 2\pi]$ are the azimuth angles of arrival and departure (AoAs/AoDs) of lth path of the receiver or transmitter, respectively. The variables $\mathbf{a}_r(\phi_l^r, \theta_l^r)$ and $\mathbf{a}_t(\phi_l^t, \theta_l^t)$ are the antenna array response vectors at the transmitter and receiver, respectively. The receive signal can be equivalently rewritten as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{2}$$

with $\tilde{\mathbf{n}}$ is the additive white Gaussian noise (AWGN) noise followed a circularly symmetric complex Gaussian distribution, $\mathscr{CN}(0, \sigma_n^2\mathbf{I})$. In this work, we adopt the SLP scheme in which the precoded transmitted signal $\tilde{\mathbf{x}} \in \mathbb{C}^{N_t\times 1}$ is designed on a symbol-by-symbol basis as a function of the instantaneous CSI, that is,

$$\tilde{\mathbf{x}} = \mathscr{P}(\tilde{\mathbf{s}}, \tilde{\mathbf{H}}), \tag{3}$$

where $\tilde{\mathbf{s}} \in \mathbb{C}^{N_r\times 1}$ is the intended symbols drawn from prespecified constellation; the operation $\mathscr{P}$ is the SLP conventional scheme.

Suppose the intended symbol $\tilde{s}_k$ is the receive constellation point, the upper-bound of SER can be further expressed as [27]

$$\mathrm{SER} \lesssim \sum_i \frac{g_i}{N^2} \mathcal{Q}\left(\frac{d - 2\widehat{d}_{ii}}{2\sigma}\right), \tag{4}$$

where g_i is the number of minimum distance neighbors of the intended symbol, $\mathcal{Q}(u) = (1/2\pi)\int_u^\infty e^{-(v^2/2)}dv$, and the constant d denotes as the minimum distance between two closest neighboring constellation points. Moreover, $\widehat{d}_{ii}$ is the distance between the intended symbol $\tilde{s}_k$ and noiseless received symbol $\widehat{\tilde{s}}_k$, that is, $\widehat{d}_{ii} \triangleq |\widehat{\tilde{s}}_k - \tilde{s}_k| = |\tilde{\mathbf{h}}_k^H \tilde{\mathbf{x}} - \tilde{s}_k|$. Because of the decreasing property of $\mathcal{Q}$-function, the SER is determined by the distance $\widehat{d}_{ii}$.Considering the hardware complexity of large antenna array systems, we strive to activate a subset of transmit antennas and minimize the achievable SER by minimizing the average Euclidean distance between the received signal by each user and the desired information symbols, that is,

$$\min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \sqrt{\gamma}\mathbf{s}\|_2^2, \tag{5a}$$

$$s.t. \|\mathbf{x}\|_0 = M_t, \tag{5b}$$

where γ is the transmit signal-to-noise ratio (SNR), and

$$\mathbf{x} = \begin{bmatrix} \Re\{\tilde{\mathbf{x}}\} \\ \Im\{\tilde{\mathbf{x}}\} \end{bmatrix}, \mathbf{s} = \begin{bmatrix} \Re\{\tilde{\mathbf{s}}\} \\ \Im\{\tilde{\mathbf{s}}\} \end{bmatrix}, \tag{6}$$

$$\mathbf{H} = \begin{bmatrix} \Re\{\tilde{\mathbf{H}}\} & -\Im\{\tilde{\mathbf{H}}\} \\ \Im\{\tilde{\mathbf{H}}\} & \Re\{\tilde{\mathbf{H}}\} \end{bmatrix}. \tag{7}$$

Note that the ℓ_0-norm constraint determines the number of activated transmit antenna; therefore, the optimum solution $\mathbf{x}^*$ works as antenna selection implicitly. However, the aforementioned optimization problem is NP hard due to the nonconvexity of the ℓ_0-norm constraint. A straightforward approach would compute the precoding vector for every possible M_t combinations by minimizing the objective function (5a). Once the optimum subset of activated transmit antennas is determined, the optimum precoded signal $\mathbf{x}^*$ can be designed based on conventional SLP schemes, such as ZF algorithm.

$$\mathbf{x}^* = \sqrt{\beta}\mathbf{H}_{\mathrm{eff}}^H \left(\mathbf{H}_{\mathrm{eff}}\mathbf{H}_{\mathrm{eff}}^H\right)^{-1}\mathbf{s}, \tag{8}$$

where $\mathbf{H}_{\mathrm{eff}}$ is the effective channel matrix. It is acquired from $\mathbf{H}$ after replacing the rest of its $N_t - M_t$ columns which correspond to the zero elements of $\mathbf{x}$ with null $N_r \times 1$ vectors. However, with N_t available antenna, there are $\binom{N_t}{M_t}$ possible subsets of M_t activated antenna. The straightforward exhaustive search method becomes computationally inefficient for the large-scale antenna array system, which motivates us to develop an efficient approach that can determine the optimum M_t activated antennas with low complexity.

3. ADMM-Net for SLP Model

3.1. ADMM Solver. Due to the ℓ_0-norm constraint, the optimization problem is NP-hard, which can be reformulated, and the optimum solution can be obtained by the recent developed algorithms, such as OMP and CD algorithms. In this work, the ℓ_1-norm regularization is introduced as one of the effective approaches to jointly optimize multiple variables, that is,

$$\min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \sqrt{\gamma}\mathbf{s}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{9}$$

where $\|\cdot\|_1$ is the ℓ_1-norm as convex surrogate of ℓ_0-norm, and the parameter $\lambda > 0$ controls the trade-off between noise sensitivity and signal sparsity. The augmented Lagrangian method (9) can be bring robustness compared with dual ascent method to achieve convergence even if the underlying problem is not strictly convex. However, the augmented Lagrangian cannot be separately minimized over the variables. Therefore, the ADMM algorithm that blends the decomposability of dual ascent with the superior convergence properties of the method of multipliers is introduced to solve this problem. By introducing an intermediate variable $\mathbf{z} \in \mathbb{R}^{N_t \times 1}$ as bridge parameter to establish consensus among $\mathbf{x}$, the problem (9) can be reformulated as

$$\begin{aligned} &\min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \sqrt{\gamma}\mathbf{s}\|_2^2 + \lambda\|\mathbf{z}\|_1, \\ &s.t. \mathbf{z} = \mathbf{x}. \end{aligned} \tag{10}$$

