面向 MU-MIMO 干扰网络波束成形的深度图展开

Arindam Chowdhury、Gunjan Verma、Ananthram Swami 和 Santiago Segarra

arXiv:2304.00446v1 [eess.SP] 2 Apr 2023

摘要—**我们为多用户多输入多输出单跳无线自组织干扰网络中的波束成形开发了一种高效且近最优的解决方案。受经典加权最小均方误差（WMMSE）方法（该方法用于解决这一问题）和算法展开原理的启发，我们提出了适用于 MU-MIMO 的展开式 WMMSE（UWMMSE）。该方法利用图神经网络（GNN）学习关键加权最小均方误差参数的参数化函数变换，其中无线网络的信道和干扰分量构成底层图。我们使用多个波束成形问题实例，通过针对网络效用指标的梯度下降来训练这些图神经网络。全面的实验分析表明，在性能、泛化能力和鲁棒性方面，UWMMSE 优于经典 WMMSE 和最先进的基于学习的方法。**

索引词—**波束成形、WMMSE、算法展开、复值图神经网络。**

I. 引言

多用户多输入多输出（MU-MIMO） [2], [3]系统在多小区 [6]和自组网 [7] 无线网络场景中，都越来越多地应用于多天线波束成形 [4], [5] 。它们尤其有助于提高频谱效率和有效网络容量，从而满足现代无线系统 [8]对服务质量（QoS）的高要求。对于无线自组织网络（WANET）而言，多天线波束成形尤其具有挑战性，因为其中的收发器可能受到严格的功率限制。例如，特定任务的军事部署可能会使用多台电池寿命有限的手持设备，构成战术无线自组织网络。此外，这些部署可能处于各种拓扑、环境和天气条件下，从而导致衰落效应和路径损耗各不相同。网络中的设备还可能相互受到干扰，这给维持所需的服务质量带来了更多挑战。总的来说，MU-MIMO 无线自组织网络中的波束成形关键任务是管理信道和干扰条件，以获得能够使给定服务质量指标达到较高水平、同时不违反整体功率约束的波束。

波束成形任务可形式化为一个优化问题，以网络效用函数（例如，和速率、平均速率或调和速率）为目标，并在每个收发器处施加资源约束，例如最大可用功率。此类优化问题已被证明是非凸且 NP-hard 的， [6], [7], [9] 因此不存在闭式解。在缺乏生成波束成形问题精确解的方法时，人们提出了多种经典方法，试图获得近似解。对于最常见的和速率最大化问题，过去几十年中普遍采用的方法大致分为以下几类：拉格朗日对偶分解 [10], [11], 逐次凸近似 [12],干扰定价 [13]以及加权最小均方误差（WMMSE）最小化 [14]。

WMMSE 是一种广泛用于多小区和自组织无线网络中波束成形的算法。它提供了闭式迭代更新规则，用于求解一种替代优化问题；已有研究表明，该问题的局部最优解与和速率优化问题的局部最优解相同 [14]。尽管 WMMSE 是一种有效的经典方法，但仍有若干缺点。首先，由于每次迭代都需要执行多次矩阵求逆和特征分解运算，其计算复杂度较高。此外，每当出现新的信道条件时，都必须从头开始运行 WMMSE，导致生成后续波束成形器所需的时间显著增加。最后，WMMSE 提供的解至多是近似最优的，因为它只能求得和速率优化问题的一个局部最优解。

人们也提出了深度学习方法来解决具有挑战性的波束成形问题。更具体地说，这些神经模型以信道状态信息（CSI）张量作为输入——该张量是一种整合了带索引节点对之间信道和干扰度量的结构——并为网络中的所有发射机生成相应的波束成形器。为此，人们采用了多种复杂程度各异的神经网络架构——多层感知机（MLP） [15], [16], 卷积神经网络（CNN） [17],循环神经网络（RNN） [18],，甚至图神经网络（GNN） [19], [20] 。这些方法的一大优势在于前馈计算复杂度较低。此外，这些方法能够并行地为网络中的所有节点生成波束成形器，因此适用于大规模网络。然而，这些方法无法泛化到训练期间未曾见过的信道条件。此外，由于缺少

本研究由美国陆军研究办公室资助，并依据合作协议编号 W911NF-19-2- 0269 开展。本文件中所述观点和结论仅代表作者，并不应被解释为美国陆军研究办公室或美国政府明示或暗示的官方政策。尽管本文含有任何版权声明，美国政府仍有权为政府用途复制和分发重印本。初步结果发表于 [1]。A. Chowdhury 和 S. Segarra 任职于莱斯大学电气与计算机工程系。G. Verma 和 A. Swami 任职于美国陆军 DEVCOM陆军研究实验室。电子邮箱：{ac131, segarra}@rice.edu, { gunjan.verma, anan- thram.swami}.civ@army.mil。

结构化的 WMMSE 更新中编码的领域专属信息，它们往往无法与 WMMSE 相媲美。

为解决这些缺点，研究者提出了一类混合算法 [1], [21]–[24]，将 WMMSE 的经典更新结构与神经模型的快速推理能力相结合。这是通过算法展开 [25]–[28]这一学习范式实现的。算法展开将迭代算法的更新步骤解耦，形成一系列混合层；这些层保留了原有的更新结构，同时引入一个或多个从数据中学习的参数。这种受领域知识启发的学习方法已在多个应用领域广受欢迎且效果显著，包括但不限于非负矩阵分解 [29], 、迭代软阈值 [30], 、语义图像分割 [31], 、盲去模糊 [32], 、杂波抑制 [33], 、粒子滤波 [34],、符号检测 [35],、链路调度 [36], 、能量感知功率分配 [37],以及无线网络中的波束成形 [21]–[23]。这些算法利用各种神经网络层来学习待展开迭代算法的一个或多个参数，或近似算法中的某些计算步骤，从而降低复杂度并加快处理速度。

在本文中，我们提出一种 WMMSE 的展开式求解方案，将两类各具特定优势的神经网络架构结合起来：一种 ReLU 多层感知机，对特定的 WMMSE 变量施加参数化函数变换，其具体方式将在第 III-A 节，特别是式（8）中进一步阐述；另一种是图神经网络，它利用无线网络的底层图结构来学习多层感知机的参数。尽管多层感知机 [15] 和图神经网络 [23] 都是常见的通用架构，但本文提出的具体展开形式完全是新颖的。所提出的方案依赖于多层感知机的通用逼近性质，使展开后的变量能够学习自身的函数变换，从而实现更好的收敛，而无需对其施加任何先验结构。同时，使用图神经网络学习变换参数，使模型能够利用嵌入在无线网络中的连接信息；对于任意网络拓扑，这些信息无法通过通用多层感知机以其他方式提取。

贡献。 本文的主要贡献如下:

i) 我们提出了一种混合算法 UWMMSE，即通过展开 WMMSE，为多用户多输入多输出无线自组织网络中的波束成形提供支持。该算法通过图神经网络学习函数变换的参数，并支持分布式实现。此外，我们通过广泛采用参数共享来降低模型复杂度，从而构建快速且轻量的模型。

ii) 我们从必要条件的角度对所提出的模型进行理论分析，即可学习变换必须满足该条件才能实现有效学习； [参见定理 1]。此外，我们证明了所提出的模型具有置换等变性； [参见命题 1]。

iii）我们通过全面的实验分析，展示所提出方法的经验优越性。首先，我们将其与最先进的连接主义方法进行比较。随后，我们展示所提出的模型对未见网络规模的泛化能力，以及其对分布外输入的鲁棒性。

符号说明。 $[\boldsymbol{\mathcal{X}}]_{ij\ldots}$, $[\mathbf{X}]_{ij}$ 和 $[\mathbf{x}]_i$ 分别表示多维张量 $\boldsymbol{\mathcal{X}}$、矩阵 $\mathbf{X}$ 和向量 $\mathbf{x}$ 的元素。通用下标 : 表示一个完整维度，例如，矩阵 $\mathbf{X}$ 的第 i 行表示为 $[\mathbf{X}]_{i:}$。 $\mathbb{E}(\cdot)$ 表示期望算子，而 $(\cdot)^H$ 表示共轭转置。全零张量和全一张量分别记为 $\mathbf{0}$ 和 $\mathbf{1}$，其维度由上下文明确给出。 $z \times z$ 单位矩阵表示为 $\mathbf{I}_z$。对角矩阵 $\mathrm{diag}(\mathbf{X})$ 存储了 $\mathbf{X}$ 的对角元素。零均值复正态分布记为 $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$。

II. 系统模型与问题表述

我们的系统是一个单跳多用户多输入多输出无线自组织网络，包含 M个不同的发射机-接收机对。每个发射机配备 T 根天线，向一个配备 R 根天线的对应接收机发送独立数据信号。令$\mathbf{V}_i \in \mathbb{C}^{T\times d}$表示发射机 i 用于发送信号$\mathbf{x}_i \in \mathbb{C}^d$的波束成形器，其中 $\mathbb{E}[\mathbf{x}_i\mathbf{x}_i^H] =\mathbf{I}_d$，发送至其指定接收机$r(i)$。假设采用线性信道模型，则接收机 $r(i)$接收到的信号$\mathbf{y}_i \in \mathbb{C}^R$ 形式如下

$$\mathbf{y}_i = \underbrace{\mathbf{H}_{ii}\mathbf{V}_i\mathbf{x}_i}_{desired\ signal} + \underbrace{\sum_{j=1\,|\,j\neq i}^{M} \mathbf{H}_{ij}\mathbf{V}_j\mathbf{x}_j + \mathbf{n}_i}_{interference\ plus\ noise}, \quad \text{for all } i, \tag{1}$$

其中$\mathbf{n}_i \in \mathbb{C}^R$ 表示从 $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_R)$采样得到的独立加性高斯白噪声。这里，$\mathbf{H}_{ii} \in \mathbb{C}^{R\times T}$表示发射机 i与其指定接收机 $r(i)$之间的通信信道，而$\mathbf{H}_{ij} \in \mathbb{C}^{R\times T}$ 对所有 $j \neq i$ 则表示 $r(i)$与其他所有发射机 j之间的干扰。最后，使用接收波束成形器$\mathbf{U}_i \in \mathbb{C}^{R\times d}$在接收机$r(i)$处估计发射信号，从而得到对所有 $i \in \{1, \ldots, M\}$的$\hat{\mathbf{x}}_i = \mathbf{U}_i^H\mathbf{y}_i$。

如果我们定义信道状态信息（CSI）张量 $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{M\times M\times R\times T}$ 使得 $[\boldsymbol{\mathcal{H}}]_{ij::} = \mathbf{H}^{ij}$ ，并定义发射机-波束成形器张量 $\boldsymbol{\mathcal{V}} \in \mathbb{C}^{M\times T\times d}$ 使得 $[\boldsymbol{\mathcal{V}}]_{i::} = \mathbf{V}_i$，那么对于每个用户 i，在假设已知完整 CSI 矩阵的情况下，其可达速率 [14] 为

$$c_i(\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{H}}) = \log_2 \det(\mathbf{I}_R + \mathbf{H}_{ii}\mathbf{V}_i\mathbf{V}_i^H\mathbf{H}_{ii}^H(\sigma^2\mathbf{I}_R + \sum_{j\neq i}\mathbf{H}_{ij}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{ij}^H)^{-1}) \tag{2}$$

我们的目标是确定能够使整个网络的和速率最大化的 $\boldsymbol{\mathcal{V}}$ ，

$$\max_{\boldsymbol{\mathcal{V}}} \sum_{i=1}^{M} \alpha_i c_i(\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{H}}) \tag{3}$$
$$\text{s.t.} \quad \mathrm{Trace}\left(\mathbf{V}_i\mathbf{V}_i^H\right) \leq P_{\max}, \quad \text{for all } i,$$

其中， $\alpha_i \in \mathbb{R}$ 表示收发机对i 的优先级，而每个发射机可用的最大功率均记为 $P_{\max} \in \mathbb{R}$。为简便起见，以下我们重点讨论目标函数中为每个用户赋予相同 $\alpha_i = 1$的情形。

式（3）中的优化问题是非凸且 NP 难的 [7], [38]。解决该问题的一种标准方法是将其重新表述为受约束的加权最小均方误差（WMMSE）优化问题 [14]。具体而言，引入接收机权重张量 $\hat{\boldsymbol{\mathcal{W}}} \in \mathbb{C}^{M\times d\times d}$ 和接收机波束成形器张量 $\boldsymbol{\mathcal{U}} \in \mathbb{C}^{M\times R\times d}$ 后，该问题可定义为

$$\begin{aligned}\min_{\hat{\boldsymbol{\mathcal{W}}},\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{V}}} & \sum_{i=1}^{M}(\mathrm{Trace}(\hat{\mathbf{W}}_i\mathbf{E}_i)-\log\det\hat{\mathbf{W}}_i), \\ \text{s.t.} \quad & \mathrm{Trace}\left(\mathbf{V}_i\mathbf{V}_i^H\right)\le P_{\max}, \quad \text{for all } i,\end{aligned} \tag{4}$$

