# Deep Learning-Assisted Beamforming Design and BER Evaluation in Multi-User Downlink Systems

Junbeom Kim[1], Hoon Lee[2], Seung-Eun Hong[3] and Seok-Hwan Park[1]
[1]Division of Electronic Engineering, Jeonbuk National University, Jeonju 54896, South Korea
[2]Department of Information and Communications Engineering, Pukyong National University, Busan 48513, South Korea
[3]Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, South Korea
Email: junbeom@jbnu.ac.kr, hlee@pknu.ac.kr, iptvguru@etri.re.kr, seokhwan@jbnu.ac.kr

***Abstract*—This paper studies deep learning-based beamforming design schemes for multi-user downlink systems. Two distinct objectives are considered: sum-rate maximization and min-rate maximization. Each of formulations is first tackled by classical majorization-minimization (MM) algorithms that find a locally optimum point iteratively. To reduce computational overheads of the MM algorithms, deep neural networks (DNNs) are introduced which yield optimized beamforming solutions from channel vector inputs. Performance of trained DNNs is evaluated in terms of bit-error rate (BER) measure. Numerical results show that deep learning approaches achieve the BER performance very close to MM algorithms with much reduced complexity. Also, it is desirable to adopt the minimum-rate criterion to achieve low BER performance rather than sum-rate.**

***Index Terms*—Deep learning, beamforming, MM algorithm.**

## I. Introduction

In space-division multiple access (SDMA) techniques, high spectral efficiency can be offered by enabling a base station (BS) to communicate with multiple user equipments (UEs) at the same time/frequency resource [1], [2]. The capacity region of multi-user downlink systems with SDMA was clearly identified in, e.g., [3]. Capacity-achieving scheme is a non-linear dirty-paper coding (DPC), which is hard to be implemented. Practical systems have instead considered linear beamforming schemes [2], [4], in which BS transmits a superposition of linearly beamformed data signals.

There have been many iterative algorithms developed to optimize the beamforming vectors. Some examples are Majorization Minimization (MM) [5], [6], Weighted Minimum Mean Squared Error (WMMSE) [4], [7], and Fractional Programming (FP) [8], [9]. However, all these algorithms require iterative process, and the number of iterations required for convergence typically increases with signal-to-noise ratio (SNR). It was reported in, e.g., [10], [11] that application of deep learning can be a potential solution to achieve a good performance, sufficiently close to those of iterative MM/WMMSE/FP algorithms, with reasonable complexity.

In this work, we provide an overview of state-of-the-art beamforming techniques, and investigate the bit-error rate (BER) performance of various beamforming schemes. In particular, we discuss the problems of sum-rate and minimum-rate maximization under transmit power constraint at the BS. To find a solution, we present a popular numerical algorithm, namely MM algorithm [5], [6], and the deep learning technique also known as *learning-to-optimize* approach [10], [11]. It is observed via numerical results that it is desirable to adopt the minimum-rate criterion to achieve low BER performance rather than sum-rate, and that deep learning approach shows performance very close to MM algorithm.

This work was supported by Basic Science Research Program through the National Research Foundation (NRF) of Korea grants funded by the Ministry of Education [NRF-2019R1A6A1A09031717, 2021R1C1C1006557, 2021R1I1A3054575]. This work was also supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-01659, 5G Open Intelligence-Defined RAN (ID-RAN) Technique based on 5G New Radio).

## II. System Model

Consider a multi-user downlink system where a BS with $N$ antennas serves $K$ single-antenna UEs. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ denote the set of UEs' indices. We model the signal received by UE $k$ as

$$y_k = \mathbf{h}_k^H \mathbf{w}_k s_k + \mathbf{h}_k^H \sum\nolimits_{l \in \mathcal{K}\setminus\{k\}} \mathbf{w}_l s_l + z_k, \qquad (1)$$

where $\mathbf{h}_k \in \mathbb{C}^N$ represents channel vector from the BS to UE $k$, $\mathbf{w}_k \in \mathbb{C}^N$ denotes the beamforming vector for UE $k$, $s_k$ indicates the data signal intended for UE $k$ with $\mathbb{E}[|s_k|^2] = 1$, and $z_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive Gaussian noise at UE $k$. In the right-hand side (RHS) of (1), the first term is the desired signal, and the second term represents the interference signals intended for the other UEs than UE $k$. The signal-to-interference-plus-noise ratio (SINR) of UE $k$, denoted as $\gamma_k(\mathbf{w})$ with $\mathbf{w} = \{\mathbf{w}_k\}_{k\in\mathcal{K}}$, is defined as

$$\gamma_k(\mathbf{w}) = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{l\in\mathcal{K}\setminus\{k\}} |\mathbf{h}_k^H \mathbf{w}_l|^2 + \sigma^2}. \qquad (2)$$