The corresponding augmented Lagrangian function over $\mathbf{x}$ and $\mathbf{z}$ can be written as

$$\mathfrak{L}(\mathbf{x}, \mathbf{z}, \alpha) = \|\mathbf{H}\mathbf{x} - \sqrt{\gamma}\mathbf{s}\|_2^2 + \lambda\|\mathbf{z}\|_1 + \frac{\rho}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 - \langle\alpha, \mathbf{x} - \mathbf{z}\rangle, \tag{11}$$

where $\alpha \in \mathbb{R}^{N_t \times 1}$ is a dual variable and $\rho > 0$ is a positive penalty parameters. Let $\beta = \alpha/\rho$ as the scaled form of dual variable, the ADMM algorithm alternately optimizes $\{\mathbf{x}, \mathbf{z}, \alpha\}$ by solving the following subproblems, such as

$$\begin{cases} \mathbf{x}^{(n+1)} = \underset{\mathbf{x}}{\mathrm{argmin}} \left\|\mathbf{H}\mathbf{x}^{(n)} - \sqrt{\gamma}\mathbf{s}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{x}^{(n)} + \beta^{(n)} - \mathbf{z}^{(n)}\right\|_2^2, \\ \mathbf{z}^{(n+1)} = \underset{\mathbf{z}}{\mathrm{argmin}} \frac{\rho}{2}\left\|\mathbf{x}^{(n)} + \beta^{(n)} - \mathbf{z}^{(n)}\right\|_2^2 + \lambda\left\|\mathbf{z}^{(n)}\right\|_1, \\ \beta^{(n+1)} \leftarrow \beta^{(n)} + \eta\left(\mathbf{x}^{(n+1)} - \mathbf{z}^{(n+1)}\right), \end{cases} \tag{12}$$

where n denotes the nth iteration and η is an update parameter for the Lagrangian multiplier. The variables $\mathbf{x}$ and $\mathbf{z}$ are iteratively updated in alternative directions to complete joint

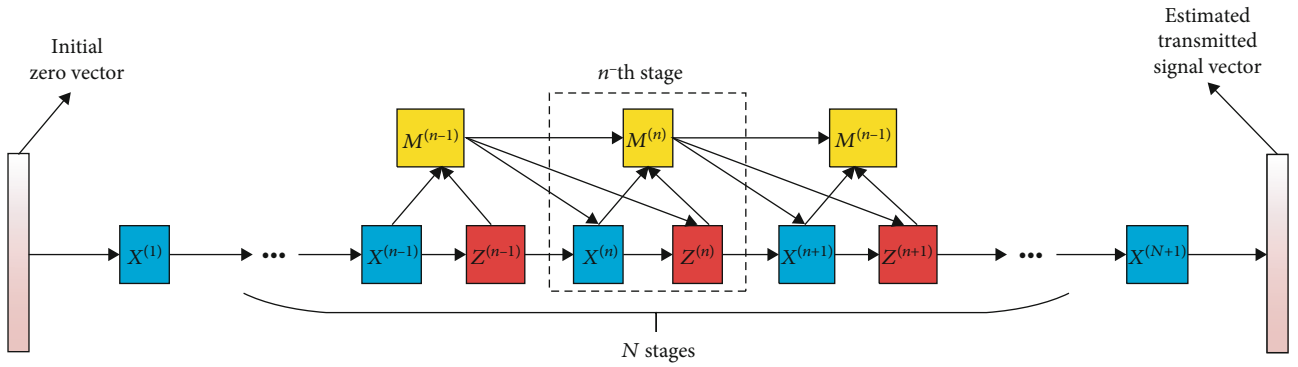


FIGURE 1: The data flow graph of ADMM-net. There are three types of layers in the nth stage, including multiplier update layer ($\mathbf{M}^{(n)}$), nonlinear transform layer ($\mathbf{Z}^{(n)}$), and reconstruction layer ($\mathbf{X}^{(n)}$).

minimization, and the dual variable β is updated according to $\mathbf{x}$ and $\mathbf{z}$.

3.2. Deep ADMM-Net Architecture. Based on the iterations in ADMM solver, we map the standard ADMM iterative process (13) into a deep ADMM-Net with a data flow graph, which is comprised of the nodes and the connecting lines shown in Figure 1. More specifically, the nth iteration of the ADMM algorithm corresponds to the data flow in the n th stage which are connected by solid lines. Three types of operations are mapped into three types of nodes, that is, reconstruction layer $\mathbf{X}^{(n)}$, nonlinear transform layer $\mathbf{Z}^{(n)}$, and multiplier update layer $\mathbf{M}^{(n)}$. Each layer is discussed as follows.

3.2.1. Reconstruction Layer $\mathbf{X}^{(n)}$. This layer reconstructs the symbol-level precoder $\mathbf{x}$ in (12). Given $\mathbf{z}^{(n-1)}$ and $\beta^{(n-1)}$ from previous layers in stage $n-1$, the output of this layer is defined as

$$\mathbf{x}^{(n)} = \left(2\mathbf{H}^T\mathbf{H} + \rho^{(n)}\mathbf{I}\right)^{-1}\left[2\sqrt{\gamma}\mathbf{H}^T\mathbf{s} + \rho^{(n)}\left(\mathbf{z}^{(n-1)} - \beta^{(n-1)}\right)\right], \tag{13}$$

where $\mathbf{s}$ is the initialize input and $\rho^{(n)}$ is the learnable penalty parameter in the nth stage. The output $\mathbf{x}^{(n+1)}$ of this layer is the input for subsequent multiplier update layer $\mathbf{M}^{(n)}$ and nonlinear transform layer $\mathbf{Z}^{(n)}$ in the nth stage. Note that, for the first stage, the reconstruction layer $\mathbf{X}^{(1)}$ is defined based on the inputs channel gain matrix $\mathbf{H}$ and the signal symbol vector $\mathbf{s}$, that is,

$$\mathbf{x}^{(1)} = \left(2\mathbf{H}^T\mathbf{H} + \rho^{(1)}\mathbf{I}\right)^{-1}\left(2\sqrt{\gamma}\mathbf{H}^T\mathbf{s}\right). \tag{14}$$

