Compressed CSI Feedback With Learned Measurement Matrix for mmWave Massive MIMO

Pengxia Wu, Zichuan Liu and Julian Cheng, Senior Member, IEEE

Abstract—A major challenge to implement compressed sensing method for channel state information (CSI) acquisition lies in the design of a well-performed measurement matrix to subsample sparse channel vectors. The widely adopted randomized measurement matrices drawn from Gaussian or Bernoulli distribution may not be optimal. To tackle this problem, we propose a fully data-driven approach to learn a measurement matrix for beamspace channel compression, and this method trains a mathematically interpretable autoencoder constructed according to the iterative solution of sparse recovery. The learned measurement matrix can achieve near perfect CSI recovery with fewer required measurements, thus the feedback overhead can be substantially reduced.

Index Terms—Compressed sensing, deep learning, massive MIMO, measurement matrix, mmWave

I. INTRODUCTION

Compressed sensing (CS) technique provides a promising alternative for channel state information (CSI) acquisition in mmWave massive multiple-input multiple-output (MIMO) systems [1], [2]. The main idea of these CS based channel acquisition approaches [3]–[6] is to exploit the beamspace sparsity and to formulate the channel estimation problem into a sparse recovery task. It is well known that the measurement matrix plays an essential role in sparse recovery [4], [7]. However, most of existing works use random matrices as measurement matrices due to simplicity.

Unfortunately, the widely adopted randomized measurement matrices drawn from Gaussian or Bernoulli distribution may not be optimal for all scenarios with time-varying channel conditions. Although it has been proved that several random measurement matrices can achieve perfect recovery with high probability when the dimension of compressed measurements is sufficiently large, but random matrices often perform unsatisfactorily in practical applications especially when the dimension of compressed measurements is insufficient [1]. Since the dimension of compressed measurements simplifies the training/feedback overhead in the CS based CSI acquisition approaches, it is meaningful to reduce the number of measurements under the accuracy guarantee of sparse recovery.

Compared with random matrices, the deterministic measurement matrix is more appealing because it requires less measurements [8], but the design of deterministic measurement matrix is complicated and lacks effective guidelines. Moreover, the deterministic measurement matrices designed in an ad hoc manner have high computational costs and can not be applied for different transmission environments. Therefore, our goal is to search for a simple and effective method to generate well-performed measurement matrices.

A good measurement matrix can be constructed by efficiently exploiting the data features [9], [10]. Many real world datasets have embedding features that can be exploited to perform dimension reduction operations. However, whether additional features beyond sparsity exist in mmWave massive MIMO channels is unknown. Fortunately, owing to state-of-the-art deep learning (DL) technology, the hidden data features of mmWave massive MIMO channels can be effectively learned by neural networks. However, because our goal is to construct a measurement matrix which is linear transformation, the conventional black-box DL architecture with fully non-linear operations are unsuitable for our problem.

In this letter, we introduce a novel autoencoder called $l_1$-minimization autoencoder ($l_1$-AE), which is mathematically interpretable because it is built by unfolding an iterative solution of sparse recovery [10]. We regard the iterations as a set of stacked neural networks parameterized with measurement matrix. By backpropagating the reconstruction error through the network, the measurement matrix is optimized based on training dataset. The proposed $l_1$-AE performs a learning linear dimensional-reduction characteristic, which inherits the data-driven property of DL as well as the clear mathematical interpretation of CS.

We train the $l_1$-AE to learn the measurement matrix from beamspace channel samples, and propose a data-driven compressed CSI feedback scheme for mmWave massive MIMO systems which applies the learned measurement matrix into CS method. Numerical experiment results show that, compared with the widely used random matrices, the learned measurement matrix can achieve higher recovery accuracy for smaller size channel vector. According to our effective achievable rate analysis, the proposed $l_1$-AE enhanced CSI feedback scheme can achieve higher achievable rate at lower feedback overhead when compared with the conventional random projection based compressed CSI feedback schemes.

II. BEAMSPACE MMWAVE MASSIVE MIMO CHANNEL

We consider a single-cell downlink mmWave massive MIMO system operating in frequency division duplexing (FDD) mode, where a BS is equipped with $N$ antennas and all user equipments (UEs) are equipped with single antenna. The channel vector for the user $k$ is given by [11]

$$ h_k = \sqrt{\frac{N}{S}} \sum_{i=1}^{S} \phi_k^{(i)} a(\phi_k^{(i)}) $$

(1)

P. Wu and J. Cheng are with the School of Engineering, The University of British Columbia, Kelowna, BC V1X 1V7, Canada (email: pengxia.wu@ubc.ca, julian.cheng@ubc.ca).