Under the assumption of the Gaussian channel codebook and sufficiently large coding blocklength, the BS can communicate with UE $k$ reliably with a data rate of $R_k$ if the following condition is satisfied:

$$R_k \leq f_k(\mathbf{w}) = \log_2(1 + \gamma_k(\mathbf{w})) \qquad (3)$$

Denoting the power budget of BS for RF transmission by $P$, the beamforming vectors $\mathbf{w}$ should satisfy the power constraint

$$\sum\nolimits_{k\in\mathcal{K}} ||\mathbf{w}_k||^2 \leq P. \qquad (4)$$

Throughout the paper, it is assumed that the BS perfectly knows all the channel vectors $\mathbf{h} = \{\mathbf{h}_k\}_{k\in\mathcal{K}}$.

## III. Problem Definition and MM Algorithm

We aim at developing beamforming optimization algorithms for two individual objectives: the sum-rate $\mathcal{R}_{\text{sum}} \triangleq \sum_{k\in\mathcal{K}} R_k$

and the minimum-rate $\mathcal{R}_{\min} \triangleq \min_{k\in\mathcal{K}} R_k$. The corresponding problems are formulated as

$$\begin{aligned} \max_{\mathbf{w},\mathbf{R}}.\ & \mathcal{R}_{\mathsf{X}} \\ \text{s.t.}\ & (3),\ k\in\mathcal{K},\ \text{and (4)}, \end{aligned} \tag{5}$$

where $\mathsf{X} \in \{\text{sum}, \min\}$ and $\mathbf{R} = \{R_k\}_{k\in\mathcal{K}}$. It has been well-known that the problem (5) is non-convex for both objectives due to the constraints (3). There have been various iterative algorithms for (5), e.g., MM [5], [6], WMMSE [4], [7], and FP [8], [9].

As a benchmark approach, we consider the MM algorithm that identifies efficient beamforming solutions. With the change of variables $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$, (5) can be restated as

$$\max_{\mathbf{W},\mathbf{R}}.\ \mathcal{R}_{\mathsf{X}} \tag{6a}$$
$$\text{s.t.}\ R_k \le F_k(\mathbf{W}),\ k\in\mathcal{K}, \tag{6b}$$
$$\sum\nolimits_{k\in\mathcal{K}} \text{tr}(\mathbf{W}_k) \le P, \tag{6c}$$
$$\mathbf{W}_k \succeq \mathbf{0},\ k\in\mathcal{K}, \tag{6d}$$
$$\text{rank}(\mathbf{W}_k) \le 1,\ k\in\mathcal{K}. \tag{6e}$$

where the rank constraint (6e) is included due to the solution structure $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$ and the function $F_k(\mathbf{W})$ in (6b) is defined as

$$\begin{aligned} F_k(\mathbf{W}) &= \log_2\left(\sigma^2 + \sum\nolimits_{l\in\mathcal{K}} \mathbf{h}_k^H\mathbf{W}_l\mathbf{h}_k\right) \\ &\quad - \log_2\left(\sigma^2 + \sum\nolimits_{l\in\mathcal{K}\setminus\{k\}} \mathbf{h}_k^H\mathbf{W}_l\mathbf{h}_k\right). \end{aligned} \tag{7}$$