3.2.2. Nonlinear Transform Layer $\mathbf{Z}^{(n)}$. This layer performs the nonlinear transform via the shrinkage function $\mathcal{S}(\mathbf{w}, \tau)$, that is,

$$\mathcal{S}(\mathbf{w}, \tau) = \begin{cases} \mathbf{w} - \tau, & \mathbf{w} > \tau, \\ \mathbf{w} + \tau, & \mathbf{w} < -\tau, \\ 0, & \text{otherwise}, \end{cases} \tag{15}$$

in which the soft-threshold function controls the sparsity of the output by adjusting the variable. The data flow of this layer in the nth stage are from two previous layers $\mathbf{X}^{(n)}$ and $\mathbf{M}^{(n-1)}$. Based on (12), given $\mathbf{X}^{(n)}$ and $\mathbf{M}^{(n-1)}$, λ and ρ, the nonlinear transform layer $\mathbf{Z}^{(n)}$ is defined as

$$\mathbf{z} = \mathcal{S}\left(\mathbf{x} + \beta, \frac{\lambda}{\rho}\right), \tag{16}$$

that is,

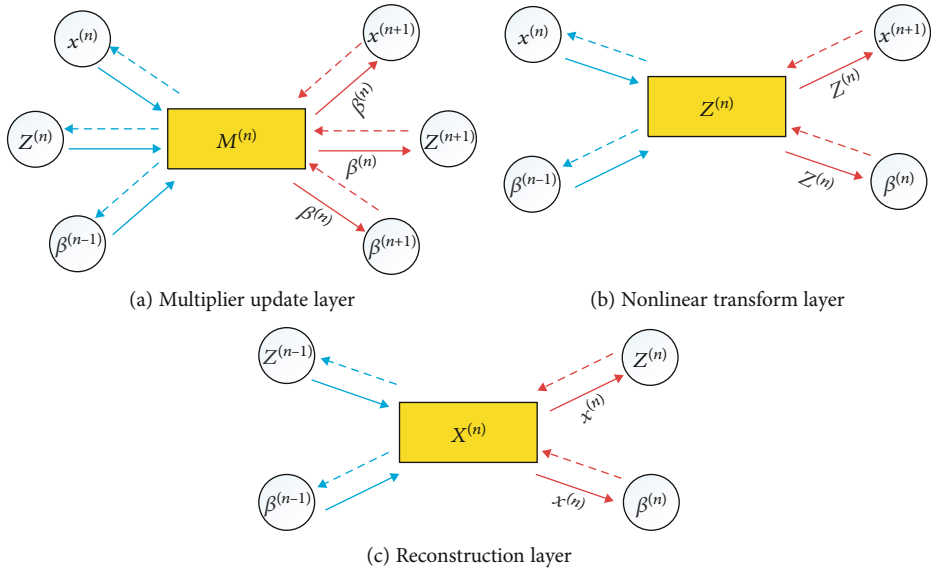
$$z_i^{(n)} = \begin{cases} z_i^{(n)} - \dfrac{\lambda^{(n)}}{\rho^{(n)}}, & z_i^{(n)} > \dfrac{\lambda^{(n)}}{\rho^{(n)}}, \\ 0, & \left|z_i^{(n)}\right| \le \dfrac{\lambda^{(n)}}{\rho^{(n)}}, \\ z_i^{(n)} + \dfrac{\lambda^{(n)}}{\rho^{(n)}}, & z_i^{(n)} < -\dfrac{\lambda^{(n)}}{\rho^{(n)}}, \end{cases} \tag{17}$$

where $z_i^{(n)}$ is the ith elements of $\mathbf{z}^{(n)}$. Each element of $\mathbf{z}^{(n)}$ is compared with $\lambda^{(n)}/\rho^{(n)}$ in sequence. Hence, the number of nonzero elements of variable $\mathbf{z}$ is depended on the value of λ/ρ.

3.2.3. Multiplier Update Layer $\mathbf{M}^{(n)}$. This layer performs an update process of the joint dual variables of $\mathbf{x}$ and $\mathbf{z}$. Given the inputs $\beta^{(n-1)}$, $\mathbf{x}^{(n)}$, and $\mathbf{z}^{(n)}$, the output of this layer ($\mathbf{M}^{(n)}$) in the nth stage is defined as

$$\beta^{(n)} = \beta^{(n-1)} + \eta^{(n)}\left(\mathbf{x}^{(n)} - \mathbf{z}^{(n)}\right), \tag{18}$$

where $\eta^{(n)}$ is the learnable parameters in the nth stage.



(a) Multiplier update layer

(b) Nonlinear transform layer

(c) Reconstruction layer

FIGURE 2: Data flows of three types of layers in network. The solid arrow is the data flow in forward paths, and the dashed arrow is the data flow in backward paths when computing gradients in back propagation.

3.2.4. Network Parameters. The deep ADMM-Net architecture is designed to update the hyperparameters, such as $\rho^{(n)}$ in the reconstruction layer, $\lambda^{(n)}$, $\rho^{(n)}$ in the nonlinear transform layer, and $\eta^{(n)}$ in the multiplier update layer. All of these parameters are taken as the weights of the neural network to be learned.

