Weighted Sum-Rate Maximization for Beamforming Design Using Minorization-Maximization: Convergence Rate and Deep Unfolding

Zhexian Yang
SIST, ShanghaiTech
Shanghai, China
zhexianyang@shanghaitech.edu.cn

Zepeng Zhang
IMOS, EPFL
Lausanne, Switzerland
zepeng.zhang@epfl.ch

Ziping Zhao
SIST, ShanghaiTech
Shanghai, China
zipingzhao@shanghaitech.edu.cn

***Abstract*—Weighted sum-rate (WSR) maximization is a fundamental yet generally NP-hard problem in communication system design. Existing optimization-based methods, such as weighted sum-minimum mean-square error (WMMSE), often suffer from high computational overhead in the subproblem, such as matrix inversion. Recent studies [1], [2] have reinterpreted WMMSE within the minorization-maximization (MM) framework, yielding matrix-inversion-free algorithms with improved efficiency. In this paper, we analyze the landscape of the WSR objective and show that all optimal solutions lie on the boundary of the beamforming constraint set. Moreover, we prove that the WSR function is globally L-smooth, which enables us to establish a global sublinear convergence rate for the MM-based algorithm proposed in [1], [2]. To further accelerate computation, we propose a neural network architecture inspired by the MM algorithm, termed MM-Net, which learns to predict a tighter upper-bound constant by exploiting a key matrix in the algorithm. Simulation results demonstrate that MM-Net achieves faster convergence than both WMMSE and MM, while also offering lower complexity compared to state-of-the-art deep learning-based methods.**

***Index Terms*—Weighted sum-rate (WSR) maximization, nonconvex beamforming, minorization-maximization (MM), convergence rate, algorithm unrolling.**

I. INTRODUCTION

Weighted sum-rate (WSR) maximization is a fundamental problem in communication system design, particularly in optimizing beamformers for multi-antenna channels [3]. However, this problem is generally NP-hard, making it challenging to solve optimally [4]. State-of-the-art optimization methods for WSR maximization typically focus on obtaining stationary point solutions. A widely used approach is a block coordinate ascent (BCA) method called weighted sum-minimum mean-square error (WMMSE) [5], [6], which exploits the equivalence between maximizing the signal-to-interference-plus-noise ratio (SINR) and minimizing the mean squared error. Another method employs fractional programming [7], [8], which addresses the fractional structure of SINR in WSR. This approach also follows the BCA principle and shares similarities with WMMSE. A key limitation of WMMSE is that it relies on solving a constrained quadratic program in each iteration. In particular, under a total power constraint, the optimization involves solving a quadratic-constrained quadratic program (QCQP), which requires matrix inversion at the transmit antenna dimension. This computational burden becomes prohibitive for large-scale antenna systems, such as massive MIMO, where efficiency is a critical concern. To address this issue, [1], [2] reinterpreted WMMSE via a constructive analysis through minorization-maximization (MM) framework and proposed a novel MM algorithm that fully leverages the advantages of the MM framework to eliminate repeated matrix inversions. Other numerical optimization approaches have also been explored to mitigate this challenge. A notable example is [9], which introduced an improved WMMSE method by exploiting the beamforming structure under a total power constraint. However, this approach is limited to WSR maximization in single-cell scenarios, whereas the MM algorithm is applicable to more general multi-cell settings.

In [2], the authors established a connection between the proposed MM algorithm and the projected gradient descent (PGD) method, highlighting its potential for analyzing the convergence rate of the MM algorithm. They further pointed out that this connection can facilitate the design of interpretable deep neural networks through the algorithm unfolding technique [10], thereby accelerating convergence. In this paper, we address both of these directions.