其中，接收机 $r(i)$ 处的权重矩阵为 $\hat{\mathbf{W}}_i = [\hat{\boldsymbol{\mathcal{W}}}]_{i::} \succeq 0$，而 $\mathbf{E}_i \in \mathbb{C}^{d\times d}$ 是发射信号与接收信号之间的均方误差 [14]。

如 [14, 定理 3] 所示，优化问题 (4) 与 (3) 等价——前者全局最优解 $\{\hat{\boldsymbol{\mathcal{W}}}^*,\boldsymbol{\mathcal{U}}^*,\boldsymbol{\mathcal{V}}^*\}$ 中的变量 $\boldsymbol{\mathcal{V}}^*$ 与后者中的最优发射机波束成形器 $\boldsymbol{\mathcal{V}}^*$ 相同。此外，问题 (4) 是三凸的，即固定任意两个变量后，目标函数关于第三个变量为凸函数。因此，问题 (4) 适合采用基于块坐标下降（BCD）的方法求解。

尽管三凸性质带来了易于处理的闭式解，WMMSE 的性能仍受限于繁琐的迭代过程，其中包含矩阵求逆、特征值分解和二分搜索等计算开销较大的步骤（WMMSE 的每次迭代复杂度为 $\mathcal{O}(M^2)$，其中 M 为用户数） [24]。显然，WMMSE 耗时会随无线网络的规模和复杂度而增加，因此在快速变化信道中尤其低效。此外，虽然 WMMSE 能够获得近似最优解，但对于给定的 $\boldsymbol{\mathcal{H}}$，它只能求解 (3) 的一个实例。若有多个 CSI 张量需要处理，例如 $\{\boldsymbol{\mathcal{H}}_i\}_{i=1}^n$——在由集中式优化代理对多个无线子网络进行优化的场景中——则必须从头开始，分别对每个 n 实例重复运行 WMMSE。

从实际角度来看，我们希望能够快速、高效且可解释地处理一组相互独立的 CSI 张量。我们提出通过基于 GNN的展开算法来实现这一目标。具体而言，我们利用经典WMMSE迭代所提供的近最优解，并借助经过训练的图神经模型提升其表达能力和计算效率。

III. 面向 MU-MIMO 的展开式 WMMSE

由于WMMSE包含计算开销较大的迭代，我们通过截断迭代次数来减少计算量，同时保留更新结构；随后利用数据驱动神经模块补偿减少的迭代次数。

A. 设计展开架构

我们定义一个 K 层的参数化函数 $\Lambda(\cdot;\Theta)$: $\mathbb{C}^{M\times M\times R\times T} \to \mathbb{C}^{M\times T\times d}$，其中 Θ 是一组可训练参数，并且 $\boldsymbol{\mathcal{V}}^{(K)} = \Lambda(\boldsymbol{\mathcal{H}};\Theta)$ 近似于给定 CSI张量 $\boldsymbol{\mathcal{H}}$ 时问题（3）的解。Λ 中的层是混合结构，利用 WMMSE 更新 [14], 并辅以可学习变换 Φ 以加速收敛。更具体地，通过设置初始波束成形器 $[\overset{(0)}{i}]_{i::} = \overset{(0)}{i} = \overset{}{\text{init}}\mathbf{1}_{T\times d}$ $\overset{}{\text{init}} = \sqrt{\frac{P_{\max}}{2Td}}\ 1+\boldsymbol{\mathcal{V}}\mathbf{v}$，其 $\sqrt{-1}$ i $\overset{(0)}{i}\ \overset{(0)}{i}^H = P_{\max}$ —— 中 v ( )，对于所有满足 Trace(VV) 的情况，我们将层 $k = 1,...K$ 定文为：

$$\mathbf{U}_i^{(k)} = \left(\sum_j \mathbf{H}_{ij}\mathbf{V}_j^{(k-1)}{\mathbf{V}_j^{(k-1)}}^H\mathbf{H}_{ij}^H + \sigma^2\mathbf{I}_R\right)^{-1}\mathbf{H}_{ii}\mathbf{V}_i^{(k-1)} \tag{5}$$

$$\hat{\mathbf{W}}_i^{(k)} = \left(\mathbf{I}_d - {\mathbf{U}_i^{(k)}}^H\mathbf{H}_{ii}\mathbf{V}_i^{(k-1)}\right)^{-1} \tag{6}$$

$$\xi^{(k)} = \Psi(\mathbf{S}, [\boldsymbol{\mathcal{U}}^{(k)}, \boldsymbol{\mathcal{V}}^{(k-1)}]; \theta) \tag{7}$$

$$\mathbf{W}_i^{(k)} = \Phi_{\xi_i^{(k)}}(\hat{\mathbf{W}}_i^{(k)}) + \hat{\mathbf{W}}_i^{(k)} \tag{8}$$

$$\bar{\mathbf{V}}_i^{(k)} = \left(\sum_j \mathbf{H}_{ij}^H\mathbf{U}_j^{(k)}\mathbf{W}_j^{(k)}{\mathbf{U}_j^{(k)}}^H\mathbf{H}_{ij} + \mu\mathbf{I}_T\right)^{-1}\mathbf{H}_{ii}^H\mathbf{U}_i^{(k)}\mathbf{W}_i^{(k)} \tag{9}$$

其中，针对每个用户 i，并行计算更新式(5)–(9)。

这里，(7)中的 Ψ 是一种复值图神经网络（CV-GNN）架构，包含一组可训练参数 θ。矩阵 $\mathbf{S} \in \mathbb{C}^{M\times M}$ 由对CSI张量进行可学习变换得到，从而可在CV-GNN Ψ 中使用。我们将在(11)中进一步详细介绍这种张量变换。可训练参数 $\mu \in \mathbb{C}$ 与原始WMMSE公式中功率约束对应的拉格朗日乘子相似。随后，我们的架构中每个混合层的输出表示为 $\bar{\boldsymbol{\mathcal{V}}}^{(k)}$，使得对于所有 i，$[\bar{\boldsymbol{\mathcal{V}}}^{(k)}]_{i::} = \bar{\mathbf{V}}_i^{(k)}$。然而，$\bar{\boldsymbol{\mathcal{V}}}^{(k)}$ 是未经处理的发射机波束成形器，不一定满足功率约束。

为对前馈架构的输出施加功率约束，我们在 $\Lambda(\cdot;\Theta)$ 的混合层中引入非线性激活函数 $\beta(\cdot)$，使模型输出在超出允许值时饱和。对于本文考虑的多天线设置，这意味着在每一层 k 中以相同方式约束 $\boldsymbol{\mathcal{V}}^{(k)}$，使其所有元素 $\mathbf{V}$ $\overset{(k)}{i} = [\boldsymbol{\mathcal{V}}^{(k)}]_{i::}$ 均满足 $\mathrm{Trace}\left(\mathbf{V}_i^{(k)}\ \mathbf{V}_i^{(k)^H}\right) \le P_{\max}$。为此，对所有 i，每一层 k 中的激活函数 β 定义为

$$\begin{aligned}\mathbf{V}_i^{(k)} &= \beta(\bar{\mathbf{V}}_i^{(k)}) \\ &= \begin{cases}\bar{\mathbf{V}}_i^{(k)}, & \text{if } \mathrm{Trace}\left(\bar{\mathbf{V}}_i^{(k)}\bar{\mathbf{V}}_i^{(k)^H}\right) \le P_{\max}, \\ \bar{\mathbf{V}}_i^{(k)}\cdot\frac{\sqrt{P_{\max}}}{\|\bar{\mathbf{V}}_i^{(k)}\|_F}, & \text{otherwise},\end{cases}\end{aligned} \tag{10}$$

其中，$||\cdot||_F$ 表示弗罗贝尼乌斯范数。我们注意到，这种形式的非线性映射曾作为投影步骤用于 [22]提出的基于投影梯度下降的波束成形策略。图1给出了所提出的 UWMMSE 中变量依赖关系的示意图。

需要特别指出的是，如果忽略 (7)，并对每一层 k 都设置 $\Phi_{\xi^{(k)}}(\cdot) = \mathbf{0}_{d\times d}$，那么 (5)-(10) 就归结为经典WMMSE 的 BCD 闭式更新。

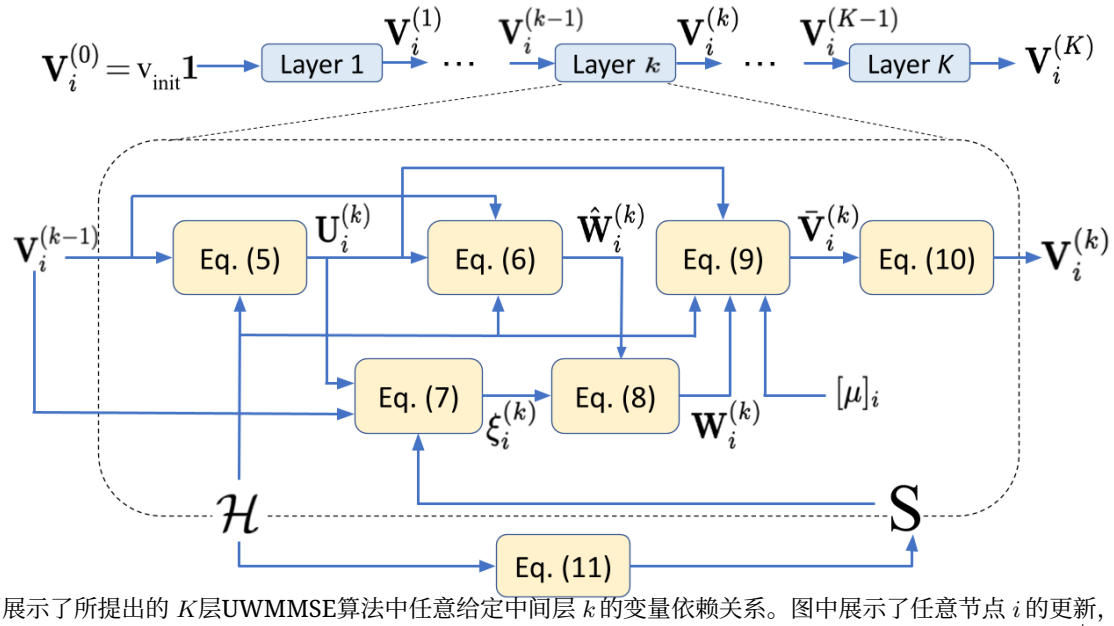


图1. 流程图展示了所提出的 K层UWMMSE算法中任意给定中间层 k 的变量依赖关系。图中展示了任意节点 i 的更新，并对所有 i并行计算。黄色模块表示式(5)-(9)以及(10)和(11)中的五个方程。分层结构的输入为$\mathbf{V}_i^{(0)} = \mathrm{v}_{\mathrm{init}}\mathbf{1}_{T\times d}$，其中$\mathrm{v}_{\mathrm{init}} = \sqrt{P_{\max 2Td}}(1+\sqrt{-1})$；输出的发射机波束成形器为$\mathbf{V}_i^{(K)}$，适用于所有 i。对所有 $j \neq i$而言，其对 σ和 $\{\mathbf{U}j, \mathbf{W}j, \mathbf{V}_j\}$的依赖关系是隐含的。

然而，通过为 UWMMSE 提供额外的灵活性，使其能够为 (7) 中的每个节点学习一组表示 $\xi^{(k)}$（这些表示在 (8) 中通过与该节点对应的 CV-MLP 参数实现），我们能够实现比经典WMMSE 更快的收敛速度和更好的性能，如第四节所示。

构建式(5)-(9)时，首要设计考量之一是选择要学习的WMMSE变量。我们选择保留$\boldsymbol{\mathcal{U}}$ 和$\boldsymbol{\mathcal{V}}$的更新结构，因为这些结构与无线网络的底层通信动态密切相关。事实上，这两个更新方程明确量化了干扰对发射机和接收机的影响。另一方面，$\boldsymbol{\mathcal{W}}$ 代表连接发射机及其对应接收机的信道质量，并在推动$\boldsymbol{\mathcal{V}}$ 和$\boldsymbol{\mathcal{U}}$ 趋近其近似最优值方面发挥关键作用。我们的假设是，若$\boldsymbol{\mathcal{W}}$ 能够通过数据驱动优化加速收敛，就能使$\boldsymbol{\mathcal{U}}$ 和$\boldsymbol{\mathcal{V}}$ 更快收敛，同时不影响无线网络的动态。确定要通过学习增强的变量后，接下来的设计考量是可学习变换 Φ 的结构。为此，我们提出使用具有单隐藏层的复值多层感知器(CV-MLP)，形式如下：$\Phi : \mathbb{C}^{d^2} \to \mathbb{C}^{d^2}$。凭借通用逼近性质(UAP)，多层感知机能够对任意复杂度的连续有界函数进行建模 [39]。因此，所提出的方法无需对变换的结构施加任何额外的归纳偏置，便具备了必要的能力，使$\boldsymbol{\mathcal{W}}$ 能够沿着比$\boldsymbol{\mathcal{W}}$ 单独更新时更优的轨迹演进。