4. Network Training

In the conventional ADMM solver, the parameter set $\Theta = \{\lambda^{(i)}, \rho^{(i)}, \eta^{(i)}\}_{i=1}^{N}$ are initialized randomly or determined empirically, where $n = 1, 2, \cdots, N$ are indexes for each stage. The different initializations would affect the accuracy of recovered transmitted signal vector $\hat{\mathbf{x}}$ and the achievable objective function in (9). The network training is introduced in which the hyperparameters become learnable variables because of back propagation with the gradient-based algorithm. In the training phase, we first sample a set of channel matrix. Each element of $\mathbf{H}$ is generated by applying (1). The noise variance was randomly sampled based on SNR requirement. Suppose the generated training data set is $\Omega = \{(\mathbf{H}^{(k)}, \mathbf{s}^{(k)}, \mathbf{x}^{(k)}), k = 1, 2, \cdots, N_s\}$, where N_s is the number of samples, $\mathbf{x}^{(k)}$ is the kth transmitted signal vector, and $\mathbf{s}^{(k)}$ is the kth observed symbol vector. After determining the loss function, the training data is employed to optimize for the parameter set Θ which is randomly initialized at the beginning of algorithm. In the testing phase, suppose the new observed signal $\mathbf{s}$ and its corresponding channel matrix $\mathbf{H}$ are known, we feed the trained ADMM-Net with learned hyperparameters to obtain the optimum $\mathbf{x}$. More details are presented as follows.

4.1. Loss Function. In this work, the AS-SLP design is obtained based on the knowledge of intended symbol and channel matrix. Hence, the normalized mean square error (NMSE) between the intended symbol and estimated symbol is chosen as the loss function. Given series of training data, the loss between the network output and intended transmitted signal vector can be defined as

$$E(\Theta) = \frac{1}{|\Omega|} \sum_{(\mathbf{H},\mathbf{s})\in\Omega} \frac{\|\mathbf{H}\hat{\mathbf{x}}(\Theta) - \mathbf{s}\|_2^2}{\|\mathbf{s}\|_2^2}, \tag{19}$$

where $\hat{\mathbf{x}}(\Theta)$ is the estimated transmitted signal vector as the network output, and it is obtained based on parameter set Θ and data set Ω. The parameters of deep ADMM-Net can be learned by minimizing the loss $E(\Theta)$ using gradient-based algorithm L-BFGS [28].

4.2. Back Propagation. To compute the gradients of loss, the parameter set is updated via back propagation. The data of the nth stage are flowed by the order of $\mathbf{X}^{(n)}$, $\mathbf{Z}^{(n)}$, and $\mathbf{M}^{(n)}$ in the forward paths, and we compute the gradients in an inverse order in the backward paths. For the nth stage, the gradients computation for each layer are briefly introduced as follows.

4.2.1. Multiplier Update Layer $\mathbf{M}^{(n)}$. Figure 2 shows that this layer has three inputs, that is, $\beta^{(n-1)}$, $\mathbf{z}^{(n)}$, and $\mathbf{x}^{(n)}$. Its output $\beta^{(n)}$ computes $\beta^{(n+1)}$, $\mathbf{z}^{(n+1)}$, and $\mathbf{x}^{(n+1)}$ in next layers. The parameter of this layer is $\eta^{(n)}$, and the gradient of the loss with respect to this parameter can be computed as

$$\frac{\partial E}{\partial \eta^{(n)}} = \frac{\partial E}{\partial \beta^{(n)T}} \frac{\partial \beta^{(n)}}{\partial \eta^{(n)}} = \frac{\partial E}{\partial \beta^{(n)T}} \left(\mathbf{x}^{(n)} - \mathbf{z}^{(n)}\right), \tag{20}$$

where $\partial E/(\partial \beta^{(n)T}) = (\partial E/(\partial \beta^{(n+1)T}))((\partial \beta^{(n+1)})/(\partial \beta^{(n)})) + (\partial E/(\partial \mathbf{z}^{(n+1)T}))((\partial \mathbf{z}^{(n+1)})/\partial \beta^{(n)}) + (\partial E/(\partial \mathbf{x}^{(n+1)T}))((\partial \mathbf{x}^{(n+1)})/(\partial$

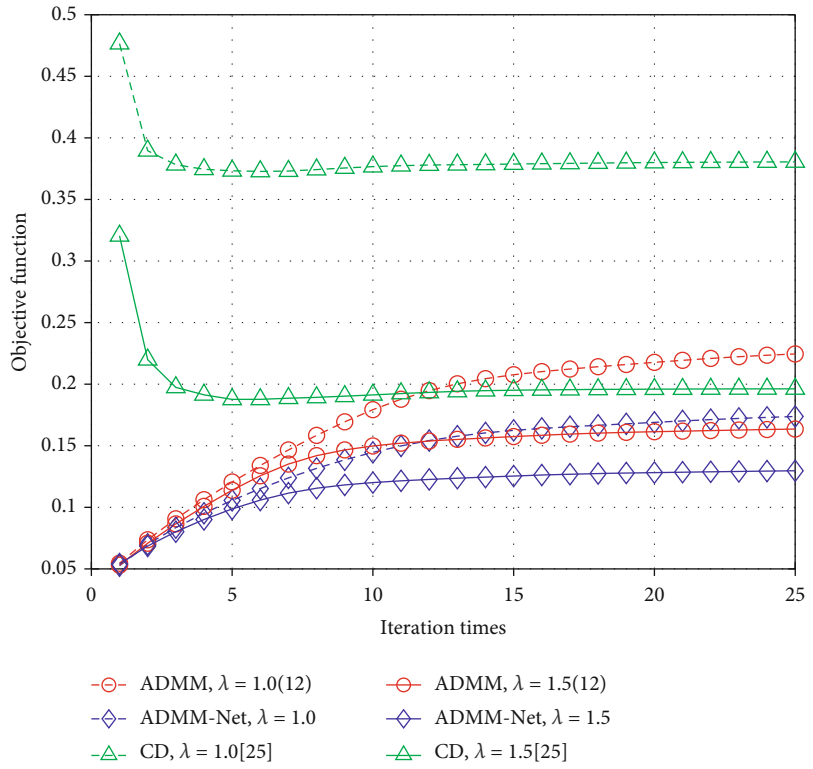


Figure 3: Convergence rate versus iteration number.