Removing the rank constraint (6e) leads to a difference-of-convex (DC) formulation. A locally optimum solution of such a rank-relaxed problem can then be addressed by the MM algorithm [5], [6]. It solves a sequence of convex approximation problems obtained by linearizing the DC term $F_k(\mathbf{W})$ in (7). The corresponding MM algorithm for solving (6) is summarized in Algorithm 1 where $z^{(t)}$ is a quantity of $z$ evaluated at the $t$-th iteration. The convex approximation of $F_k(\mathbf{W}^{(t)})$ for a given previous solution $\mathbf{W}^{(t-1)}$, denoted by $\tilde{F}_k(\mathbf{W}^{(t)}, \mathbf{W}^{(t-1)})$, is written by

$$\begin{aligned} \tilde{F}_k\left(\mathbf{W}^{(t)}, \mathbf{W}^{(t-1)}\right) &= \log_2\left(\sigma^2 + \sum\nolimits_{l\in\mathcal{K}} \mathbf{h}_k^H\mathbf{W}_l^{(t)}\mathbf{h}_k\right) \\ - \log_2 \omega_k^{(t-1)} &- \frac{1}{\ln 2}\sum_{l\in\mathcal{K}\setminus\{k\}} \frac{\mathbf{h}_k^H\left(\mathbf{W}_l^{(t)} - \mathbf{W}_l^{(t-1)}\right)\mathbf{h}_k}{\omega_k^{(t-1)}}, \end{aligned}$$

with $\omega_k^{(t-1)} = \sigma^2 + \sum_{l\in\mathcal{K}\setminus\{k\}} \mathbf{h}_k^H\mathbf{W}_l^{(t-1)}\mathbf{h}_k$.

## IV. Deep Learning Assisted Design

We discuss a deep learning-based beamforming optimization methods for problem (5). Unlike [11] confined to the sum-rate maximization task, this paper considers a more general setup including the minimum-rate maximization problem. We construct a DNN accepting the channel vector $\mathbf{h}$ as an input feature. System parameters, i.e., $\{P, \sigma^2\}$, are also treated as additional inputs. The resulting outputs denoted by $\mathbf{p} \triangleq \{p_k\}_{k\in\mathcal{K}} \in \mathbb{R}_+^K$ and $\mathbf{q} \triangleq \{q_k\}_{k\in\mathcal{K}} \in \mathbb{R}_+^K$ act as key feature parameters retrieving efficient beamforming vectors. In particular, we adopt the optimum beamforming structure expressed by [12]

$$\mathbf{w}_k = \sqrt{p_k}\frac{\left(\sigma^2\mathbf{I} + \sum_{l\in\mathcal{K}} q_l\mathbf{h}_l\mathbf{h}_l^H\right)^{-1}\mathbf{h}_k}{||\left(\sigma^2\mathbf{I} + \sum_{l\in\mathcal{K}} q_l\mathbf{h}_l\mathbf{h}_l^H\right)^{-1}\mathbf{h}_k||}. \tag{9}$$

Here, $p_k$ stands for the power allocated to UE $k$, thereby leading to $\sum_{k\in\mathcal{K}} p_k = P$ for satisfying (4). Also, $q_k$ indicates the transmit power of UE $k$ for a dual uplink channel with the identical power constraint $\sum_{k\in\mathcal{K}} q_k = P$. It has been proved that any Pareto-boundary achieving beamforming vectors, e.g., the sum-rate and minimum-rate maximizing points, can be identified by adjusting $\mathbf{p}$ and $\mathbf{q}$.

Fig. 1 illustrates the proposed DNN structure which outputs only the feature variables $\mathbf{p}$ and $\mathbf{q}$. The final beamforming vectors $\mathbf{w}$ are obtained by the recovery process in (9). The input-output relationship of the DNN is expressed as $\{\mathbf{p}, \mathbf{q}\} = \mathcal{F}(\mathbf{h}, P, \sigma^2; \mathbf{\Theta})$, where $\mathbf{\Theta}$ represents learnable parameters of the DNN. For notational simplicity, let $\mathbf{w} = \mathcal{G}(\mathbf{p}, \mathbf{q})$ be a collection of the beam recovery process (9) for $k \in \mathcal{K}$. It is not difficult to see that $R_k = f_k(\mathbf{w})$ holds at the optimum of (5). Thus, the sum-rate and the minimum-rate objectives are rewritten by $\mathcal{R}_{\text{sum}}(\mathbf{\Theta}) = \sum_{k\in\mathcal{K}} f_k(\mathcal{G}(\mathcal{F}(\mathbf{h}, P, \sigma^2; \mathbf{\Theta})))$ and $\mathcal{R}_{\min}(\mathbf{\Theta}) = \min_{k\in\mathcal{K}} f_k(\mathcal{G}(\mathcal{F}(\mathbf{h}, P, \sigma^2; \mathbf{\Theta})))$, respectively. Both objectives are given by functions of the DNN parameter $\mathbf{\Theta}$. The corresponding DNN training task is expressed as