在任意给定的层 k中，对应于定义在节点 i上的 CV-MLP $\Phi_{\xi_i^{(k)}}$ 的参数 $\xi_i^{(k)}$ 由 CV-GNN Ψ学习。除了复值参数 ξ， Φ 还在隐藏层和输出层使用笛卡尔非线性函数 [40]，能够处理复值输出。具体而言，它使用 ReLU 族激活函数 [41]，分别作用于 Φ各层输出的实部和虚部，从而同时变换幅度和相位。本质上，我们在每个展开层 k 中提出了一种分两个层次进行的可学习变换。首先，我们利用 CV-MLP Φ的通用逼近性质，为接收机权重构建一种通用函数变换；其次，变换 $\Phi_{\xi_i^{(k)}}$ 的节点特定参数 $\xi_i^{(k)}$ 则通过 CV-GNN Ψ 学习，并作为所有节点 i的表示。尽管通常直接学习 CV-MLP 的参数，但我们采用上述方法，因为独立的 CV-MLP 无法泛化到无线网络图中的任意连接模式。例如，可学习模型必须能够利用嵌入在 S中的无线网络局部连接结构来生成节点表示。通用图神经网络架构通过一系列聚合与变换操作 [42], [43], ，能够充分捕获这种结构。此外，图神经网络通常具有置换等变性 [42],，因此对于节点排序变化具有更好的泛化性能。不过，具体图神经网络架构的选择和可训练参数的规模并无定式，可根据问题的性质确定。在本例中，为了确保计算简洁性 [44], ，我们选择了一种受图卷积网络启发的复值架构（CV-GCN） [42], 并在 $\Psi(\cdot;\theta)$上使用笛卡尔非线性函数 [40]，如式（12）–（13）所述。

无论采用何种架构 Ψ，我们都将 CSI 张量视为有向图的加权邻接结构，用于聚合来自邻居节点 [45] 的信息。在我们所讨论的多天线场景中，

任意发射机 i 与接收机 $r(j)$之间的链路（无论是信道还是干扰）都由一个 $R \times T$ 矩阵描述，该矩阵取决于发射天线和接收天线的数量。然而，由于 CV-GCN Ψ 要求将 i 与 $r(j)$之间的信道表示为标量系数 [42], ，我们提出采用单层1×1 深度卷积 [46] 操作，并使用共享滤波器参数，将$\boldsymbol{\mathcal{H}}$ 转换为适用的结构。本质上，我们定义了一个额外的全连接神经层 $\Gamma(\boldsymbol{\mathcal{H}};\boldsymbol{\omega})$: $\mathbb{C}^{M\times M\times R\times T} \to \mathbb{C}^{M\times M\times 1}$, $\boldsymbol{\omega} \in \mathbb{C}^{R\times T}$，以生成矩阵 $\mathbf{S}$，使得

$$[\mathbf{S}]_{ij} = \sum_{p,q} \omega_{pq} H_{ijpq} \quad \text{for all } i,j \tag{11}$$

– 这构成了式 (7) 中 [1] 到 $\Psi(\cdot;\theta)$ 的输入。事实上，这一操作可以解释为对每个信道矩阵 $\mathbf{H}_{ij} \in \mathbb{C}^{R\times T}$ 的 RT 个天线系数进行可学习的加权组合，从而为该信道生成标量表示 $S_{ij} = [\mathbf{S}]_{ij}$ 。

除了捕捉图中的局部连接结构外，图神经网络也非常适合处理定义在图节点上的特征或信号 [42], [47]。更具体地，令聚合矩阵为 $\mathbf{S} \in \mathbb{C}^{M\times M}$，并定义特征矩阵 $\mathbf{Q} \in \mathbb{C}^{M\times F'}$，则式 (7) 中的图神经网络具有如下函数形式： $\Psi(\mathbf{S},\mathbf{Q};\theta)$，其中 $\mathbf{Q} = [\boldsymbol{\mathcal{U}}^{(k)},\boldsymbol{\mathcal{V}}^{(k-1)}]$ 由 $\boldsymbol{\mathcal{U}}$ 和 $\boldsymbol{\mathcal{V}}$ 的当前迭代值拼接而成。具体而言，本文其余部分考虑 $d = 1$ 的情形，因此得到 $F' = R + T$。不过请注意，只需对 $\boldsymbol{\mathcal{U}}$ 和 $\boldsymbol{\mathcal{V}}$ 的最后几个维度进行简单的池化变换，便可将模型扩展到 $d > 1$ 的情形。这种表述将 CV-GNN 与 WMMSE 变量的当前状态显式耦合起来。这对于确保 CV-GNN 的输出在各层上依赖于优化轨迹至关重要。此外，在这种情况下，某些服务质量指标（如业务速率、用户优先级、队列长度等）也可作为相关的节点特征。虽然将上述特征或其他节点特征纳入模型十分直接，但对其模型性能影响的详细理论与实验分析超出了本文的讨论范围。

因此，在介绍了聚合器矩阵$\mathbf{S}$ 和特征矩阵$\mathbf{Q}$之后，我们现在给出所提出的模型中$\Psi(\cdot;\theta)$的具体架构。

$$\Psi(\mathbf{S},\mathbf{Q};\theta) = \alpha_2 \left(\operatorname{diag}(\mathbf{S})\,\mathbf{Z}\,\boldsymbol{\theta}_{21}^H + \mathbf{S}\,\mathbf{Z}\,\boldsymbol{\theta}_{22}^H\right), \tag{12}$$

$$\mathbf{Z} = \alpha_1 \left(\operatorname{diag}(\mathbf{S})\,\mathbf{Q}\,\boldsymbol{\theta}_{11}^H + \mathbf{S}\,\mathbf{Q}\,\boldsymbol{\theta}_{12}^H\right), \tag{13}$$

其中， $\theta = \{\boldsymbol{\theta}_{11},\boldsymbol{\theta}_{12},\boldsymbol{\theta}_{21},\boldsymbol{\theta}_{22}\}$，且 α_1和 α_2均为笛卡尔 ReLU 激活函数。请注意，在 $\Psi(\cdot;\theta)$的公式中，我们为对角元素额外设置了一组权重。这主要是为了强调传输信道元素相对于可学习模型中的非对角干扰元素的重要性。

最后，所有节点共享的可训练参数 μ 取代了原始WMMSE公式中的拉格朗日乘子。它的主要作用是纳入节点功率约束。直观而言，式（9）中的 μ 越大，即使干扰分量较小，也会使 $\bar{\mathbf{V}}$满足 Trace($\bar{\mathbf{V}}$ $\bar{\mathbf{V}}^H$)较小。这可以确保在式（10）明确施加功率约束之前，$\bar{\mathbf{V}}$ 就不太可能偏离功率约束太远。这使模型更常在线性区域内运行，即激活函数 β的线性区域，从而改善跨层梯度传播的数值条件。参数 μ 通过式（14）的梯度反馈直接训练。

[1]在实践中，我们发现，在将$\mathbf{S}$ 输入式（7）之前对其进行行归一化操作，可以稳定训练，从而提升整体性能。实现代码请参见脚注 2。

B. UWMMSE 的置换等变性

置换等变性是一种使图神经网络模型的性能在节点身份发生变化时仍保持一致的性质。对于动态无线自组网，这一性质尤为重要，因为节点可能移入或移出网络，甚至在不同的拓扑结构下运行。此外，信道传感器可能以不同顺序传递其估计值，从而导致$\mathbf{S}$的行和列重新排列。在所有这些情况下，图神经网络都必须能够在各种变化下始终保持其预测质量。由于所提出模型 $\Lambda(\boldsymbol{\mathcal{H}};\Theta)$中的基本可学习模块 $\Psi(\mathbf{S},\mathbf{Q};\theta)$（其中 $\Theta = \{\theta,\boldsymbol{\omega},\mu\}$，且$\boldsymbol{\omega}$ 和 μ 均与节点排序无关）具有置换等变性 [42]，因此有必要证明这一关键性质也为整个 UWMMSE架构所继承。

类似于 [24], [48] ，我们现在形式化定义置换等变性。考虑一个通用的置换矩阵 $\boldsymbol{\Pi} \in \{0,1\}^{M\times M}$。此外，令 $\mathcal{F}$ 表示所有函数 $f\colon \mathbb{C}^{M\times M} \to \mathbb{C}^{M}$的集合。

定义 1 如果对于所有置换 $\boldsymbol{\Pi}$ 和所有矩阵$\mathbf{S}$，函数 $f \in \mathcal{F}$ 满足$f(\boldsymbol{\Pi}\mathbf{S}\boldsymbol{\Pi}^\top) = \boldsymbol{\Pi} f(\mathbf{S})$ ，则称该函数为置换等变函数。

本质上，将图的节点索引进行某种置换并输入置换等变函数 f ，会对相应节点输出的索引应用相同的置换，同时不改变其值。接下来，我们将正式说明所提出的UWMMSE模型在式 (7) 中特定 Ψ 条件下的置换等变性。

命题 1 *UWMMSE*模型 $\Lambda(\cdot;\ \Theta)$ 是置换等变的。

Proof: The proof is relegated to Appendix A. ■

C. 训练过程

在说明了针对给定一组可学习参数 $\Theta = \{\theta,\omega,\mu\}$ 时，所提出的 UWMMSE 的内部工作原理后，我们现在将重点转向该架构的训练。给定固定的 Θ，模型 $\Lambda(\boldsymbol{\mathcal{H}};\Theta)$ 会生成与 CSI $\boldsymbol{\mathcal{H}}$ 对应的发射机波束成形器，并据此获得网络和速率效用 $\sum_{i=1}^{M} c_i(\Lambda(\boldsymbol{\mathcal{H}};\Theta),\boldsymbol{\mathcal{H}})$;[, 关于 $\alpha_i = 1$], 参见（3）。因此，损失函数可定义为

$$\ell(\Theta) = -\mathbb{E}_{\boldsymbol{\mathcal{H}}\sim\mathcal{D}}\left[\sum_{i=1}^{M} c_i(\Lambda(\boldsymbol{\mathcal{H}};\Theta),\boldsymbol{\mathcal{H}})\right], \tag{14}$$

其中， $\mathcal{D}$ 是我们关注的信道状态分布。即使已知$\mathcal{D}$，最小化 $\ell(\Theta)$ 关于 $\Theta = \{\theta, \omega, \mu\}$ 的函数仍是一个非凸问题。不过请注意，（7）中的 $\Psi(\cdot;\theta)$ 和（11）中的 $\Gamma(\cdot;\omega)$ 分别关于 θ 和 ω可微。因此，给定从 $\mathcal{D}$中抽取的一组 $\boldsymbol{\mathcal{H}}$，我们采用随机梯度下降来最小化（14）。因此，UWMMSE 本质上是一种无监督学习算法，只需要 CSI $\boldsymbol{\mathcal{H}}$ 的样本，而不 需要这些信道对应的真实最优发射机波束成形器（标签）；生成这些标签的成本可能极其高昂。另一种做法是使用 WMMSE 功率分配作为训练标签，但这实际上会使所提出的 UWMMSE 的学习能力受限于 WMMSE 输出的近似最优性。

注释 1（在 SISO 系统中的应用）虽然我们的方法旨在适用于更一般的 MIMO场景中的波束成形，但也可以无缝用于 SISO无线网络中的功率分配。不过，在 SISO 场景中，无需使用式 (11) 中的神经网络，因为当 $R = T = 1$ 时，CSI 矩阵可直接用作图神经网络中的广义邻接矩阵 Ψ，从而使可训练参数仅减少至 $\Theta = \{\theta, \mu\}$。同样，中间变量 $\mathbf{U}_i^{(k)}$、$\mathbf{W}_i^{(k)}$、$\mathbf{V}_i^{(k)}$ 会简化为标量，但其在式 (5)–(9) 中的更新方程仍然成立，且自然带来更低的计算负载。从这个意义上说，对于 SISO 情形，我们提出的 UWMMSE 类似于 [24] 中介绍的展开方案。然而，在 [24]，中，$\boldsymbol{\mathcal{W}}$ 上的嵌入变换 Φ 在每个混合层中都具有固定仿射结构，不具备 CV-MLP 的通用逼近性质。此外， [24]中的作者仅考虑实值信道实现，这些实现由复值信道的幅度构造而成。因此，他们的方法完全忽略了信道系数中蕴含的相位信息，从而过度简化了问题。此外， [24]中提出的模型的每个层 k 都有各自独立的一组图神经网络 $\Psi(\cdot;\theta^{(k)})$，导致其方案的复杂度随层数增加而上升。更重要的是，这种形式的模型要求训练和推理时的层数固定，因此在部署时缺乏灵活性。就此而言，即使聚焦于 SISO 场景，我们更通用的复值 MIMO UWMMSE 框架相较于现有工作仍具有优势。