$\beta^{(n)}$)) is the summation of gradients along three dashed red arrows. The gradients of the loss with respect to three inputs $\beta^{(n-1)}$, $\mathbf{z}^{(n)}$, and $\mathbf{x}^{(n)}$ in this layer can be presented as

$$\frac{\partial E}{\partial \beta^{(n-1)}} = \frac{\partial E}{\partial \beta^{(n)T}} \frac{\partial \beta^{(n)}}{\partial \beta^{(n-1)}} = \frac{\partial E}{\partial \beta^{(n)T}}, \tag{21a}$$

$$\frac{\partial E}{\partial \mathbf{x}^{(n)}} = \eta^{(n)} \frac{\partial E}{\partial \beta^{(n)T}} = -\frac{\partial E}{\partial \mathbf{z}^{(n)}}. \tag{21b}$$

4.2.2. Nonlinear Transform Layer $\mathbf{Z}^{(n)}$. Figure 2 shows that the nonlinear transform layer has two inputs $\beta^{(n-1)}$ and $\mathbf{x}^{(n)}$ with two parameters $\lambda^{(n)}$ and $\rho^{(n)}$. The gradient of the loss with respect to two parameters can be computed as

$$\frac{\partial E}{\partial \lambda^{(n)}} = \frac{\partial E}{\partial \mathbf{z}^{(n)T}} \frac{\partial \mathbf{z}^{(n)}}{\partial \lambda^{(n)}}, \frac{\partial E}{\partial \rho^{(n)}} = \frac{\partial E}{\partial \mathbf{z}^{(n)T}} \frac{\partial \mathbf{z}^{(n)}}{\partial \rho^{(n)}}, \tag{22}$$

where $(\partial E/\partial \mathbf{z}^{(n)T}) = (\partial E/(\partial \beta^{(n)T}))(\partial \beta^{(n)}/\partial \mathbf{z}^{(n)}) + (\partial E/(\partial \mathbf{x}^{(n+1)T}))(\partial \mathbf{x}^{(n+1)}/\partial \mathbf{z}^{(n)})$, $\partial \mathbf{z}^{(n)}/\partial \lambda^{(n)}$, and $\partial \mathbf{z}^{(n)}/\partial \rho^{(n)}$ are calculated by following method. Let $z_i^{(n)}$ be the ith element of $\mathbf{z}^{(n)}$, it yields

$$\frac{\partial z_i^{(n)}}{\partial \lambda^{(n)}} = \begin{cases} -\dfrac{1}{\rho^{(n)}} & z_i^{(n)} > \dfrac{\lambda^{(n)}}{\rho^{(n)}}, \\ 0 & \left| z_i^{(n)} \right| \le \dfrac{\lambda^{(n)}}{\rho^{(n)}}, \\ \dfrac{1}{\rho^{(n)}} & z_i^{(n)} < -\dfrac{\lambda^{(n)}}{\rho^{(n)}}. \end{cases} \tag{23}$$

Similarly, the term $\partial z_j^{(n)}/\partial \rho^{(n)}$ can be illustrated as

$$\frac{\partial z_j^{(n)}}{\partial \rho^{(n)}} = \begin{cases} \dfrac{1}{\left[\rho^{(n)}\right]^2} & z_i^{(n)} > \dfrac{\lambda^{(n)}}{\rho^{(n)}}, \\ 0 & \left| z_i^{(n)} \right| \le \dfrac{\lambda^{(n)}}{\rho^{(n)}}, \\ -\dfrac{1}{\left[\rho^{(n)}\right]^2} & z_i^{(n)} < -\dfrac{\lambda^{(n)}}{\rho^{(n)}}. \end{cases} \tag{24}$$

The gradients of the loss with respect to two inputs $\beta^{(n-1)}$ and $\mathbf{x}^{(n)}$ in this layer are

$$\frac{\partial E}{\partial \beta^{(n-1)}} = \frac{\partial E}{\partial \mathbf{z}^{(n)T}} \frac{\partial \mathbf{z}^{(n)}}{\partial \beta^{(n-1)}}, \tag{25a}$$

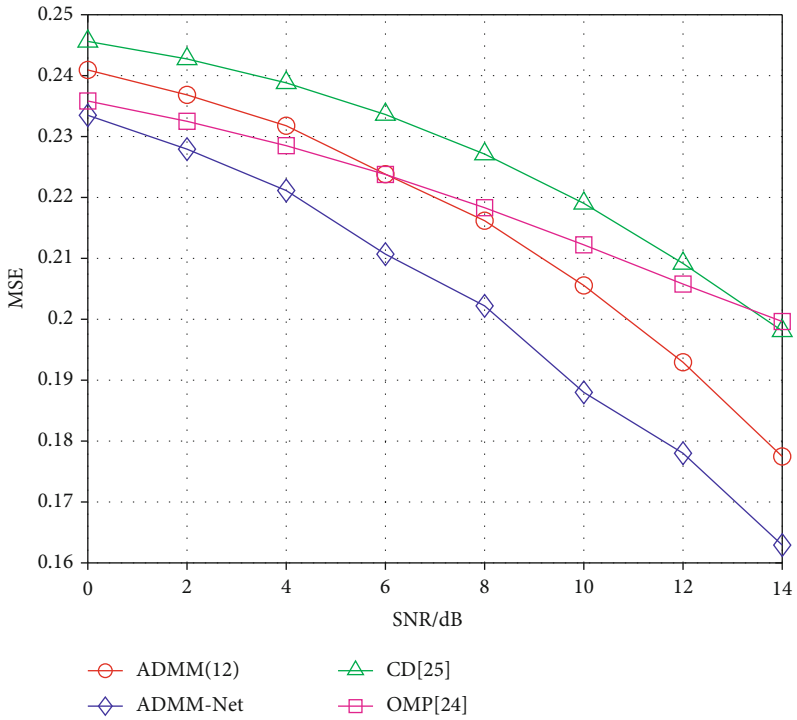


FIGURE 4: MSE performance versus SNR with $N_t = 16$ transmit antennas, $M_t = 4$ activated antennas and 4-QAM.