Regarding the convergence rate analysis, it is important to note that in [11], a convergence rate result has been established for the MM algorithm in [1], [2]. However, this analysis relies on the assumption that the optimal point is an interior point and that the WSR function is strongly concave within a neighborhood around it. In practice, the validity of these assumptions is questionable. In this paper, by analyzing the landscape of the WSR function, we prove that all stationary points of the WSR maximization problem lie on the boundary of the beamforming constraint set. This result implies that the assumption of an interior stationary point made in [11] is never satisfied. To illustrate this result, we numerically examine a WSR maximization problem in a SISO interference channel, formulated as in (1). We consider a system with two base stations (BSs), each with transmit powers p_1 and p_2, respectively. The channels $h_{i,j}$ are generated from $\mathcal{CN}(0,1)$, for $i, j = 1, 2$, and the noise powers $\sigma_1^2 = \sigma_2^2 = 1$. The MM

$$
\begin{aligned}
\max_{p_1,\, p_2} \quad & \frac{1}{2}\log\left(1+\frac{|h_{1,1}|^2 p_1}{|h_{1,2}|^2 p_2+\sigma_1^2}\right) \\
& +\frac{1}{2}\log\left(1+\frac{|h_{2,2}|^2 p_2}{|h_{2,1}|^2 p_1+\sigma_2^2}\right) \\
\text{s.t.} \quad & 0 \le p_1, p_2 \le 10
\end{aligned} \tag{1}
$$

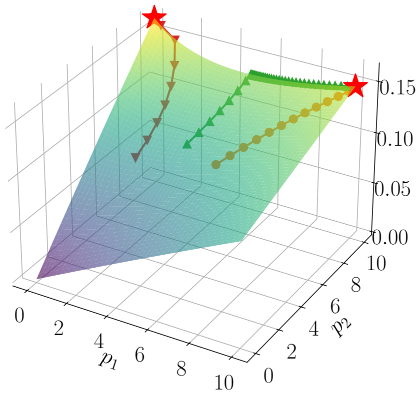


Fig. 1: The WSR landscape.

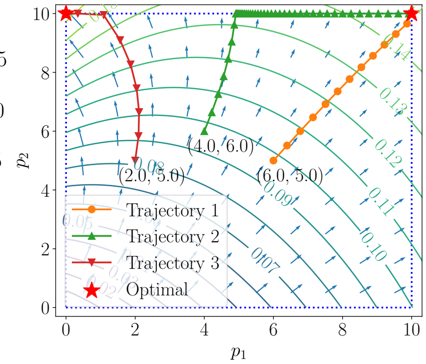


Fig. 2: MM convergence trajectories.

method is applied to solve this problem. Figures 1 and 2 depict the optimization landscape along with convergence trajectories from three different initializations. It can be observed that the points $(0, 10)$ and $(10, 0)$ are two stationary solutions, both located on the boundary of the constraint set. Moreover, at $(0, 10)$, the objective function is convex in the p_1 direction, violating the strong local concavity assumption in [11], where the Hessian is assumed to be negative definite. In addition, we show that the WSR function is globally L-smooth, with the smoothness constant L determined solely by the WSR weights, channel information, and noise power. This result enables us to establish a global sublinear convergence rate for the MM algorithm.

The core idea of the MM algorithm proposed in [1] is to construct surrogate functions for the WSR objective. To generate such a surrogate at each iteration, an upper-bound constant must be computed, which depends on the spectral radius of an intermediate matrix whose dimension is determined by the number of transmit antennas. Although fast approximation methods exist for estimating the spectral radius, they often lead to slower convergence. To address this limitation, we propose an unfolded neural network inspired by the MM algorithm, termed MM-Net, which serves as a data-driven approximation to its numerical counterpart. MM-Net is designed to predict a tighter upper-bound constant by learning the structure of a key matrix in the MM algorithm, thereby accelerating convergence.

Several studies have explored deep learning-based approaches for WSR maximization. In [12], the authors proposed an end-to-end multilayer perceptron (MLP) to approximate the full mapping of the WMMSE algorithm. While the MLP effectively approximates WMMSE in SISO channels, its parameter size becomes prohibitively large in general MIMO interference broadcast channels, rendering training difficult. To alleviate computational burden, [13] proposed an unfolded WMMSE algorithm (IAIDNN), which replaces matrix inversions with simple nonlinear operations during forward propagation. Similarly, [14] introduced a matrix-inversion-free WMMSE method (WMMSE-Net), which approximates the QCQPs in WMMSE via a fixed number of PGD iterations in the MISO setting; this method was later extended to MIMO systems in [15]. In contrast to these approaches, the input dimension of the proposed MM-Net depends only on the number of transmit antennas, making it more scalable and adaptable to transferable system designs. Simulation results show that MM-Net converges faster than both WMMSE and MM algorithms, while achieving lower computational complexity than IAIDNN and WMMSE-Net.