D. UWMMSE 收敛性：一个必要条件

我们从理论上确立了 UWMMSE 架构在层数任意多时的行为。

定理 1 考虑一个无限深度的 *UWMMSE*架构(5)-(11)，其中 $\Phi_{\xi_i^{(k)}}(\cdot)$ 对所有 i, k 都是连续函数，并将其用于波束成形以传输信号$\mathbf{x}_i \in \mathbb{C}^1$，即 $d = 1$。考虑极低噪声区域，使得 $\sigma = 0$，并令 $\mu = 0$。用 * 表示问题(4)的最优解。如果$\mathbf{V}_i^{(k)^H}\mathbf{V}_i^{(k)} \to P_i^*$ 满足$0 < P_i^* < P_{\max}$ 对所有 i 均成立，且当 $k \to \infty$ 时一致成立，那么它必须满足

$$\mathbf{A}_i^{-1}\left[\sum_{j\neq i}\mathbf{H}_{ij}^H\mathbf{U}_j^{(k)}(w_j^*\Phi_{\xi_i^{(k)}}(w_i^{(k)}) - w_i^*\Phi_{\xi_j^{(k)}}(w_j^{(k)}))\right.$$
$$\left.\mathbf{U}_j^{(k)^H}\mathbf{H}_{ij}\right]\mathbf{C}_i^{-1}\bar{\mathbf{B}}_i \to \mathbf{0} \ \textit{for all } i \textit{ as } k \to \infty \quad (15)$$

其中 $\mathbf{W} \in \mathbb{C}^{1\times 1}$表示为标量 $w_{ij} \in \mathbb{C}$，$\mathbf{A}_i = \sum_j \mathbf{H}_{ij}^H\mathbf{U}_j^{(k)}w_j^{(k)}\mathbf{U}_j^{(k)^H}\mathbf{H}_{ij}$、$\bar{\mathbf{B}} = \sum_j \mathbf{H}_{ij}^H\mathbf{U}_j^{(k)}w_j^*w_j^{*H}\mathbf{U}_j^{(k)^H}\mathbf{H}_{ij}$、$\mathbf{C}_i = \sum_j \mathbf{H}_{ij}^H\mathbf{U}_j^{(k)}\mathbf{U}_j^{(k)^H}\mathbf{H}_{ij}$，以及 $\mathbf{CHUUH}$。

证明： 证明见附录 B。 ■

定理 1 指出，如果 UWMMSE 在更深层均匀地学习最优发射端波束成形器 $\mathbf{V}_i^*$–其中，$\mathbf{V}_i^{*H}\mathbf{V}_i^* = P_i^*$–，那么学习到的变换 $\Phi_{\xi_i}(\cdot)$ 必须渐近地对所有 i 满足式 (15)。注意，$\Phi_{\xi_i^{(k)}}(w_i^{(k)}) \to 0$，或者更一般地，对于某个常数 δ，$\Phi_{\xi_i^{(k)}}(w_i^{(k)}) \to \delta w_i^{(k)}$，都满足式 (15) 的要求。这一直观上令人满意，因为 $\Phi_{\xi_i^{(k)}}(w_i^{(k)})$ 的这种极限行为意味着 UWMMSE 的更深层将类似于 WMMSE 的经典迭代，而我们知道，最优波束成形器是不动点 [14]。换言之，所提出的可学习模块具有足够的表达能力，能够调整前几层的更新以加速收敛，同时在更深的展开层恢复经典 WMMSE 算法的最优渐近保证。最后请注意，定理 1 与式 (7) 中具体 CV-GNN $\Psi(\cdot;\theta)$ 的选择无关。因此，我们可以有把握地断言，上述结果是式 (5)–(10) 中混合模型的属性，与用于学习参数 ξ的具体 CV-GNN 架构无关。

E. 提高计算效率的参数共享

所有展开层共享 CV-GNN $\Psi(\cdot;\theta)$，即， θ 不依赖式 (7) 中的层索引 k。这种方案确保使用跨层累积、并取决于整体优化轨迹的梯度反馈来训练 θ。此外，这种形式的建模使得在部署时（训练完成后）灵活添加或移除展开层成为可能。另一个直接优势是，相较于依赖层的方案，可训练参数数量减少了 $\mathcal{O}(K)$，从而降低了训练过程的计算开销和耗时。此外，式 (11) 中张量变换 $\Gamma(\cdot;\omega)$ 的可训练参数 ω 对于所有信道元素均相同。这是合理的，因为所有信道表示都必须以相同的函数映射方式，将各自的天线系数映射到输出，这类似于图像像素共享 1×1 卷积。此外，共享滤波器核可使可训练参数数量减少 $\mathcal{O}(M^2)$。最后请注意， μ 也在各层之间共享，从而进一步减少可训练参数的数量。

F. 复杂度分析、可扩展性与分布式实现

WMMSE 每次迭代的计算复杂度为 [14] 是$\mathcal{O}($ $M^2TR^2 + M^2RT^2 + M^2T^3 + M^2R^3$)，适用于 M 用户干扰网络，其中

接收天线数为 R，发射天线数为 T。该复杂度可改写为 $\mathcal{O}(M^2[\max\{R,T\}]^3)$。UWMMSE 中的每个展开层都直接沿用这一复杂度，因为它们执行的更新与 WMMSE 相同。此外，每个展开层中训练好的 CV-GCN 的前馈复杂度为 $\mathcal{O}(M^2F)$，其中隐藏层规模为 F[42]。最后，每个训练好的单隐藏层复值多层感知机的前馈复杂度为 $\mathcal{O}(MG)$，其中隐藏层规模为 G。因此，训练好的 UWMMSE 的总复杂度为 $\mathcal{O}(M^2[[\max\{R,T\}]^3+F]+MG)$。显然，对于固定的天线规模集合 $\{R,T\}$ 以及预先设计的隐藏层维度 F 和 G，UWMMSE 的每层复杂度随网络规模 M变化为 $\mathcal{O}(M^2)$。这与天线规模固定时 WMMSE 每次迭代的复杂度相同。因此，与 WMMSE 的迭代次数相比，通过截断 UWMMSE 的展开层数 K，可以显著缩短推理时间。第四节 C 将通过实验验证这一点。

除计算复杂度外，还需注意由可训练参数数量所决定的模型规模。所提架构的可训练参数非常少，因此易于训练，并且很可能具有良好的泛化能力，如第 IV 节所示。两层复值图卷积网络中的参数数量 θ为 Ψ $\mathcal{O}(F[F'+G])$，其中 F' 为输入规模。此外，线性层 Γ 包含$\mathcal{O}RT$个可训练参数，而 μ 的规模仅为 1。因此，UWMMSE 的可训练参数总数为$\mathcal{O}RT+F[F'+G]$，且与用户数量 M无关。这样，在底层信道模型相同的假设下，同一模型便可用于处理规模不同的无线网络。

虽然在集中式训练器能够访问完整 CSI张量的假设下，有必要以集中式方式训练所提 UWMMSE，但训练完成的 UWMMSE 可以支持分布式部署，使每个节点仅使用本地信息。这主要得益于以下事实：给定节点的功率分配 i 仅取决于 CSI张量的行切片$\boldsymbol{\mathcal{H}}^{r(i)\cdots}$ 和列切片$\boldsymbol{\mathcal{H}}_{:i:}$。然而，要实现完全分布式部署，还需要三个重要假设，其中两个直接继承自 WMMSE 的分布式版本 [14]。首先，对于所有接收机 $r(j)$，其中 $j=1\ldots M$，每个发射机i都应能够获得本地信道状态估计$\mathbf{H}^{ji}$。其次，需要一种机制，以便将信息从接收机反馈给所有发射机。这些假设使发射机能够在所有展开层 k中从每个接收机$r(i)$接收到相应的$\mathbf{U}_i^{(k)}$ 和 $\mathbf{W}_i^{(k)}$ 后，计算$\mathbf{V}_i^{(k)}$。最后，每个节点都必须具备一份完整的已训练参数集合 $\Theta=\{\theta,\omega,\mu\}$。需要注意的是，尽管我们可以通过这种方式实现分布式部署，但反馈链路会增加通信开销，因此推理时间将高于集中式版本。

IV. 数值实验

本节通过全面的数值实验，展示所提出的 UWMMSE模型在各种衰落条件和拓扑结构下，为复值 MU-MIMO无线自组织网络分配功率时的性能。[2] 数据集的详细说明见 IV-A 节，模型架构、超参数和系统设置见 IV-B 节。在 IV-C 节中，我们从实现的总速率和分配时间两个方面，将我们的模型性能与 WMMSE 及其截断版本进行比较。在 IV-D 节和 IV-E 节中，我们评估模型在训练和推理的不同运行条件下的泛化性能。此外，在 IV-F 节中，我们展示了模型收敛行为的示例。最后，在 IV-G 节中，我们研究所提出的模型面对输入信道状态信息张量中范数有界失真时的鲁棒性。

A. 数据集

我们使用随机生成的几何信道实现来评估模型性能。几何信道模型由路径损耗和衰落分量构成。为模拟这一模型，我们构建一个二维几何图，其中包含 M 随机采样的收发机对。所有发射机和接收机均在位置$\mathbf{t}_i\in[0,\sqrt{M}]^2$和$\mathbf{r}_i\in[0,\sqrt{M}]^2$处均匀随机分布。随后，将发射机 i 与接收机 r之间的路径增益计算为其对应物理距离 l_{ij}的逆函数。所有实验中，天线数设为 $R=3$ 和 $T=5$。为简单起见，我们假设传输的是标量复数信号，即 $d=1$。

与 [1], [15], [16], [49], 类似，我们选择以下衰落信道模型：

瑞利：对于与收发机对 ij对应的每个信道矩阵$\mathbf{H}^{ij}$，我们为所有天线对（r, t）独立生成瑞利信道系数$[\mathbf{H}_{ij}]_{rt}$，其各实部和虚部均独立采样自标准正态分布。结合路径损耗分量，信道矩阵$\mathbf{H}$ $[_{ij}]_{rt}$ 的元素表示为

$$[\mathbf{H}_{ij}]_{rt}=\frac{1}{\sqrt{2}(1+l_{ij}^3)}(a+\sqrt{-1}b)\quad\text{for all } r,t.$$
$$\text{where, } a\sim\mathcal{N}(0,1), b\sim\mathcal{N}(0,1)$$

莱斯：对于与收发机对 ij 对应的每个信道矩阵 $\mathbf{H}^{ij}$，我们为每个天线对 (r,t) 独立生成莱斯信道系数$[\mathbf{H}_{ij}]_{rt}$，其莱斯因子为 20 dB *K-factor* [50]；这些系数的实部和虚部分别独立采样自正态分布。考虑路径损耗分量后，信道矩阵 $[\mathbf{H}_{ij}]_{rt}$的各元素表示为

$$[\mathbf{H}_{ij}]_{rt}=\frac{1}{(1+l_{ij}^3)}(a+\sqrt{-1}b)\quad\text{for all } r,t.$$
$$\text{where, } a\sim\mathcal{N}(\mu_{ric},\sigma_{ric}), b\sim\mathcal{N}(\mu_{ric},\sigma_{ric})$$

where $\mu_{ric}=\sqrt{\frac{k}{2(k+1)}}$ and $\sigma_{ric}=\sqrt{\frac{1}{2(k+1)}}$ with $k=100$.