Z. Liu is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798 Singapore.
where $i = 1$ is the index for the line-of-sight path, and $2 \leq i \leq S$ is the index for non-line-of-sight paths; $\beta^{(i)}_k$ is the complex path gain; $\phi^{(j)}_k$ denotes the spatial direction of the $j$th path, and $\alpha(\phi^{(j)}_k)$ is the corresponding array steering vector that contains a list of complex spatial sinusoids. The spatial direction $\phi^{(j)}_k$ relates to the physical angle $\theta^{(j)}_k$ by $\phi^{(j)}_k = \frac{\pi}{d} \sin \theta^{(j)}_k$, for $-1/2 \leq \phi^{(j)}_k \leq 1/2$, $-\pi/2 \leq \theta^{(j)}_k \leq \pi/2s$ [11], where $\lambda$ is the wavelength of mmWave, $d = \lambda/2$ is the antenna spacing distance.

For uniform linear array (ULA) with $N$ antennas, the array steering vector is expressed as $\alpha(\phi^{(j)}_k) = \frac{1}{\sqrt{N}} [e^{-j2\pi\phi^{(j)}_k}]$, where $n \in \mathcal{I}(N) \triangleq \{n - (N - 1)/2 : n = 0, 1, ..., N - 1\}$, and where $\mathcal{I}(N)$ is a set of indices symmetrically arranged around zero [11].

The spatial channel vector $h_k$ in [1] can be transformed into the beamspace channel representations $\tilde{h}_k$ with a set of orthogonal basis $\{u(\phi_j)\}$, which is [11]

$$h_k = U_b \tilde{h}_k = \sum_{j=1}^{M} u(\phi_j) \tilde{h}_k(j)$$

where $\tilde{h}_k(j)$ is the $j$th element of the beamspace channel vector $\tilde{h}_k$; $M$ indicates the beamspace dimension; $U_b^{N \times M}$ is the transformation matrix. The columns of $U_b$ indicate $M$ orthogonal beams and form an optimal basis set for spatial space: each of the beamspace channels $\tilde{h}_k(j)$ is mapped onto a corresponding orthogonal beam represented by a column $u(\phi_j)$ of $U_b$. Usually, $U_b$ is constructed as an $N \times N$ DFT matrix, and can be expressed as $U_b = [\alpha(\tilde{\phi}_1), \alpha(\tilde{\phi}_2), ..., \alpha(\tilde{\phi}_N)]^H$ [11], where $\tilde{\phi}_m = \frac{1}{m-N+1}$ for $m = 1, 2, ..., N$ is the spatial directions predefined by antenna array.

Beamspace sparsity due to limited multipath scatters is an important feature for mmWave channel. The small number of scatters, $S$, results in the limited number of spatial directions $\phi^{(j)}_k$ in [1]. As a result, limited number of non-zero coefficients $\{h_k(j), 1 \leq j \leq M\}$ exist in [2], thus the beamspace channel vector $\tilde{h}_k$ is sparse. In practice, due to power leakage, the beamspace channel vector $\tilde{h}_k$ is approximately sparse or compressive due to the fact that only few large-value elements exist while most of the elements are close to zero [12].

III. $l_1$-AE ENHANCED COMPRESSED CSI FEEDBACK

This work focuses on CSI feedback under the assumption that downlink channel estimation has been completed and feedback links are perfect. The task is to feedback the beamspace channel vector $h \in \mathbb{C}^{N \times 1}$ for one UE without loss of generalization. Because the neural network often works with real numbers, we convert the complex channel vector $\tilde{h} \in \mathbb{C}^{N \times 1}$ into the corresponding real form $h \in \mathbb{R}^{2N \times 1}$ by stacking its real part and its complex part.

The compressed feedback measurement vector $y$ of channel vector $h$ is obtained by

$$y = \Phi h$$

where $h \in \mathbb{R}^{2N \times 1}$ is the real form sparse or compressive channel vector; $y \in \mathbb{R}^{m \times 1}$ is the compressed measurement vector where $m$ denotes the dimension of compressed measurements; $\Phi \in \mathbb{R}^{m \times 2N}$ is the measurement matrix where $m < 2N$. In the compressed CSI feedback scheme, the UE sends the channel vector $y$ with much reduced dimension to the BS; the BS reconstructs the channel vector $h$ from the received compressed measurement vector $y$ and the measurement matrix $\Phi$. The required number of feedback parameters is compressed from $2N$ to $m$, and the value of $m$ indicates the feedback overhead. We desire to make $m$ as small as possible while guaranteeing the recovery accuracy. This desired reconstruction performance highly depends on the measurement matrix $\Phi$, which is applied to project the high dimensional vector $h$ into the lower dimensional subspace spanned by its columns.