II. WSR Maximization using MM Algorithm

This section provides a review of the MM algorithm for WSR maximization, as developed in [1], [2]. Consider a K-cell interfering broadcast channel with K BSs and I_k users in cell k, where $k = 1, \dots, K$. Let N_k^t and $N_{i_k}^r$ denote the number of antennas at BS k and user i_k, respectively, where $i = 1, \dots, I_k$ and $k = 1, \dots, K$. The beamforming matrix from BS k to user i_k is denoted by $\mathbf{W}_{i_k} \in \mathbb{C}^{N_k^t \times N_{i_k}^s}$, where $N_{i_k}^s$ is the dimension of the transmit signal. The channel between BS k and user i_k is represented by $\mathbf{H}_{i_k,k} \in \mathbb{C}^{N_{i_k}^r \times N_k^t}$. We assume that the data streams have covariance $\mathbf{I}$ and that the noise for user i_k follows a circularly symmetric complex Gaussian distribution with mean $\mathbf{0}$ and covariance $\sigma_{i_k}^2 \mathbf{I}$. The WSR maximization problem can be formulated as follows:

$$
\begin{aligned}
\max_{\{\mathbf{W}_{i_k}\}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k} \alpha_{i_k} \log\det\left(\mathbf{I}+\mathbf{W}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k,k}^{\mathsf{H}}\mathbf{F}_{i_k}^{-1}\mathbf{H}_{i_k,k}\mathbf{W}_{i_k}\right) \\
\text{s.t.} \quad & \{\mathbf{W}_{i_k}\}_{i=1}^{I_k} \in \mathcal{C}_k, \quad \forall\, k = 1, \dots, K,
\end{aligned} \tag{2}
$$

where $\mathbf{F}_{i_k} = \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k,l}\mathbf{W}_{j_l}\mathbf{W}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k,l}^{\mathsf{H}} + \sigma_{i_k}^2\mathbf{I}$, for $i = 1, \dots, I_k$ and $k = 1, \dots, K$, and $\mathcal{C}_k$ is a convex and bounded beamforming constraint set for the k-th BS, which can represent various practical constraints, such as a total power constraint or per-antenna power constraints.

Denote the objective in (2) as $f(\{\mathbf{W}_{i_k}\})$. To apply the MM algorithm, the key step is to construct a lower-bound surrogate function such that the resulting subproblems are computationally efficient to solve. Based on [16, Proposition 7], we obtain the surrogate function $f_1(\{\mathbf{W}_{i_k}\})$ at iterate

$$f_1\left(\{\mathbf{W}_{i_k}\}\right)=\sum_{(i,k)}\left[-\operatorname{tr}\left(\left(\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^t\right)^{\mathsf{H}}\mathbf{A}_k^t\left(\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^t\right)\right)+2\Re\left\{\operatorname{tr}\left(\left(\mathbf{G}_{i_k}^t\right)^{\mathsf{H}}\left(\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^t\right)\right)\right\}\right]+f\left(\{\mathbf{W}_{i_k}^t\}\right) \quad (3)$$

$$f_2\left(\{\mathbf{W}_{i_k}\}\right)=\sum_{(i,k)}\left[-\eta_k^t\left\|\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^t\right\|_F^2+2\Re\left\{\operatorname{tr}\left(\left(\mathbf{G}_{i_k}^t\right)^{\mathsf{H}}\left(\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^t\right)\right)\right\}\right]+f\left(\{\mathbf{W}_{i_k}^t\}\right) \quad (4)$$

Algorithm 1 WSR Maximization via MM

Input: The channel $\{\mathbf{H}_{i_k,k}\}$ and noise power $\sigma_{i_k}^2$.