[2]我们已在 https://github.com/ArCho48/Unrolled-WMMSE-for-MU-MIMO 发布了本研究的代码。

B. 模型架构

我们提出的前馈式 UWMMSE 架构由 3 个展开式 WMMSE 层组成，并采用一个由展开层共享的 2 层 CV-GCN 来建模式 (7) 中的函数Ψ 。我们将 CV-GNN 的隐藏层设为 $F = 32$，CV-MLP 的隐藏层设为 $G = 16$。该模型包含 3302 个可训练参数。训练时采用 NovoGrad [51]优化器，最多迭代 15000 次，每批使用 64 个随机采样的信道实现，并启用提前停止。初始学习率设为 1×10^{-2}。一个有趣的观察结果是，虽然 3 个展开层在推理时的和速率与耗时之间取得了最佳性能折衷，但仅使用 1 个展开层时的训练性能最佳。鉴于我们的模型具有灵活性，可利用参数共享在训练和推理时采用不同数量的展开层，我们设计了一种模型设置：先用单层训练 UWMMSE，再在推理时追加 2 层，并让它们共享已学习的组件。本节给出的实验结果证明了这种设置的有效性。在所有实验中，推理时的和速率均取 10000 个信道实现上的平均值。由于模型轻量，它非常适合在中央处理器和图形处理器上运行。不过，为确保一致性和可复现性，本文所有实验结果均在 Nvidia GeForce RTX 2080 GPU 上生成。

C. 性能比较

将所提出的 UWMMSE 的和速率性能与现有基线方法进行比较，包括经典 WMMSE 和最先进的连接主义方法。为了进行这些比较，我们需要根据加性信道噪声，在两种工作区域之间选择一个工作点。首先，在低噪声 区域，信道噪声功率设为不高于 -90dB。这是一个更具挑战性的场景，因为干扰的影响往往占主导地位，而非信道噪声，因此所能达到的和容量很大程度上取决于各发射机的精确波束成形。另一方面，高噪声 区域（噪声功率设为 0 dB）则提供了一个简化的设置：由于噪声较大，可达到的和容量本身较低，因此发射机的精确波束成形并不十分重要。事实上，我们观察到，在此区域中，发射机通常表现出二元特征：要么以全功率发射，要么完全不发射。在实验中，我们选择在低噪声区域运行，具体为 $\sigma = 2.6 \times 10^{-5}$ （-114dB），因为这使我们能够评估所提出模型的全部潜力。

现在，我们列出用于性能比较的方法。这些方法都针对 CV 多用户多输入多输出无线自组网中的波束成形这一共同问题。

1) *WMMSE* [14] 是我们实验中的经典基线，因为我们的方法是对其进行展开扩展，并且有望取得更好的效果。WMMSE 的每个样本最大迭代次数设为 100。2) 截断加权最小均方误差 (Tr-WMMSE)为 UWMMSE 提供了一个经验下界，其性能可达到与展开层数相同的 WMMSE 迭代次数，且不进行任何学习。为与 UWMMSE 的层数匹配，Tr-WMMSE 允许进行 3 次迭代。

3) *IAIDNN* [21]是一种深度展开框架，用于解决多用户多输入多输出系统预编码设计中的和速率最大化问题。

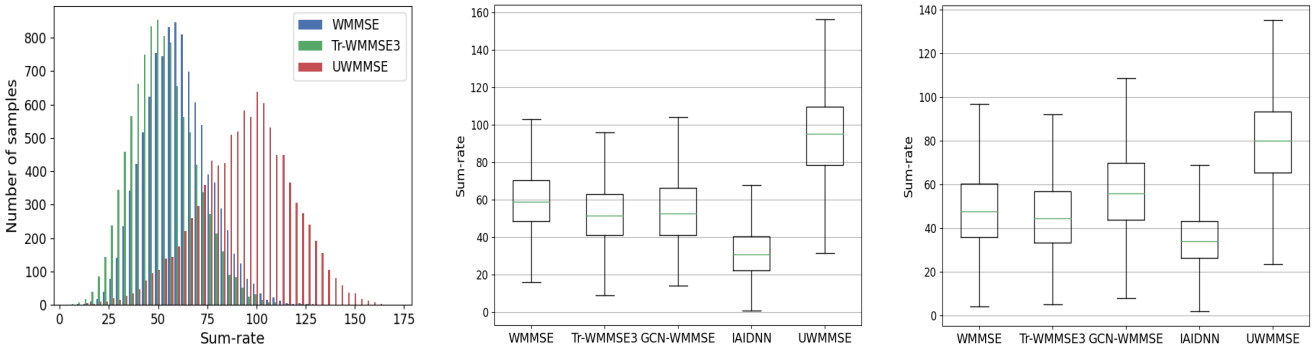
4) *GCN-WMMSE* [23] 是一种基于图的展开框架，利用局部信道状态信息进行多小区 MU-MIMO 干扰信道中的收发器设计。

需要指出的是，IAIDNN [21]和 GCN-WMMSE [23] 都是 WMMSE 的展开扩展[3]，因此属于一类非常特定的混合算法，所提出的 UWMMSE 也属于这一类。UWMMSE 与这些方法的主要区别在于各自展开组件的结构以及相应可学习模块的选择（见第三节）。

比较结果如图2所示。在任一给定时刻，信道条件均从某种衰落分布中随机采样。因此，取决于用于评估不同波束成形算法的具体样本，信道状态信息所决定的和速率效用可能会有显著差异。根据干扰条件及其与信道强度的对比情况，某个特定的信道状态信息实例可能较容易或较难求解。然而，在大量测试样本上的总体表现应能体现某种波束成形算法相较于其他算法的优越性。图2(a)以直方图形式展示了在瑞利信道设置下，完整测试集所达到的和速率。多个信道实现上的和速率值经验分布清楚地表明，对于大多数实现，UWMMSE 的表现显著优于 WMMSE，二者仅有少量重叠。与 Tr-WMMSE 相比，完整 WMMSE 的性能有所提升；考虑到迭代次数不同，这一提升虽在预期之中，但并不显著。直观而言，这或许表明 WMMSE 的迭代过程很快就接近和速率目标函数的某个局部最优解，随后只是在剩余的梯度步中逐渐收敛至该解。另一方面，得益于嵌入其中的可学习组件，UWMMSE 的迭代过程采用数据驱动的修正梯度步，在更少的步数内收敛至更优的局部最优解。

在确立 UWMMSE 相较于 WMMSE 的优越性后，我们现在将研究扩展到瑞利信道设置下的一类混合算法；见图2(b)。IAIDNN [21],使用可学习参数来近似加权最小均方误差更新规则中的矩阵乘法和求逆步骤，其表现显著不及 WMMSE，平均和速率为 32.17。我们推测，IAIDNN 的连接主义组件未能利用无线网络固有的图结构；尽管其当前形式在不含路径损耗分量（本质上缺乏几何结构）的衰落信道以及简化的高噪声环境下表现良好，但它并未

[3]为了测试 IAIDNN 和 GCN-WMMSE，我们使用了*Schynol* 等人在 https://github.com/lsky96/gcnwmmse.git 上提供的实现。



(a) (b) (c)

图2. 所提出的 UWMMSE 与完整 WMMSE [14], 、其截断版本、IAIDNN [21], 和 GCN-WMMSE [23]所达到的和速率对比。(a) 低噪声区域（$\sigma = 2.6 \times 10^{-5}$）下 ~ 10000 个瑞利信道状态信息样本的和速率直方图。UWMMSE 的性能显著优于 WMMSE，同时所需迭代次数与其截断版本相同。(b) 与 (a) 中直方图相对应的箱线图，并在相同的信道状态信息上额外比较了 IAIDNN 和 GCN-WMMSE。(c) 莱斯信道下 (b) 的对应结果。

具备在具有挑战性的低噪声场景中处理完整几何信道模型的能力。较新的 GCN-WMMSE [23], 通过结合图滤波器和基于图卷积网络 [42] 的图学习器来近似某些 WMMSE 变量，从而弥补了这一不足。显然，利用图信息显著提升了其性能（平均和速率为 55.19），相较于 IAIDNN 有了很大提升。尽管如此，在这一特定设置下，它的表现仍略逊于 WMMSE。显然，在所有比较方法中，所提出的 UWMMSE 是更优的算法。这种优势很大程度上归因于两步学习方案：第一步利用几何无线网络中的底层图结构学习一组参数，第二步则基于这些学习到的参数，对一个关键的 WMMSE 变量施加一般函数变换（参见第三节-A），从而比其他方法实现更好的收敛。

现在我们将重点转向莱斯信道设置。选择两种不同信道模型进行比较，旨在证明我们方法的优越性以及各种波束成形算法的一般性能趋势并不依赖于特定的信道类型；见图2(c)。虽然我们观察到类似的性能趋势，即所提出的 UWMMSE 轻松胜过所有其他方法，但在这种情况下，GCN-WMMSE [23] 略微优于 WMMSE，与 U WMMSE 的差距也略有缩小。IAIDNN [21] 仍然落后于所有其他方法，但表现略好于其瑞利信道对应方法。

表 I 各方法的时间比较

算法	训练 时间（分钟）	Test 时间（秒）
WMMSE [14]	-	1.305
Tr-WMMSE	-	0.047
IAIDNN [15]	~ 10	0.64
GCN-WMMSE [21]	~ 21	1.365
UWMMSE	~ 35	0.054

尽管实现较高的和速率是混合波束成形算法的首要目标，但这些方法还必须具备快速推理能力，因为波束成形过程必须能够适应可能快速变化的信道所对应的时间尺度。因此，我们认为有必要比较上述算法在给定任意 CSI 输入时生成波束成形输出的时间复杂度。表 I 给出了计算时间的比较。UWMMSE 的单样本推理时间为 54 ms，比 WMMSE 的推理时间低 24 倍；WMMSE 的推理时间约为 1.30 秒。由于迭代次数相同，UWMMSE 的推理时间与 Tr-WMMSE（每个样本 47 ms）相近，这在预料之中；不过，UWMMSE 的可学习组件会略微增加其时间复杂度。IAIDNN [21] 的推理时间比 UWMMSE 高一个数量级，而 GCN-WMMSE [23] 的推理时间最长（每个样本 1.36 秒）。此外，IAIDNN 和 GCN-WMMSE 在各自的推理时间内，平均性能均不及所提出的 UWMMSE。再者，与经典方法不同，所有混合方法都需要训练，这通常十分耗时 [参见表 I]。显然，IAIDNN 的训练速度最快，而所提出的 UWMMSE 在所有混合方法中训练时间最长。不过，训练只需进行一次，且通常在部署前完成，因此对大多数应用而言，这通常不是主要问题。在瑞利和莱斯衰落情况下，我们观察到类似趋势；但为简洁起见，表 I 中仅展示瑞利衰落的情况。

D. 不同网络规模和衰落类型下的泛化

无线网络在规模、衰落条件、信道噪声功率以及其他诸多方面通常会随时间动态变化。因此，我们关注能够适应多种网络条件的模型。为量化这一特性，我们评估模型在推理阶段面对一组不同于训练阶段的运行条件时的泛化性能。具体而言，我们针对网络规模和衰落信道类型进行评估，如图 3 所示。选择网络规模作为动态变化的因素之一，是因为无线网络很常见的一种情况是

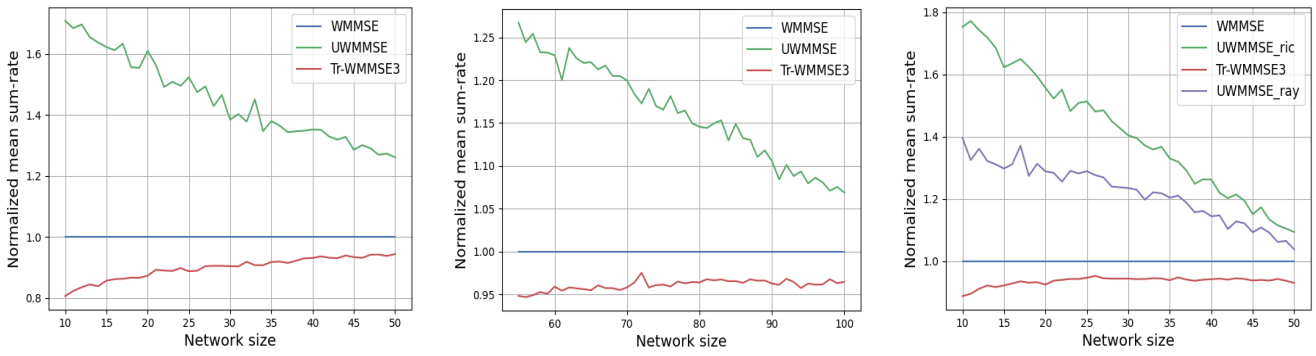


图 3. UWMMSE 的泛化性能——相对于对应的 WMMSE 性能归一化——在不同网络规模和衰落模型下的表现。(a) UWMMSE 在瑞利信道实现上的归一化平均和速率，测试规模范围为 $\{10, 11, \ldots, 50\}$，训练规模范围为 $\{10, 12, 14, \ldots, 50\}$。同时展示了 Tr-WMMSE 在相同网络规模范围内实现的归一化平均和速率，以供比较。(b) 与 (a) 对应，但测试规模范围为 $\{55, \ldots, 100\}$，训练规模仍为 $\{10, 12, 14, \ldots, 50\}$。(c) 与 (a) 对应，但采用莱斯信道实现；此外，还绘制了在瑞利信道实现上训练、在莱斯信道实现上测试的 UWMMSE 模型所实现的归一化和速率。

自组网中会新增节点或移除现有节点。选择衰落信道类型作为第二种动态设置，是因为自组网可以根据不同的地理环境（农村、城市、郊区）和气候条件进行组网和重新组网，因此发射机与相应接收机之间可能存在，也可能不存在占主导地位的视距路径。