$$\max_{\mathbf{\Theta}}.\ \mathbb{E}_{\mathbf{h},P,\sigma^2}\left[\mathcal{R}_{\mathsf{X}}(\mathbf{\Theta})\right]. \tag{10}$$

Thanks to the optimum beam structure (9), the power constraint (4) can be ignored in (10) without loss of the optimality.

The training problem (10) is addressed by the mini-batch stochastic gradient descent (SGD) algorithms, e.g., the Adam optimizer [13]. Defining $\mathcal{B} \triangleq \{\mathbf{h}, P, \sigma^2\}$ as a mini-batch set containing $|\mathcal{B}|$ independently generated training samples $\{\mathbf{h}, P, \sigma^2\}$, the SGD update rule for DNN parameters $\mathbf{\Theta}$ at the $t$-th epoch is written by

$$\mathbf{\Theta}^{[t]} = \mathbf{\Theta}^{[t-1]} + \eta\mathbb{E}_{\mathcal{B}}\left[\mathcal{R}_{\mathsf{X}}(\mathbf{\Theta})\right], \tag{11}$$

where $\eta$ represents the learning rate, and $\nabla_{\mathbf{\Theta}}$ stands for the gradient with respect to $\mathbf{\Theta}$. The training rule in (11) implies that the proposed approach adopts the unsupervised learning strategy which does not require labels, i.e., the optimal beamforming solution to (5).

---

**Algorithm 1** MM algorithm for tackling the problem (6)

---

**1.** Initialize the matrices $\mathbf{W}^{(0)}$ such that the constraints (6c), (6d), and (6e) are satisfied, and set $t \leftarrow 1$.
**2.** Set $\mathbf{W}^{(t)}$ as a solution of the convex problem

$$\begin{aligned} \max_{\mathbf{W}^{(t)},\mathbf{R}}.\ & \mathcal{R}_{\mathsf{X}} \\ \text{s.t.}\ & R_k \le \tilde{F}_k\left(\mathbf{W}^{(t)}, \mathbf{W}^{(t-1)}\right),\ k\in\mathcal{K}, \\ & \sum\nolimits_{k\in\mathcal{K}} \text{tr}\left(\mathbf{W}_k^{(t)}\right) \le P, \\ & \mathbf{W}_k^{(t)} \succeq \mathbf{0},\ k\in\mathcal{K}. \end{aligned} \tag{8}$$

**3.** If $\sum_{k\in\mathcal{K}} ||\mathbf{W}_k^{(t)} - \mathbf{W}_k^{(t-1)}||_F^2 \le \delta$, stop. Otherwise, set $t \leftarrow t+1$, and go back to Step 2.

---

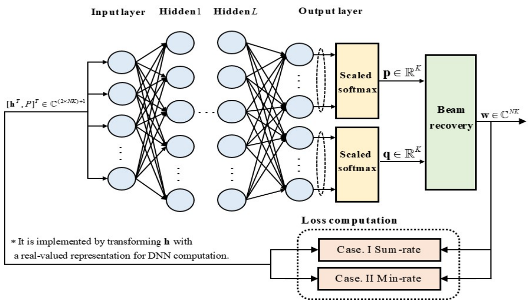


Fig. 1. Deep learning framework for beamforming optimization [11].

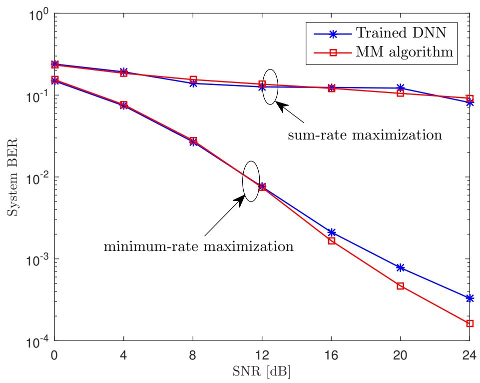


Fig. 2. System BER versus SNR with the BPSK modulation.