Initialize $t \leftarrow 0$ and $\left\{\mathbf{W}_{i_k}^0\right\}$.
while not convergence **do**
  Compute η_k^t
  $\mathbf{W}_{i_k}^{t+1} \leftarrow \Pi_{\mathcal{C}_k}\left(\mathbf{W}_{i_k}^t+\frac{1}{\eta_k^t}\mathbf{G}_{i_k}^t\right)$.
  $t \leftarrow t+1$.
end while

Output: Beamforming matrices $\left\{\mathbf{W}_{i_k}^t\right\}$.

$\left\{\mathbf{W}_{i_k}^t\right\}$ in (3), where

$$\mathbf{A}_k^t=\sum_{(j,l)}\alpha_{j_l}\mathbf{H}_{j_l,k}^{\mathsf{H}}\left(\sum_{(n,m)}\mathbf{H}_{j_l,m}\mathbf{W}_{n_m}^t\left(\mathbf{W}_{n_m}^t\right)^{\mathsf{H}}\mathbf{H}_{j_l,m}^{\mathsf{H}}+\sigma_{j_l}^2\mathbf{I}\right)^{-1}\mathbf{H}_{j_l,l}\mathbf{W}_{j_l}^t\left(\mathbf{W}_{j_l}^t\right)^{\mathsf{H}}\mathbf{H}_{j_l,l}^{\mathsf{H}}\left(\mathbf{F}_{j_l}^t\right)^{-1}\mathbf{H}_{j_l,k},$$

and

$$\mathbf{G}_{i_k}^t=\left(\alpha_{i_j}\mathbf{H}_{i_k,k}^{\mathsf{H}}\left(\mathbf{F}_{i_k}^t\right)^{-1}\mathbf{H}_{i_k,k}-\mathbf{A}_k^t\right)\mathbf{W}_{i_k}^t$$

for $i=1,\ldots,I_k$ and $k=1,\ldots,K$. According to [2, Proposition 16], we further get the surrogate function $f_2\left(\{\mathbf{W}_{i_k}\}\right)$ in (4), where $\eta_k^t \geq \lambda_{\max}\left(\mathbf{A}_k^t\right)$. Within the MM framework, at the t-th iteration, we solve the following optimization problem:

$$\begin{aligned}\min_{\{\mathbf{W}_{i_k}\}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}\left\|\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^t-\frac{1}{\eta_k^t}\mathbf{G}_{i_k}^t\right\|_F^2 \\ \text{s.t.} \quad & \{\mathbf{W}_{i_k}\}_{i=1}^{I_k}\in\mathcal{C}_k, \quad \forall\, k=1,\ldots,K,\end{aligned} \quad (5)$$

whose solution is given by

$$\mathbf{W}_{i_k}^{t+1}=\Pi_{\mathcal{C}_k}\left(\mathbf{W}_{i_k}^t+\frac{1}{\eta_k^t}\mathbf{G}_{i_k}^t\right), \quad (6)$$

where $\Pi_{\mathcal{C}_k}(\cdot)$ is the orthogonal projection onto $\mathcal{C}_k$. We summarize the overall MM algorithm in Algorithm 1.

III. Convergence Rate Analysis

We examine the first-order optimality condition:

$$\nabla f\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)-\nabla f_1\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)=\nabla f\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)-\mathbf{G}_{i_k}^t=\mathbf{0}.$$

Thus, the MM algorithm can be viewed as a PGD method with iteration-dependent stepsize $1/\eta_k^t$, i.e.,

$$\mathbf{W}_{i_k}^{t+1}=\Pi_{\mathcal{C}_k}\left(\mathbf{W}_{i_k}^t+\frac{1}{\eta_k^t}\nabla f\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)\right), \quad (7)$$

for $i=1,\ldots,I_k$ and $k=1,\ldots,K$. We further consider the second-order information at $\left\{\mathbf{W}_{i_k}^t\right\}$:

$$\begin{aligned}&\nabla^2 f\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)-\nabla^2 f_1\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)\\&=\nabla^2 f\left(\left\{\mathbf{W}_{i_k}^t\right\}\right)+2\mathbf{A}_{i_k}^t\succeq\mathbf{0},\end{aligned}$$

Therefore, the Hessian of $f(\{\mathbf{W}_{i_k}\})$ can be bounded.