The better measurement matrix will involve more priors such as the data structural information of beamspace channel vectors. In order to extract the hidden additional features beyond sparsity in beamspace channels, we propose the $l_1$-AE to directly learn through vast beamspace channel vector samples. Then we propose the $l_1$-AE enhanced CSI feedback scheme which can significantly reduce the feedback overhead for downlink CSI acquisition at the BS.

A. $l_1$-AE Neural Network

As shown in Fig. 1 [10], the $l_1$-AE neural network contains a linear encoder and a dedicated multi-layer non-linear decoder, which are jointly trained together to minimize the difference between input $h$ and output $\hat{h}$.

**Compressive sensing linear encoder:** The encoder of $l_1$-AE is simply a matrix-vector multiplication $y = \Phi h$, in which the sparse channel vector $h$ is subsampled by the measurement matrix $\Phi$, and the dimensional-reduced measurement vector $y$ is the output of encoder, which is also the input of the decoder.

**Projection subgradient descent unfolded decoder:** The decoder is designed mainly by unfolding the projection subgradient descent iterations of the $l_1$-minimization optimization for
sparse recovery. Thus, the decoder reconstructs sparse vector $h$ by forward propagation, while it can update the measurement matrix $\Phi$ by back propagation.

The sparse recovery problem can be formulated into the $l_1$-minimization optimization as

$$\min_{h} ||h||_1 \quad \text{s.t.} \quad \Phi h = y$$  \quad (4)$$

where $||h||_1$ represents the $l_1$-norm of vector $h$.

In order to ensure the gradient with respect to $\Phi$ computable to perform back propagation, the projection subgradient update of the $l_1$-minimization optimization in (4) is adopted to build the multi-layer decoder. The projection subgradient method is given by [13]

$$h^{(t+1)} = \mathcal{P}[h^t - \alpha_t \cdot \text{sign}(h^t)]$$  \quad (5)$$

where $t$ indicates the $t$th update; $\alpha_t$ is the step size; $\text{sign}(h^t)$ is the subgradient of $||h^t||_1$; $\mathcal{P}$ indicates the projection onto the convex set $\{h : \Phi h = y\}$. The projection of the vector $x$ on the set $\{h : \Phi h = y\}$ is computed to be

$$\mathcal{P}[x] = x + \Phi^T(y - \Phi x)$$  \quad (6)$$

where $\Phi^T = \Phi^T(\Phi \Phi^T)^{-1}$ is the pseudoinverse of $\Phi$.

According to the projection subgradient descent in (5), we can obtain the $t$th-step update $h^{(t+1)}$ by substituting $x = h^t - \alpha_t \cdot \text{sign}(h^t)$ into (6) and set the step size as $\alpha_t = 2$. In this way, the $t$th ($1 \leq t \leq L$) layer decoder can be expressed as

$$h^{(t+1)} = h^t - \frac{\alpha}{t} (I - \Phi^T \Phi) \text{sign}(h^t).$$  \quad (7)$$

It is worth mentioning that the pseudoinverse of $\Phi^T$ in (6) can be replaced by the simple transpose operation $\Phi^T$ without performance degradation [10], so that the computations of back propagation can be simplified. The first layer of decoder is set to be $h^{(1)} = \Phi^T y$. Additionally, each layer is added by a batch normalization (BN) modual to empirically enhance the neural network performance.

The output layer adopts a ReLU activation function, so the reconstructed channel vector $\hat{h}$ is

$$\hat{h} = \text{ReLU}(h^{(L+1)}) = \max(h^{(L+1)}, 0).$$  \quad (8)$$

Loss function: The loss function is defined as the mean square $l_2$-norm error between $h$ and $\hat{h}$ samples

$$\text{loss} = \frac{1}{n} \sum_{i=1}^{n} ||h - \hat{h}||_2^2$$  \quad (9)$$

where $n$ is the number of training samples.

Computational Complexity: The network complexity of $l_1$-AE concentrates in the the computations of weight matrices $I - \Phi^T \Phi$ from the second layer decoder to the $(L+1)$th layer decoder. So the complexity of $l_1$-AE is about $O(mN^2L)$. Note that for the structured weight matrix $I - \Phi^T \Phi$, the number of independent parameters is only $2mN$. This design of structured weight matrix reduces computation complexity significantly. Because a fully-connected layer requires $2N \times 2N$ independent parameters in the weight matrix, which is much more computational-complex when $N$ is large.