首先，为了评估 UWMMSE 在不同网络规模下的插值行为，我们在 $\{10, \ldots, 50\}$ 之间的偶数规模上对其进行训练，然后在瑞利信道设置下测试 10 到 50 之间的所有规模。如图 3(a) 所示，对于所有规模（包括训练期间未出现的奇数规模），模型的表现都显著更好，和速率超过相应 WMMSE 的 1.2-倍 。随着网络规模增大，模型性能有所下降，这是意料之中的，因为更大的网络会使每个收发器受到更多干扰，本质上也给波束成形算法带来了更具挑战性的问题。接下来，为了评估 UWMMSE 的外推行为，我们使用上一个实验中以相同方式训练的模型，测试 55 到 100 之间的所有规模。如图 3(b) 所示，对于所有未见过的规模，模型表现尚可，和速率超过相应 WMMSE 的 1.05-倍 。与插值设置类似，在外推设置下，平均性能稳步下降，主要是由于网络规模增大以及与训练设置的差异所致。然而，对于 10 到 100 之间的所有规模，无论训练条件如何，UWMMSE 的性能都严格优于 WMMSE，这一事实证明了我们的模型对网络规模变化具有泛化能力。

图 3(c) 展示了 UWMMSE 对衰落条件变化的泛化表现。一个在瑞利设置下训练的 UWWMSE 模型（UWMMSE r ay）——与插值实验中使用的模型相同——被用于在莱斯设置下进行推理。尽管 UWMMSE ray 未能胜过 UWMMSE ric（该模型专门在莱斯信道设置下训练），但对于所有网络规模，它仍能轻松胜过相应的 WMMSE 性能。显然，所提出的 UWMMSE 能够合理地泛化到不同的衰落信道设置——即使只在固定的衰落条件下训练，也无需针对其他衰落条件重新训练。

E. 不同空间分布下的泛化能力

到目前为止，我们考虑的是一种无线自组织网络（WANET）场景，其中收发器在面积为 M 的正方形区域内均匀随机分布，如 IV-A 节所述。这种分布选择并无特殊依据，只是为了简化问题。根据拓扑条件或特定任务的要求，不同的现实场景可能会呈现其他空间分布。例如，某陆军部队可能以这样的方式部署：更多士兵驻扎在某个特定兴趣点，且其聚集程度随着与该点距离的增加而降低。此类部署更适合用高斯分布建模。重要的是，所提出的 UWMMSE 能够在推理时处理此类 CSI 张量，即使其训练所用的分布不同。为此，我们采用一个在收发器均匀分布的条件下训练的 UWMMSE 模型，并评估其在高斯分布收发器场景下的泛化性能，同时控制标准差参数。如图 4(a) 所示，只有当标准差足够大、使高斯分布在整个正方形区域内足够宽，从而基本模拟均匀分布时，UWMMSE 在高斯测试集上的性能才与在均匀测试集上的性能相当。随着节点间距缩小、干扰增强，性能明显下降。不过，需要注意的是，归一化为相应 WMMSE 和速率后的平均和速率始终高于 1.0。因此，尽管 UWMMSE 在训练和推理时遇到相同分布时性能最佳，但即使面对推理时未见过的分布，其相较于 WMMSE 仍具有良好的泛化能力。

F. 收敛

本节对所提出的 UWMMSE 模型进行经验性收敛性分析，并在相同框架下与 WMMSE 算法进行比较。我们的分析聚焦于变换后的接收机权重 W 的变化，以及其与最终发射机波束成形器输出 V 的相似性。从本质上说，W 表示给定收发机对之间的信道条件。直观而言，信道条件较好的收发器应以较高功率传输，而信道条件较差的收发器则应停止传输以节省功率。因此，W 在推动 V 达到其近似最优值方面发挥着关键作用。此外，W 是 V 的函数，并依赖 V 来衡量信道质量。为强调这种相互依赖关系如何促使 UWMMSE 模型收敛，我们提取所有 10000 个测试样本在所有层 $k \in 1, 2, 3$ 中的所有节点 $i \in 1, \ldots, M$ 的 $\mathrm{W}_i^{(k)}$ 和 $\mathrm{V}_i^{(k)}$。通常，较大的 $\|\mathrm{W}_i\|_F$ 值表示信道较强、适合传输，因此应分配较高功率。相反，较小的值表示信道较差，不宜进行传输。为验证这一假设，我们对所有 i 的 W_i进行阈值处理：将 $\|\mathrm{W}_i\|_F > 1.0$ 的所有发射机赋值为标量 1.0，将 $\|\mathrm{W}_i\|_F \leq 1.0$ 的所有发射机赋值为 0.0。此外，我们计算整个网络的实值最终功率分配向量 p，即所有 i 对应的 $\|\mathrm{V}_i\|_F$，并对其进行阈值处理，以得到二元功率分配向量。与二元分类类似，我们将 p 视为真实值，将（经过阈值处理的）w 视为预测值，然后计算二者向量之间的 F1分数以衡量匹配程度。选择这一指标是为了确保假阳性和假阴性都会影响得分。随后，我们对所有测试样本的得分取平均，以考察这两个变量整体上的匹配趋势。同样，对于 WMMSE 算法，也计算 W 与 V 之间的匹配得分^。由于 WMMSE 需要 100 次迭代才能达到最佳性能，我们计算第 1、2 和 3 次迭代的得分，以便与 UWMMSE 直接比较；此外还计算第 50 次和第 100 次迭代的得分，以进行更全面的比较。需要指出的是，在本实验中，我们分别使用 UWMMSE 和 WMMSE 最终的功率分配。这是因为我们已在 IV-C 节中证明，UWMMSE 在和速率方面比 WMMSE 收敛得更好，因此两种算法不太可能达到相同的最终 p。不过，本实验的主要目的是评估与经典WMMSE 相比，所提出的 UWMMSE 达到各自收敛状态的速度。比较结果见图 4(b)。显然，在全部 3 个层中，UWMMSE 所得到的 W 与 V 的匹配度均优于 WMMSE 前 3 次迭代的结果。尽管存在误差棒，我们仍观察到，所提出的 UWMMSE 比 WMMSE 更早学到近似最优的 W。

G. 鲁棒性

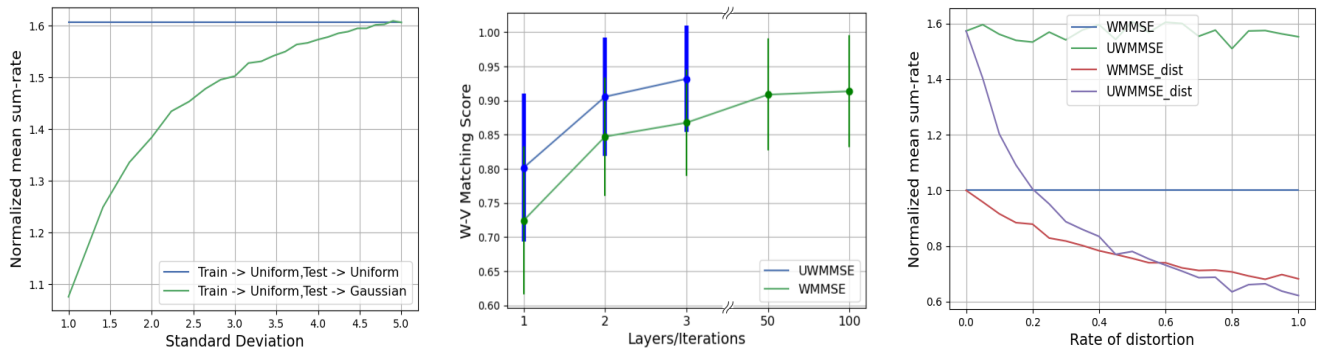
我们分析所提出的 UWMMSE 模型对信道状态信息失真的鲁棒性。进行这类分析至关重要，因为在真实场景中，信道估计过程并不完美 [52]。然而，波束成形算法必须具备足够的鲁棒性，才能在这些变化的影响下维持稳定的性能。为此，我们在 CSI 张量中的信道系数上添加方差有界的随机高斯噪声。不过，和速率是根据未失真的 CSI 张量$\boldsymbol{\mathcal{H}}$计算的。例如，对于¯瑞利信道模型，输入张量$\boldsymbol{\mathcal{H}}$ 中的失真元素 $[\mathbf{H}_{ij}]_{rt}$ 为

$$[\bar{\mathbf{H}}_{ij}]_{rt} = [\mathbf{H}_{ij}]_{rt} + (c + \sqrt{-1}d) \quad \text{for all } r, t.$$
$$\text{where, } c \sim \mathcal{N}(0, \sigma_r), d \sim \mathcal{N}(0, \sigma_r)$$

其中， $\sigma_r = 0.001$。从张量的所有元素中均匀随机地选取发生失真的元素。图 4(c) 显示，在失真率受控且介于 0.0（$\boldsymbol{\mathcal{H}}$ 中没有元素发生失真）与 1.0（$\boldsymbol{\mathcal{H}}$ 中的所有元素均发生失真）之间时，所提出的 UWMMSE 在约 20% 的信道系数发生失真之前，其和速率始终优于使用未失真输入的 WMMSE。尽管性能在超过 20% 后有所下降，但在约 60% 之前仍优于使用失真输入的 WMMSE；超过这一比例后，两种应用于失真 CSI 的算法性能相近。显然，在信道估计失真的情况下，只要仍有约 40% 的系数估计正确，UWMMSE 就具有相当好的鲁棒性，其和速率优于经典方法。尽管这一鲁棒性程度是根据实验观察得出的，但它是模型固有的，因为训练时并未施加任何鲁棒性准则。我们有充分理由认为，对抗训练和噪声正则化训练等方法将进一步增强所提出模型的鲁棒性。然而，这方面的分析超出了本文的范围。

V. 结论

我们提出了 UWMMSE，这是一种混合算法，通过使用复值神经模型展开经典 WMMSE 算法的迭代过程，实现复杂值 MU-MIMO 无线自组织网络中的快速、高效且近乎最优的波束成形。该方法的主要贡献在于，将基于多层感知机的参数化函数变换与适用于处理无线网络图的基于图神经网络的学习器协同结合。大量实验验证了该方法的优越性。此外，所提出的模型得益于广泛的参数共享，因而轻量且易于以分布式方式实现。UWMMSE 每层的计算复杂度与 WMMSE 每次迭代的计算复杂度相当。不过，主要优势在于，与 WMMSE 收敛所需的迭代次数相比，UWMMSE 所需的层数更少。未来的工作将考虑时变信道，以及波束成形选择中的长期约束，例如公平性或电池约束。在真实世界数据集上评估模型性能也是下一步的重要工作。



(a) (b) (c)

图 4. 不同空间分布下的性能泛化、收敛性和鲁棒性结果。(a) 归一化和速率效用的变化：UWMMSE模型在均匀采样的收发器位置上训练，并在从标准差各异的高斯分布中采样的收发器位置上测试。(b) 所提出的 UWMMSE 在第 1、2、3 层中变量 $\boldsymbol{\mathcal{W}}$ 和 $\boldsymbol{\mathcal{V}}$ 的匹配分数，与经典 WMMSE 在第 1、2、3、50 和 100 次迭代中的匹配分数。(c) 在输入信道状态信息张量 $\boldsymbol{\mathcal{H}}$ 受到受控失真时，UWMMSE 和 WMMSE 所达到的归一化平均和速率变化，并与相同信道状态信息张量未受失真时达到的理想性能进行比较。

附录 A 命题 I 的证明

证明：设 $\mathrm{V}^{(k)} = \Lambda_k(\mathrm{H}, \mathrm{V}^{(k-1)}; \Theta)$ 表示式 (5)-(10) 中 UWMMSE架构在第 k 层的输出。此外，$\tilde{\mathrm{H}} = \boldsymbol{\Pi}\mathrm{H}\boldsymbol{\Pi}^\top$ 表示信道张量经过任意置换后的版本，其中置换仅作用于张量的前两个维度。并且，$\tilde{\mathrm{V}}^{(k-1)} = \boldsymbol{\Pi}\mathrm{V}^{(k-1)}$表示作为第 k 层输入的 $\mathrm{V}^{(k-1)}$，其中置换仅作用于张量的第一个维度。此外，$\tilde{\mathrm{S}} = \boldsymbol{\Pi}\mathrm{S}\boldsymbol{\Pi}^\top$ 是变换后的$\tilde{}$信道状态信息矩阵经过置换后的版本，而 $\mathrm{Q} = \boldsymbol{\Pi}\mathrm{Q}$ 是节点特征向量经过相应置换后的结果。首先，我们要证明 $\Lambda_k(\tilde{\mathrm{H}}, \tilde{\mathrm{V}}^{(k-1)}, \Theta) = \boldsymbol{\Pi}\xi^{(k)}(\mathrm{H}, \mathrm{V}, \mathrm{S}, \mathrm{Q}; \theta) = \boldsymbol{\Pi}\xi^{(k)}(\tilde{\mathrm{H}}, \tilde{\mathrm{V}}, \tilde{\mathrm{S}}, \tilde{\mathrm{Q}}; \theta) = \boldsymbol{\Pi}\mathrm{V}^{(k)}$。由于 $\xi^{(k)}$ 是置换等变的，我们知道 $\xi^{(k)}(\mathrm{H}, \mathrm{V}, \mathrm{S}, \mathrm{Q}; \theta) = \boldsymbol{\Pi}\xi^{(k)}(\tilde{\mathrm{H}}, \tilde{\mathrm{V}}, \tilde{\mathrm{S}}, \tilde{\mathrm{Q}}; \theta) = \boldsymbol{\Pi}\mathrm{V}^{(k)}$。设节点 i 在置换 $\boldsymbol{\Pi}$后被赋予新索引 $\pi(i)$。$\tilde{\mathrm{H}}$ 然后令 $\mathrm{H}_{ij} = [\mathrm{H}]_{ij::}$，由式 (5) 可得

$$\begin{aligned}\tilde{\mathbf{U}}_i^{(k)} &= \left(\sum_{j\neq i}\tilde{\mathbf{H}}_{ij}\tilde{\mathbf{V}}_j^{(k-1)}\tilde{\mathbf{V}}_j^{(k-1)^H}\tilde{\mathbf{H}}_{ij}^H + \sigma^2\mathbf{I}_R\right)^{-1}\tilde{\mathbf{H}}_{ii}\tilde{\mathbf{V}}_i^{(k-1)} \\ &= \left(\sum_{j\neq i}\mathbf{H}_{\pi(i)\pi(j)}\mathbf{V}_{\pi(j)}^{(k-1)}\mathbf{V}_{\pi(j)}^{(k-1)^H}\mathbf{H}_{\pi(i)\pi(j)}^H + \sigma^2\mathbf{I}_R\right)^{-1} \\ &\qquad \mathbf{H}_{\pi(i)\pi(i)}\mathbf{V}_{\pi(i)}^{(k-1)} = \mathbf{U}_{\pi(i)}^{(k)},\end{aligned}$$