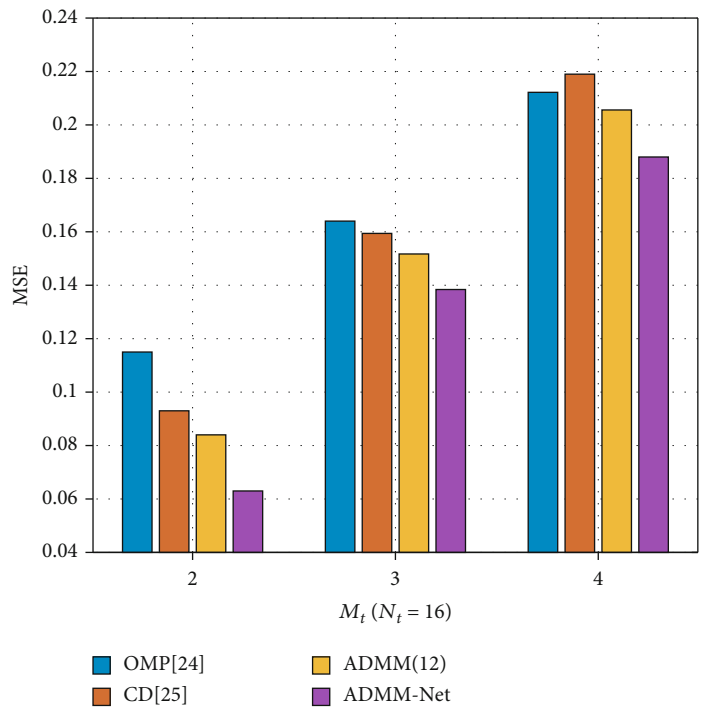


FIGURE 5: MSE performance versus number of activated antennas with $N_t = 16$ transmit antennas and 4-QAM under SNR = 10dB.

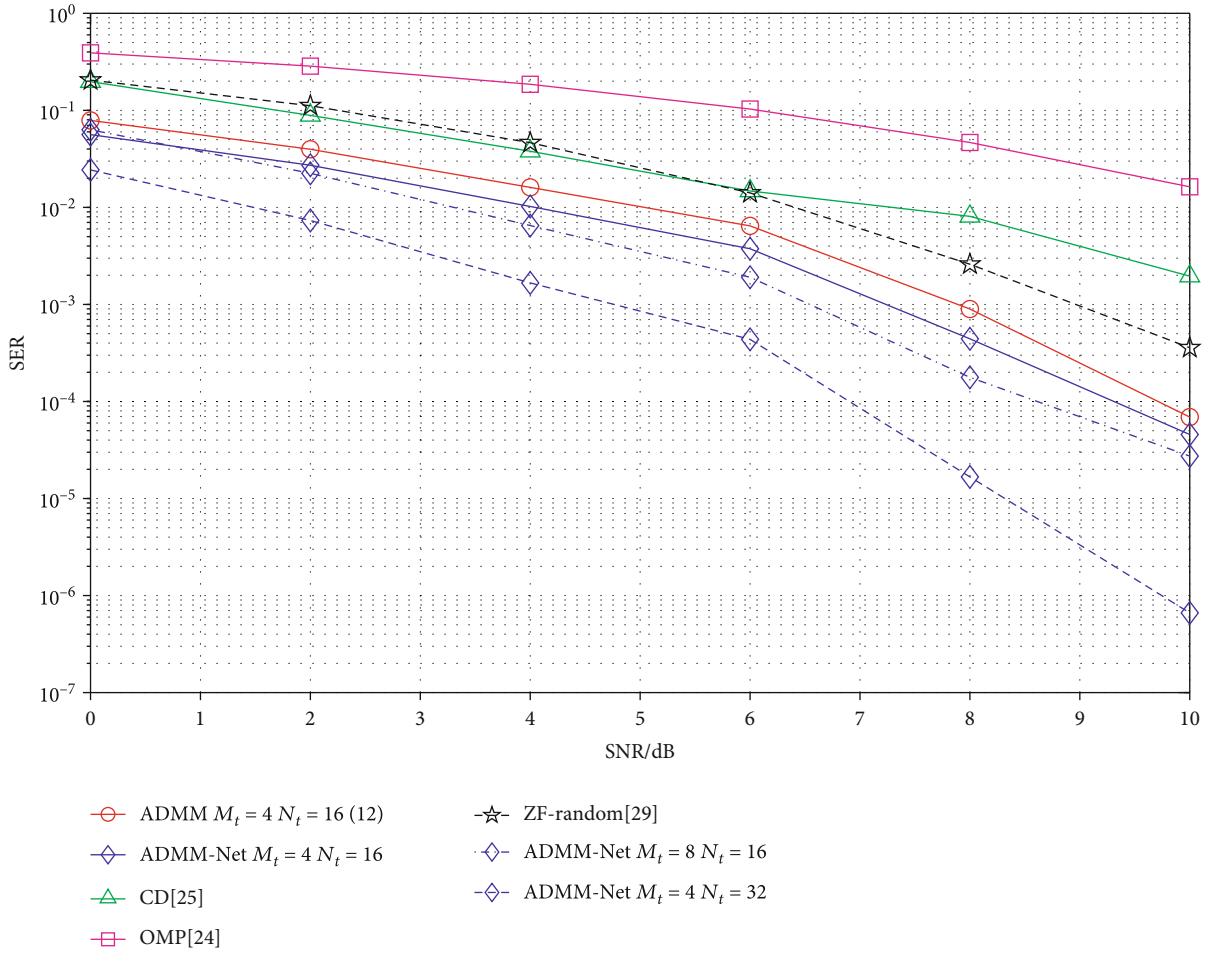


Figure 6: SER performance versus SNR with different number of transmit antennas and activated antennas and 4-QAM.

$$\frac{\partial E}{\partial \mathbf{x}^{(n)}} = \frac{\partial E}{\partial \mathbf{z}^{(n)T}} \frac{\partial \mathbf{z}^{(n)}}{\partial \mathbf{x}^{(n-1)}}. \tag{25b}$$