## V. Numerical Results

The performance of trained DNNs are evaluated in terms of the average BER. The number of the BS antennas $N$ and UEs $K$ are set to $N = K = 4$. The Rayleigh fading setup is considered. The DNN consists of five hidden layers having 320 output dimension with a rectified linear unit (ReLU). An output layer is designed with a softmax function for generating feasible $\mathbf{p}$ and $\mathbf{q}$. We focus on the uncoded system with the BPSK and QPSK modulations. The "system BER" is defined as the average BER over all UEs.

Figs. 2 depicts the system BER versus the SNR for a multi-user system with $N = K = 4$ and BPSK modulation. We compare the performance of the beamforming schemes optimized with MM algorithm and the DNN with the sum-rate and minimum-rate maximization criteria. Since the overall BER is dominated by that of the worst-channel UE, the minimum-rate objective significantly outperforms the sum-rate maximizing schemes. The DNN schemes achieve BER performance close to those of MM algorithms, validating the effectiveness of the proposed deep learning approach. Also, the DNN approach is a more computationally efficient scheme compared to the MM which requires an iterative optimization process.

Fig. 3 compares the system BER performance of various beamforming schemes obtained for the minimum-rate maximizing problem. The BER of traditional ZF solution [2] is also plotted. Regardless of the modulation schemes, the MM and DNN method achieve notable gains over the ZF baseline. The performance gain is more pronounced for a higher modulation level for which the error performance is more sensitive to interfering signals.

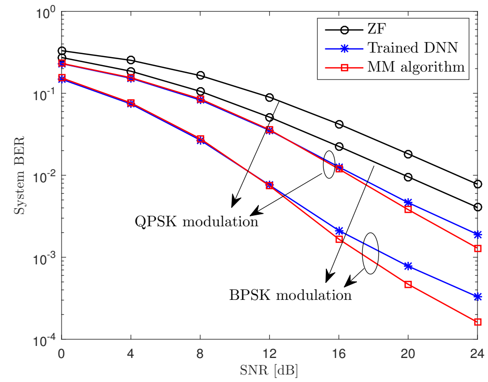


Fig. 3. System BER versus SNR for minimum-rate maximization criteria.

## VI. Conclusion

We have studied DNN-based beamforming design techniques for sum-rate and minimum-rate maximizing tasks. Numerical results have verified the effectiveness of the DNN approaches over classical MM optimization algorithms. We have observed that the minimum-rate maximizing beamforming schemes achieve much lower BERs than those of the sum-rate maximizing schemes.

## References


[1] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.

[2] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003,

[4] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[5] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[6] H. Lee, C. Song, J. Moon, and I. Lee, "Precoder designs for MIMO Gaussian multiple access wiretap channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8563–8568, Sept. 2017.

[7] D. Yu, J. Kim, and S.-H. Park, "An efficient rate-splitting multiple access scheme for the downlink of C-RAN systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1555–1558, Dec. 2019.

[8] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization–maximization approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2164–2177, Oct. 2019,

[9] S.-H. Park, S. Jeong, J. Na, O. Simeone, and S. Shamai (Shitz), "Collaborative cloud and edge mobile computing in C-RAN systems with minimal end-to-end latency," to appear, *IEEE Trans. Signal Inf. Process. Netw.*, 2021.

[10] W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang, and A. P. Petropulu, "A deep learning framework for optimization of MISO downlink beamforming," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1866–1880, Mar. 2020.

[11] J. Kim, H. Lee, S.-E. Hong, and S.-H. Park, "Deep learning methods for universal MISO beamforming," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1894–1898, Nov. 2020.

[12] H. Boche and M. Schubert, "Duality theory for uplink and downlink multiuser beamforming," in *Smart Antennas: State of the Art*, T. Kaiser et al., Ed. New York, NY, USA: Hindawi Publishing Corp., 2005, Ch. 27, pp. 545–576.

[13] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–41.