Lemma 1. *The Hessian of the WSR can be bounded as*

$$\nabla^2 f\left(\{\mathbf{W}_{i_k}\}\right)\succeq-2\mathbf{A}\succeq-L\mathbf{I}, \quad \forall\,\{\mathbf{W}_{i_k}\} \quad (8)$$

where constant $L=\max_k\sum_{(j,l)}\frac{2\alpha_{j_l}}{\sigma_{j_l}^2}\left\|\mathbf{H}_{j_l,k}\right\|_2^2$, *matrix* $\mathbf{A}=\operatorname{blkdiag}(\tilde{\mathbf{A}}_{1_1},\ldots,\tilde{\mathbf{A}}_{I_{1_1}},\ldots,\tilde{\mathbf{A}}_{1_K},\ldots,\tilde{\mathbf{A}}_{I_{K}K})$, *and* $\tilde{\mathbf{A}}_{i_k}=\operatorname{blkdiag}(\mathbf{A}_k,\ldots,\mathbf{A}_k)\in\mathbb{C}^{N_k^tN_{i_k}^s\times N_k^tN_{i_k}^s}$.

According to Lemma 1, the WSR objective function is globally L-smooth. While a similar result is presented in [14], our definition of L is independent of the beamforming constraints, thereby guaranteeing global L-smoothness over the whole domain, which constitutes a stronger property. Based on [17], we further have the following theorem.

Theorem 2. *Let* $\left\{\left\{\mathbf{W}_{i_k}^t\right\}\right\}_{t\geq 0}$ *be the beamformer sequence generated by Algorithm 1 for solving the WSR problem* (2). *Then, all limit points of* $\left\{\left\{\mathbf{W}_{i_k}^t\right\}\right\}_{t\geq 0}$ *are stationary points of problem* (2) *and*

$$\min_{\tau=0,\ldots,t}\sum_{(i,k)}\left\|\mathbf{W}_{i_k}^{\tau+1}-\mathbf{W}_{i_k}^{\tau}\right\|_F^2\leq\frac{f^\star-f\left(\left\{\mathbf{W}_{i_k}^0\right\}\right)}{M(t+1)}, \quad (9)$$

where $M=\eta-\frac{L}{2}$ *with* $\eta_k^t\equiv\eta>\frac{L}{2}$, *and* $f^\star$ *is the optimal WSR of problem* (2).

In contrast to the convergence rate analysis in [11], Theorem 2 establishes a rate that does not rely on the assumption that the stationary point is a strict local optimum residing in a neighborhood where the objective function is concave. Next, we establish a key theoretical result, which characterizes the relationship between the gradient of the WSR function and the beamformers.

Lemma 3. *Given any beamformers* $\{\mathbf{W}_{i_k}\}$, *the following relation holds*

$$\sum_{(i,k)}\operatorname{tr}\left(\mathbf{W}_{i_k}^{\mathsf{H}}\mathbf{G}_{i_k}\right)\geq 0,$$

and the equality holds if and only if $\mathbf{H}_{i_k,k}\mathbf{W}_{i_k}=\mathbf{0}$ *for all* $i=1,\ldots,I_k$ *and* $k=1,\ldots,K$.

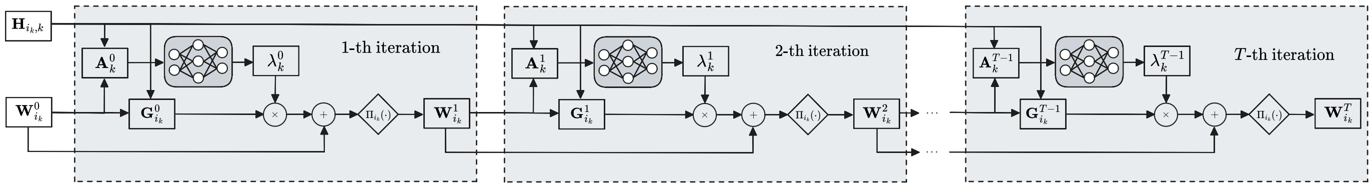


Fig. 3: The architecture of the MM-Net for the k-th cell

Lemma 3 shows that the gradient of the WSR function and the beamforming matrices are aligned in the same direction. Based on this, we establish the following theorem.