It is worth pointing out that the training of $l_1$-AE is conducted offline, i.e., before the process of CSI compression and feedback. Moreover, the offline training is only required once. Hence the training of $l_1$-AE does not occupy any time or spectrum resource of the communication system.

B. $l_1$-AE Enhanced Compressed CSI Feedback

Once the training of $l_1$-AE is completed, a learned measurement matrix $\Phi^*$ as the optimized weight parameters can be extracted from the trained $l_1$-AE network. Then the learned measurement matrix $\Phi^*$ is applied to perform the compressed CSI feedback scheme. The process of $l_1$-AE enhanced compressed CSI feedback can be described in three steps. First, the training process is performed at the BS, which has enough computation powers and data. The BS shares the learned measurement matrix $\Phi^*$ to its UEs. Then, each UE uses $\Phi^*$ to compress its beamspace channel vectors by the simple multiplication $y^{m \times 1} = \Phi^* \times h^{2N \times 1}$, where $m \ll 2N$. The compressed channel vector $y$ is sent to the BS. Finally, based on the knowledge of measurement matrix $\Phi^*$ and the feedback vector $y$, the sparse channel vector $h$ can be recovered by various sparse recovery algorithms at the BS.

IV. EXPERIMENTS AND RESULTS

A. Experiment and Training Parameters

We consider the massive MIMO system with 256 antennas for the BS and single antenna for the UE. The channel is generated from the channel model [1], and the number of paths is set to be 3. We randomly generate 20,000 channel vector samples and then split them into training, development, and test dataset by the ratio of 0.8/0.1/0.1. Stochastic gradient descent (SGD) method is used to train the $l_1$-AE, and the training parameters are set as follows: learning rate is 0.01; batch size is 128; the maximum number of epochs is 1000. Measurement matrix $\Phi$ is initialized by the truncated normal distribution with standard deviation $\sigma = 1/\sqrt{512}$. The number of decoder layers is 10, i.e. $L = 9$; the step size parameter of decoder is initialized as $\alpha = 1.0$, and the value of $\alpha$ will automatically update to the appropriate value during training.

We preprocess data to adapt with the valid input-output range of neural network by scaling and shifting the nonzero entries to $[0, 1]$ range for all samples. Thus, the original data formation can be easily recovered by taking corresponding inverse processing on the outputs. The training only takes a few minutes. After that, we obtain the learned measurement matrix $\Phi^*$.

B. Analysis of Experimental Results

We set five baselines to compare with the learned measurement matrix $\Phi^*$, which are random Gaussian matrix $G$, random Bernoulli matrix $B$, partial Fourier matrix $F$, random selection matrix $S^R_1$ and random phase shifter matrix $F^R_1$ respectively. We use linear programming to solve sparse recovery. The recovery performance is evaluated over the test dataset.

1 For random selection matrix, entries are 0 or 1 with equal probability [1].

2 For random phase shifter matrix, each entry is in the form of $\epsilon e^{i\xi}$, where $\xi$ is randomly selected from a set of quantized angles [4].
Table I: Exact recovery percentage.

| Matrix       | $m = 20$ | $m = 25$ | $m = 30$ | $m = 35$ | $m = 40$ |
|--------------|----------|----------|----------|----------|----------|
| $\Phi^*$     | 95.90%   | 98.70%   | 99.66%   | 100%     | 100%     |
| F            | 0.90%    | 7.85%    | 89.20%   | 99.80%   | 99.75%   |
| S            | 5.30%    | 30.15%   | 72.70%   | 90.00%   | 98.45%   |
| B            | 5.90%    | 26.80%   | 63.10%   | 87.70%   | 99.10%   |
| G            | 2.15%    | 14.40%   | 58.50%   | 84.75%   | 97.70%   |
| $P^*$        | 0.00%    | 0.00%    | 0.45%    | 1.00%    | 6.85%    |

Fig. 2: Normalized root square error (NRSE).

Fig. 3: Effective achievable rate.

Table I shows the exact recovery percentages over test dataset for different measurement matrices, where one sample is counted as exactly recovered if $\|h - \hat{h}\|_2 \leq 10^{-8}$. When $m = 20$, the learned matrix $\Phi^*$ achieves 95.9% exact recovery, whereas for random matrices the recovery percentages are all less than 5%. When the learned matrix $\Phi^*$ achieves 98.7% recovery percentage at $m = 25$, the highest percentage of random matrix is only 30.15% for random selection matrix S.