4.2.3. Reconstruction Layer $\mathbf{X}^{(n)}$. In Figure 2, the reconstruction layer contains two inputs $\beta^{(n-1)}$ and $\mathbf{z}^{(n-1)}$ and one parameters $\rho^{(n)}$. Let $Q = (2\mathbf{H}^T\mathbf{H} + \rho^{(n)}\mathbf{I})^{-1}$, the gradient of the loss with respect to this parameter can be computed as

$$\frac{\partial E}{\partial \rho^{(n)}} = \frac{\partial E}{\partial \mathbf{x}^{(n)T}} Q\left\{\left(\mathbf{z}^{(n-1)} - \beta^{(n-1)}\right) - Q\left[\mathbf{H}^T\mathbf{s} + \rho\left(\mathbf{z}^{(n-1)} - \beta^{(n-1)}\right)\right]\right\}, \tag{26}$$

where $\partial E/(\partial \mathbf{x}^{(n)T}) = (\partial E/(\partial \beta^{(n)T}))(\partial \beta^{(n)}/\partial \mathbf{x}^{(n)}) + (\partial E/(\partial \mathbf{z}^{(n)T}))(\partial \mathbf{z}^{(n)}/\partial \mathbf{x}^{(n)})$. The gradients of the loss with respect to the inputs $\beta^{(n-1)}$ and $\mathbf{z}^{(n-1)}$ in this layer can be written as

$$\frac{\partial E}{\partial \mathbf{z}^{(n-1)}} = \rho^{(n)} Q \frac{\partial E}{\partial \mathbf{x}^{(n)}} = -\frac{\partial E}{\partial \beta^{(n-1)}}. \tag{27}$$

5. Simulation

In this work, we consider the massive MIMO system over mmWave channel which includes a single BS and multiple user terminals. For each channel matrix realization, it is assumed that $L = 10$, $\phi_l^{r(t)}$ and $\theta_l^{r(t)}$ are both uniform randomly selected from the interval $[0, 2\pi]$. The 4-order quadrature amplitude modulation (4-QAM) is considered. The proposed AS-SLP design via deep ADMM-Net with back propagation (abbr. as ADMM-Net) is compared with the AS-SLP design via basic ADMM solver (abbr. as ADMM) (12), the AS-SLP design via OMP (abbr. as OMP), the AS-SLP design via CD (abbr. as CD), and the ZF SLP with random antenna selection (abbr. as ZF-random) [29]. Note that all the SLP schemes are designed by ZF method once the sparse pattern of $\mathbf{x}$ is determined. In the ADMM algorithm, without specified indication, the initialized hyperparameters are set as $\text{iter}_{\max} = 100$, $\lambda = 1.5$, $\rho = 3$, $\eta = 1$, and $\varepsilon = 10^{-4}$, respectively.

Figure 3 illustrates the convergence performance of ADMM, ADMM-Net, and CD algorithms with $N_t = 32$ transmit antennas and $M_t = 8$ activated antennas under SNR= 10dB. The group of solid curves and dash curves were initialized with the parameters $\lambda = 1.5$ and $\lambda = 1.0$, respectively. The achievable objective function in (5a) is defined as the average Euclidean distance between the received symbols and the desired symbols, that is $\|\mathbf{Hx} - \sqrt{\gamma}\mathbf{s}\|_2^2$. Given the same input parameter set Θ, the objective function via the

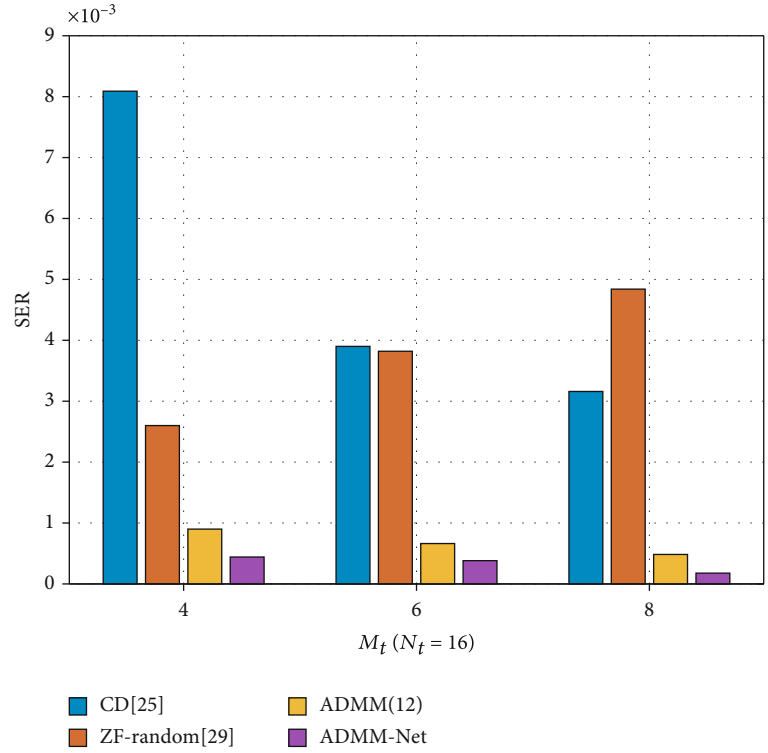


Figure 7: SER performance versus number of activated antennas with $N_t = 16$ transmit antennas and 4-QAM under SNR = 8dB.

proposed deep ADMM-Net algorithm converges to the lowest value. Because of back propagation process, the learnable hyperparameters provide the improvement on the objective function. On the other hand, the convergence speed of CD algorithm is faster than that of ADMM framework. However, the CD algorithm is sensitive to the parameter initializations, compared with the proposed ADMM-Net which is more stable.

Figures 4 and 5 demonstrate the MSE performance with $N_t = 16$ transmit antennas and different number of activated antennas. The MSE between the intended transmitted signal and the estimated objective is defined as

$$\text{MSE} = \frac{1}{N_t} \|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2. \tag{28}$$

As we expected, the MSE value decreases with increasing SNR (as shown in Figure 4). Because of the ℓ_0-norm constraint, the OMP algorithm achieves higher MSE than that obtained by the ADMM algorithm in the low SNR regime. By learning the hyperparameters, the ADMM-Net algorithm achieves the lowest MSE among the ADMM, CD, and OMP algorithms over all SNR domains. Moreover, Figure 5 shows that when the number of activated antennas increases, the MSE under SNR = 10dB obtained by all algorithms becomes larger, which indicates that the sparse signal is in favor of the AS-SLP design.