Theorem 4. *The gradient of the WSR objective is zero if and only if* $\mathbf{H}_{i_k,k}\mathbf{W}_{i_k} = \mathbf{0}$ *for all* $i = 1, \ldots, I_k$ *and* $k = 1, \ldots, K$.

It is worth noting that Lemma 2 in [9] corresponds to a special case of Theorem 4 when restricted to the single-cell scenario. Leveraging the insights provided by Theorem 4, we derive the following result.

Corollary 5. *The optimal beamformers* $\left\{\mathbf{W}_{i_k}^{\star}\right\}$ *of the problem* (2) *necessarily lie on the boundary of the constraint set* $\mathcal{C}$.

IV. MM-Net: A Deep Learning Adaptation

The MM algorithm requires computing the step size $1/\eta_k^t$, where η_k^t is associated with the maximum eigenvalue of the matrix $\mathbf{A}_k$. However, explicitly computing this eigenvalue can be computationally intensive in massive MIMO systems. To address this issue, [1], [2] propose approximating η_k^t using the Frobenius norm, i.e., $\eta_k^t = \|\mathbf{A}_k\|_F$. This substitution significantly reduces computational complexity, but may also affect the convergence speed of the algorithm.

In this section, we propose a deep learning model based on the deep unfolding technique [10], inspired by the MM algorithm and referred to as MM-Net. Instead of explicitly computing η_k^t, we employ a fully connected neural network $\mathrm{Net}(\cdot)$ to predict the step size. The input to the network is $\mathrm{vec}([\Re\{\mathbf{A}_k\}, \Im\{\mathbf{A}_k\}]) \in \mathbb{R}^{2(N_k^t)^2}$ and output λ_k is a non-negative scalar representing the predicted step size. The hidden layers of the network use the $\mathrm{ReLU}(\cdot)$ activation function, while the output layer adopts the softplus activation ($\mathrm{softplus}(x) = \log(1 + e^x)$), which ensures non-negativity and avoids vanishing gradients. We train the network by minimizing the negative WSR as the loss function:

$$\mathcal{L}(\{\lambda_k\}) = -\sum_{t=1}^{T} w_t f\left(\left\{\Pi_{\mathcal{C}_k}\left(\mathbf{W}_{i_k}^t + \lambda_k \mathbf{G}_{i_k}^t\right)\right\}\right),$$

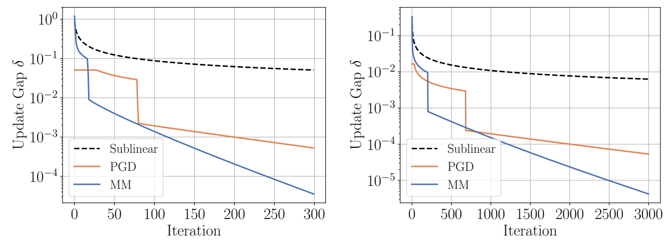
where $w_t \geq 0$ is the weight of the t-th layer. The architecture of the proposed MM-Net is illustrated in Fig. 3.

V. Numerical Results

The numerical experiment was conducted on a desktop equipped with an Intel i5-11500 CPU. The proposed MM-Net was implemented in Python 3.12.8 using PyTorch 2.5.1, with both training and inference performed on the CPU. In the following experiments, we focus on a special case of the broadcast interference channel, namely the single-cell setting with $K = 1$. We adopt a total power constraint, i.e.,

$$\mathcal{C} = \left\{\{\mathbf{W}_i\}_{i=1}^{I} \mid \sum_{i=1}^{I} \|\mathbf{W}_i\|_F^2 \leq P\right\}.$$