将其写成张量形式，可得 $\tilde{\boldsymbol{\mathcal{U}}}^{(k)} = \boldsymbol{\Pi}\boldsymbol{\mathcal{U}}_{(k)}$。同样，由 (6) 和 (9) 可推导出 $\tilde{\boldsymbol{\mathcal{W}}}^{(k)} = \boldsymbol{\Pi}\boldsymbol{\mathcal{W}}^{(k)}$ 和$\tilde{\boldsymbol{\mathcal{V}}}^{(k)} = \Lambda_k(\tilde{\mathbf{H}}, \tilde{\boldsymbol{\mathcal{V}}}^{(k-1)}; \Theta) = \boldsymbol{\Pi}\boldsymbol{\mathcal{V}}^{(k)}$，证毕。

利用这些恒等式，我们现在要证明$\Lambda(\cdot;\Theta)$是等变的。具体而言，对于$K = 1$这一特殊情况，根据 $\Lambda(\cdot;\Theta)$的定义可得，

$$\begin{aligned}\Lambda(\tilde{\boldsymbol{\mathcal{H}}};\Theta) &= \Lambda_1(\tilde{\boldsymbol{\mathcal{H}}}, \boldsymbol{\mathcal{V}}^{(0)};\Theta) = \Lambda_1(\tilde{\boldsymbol{\mathcal{H}}}, \boldsymbol{\Pi}\boldsymbol{\mathcal{V}}^{(0)};\Theta) \\ &= \boldsymbol{\Pi}\Lambda_1(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^{(0)};\Theta) = \boldsymbol{\Pi}\Lambda(\boldsymbol{\mathcal{H}};\Theta),\end{aligned}$$

$\boldsymbol{\mathcal{V}}^{(0)}$ 是一个常量张量这一事实得出了第二个等式，而第三个等式则是将前一个恒等式用于 $k = 1$ 后得到的特例。至此，单层 UWMMSE 的置换等变性证明完成。对于具有 $K > 1$ 的 UWMMSE，可以通过一个简单的归纳论证证明其置换等变性，此处略去。 ■

附录 B 定理 1 的证明

证明：本证明受到 [24], 中收敛结果的证明以及 [53] 中给出的展开 ISTA 理论线性收敛结果 [30], 的启发。我们假设对所有 i，$\mathrm{Trace}\left(\mathrm{V}_i^{(k)}\mathrm{V}_i^{(k)^H}\right) \to P_i^* = \mathrm{Trace}\left(\mathrm{V}_i^*\mathrm{V}_i^{*^H}\right)$ 均一致成立。这意味着，对所有 $\eta > 0$，存在一个层索引 K_1，使得对所有 $k > K_1$，$\mathrm{V}^{(k-1)} = \mathrm{V}^* + \mathrm{E}_V$ 且 $\mathrm{V}^{(k)} = \mathrm{V}^* + \mathrm{E}_V'$，其中 $\|[\mathrm{E}_V]_i\|_F < \eta$ 且 $\|[\mathrm{E}_V']_i\|_F < \eta$ 对所有 i 均成立。沿用附录 A 中命题 1 证明的相同记号，我们有 $\mathrm{V}^{(k)} = \Lambda_k(\mathrm{H}, \mathrm{V}^{(k-1)}; \Theta)$。由一致收敛可得，

$$\boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V' = \boldsymbol{\mathcal{V}}^{(k)} = \Lambda_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta). \tag{16}$$

因此，我们需要得到一个表达式 $g_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta)$，使得 $\Lambda_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta) = \boldsymbol{\mathcal{V}}^* + g_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta)$，并可将其代入式 (16) 得到

$$\boldsymbol{\mathcal{E}}_V' = g_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta). \tag{17}$$

式 (17) 对正的 $\eta \to 0$ 成立，这意味着当$k \to \infty$时，$g_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta) \to 0$。接下来，我们将致力于确定 $g_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{E}}_V; \Theta)$的显式形式，从而证明式 (15)。

首先，针对任意的 i，我们将式 (9) 中的 $\mathbf{V}_i^{(k-1)}$ 替换为 $\mathbf{V}_i^* + \mathcal{E}_{V_i}$，其中 $\mathcal{E}_{V_i} = [\boldsymbol{\mathcal{E}}_V]_i$，得到

$$\mathbf{U}_i^{(k)} = \left(\sum_j \mathbf{H}_{ij}(\mathbf{V}_j^* + \mathcal{E}_{V_j})(\mathbf{V}_j^* + \mathcal{E}_{V_j})^H\mathbf{H}_{ij}^H + \sigma^2\mathbf{I}_R\right)^{-1} \mathbf{H}_{ii}(\mathbf{V}_i^* + \mathcal{E}_{V_i})^H$$

由于我们工作在低噪声区域，因此在逆矩阵中，噪声项相较于干扰项可以忽略，从而得到

$$\mathbf{U}_i^{(k)} = \mathbf{A}_i^{-1}\mathbf{B}_i$$

其中，我们定义两个哑变量 $\mathbf{A}_i = \sum_j \mathbf{H}_{ij}(\mathbf{V}_j^* + \mathcal{E}_{V_j})^H\mathbf{H}_{ij}^H$ 和 $\mathbf{B}_i = \mathbf{H}_{ii}(\mathbf{V}_i^* + \mathcal{E}_{V_i})$。定义

第三个哑变量 $\mathbf{C}_i = \sum_j \mathbf{H}^{ij}\mathbf{V}_j^{*}{}^{H}\mathbf{H}_{ij}^{H}$，我们有，

$$\begin{aligned}\mathbf{U}_i^{(k)} &= \mathbf{A}_i^{-1}\mathbf{C}_i\mathbf{C}_i^{-1}\mathbf{B}_i \\ &= \mathbf{C}_i^{-1}\mathbf{B}_i + \mathbf{A}_i^{-1}(\mathbf{C}_i - \mathbf{A}_i)\mathbf{C}_i^{-1}\mathbf{B}_i \\ &= \mathbf{C}_i^{-1}\hat{\mathbf{B}}_i + \mathbf{C}_i^{-1}\mathbf{H}_{ii}\mathcal{E}_{V_i}^{H} - \mathbf{A}_i^{-1}\Bigg[\sum_j \mathbf{H}_{ij}(\mathbf{V}_j^{*}\mathcal{E}_{V_j}^{H} \\ &\qquad + \mathcal{E}_{V_j}\mathbf{V}_j^{*}{}^{H} + \mathcal{E}_{V_j}\mathcal{E}_{V_j}^{H})\mathbf{H}_{ij}^{H}\Bigg]\mathbf{C}_i^{-1}\mathbf{B}_i \\ &\qquad\qquad \text{(where, } \hat{\mathbf{B}}_i = \mathbf{H}_{ii}\mathbf{V}_i^{*}{}^{H}\text{)}\end{aligned}$$

显然，第一项表示$\mathbf{U}_i^{*}$。所有其余依赖于 $\mathcal{E}_V$的项构成 $\mathcal{E}_U$，使得对于所有 i，当 $\eta \to 0$ 时，$||\mathcal{E}_{U_i}||_F \to 0$。因此，$\mathbf{U}_i^{(k)}$ 的形式为，

$$\mathbf{U}_i^{(k)} = \mathbf{U}_i^{*} + \mathcal{E}_{U_i}, \text{ for all } i. \tag{18}$$

接下来考虑式（6），其中分别将$\mathbf{U}^{(k)}$ 和$\mathbf{V}^{(k-1)}$替换为$\mathbf{U}^{*} + \mathcal{E}_U$ 和$\mathbf{V}^{*} + \mathcal{E}_V$。与式（18）中的步骤类似，并利用 Φ_ξ的连续性，可得

$$w_i^{(k)} = w_i^{*} + \Phi_{\xi_i^{(k)}}(w_i^{(k)}) + \epsilon_{w_i}, \text{ for all } i \tag{19}$$

其中，对于所有 i，$|\epsilon_{w_i}| \to 0$ 当 $\eta \to 0$ 时。接下来，我们对式（9）重复这一过程。为此，我们将哑变量 A_i、B_i和 C_i重新定义为：$\mathrm{A}_i = \sum_j \mathrm{H}_{ij}^{H}\mathrm{U}_j^{(k)}w_j^{(k)}\mathrm{U}_j^{(k)}{}^{H}\mathrm{H}_{ij}$、$\mathrm{B}_i = \mathrm{H}_{ii}^{H}\mathrm{U}_i^{(k)}w_i^{(k)}$和 $\mathrm{C}_i = \sum_j \mathrm{H}_{ij}^{H}\mathrm{U}_j^{*}w_j^{*}\mathrm{U}_j^{*}{}^{H}\mathrm{H}_{ij}$。此外，我们用式（18）和式（19）中得到的相应形式替换 $\mathrm{U}_i^{(k)}$和 $w_i^{(k)}$。显然，对于 $k \to \infty$，非线性 $\beta(\cdot)$已不再显著，因为 Trace $\left(\bar{\mathrm{V}}_i^{(k)}\bar{\mathrm{V}}_i^{(k)H}\right) \to P_i^{*}$，且对于所有 i，$0 < P_i^{*} < P_{\max}$。因此，由式（9）—（10）可得，

$$\begin{aligned}\mathbf{V}_i^{(k)} &= \mathbf{C}_i^{-1}\mathbf{B}_i + \mathbf{A}_i^{-1}(\mathbf{C}_i - \mathbf{A}_i)\mathbf{C}_i^{-1}\mathbf{B}_i \\ &= \mathbf{C}_i^{-1}\hat{\mathbf{B}}_i + \mathcal{E}_{V_i} + \mathbf{A}_i^{-1}\left[\mathbf{A}_i\Phi_{\xi_i^{(k)}}(w_i^{(k)})\right]\mathbf{C}_i^{-1}\bar{\mathbf{B}}_i \\ &\quad + \mathbf{A}_i^{-1}\left[(\mathbf{C}_i - \mathbf{A}_i)(w_i^{*} + \Phi_{\xi_i^{(k)}}(w_i^{(k)}))\right]\mathbf{C}_i^{-1}\bar{\mathbf{B}}_i, \\ &\qquad \text{(where, } \hat{\mathbf{B}}_i = \mathbf{H}_{ii}^{H}\mathbf{U}_i^{*}w_i^{*} \text{ and } \bar{\mathbf{B}}_i = \mathbf{H}_{ii}^{H}\mathbf{U}_i^{*}\text{)} \\ &= \mathbf{C}_i^{-1}\hat{\mathbf{B}}_i + \mathcal{E}_{V_i} + \mathbf{A}_i^{-1}\Bigg[\sum_{j\neq i}\mathbf{H}_{ij}^{H}\mathbf{U}_j^{(k)}(w_j^{*}\Phi_{\xi_i^{(k)}}(w_i^{(k)}) \\ &\quad - w_i^{*}\Phi_{\xi_j^{(k)}}(w_j^{(k)}))\mathbf{U}_j^{(k)}{}^{H}\mathbf{H}_{ij}\Bigg]\mathbf{C}_i^{-1}\bar{\mathbf{B}}_i \text{ for all } i\end{aligned} \tag{20}$$