Figure 2 compares the normalized root square error (NRSE) of sparse recoveries. The learned matrix $\Phi^*$ achieves lower NRSE for the same dimension of measurements when compared with random matrices. In another words, the learned matrix $\Phi^*$ can achieve the same level of recovery accuracy with much fewer measurements when compared with random matrices. We also find that the partial Fourier matrix has unstable performance. As shown in Fig. 2, the curves of partial Fourier matrix show several irregular turning points such as at $m = 20, 45$.

Larger dimension of compressed measurements will lead to better recovery. However, the more required measurements means more spectrum occupancy and lower spectrum efficiency. In order to analyze the trade-off between the number of measurements $m$ and the recovery accuracy, we define the effective achievable rate as $R_e = R_0(1 - \frac{n_b}{B})P$, where $R_0$ is the maximal achievable rate for one user, $\frac{n_b}{B}$ is the pilot occupation ratio of one transmission block, $B$ is the block length which is set as 200 symbols, $P$ is the probability of successful recoveries. As shown in Fig. 3, the effective achievable rate achieves maximum at $m = 20$ when using the learned matrix $\Phi^*$, while for random matrices $S, B, G$ the maximum effective achievable rates are achieved at $m = 35$ or $m = 40$. Moreover, the maximal effective achievable rate for learned matrix $\Phi^*$ is higher than these with random matrices. The tremendous performance gain obtained by the learned measurement $\Phi^*$ over random matrices reveals that the sparse beamspace channels have structural features which can be efficiently learned by the DL techniques.

V. Conclusion

We proposed a data-driven compressed CSI feedback approach for downlink CSI acquisition of FDD systems. In such a scheme, a fully data-adaptive measurement matrix was constructed by the $l_1$-AE to enhance the CS method. Compared with the conventional CS methods with random projections, the proposed $l_1$-AE can fully exploit the hidden data structures of beamspace channel datasets, hence the channel vectors can be compressed into smaller size at the UE and can still be perfectly recovered at the BS.

References

[1] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, “Compressed sensing for wireless communications: Useful tips and tricks,” IEEE Commun. Surveys Tut., vol. 19, no. 3, pp. 1527–1550, Third Quart. 2017.

[2] Z. Gao, L. Dai, S. Han, C. I. Z. Wang, and L. Hanzo, “Compressive sensing techniques for next-generation wireless communications,” IEEE Wireless Commun., vol. 25, no. 3, pp. 144–153, June 2018.

[3] J. W. Choi, B. Shim, and S. Chang, “Downlink pilot reduction for massive MIMO systems via compressed sensing,” IEEE Commun. Lett., vol. 19, no. 11, pp. 1889–1892, Nov. 2015.

[4] A. Alkhateeb, G. Leus, and R. W. Heath, “Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Apr. 2015, pp. 2909–2913.

[5] X. Gao, L. Dai, S. Han, C. I, and X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 6010–6021, Sept. 2017.

[6] M. E. Eltayeb, T. Y. Al-Naffouri, and H. R. Bahrami, “Compressive sensing for feedback reduction in MIMO broadcast channels,” IEEE Trans. Commun., vol. 62, no. 9, pp. 3209–3222, Sept. 2014.

[7] 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 J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 436–453, Apr. 2016.

[8] M. Lotfi and M. Vidyasagar, “A fast noniterative algorithm for compressive sensing using binary measurement matrices,” IEEE Transactions on Signal Processing, vol. 66, no. 15, pp. 4079–4089, Aug 2018.

[9] S. Arora, M. Khodak, N. Saunshi, and K. Vodrahalli, “A compressed sensing view of unsupervised text embeddings, bag-of-n-grams, and LSTMs,” in Proc. ICLR, 2018.

[10] M. Lotfi and M. Vidyasagar, “A fast noniterative algorithm for compressive sensing using binary measurement matrices,” IEEE Transactions on Signal Processing, vol. 66, no. 15, pp. 4079–4089, Aug 2018.

[11] J. Brady, N. Behdad, and A. M. Sayeed, “Beamspaces MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements,” IEEE Trans. Antennas Propag., vol. 61, no. 7, pp. 3814–3827, July 2013.

[12] S. Boyd, L. Xiao, and A. Mutapcic, “Subgradient methods,” Notes for EE392o Stanford University Autumn, 2003–2004.