The channel coefficients are independently generated from the complex Gaussian distribution $\mathcal{CN}(0, 1)$, and the noise variances are set uniformly as $\sigma_i^2 = \sigma^2$ for all $i = 1, \ldots, I$. The signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = P/\sigma^2$. The beamforming matrices are initialized randomly within the constraint set $\mathcal{C}$. The methods compared in this section include the WMMSE algorithm (WMMSE) [6], the MM algorithm (MM) [2], the iterative algorithm-induced deep unfolding neural network (IAIDNN) [13], the matrix-inverse-free WMMSE-based deep unfolding neural network (WMMSE-Net) [15], and the proposed MM-Net. All reported results are averaged over 1000 independent Monte Carlo simulations.



(a) SNR = 10 dB (b) SNR = 20 dB

Fig. 4: Numerical verification of convergence rate.

A. Numerical verification of convergence rate

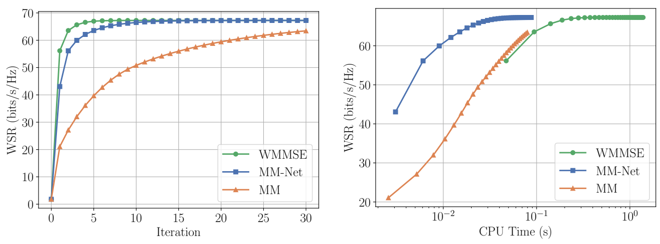
We first empirically evaluate the convergence behavior. We set $I = 2$, $N^t = 4$, and $N^r = N^s = 2$. Fig. 4 illustrates the sublinear convergence of both the MM algorithm and a projected gradient descent (PGD) method with a fixed step size $\lambda = 1/L$. The iteration update gap is defined as

$$\delta = \min_{\tau=0,\ldots,t-1} \sqrt{\sum_{i=1}^{I} \left\|\mathbf{W}_i^{(\tau+1)} - \mathbf{W}_i^{(\tau)}\right\|_F^2}.$$

While the PGD method exhibits significantly slower convergence than MM, it still maintains a sublinear rate.

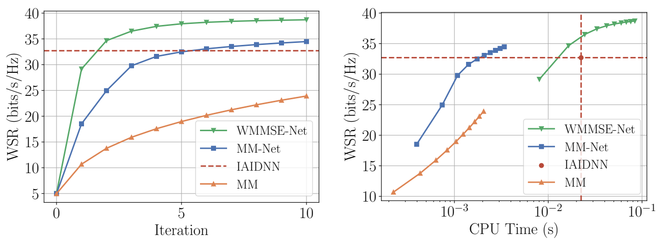
B. Comparison with iterative algorithms

To demonstrate the performance of the unfolding algorithm in a massive MIMO scenario, we set $I = 8$, $N^t = 128$,



(a) Convergence over iterations (b) Convergence over CPU time

Fig. 5: Performance comparison with iterative algorithms.



(a) Convergence over iterations (b) Convergence over CPU time

Fig. 6: Performance comparison with unfolding algorithms.

$N^r = N^s = 4$, and SNR $= 0\,\text{dB}$ in this experiment. The training set consists of 1×10^3 samples with a batch size of 10. The learning rate is set to 1×10^{-3}, and the loss function weight is uniformly defined as $w_t \equiv 1/T$. The number of unfolded layers is chosen as $T = 30$. In Fig. 5(a), the unfolded algorithm MM-Net converges faster than the MM algorithm but slower than the WMMSE algorithm, indicating that MM-Net successfully accelerates the convergence of MM. In Fig. 5(b), MM-Net achieves the same level of WSR with a shorter runtime compared to both MM and WMMSE, demonstrating its computational efficiency. These results highlight the effectiveness of MM-Net in learning upper-bound parameters and improving the performance of WSR maximization.