其中，对于所有 i，$||\mathcal{E}_{V_i}||_F \to 0$as $\eta \to 0$。由于（$\boldsymbol{\mathcal{U}}^{*}$，$\boldsymbol{\mathcal{W}}^{*}$，$\boldsymbol{\mathcal{V}}^{*}$）构成 WMMSE 更新的不动点，因此，对于所有 i，$_i^{*}$由式（20）右端的第一项给出。因此，式（20）右端的最后两项逐元素构成函数 $g_k(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{V}}^{*} + \boldsymbol{\mathcal{E}}; \Theta)$。此外，由于 $||\mathcal{E}_{V_i}||_F \to 0$ as$\eta \to 0$（等价地，as $k \to \infty$），当 k 趋于无穷大时，式（20）中的最后一项必须趋于 0，从而完成证明。 ■

参考文献

[1] A. Chowdhury、G. Verma、C. Rao、A. Swami 和 S. Segarra，“用于战术 MIMO 的机器学习辅助功率分配”，载于 *IEEE* 军事通信会议 IEEE，2021 年，第 273–278 页。

[2] R. W. Heath Jr 和 A. Lozano，《MIMO 通信基础》。剑桥大学出版社，2018 年。[3] E. Bj¨ornson、J. Hoydis、L. Sanguinetti 等，“大规模 MIMO 网络：频谱效率、能效与硬件效率”，《信号处理研究与趋势》，第 11 卷，第 3–4 期，第 154–655 页，2017年。[4] E. Bj¨ornson、L. Sanguinetti、J. Hoydis 和 M. Debbah，“高能效多用户 MIMO 系统的最优设计：大规模 MIMO 是否就是答案？”，IEEE 无线通信汇刊，第 14 卷，第 6 期，第 3059–3075 页，2015 年。[5] H. Q. Ngo、E. G. Larsson 和 T. L. Marzetta，“超大规模多用户 MIMO 系统的能效与频谱效率”，IEEE 通信汇刊，第 61 卷，第 4 期，第 1436–1449 页，2013 年。[6] M. Razaviyayn、M. Hong 和 Z.-Q. Luo，“实现最大最小公平性的 MIMO 干扰广播信道线性收发器设计”，《信号处理》，第 93 卷，第 12 期，第 3327–3340 页，2013 年。[7] Z.-Q. Luo 和 S. Zhang，“动态频谱管理：复杂度与对偶性”，IEEE 信号处理精选主题期刊，第 2 卷，第 1 期，第 57–73 页，2008 年。[8] L. Liang、H. Ye、G. Yu 和 G. Y. Li，“基于深度学习的无线资源分配及其在车联网中的应用”，《IEEE 会议录》，第 108 卷，第 2 期，第 341–356 页，2019 年。[9] Y.-F. Liu、Y.-H. Dai 和 Z.-Q. Luo，“MISO 干扰信道的协同波束成形：复杂度分析与高效算法”，IEEE 信号处理汇刊，第 59 卷，第 3 期，第 1142–1157 页，2011 年。[10] W. Yu 和 R. Lui，“多载波系统非凸频谱优化的对偶方法”，IEEE 通信汇刊，第 54 卷，第 7 期，第 1310–1322 页，2006 年。[11] S. Hayashi 和 Z.-Q. Luo，“干扰受限多用户通信系统的频谱管理”，IEEE 信息论汇刊，第 55 卷，第 3 期，第 1153–1175 页，2009 年。[12] J. Papandriopoulos 和 J. S. Evans，“SCALE：用于多用户 DSL 网络频谱均衡的低复杂度分布式协议”，IEEE 信息论汇刊，第 55 卷，第 8 期，第 3711–3724 页，2009 年。[13] F. Wang、M. Krunz 和 S. Cui，“认知无线电网络中基于价格的频谱管理”，IEEE 信号处理精选主题期刊，第 2 卷，第 1 期，第 74–87 页，2008 年。[14] Q. Shi、M. Razaviyayn、Z.-Q. Luo 和 C. He，“面向 MIMO 干扰广播信道分布式总效用最大化的迭代加权 MMSE 方法”，IEEE 信号处理汇刊，第 59 卷，第 9 期，第 4331–4340 页，2011 年。[15] H. Sun、X. Chen、Q. Shi、M. Hong、X. Fu 和 N. D. Sidiropoulos，“学习优化：训练深度神经网络以进行干扰管理”，IEEE 信号处理汇刊，第 66 卷，第 20 期，第 5438–5453 页，2018 年。[16] H. Sun、W. Pu、M. Zhu、X. Fu、T.-H. Chang 和 M. Hong，“在分阶段动态环境中学习连续优化无线资源”，载于 IEEE 声学、语音与信号处理国际会议（ICASSP），2021 年，第 4945–4949 页。[17] W. Lee、M. Kim 和 D.-H. Cho，“深度功率控制：一种基于卷积神经网络的发射功率控制方案”，IEEE 通信快报，第 22 卷，第 6 期，第 1276–1279 页，2018 年。[18] Z. Deng、Q. Sang、Y. Pan 和 Y. Xin，“深度学习在干扰信道功率控制中的应用：一种基于 RNN 的方法”，载于《自适应与收敛系统研究会议论文集》，2019 年，第 96–100 页。[19] M. Eisen、C. Zhang、L. F. Chamon、D. D. Lee 和 A. Ribeiro，“学习无线系统中的最优资源分配”，IEEE 信号处理汇刊，第 67 卷，第 10 期，第 2775–2790 页，2019 年。[20] M. Eisen 和 A. R. Ribeiro，“利用随机边图神经网络实现最优无线资源分配”，IEEE 信号处理汇刊，2020 年。[21] Q. Hu、Y. Cai、Q. Shi、K. Xu、G. Yu 和 Z. Ding，“迭代算法诱导的深度展开神经网络：多用户 MIMO 系统的预编码设计”，IEEE 无线通信汇刊，第 20 卷，第 2 期，第 1394–1410 页，2020 年。[22] L. Pellaco、M. Bengtsson 和 J. Jald´en，“加权 MMSE 波束成形算法的深度展开”，arXiv 预印本 arXiv:2006 08448，2020 年。[23] L. Schynol 和 M. Pesavento，“采用深度展开的多小区 MU-MIMO 协同和速率最大化”，arXiv 预印本 arXiv:2202.10371，2022年。[24] A. Chowdhury、G. Verma、C. Rao、A. Swami 和 S. Segarra，“利用图神经网络展开 WMMSE，实现高效功率分配”，IEEE 无线通信汇刊，第 20 卷，第 9 期，第 6004–6017 页，2021年。[25] V. Monga、Y. Li 和 Y. C. Eldar，“算法展开：用于信号与图像处理的可解释、高效深度学习”，arXiv 预印本 arXiv:1912.10557，2019 年。

[26] A. 通信系统的深度Learning算法与算法研究方向”，载于 IEEE 信号处理系统研讨会（SiPS）。IEEE，2019年，第 266–271 页。[27] R. Liu、S. Cheng、L. Ma、X. Fan 和 Z. Luo，“深度近端展开：算法框架、收敛性分析及应用”，IEEE 图像处理汇刊，第 28 卷，第 10 期，第 5013–5026 页，2019年。[28] N. Farsad、N. Shlezinger、A. J. Goldsmith 和 Y. C. Eldar，“基于模型的机器学习驱动的数据驱动符号检测”，arXiv 预印本 arXiv:2002.07806，2020 年。[29] J. R. Hershey、J. L. Roux 和 F. Weninger，“深度展开：基于模型的灵感启发新型深度架构”，arXiv 预印本 arXiv:1409.2574，2014年。[30] K. Gregor 和 Y. LeCun，“学习稀疏编码的快速近似方法”，载于国际机器学习会议（ICML），2010年，第 399–406 页。[31] Z. Liu、X. Li、P. Luo、C. C. Loy 和 X. Tang，“用于语义分割的深度学习马尔可夫随机场”，IEEE 模式分析与机器智能汇刊，第 40 卷，第 8 期，第 1814–1828 页，2017年。[32] Y. Li、M. Tofighi、V. Monga 和 Y. C. Eldar，“深度图像去模糊的算法展开方法”，载于 IEEE 国际声学、语音与信号处理会议（ICASSP）。IEEE，2019年，第 7675–7679 页。[33] O. Solomon、R. Cohen、Y. Zhang、Y. Yang、Q. He、J. Luo、R. J. van Sloun 和 Y. C. Eldar，“应用于超声杂波抑制的深度展开鲁棒主成分分析”，IEEE 医学影像汇刊，第 39 卷，第 4 期，第 1051–1063 页，2019年。[34] F. Gama、N. Zilberstein、M. Sevilla、R. Baraniuk 和 S. Segarra，“粒子滤波器采样分布的无监督学习”，arXiv 预印本 arXiv:2302.01174，2023年。[35] N. Zilberstein、C. Dick、R. Doost-Mohammady、A. Sabharwal 和 S. Segarra，“用于大规模 MIMO 检测的退火朗之万动力学”，IEEE 无线通信汇刊，2022年。[36] Z. Zhao、G. Verma、C. Rao、A. Swami 和 S. Segarra，“基于图神经网络的链路调度”，IEEE 无线通信汇刊，2022年。[37] B. Li、G. Verma 和 S. Segarra，“面向无线网络能量感知功率分配的基于图的算法展开”，IEEE 无线通信汇刊，第 22 卷，第 2 期，第 1359–1373 页，2023年。[38] M. Hong 和 Z.-Q. Luo，“第 8 章——干扰信道的信号处理与最优资源分配”，载于《Academic Press信号处理文库：第 2 卷》，丛书《Academic Press信号处理文库》，N. D. Sidiropoulos、F. Gini、R. Chellappa 和 S. Theodoridis 编。Elsevier，2014年，第 2 卷，第 409–469 页。[39] M. Leshno、V. Y. Lin、A. Pinkus 和 S. Schocken，“带有非多项式激活函数的多层前馈网络可以逼近任意函数”，《神经网络》，第 6 卷，867，1993年 6，861第。– 页。[40] Y. Kuroe、M. Yoshid 和 T. Mori，“关于复值神经网络的激活函数——能量函数的存在性——”，载于《人工神经网络与神经信息处理》。Springer，2003年，第 985–992 页。[41] A. L. Maas、A. Y. Hannun、A. Y. Ng 等，“整流非线性改善神经网络声学模型”，载于国际机器学习会议（ICML），第 30 卷，第 1 期，2013 年，第 3 页。[42] T. N. Kipf 和 M. Welling，“基于图卷积网络的半监督分类”，载于国际学习表征会议（ICLR），2017年。[43] F. Gama、A. G. Marques、G. Leus 和 A. Ribeiro，“面向图上信号的卷积神经网络架构”，IEEE 信号处理汇刊，第 67 卷，第 4 期，第 1034–1049 页，2018 年。[44] V. P. Dwivedi、C. K. Joshi、A. T. Luu、T. Laurent、Y. Bengio 和 X. Bresson，“图神经网络基准测试”，《机器学习研究杂志》，第 24 卷，第 43 期，第 1–48 页，2023 年。[45] S. Segarra、A. G. Marques 和 A. Ribeiro，“最优图滤波器设计及其在分布式线性网络算子中的应用”，IEEE 信号处理汇刊，第 65 卷，第 15 期，第 4117–4131 页，2017年。[46] M. Lin、Q. Chen 和 S. Yan，“网络中的网络”，arXiv 预印本 arXiv:1312.4400，2013 年。[47] A. G. Marques、S. Segarra、G. Leus 和 A. Ribeiro，“通过连续局部聚合进行图信号采样”，IEEE 信号处理汇刊，第 64 卷，第 7 期，第 1832–1843 页，2016年。[48] A. Chowdhury、G. Verma、C. Rao、A. Swami 和 S. Segarra，“利用图神经网络和深度算法展开实现高效功率分配”，载于 ICASSP 2021——2021年 IEEE 声学、语音与信号处理国际会议（ICASSP）。IEEE，2021年，第 4725–4729 页。[49] F. Liang、C. Shen、W. Yu 和 F. Wu，“通过集成深度神经网络实现最优功率控制”，IEEE 通信汇刊，第 68 卷，第 3 期，第 1760–1776 页，2019 年。[50] D. Tse 和 P. Viswanath，《无线通信基础》。剑桥大学出版社，2005年。

[51] B. Ginsburg, P. Castonguay, O. Hrinchuk, O. Kuchaiev, V. Lavrukhin, R. Leary, J. Li, H. Nguyen, Y. Zhang, and J. M. Cohen，《采用逐层自适应矩训练深度网络的随机梯度方法》，*arXiv*预印本 *arXiv:1905.11286*，2019 年。[52] A. Chowdhury, F. Gama, and S. Segarra，《用于功率分配的展开式 WMMSE的稳定性分析》，载于*IEEE* 国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2022年，第5298–5302页。[53] X. Chen, J. Liu, Z. Wang, and W. Yin，《展开 ISTA 的理论线性收敛及其实用权重和阈值》，载于神经信息处理系统进展（*NeurIPS*），2018 年，第9061–9071页。