Figure 6 presents the performance of average SER of ADMM-Net, ADMM, CD, OMP, and ZF-random algorithms with $N_t = 16$ and $M_t = 4$ and 4-QAM modulation. The proposed ADMM-Net algorithm can achieve the lowest SER among the ADMM, ZF-random, CD, and OMP algorithms in all SNR domains. The improvement of ADMM-Net comes from the learnable hyperparameters trained via back propagation. Moreover, under the same number of available transmit antennas $N_t = 16$, the more transmit antenna is activated, the better performance can be achieved in high SNR regime. Furthermore, under the same number of activated antennas $M_t = 4$, the ADMM framework with $N_t = 32$ transmit antennas can obtain lower SER than that with $N_t = 16$, because of the high degree-of-freedom for antenna selection.

Figure 7 demonstrates the average SER performance versus number of activated transmit antennas under $N_t = 16$ and 4-QAM under SNR = 8dB. It explores the effect of the different number of activated transmit antennas M_t on average SER performance. Generally, the achievable SER is reduced with the increased M_t by most for aforementioned algorithms, except the ZF-random approach. It is because that the probability of selecting the improper activated subset by the ZF-random approach would increase as M_t increases. Moreover, because of the back propagation process, the ADMM-Net scheme can provide the lowest SER among all other algorithms with different settings on N_t and M_t.

6. Conclusion

In this work, we develop the AS-SLP design to reduce the power consumption of the RF chains by jointly minimizing the achievable SER and the number of activated transmit antennas. In the SLP scheme, the optimization problem is formulated as the minimization of the average Euclidean distance between the received symbols and desired symbols

with the constraint on the number of activated transmit antennas. Due to the nonconvex ℓ_0-norm constraint, the underlying optimization problem is further transferred into the regularized ℓ_1-norm problem and solved via ADMM algorithm effectively. By taking the deep learning network as building block, the conventional ADMM process is mapped into the iteratively constructive process which is compose of the reconstruction layer, nonlinear transform layer, and multiplier update layer. Furthermore, considering the effect of parameters initialization on SER performance, the hyperparameters are learned from the ADMM-Net via back propagation. Simulation results demonstrate that the proposed deep ADMM-Net scheme can achieve the considerably low MSE performance and reduce the SER significantly with low power consumption of RF chains.

Data Availability

All experiments of our work were conducted in MATLAB (2017b) and all parameters setting and results of data were presented in the simulation part of the paper, so the data were not additionally provided.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[2] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog-digital transceivers for single-and multi-carrier large antenna array systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1980–1995, 2017.

[3] J. Noh, T. Kim, J.-Y. Seol, and C. Lee, "Zero-forcing based hybrid beamforming for multi-user millimeter wave systems," *IET Communications*, vol. 10, no. 18, pp. 2670–2677, 2016.

[4] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, 2014.

[5] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[6] S. Park, J. Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3818–3832, 2017.

[7] Q. Shi and M. Hong, "Spectral efficiency optimization for millimeter wave multiuser MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 455–468, 2018.

[8] A. A. Nasir, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Hybrid beamforming for multi-user millimeter-wave networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2943–2956, 2020.

[9] H. Tuy, "Convex analysis and global optimization," in *Springer Optimization and Its Applications*, Springer, 2016.

[10] A. Li, D. Spano, J. Krivochiza et al., "A tutorial on interference exploitation via symbol-level precoding: overview, state-of-the-art and future directions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 796–839, 2020.

[11] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1396–1404, 2009.

[12] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 252–262, 2010.

[13] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the MISO downlink channel," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2239–2252, 2015.

[14] M. Alodeh, D. Spano, A. Kalantari et al., "Symbol-level and multicast precoding for multiuser multiantenna downlink: a state-of-the-art, classification, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1733–1757, 2018.

[15] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3628–3640, 2015.

[16] J. W. Lee and C. G. Kang, "Constructive interference optimization for hybrid beamforming with data-aided symbol-level precoding in multi-user MISO system," in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 510–513, Zagreb, Croatia, July 2019.

[17] V.-K. Dinh, M.-T. Le, V.-D. Ngo, and C.-H. Ta, "PCA-aided linear precoding in massive MIMO systems with imperfect CSI," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 3425952, 9 pages, 2020.

[18] A. Li, C. Masouros, B. Vucetic, Y. Li, and A. L. Swindlehurst, "Interference exploitation precoding for multi-level modulations: closed-form solutions," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 291–308, 2021.

[19] A. Li and C. Masouros, "Interference exploitation precoding made practical: optimal closed-form solutions for PSK modulations," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7661–7676, 2018.

[20] C.-E. Chen, "Computationally efficient constructive interference precoding for PSK modulations under per-antenna power constraint," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9206–9211, 2020.

[21] K. L. Law and C. Masouros, "Constructive interference exploitation for downlink beamforming based on noise robustness and outage probability," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3291–3295, Shanghai, China, March 2016.

[22] K. L. Law and C. Masouros, "Symbol error rate minimization precoding for interference exploitation," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5718–5731, 2018.

[23] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding for the nonlinear multiuser MISO downlink channel," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1331–1345, 2017.

[24] M. O. K. Mendonça, P. S. R. Diniz, T. N. Ferreira, and L. Lovisolo, "Antenna selection in massive MIMO based on

greedy algorithms," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1868–1881, 2019.

[25] S. Domouchtsidis, C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding for low complexity transmitter architectures in large-scale antenna array systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 852–863, 2018.

[26] A. M. A. Abdo, X. Zhao, R. Zhang et al., "MU-MIMO downlink capacity analysis and optimum code weight vector design for 5G big data massive antenna millimeter wave communication," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7138232, 12 pages, 2018.

[27] F. Sohrabi, Y.-F. Liu, and W. Yu, "One-bit precoding and constellation range design for massive MIMO with QAM signaling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 557–570, 2018.

[28] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: a deep learning approach for image compressive sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 521–538, 2020.

[29] L.-U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 20–24, 2004.