C. Comparison with unfolding algorithms

We compare several unfolding algorithms under the setting $I = 4$, $N^t = 16$, $N^r = N^s = 4$, and SNR $= 10\,\text{dB}$. All unfolding algorithms are configured with 10 layers. As shown in Fig. 6(a), WMMSE-Net achieves the fastest convergence, followed by MM-Net and then MM. However, Fig. 6(b) shows that, to attain the same WSR level, MM-Net achieves a shorter runtime compared to both IAIDNN and WMMSE-Net. This advantage may be attributed to the fact that WMMSE-Net employs a double-loop structure which, although it accelerates convergence in terms of iterations, introduces higher computational overhead than the single-loop design of MM-Net.

VI. Conclusion

In this paper, we have investigated the WSR maximization problem, a fundamental yet generally NP-hard task in communication system design. We have analyzed the optimization landscape of the WSR objective and have established that all optimal solutions necessarily lie on the boundary of the beamforming constraint set. Furthermore, we have proved that the WSR objective is globally L-smooth, which has enabled the derivation of a global sublinear convergence rate for the MM algorithm proposed in [1], [2]. To enhance computational efficiency, we have proposed a deep unfolded neural network inspired by the MM algorithm, termed MM-Net, which predicts tighter upper-bound constants by learning from key matrices in the MM updates. Simulation results have validated the theoretical convergence rate and demonstrated that MM-Net achieves superior convergence compared to both the state-of-the-art iterative algorithms and deep unfolding methods.

References

[1] Z. Zhang, Z. Zhao, and K. Shen, “Enhancing the Efficiency of WMMSE and FP for Beamforming by Minorization-Maximization,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.
[2] Z. Zhang, Z. Zhao, K. Shen, D. P. Palomar, and W. Yu, “Discerning and Enhancing the Weighted Sum-Rate Maximization Algorithms in Communications,” *Availabel at arXiv:2311.04546*, 2023.
[3] M. Chiang, P. Hande, T. Lan, C. W. Tan *et al.*, “Power Control in Wireless Cellular Networks,” *Foundations and Trends® in Networking*, vol. 2, no. 4, pp. 381–533, 2008.
[4] Z.-Q. Luo and S. Zhang, “Dynamic Spectrum Management: Complexity and Duality,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb. 2008.
[5] 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 Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.
[6] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel,” *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.
[7] A. Zappone, E. Jorswieck *et al.*, “Energy Efficiency in Wireless Networks via Fractional Programming Theory,” *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.
[8] K. Shen and W. Yu, “Fractional Programming for Communication Systems—Part I: Power Control and Beamforming,” *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, May 2018.
[9] X. Zhao, S. Lu, Q. Shi, and Z.-Q. Luo, “Rethinking WMMSE: Can Its Complexity Scale Linearly With the Number of BS Antennas?” *IEEE Transactions on Signal Processing*, vol. 71, pp. 433–446, 2023.
[10] V. Monga, Y. Li, and Y. C. Eldar, “Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing,” *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, Mar. 2021.
[11] K. Shen, Z. Zhao, Y. Chen, Z. Zhang, and H. Victor Cheng, “Accelerating Quadratic Transform and WMMSE,” *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 11, pp. 3110–3124, 2024.
[12] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, “Learning to Optimize: Training Deep Neural Networks for Interference Management,” *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.
[13] Q. Hu, Y. Cai, Q. Shi, K. Xu, G. Yu, and Z. Ding, “Iterative Algorithm Induced Deep-Unfolding Neural Networks: Precoding Design for Multiuser MIMO Systems,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1394–1410, Feb. 2021.
[14] L. Pellaco, M. Bengtsson, and J. Jaldén, “Matrix-Inverse-Free Deep Unfolding of the Weighted MMSE Beamforming Algorithm,” *IEEE Open Journal of the Communications Society*, vol. 3, pp. 65–81, 2022.
[15] L. Pellaco and J. Jaldén, “A Matrix-Inverse-Free Implementation of the MU-MIMO WMMSE Beamforming Algorithm,” *IEEE Transactions on Signal Processing*, vol. 70, pp. 6360–6375, 2022.
[16] Z. Zhang and Z. Zhao, “Rate Maximizations for Reconfigurable Intelligent Surface-Aided Wireless Networks: A Unified Framework via Block Minorization-Maximization,” *Availabel at arXiv:2105.02395*, 2021.
[17] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2017.