

模型驱动深度学习增强的马尔可夫链蒙特卡罗 MIMO 检测器：设计、仿真与原型验证

曹益泉¹, 周星宇¹, 张 静¹, 梁 乐¹, 李 勇^{2,3}, 金 石^{1,2*}

(1. 东南大学信息科学与工程学院, 江苏南京 211189; 2. 先进通信网全国重点实验室, 河北石家庄 050081;
3. 中国电科网络通信研究院, 河北石家庄 050081)

摘 要: 多输入多输出 (Multiple-Input Multiple-Output, MIMO) 系统规模日益增长, 导致接收机信号检测计算复杂度急剧上升, 传统检测算法难以在误码性能和复杂度之间取得良好平衡。基于马尔可夫链蒙特卡罗 (Markov Chain Monte Carlo, MCMC) 的检测算法能以多项式量级的复杂度实现近最优的检测性能, 然而该方法在低采样数下性能损失严重。因此, 本文引入了基于模型驱动深度学习技术, 将 MCMC 迭代过程展开为级联网络结构, 向网络中引入可训练参数, 通过深度学习优化参数设置。根据复杂度分析与仿真验证, 所提方案在编码场景下的误码性能优于原始算法约 1 dB, 同时计算复杂度显著低于原始算法。为验证模型驱动深度学习方案在实际传输中的性能, 搭建 2 × 2 MIMO 智能通信原型验证平台, 并进行端到端口传输测试。测试结果表明, 模型驱动深度学习增强的 MCMC 检测算法可以以更低的计算复杂度实现误码性能优势, 从而证实了所提方案在实际传输环境中的有效性和鲁棒性。

关键词: MIMO 检测; 马尔可夫链蒙特卡罗; 模型驱动; 深度学习; 原型验证平台

基金项目: 国家自然科学基金 (No.62261160576, No.62301154, No.623B2019, No.62231019); 先进通信网全国重点实验室基金课题 (No.SCX23641X011)

中图分类号: TN929.5

文献标识码: A

文章编号: 0372-2112(2025)04-1142-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240798

Model-Driven Deep Learning-Enhanced Markov Chain Monte Carlo MIMO Detector: Design, Simulation and Prototyping

CAO Yi-xiao¹, ZHOU Xing-yu¹, ZHANG Jing¹, LIANG Le¹, LI Yong^{2,3}, JIN Shi^{1,2*}

(1. School of Information Science and Engineering, Southeast University, Nanjing, Jiangsu 211189, China;

2. National Key Laboratory of Advanced Communication Networks, Shijiazhuang, Hebei 050081, China;

3. Academy of Network & Communications of CETC, Shijiazhuang, Hebei 050081, China)

Abstract: The scale of multiple-input multiple-output (MIMO) systems is growing rapidly, leading to a dramatic increase in the computational complexity of receiver signal detection. Traditional detection algorithms struggle to achieve a good balance between bit error rate (BER) performance and computational complexity. Markov chain Monte Carlo (MCMC)-based detection algorithms can achieve near-optimal detection performance with polynomial complexity, but their performance deteriorates significantly with low sampling rates. To address this issue, this paper introduces a model-driven deep learning approach, which transforms the MCMC iterative process into a cascade network structure. Trainable parameters are incorporated into the network, and deep learning techniques are employed to optimize their settings. Based on complexity analysis and simulation results, the proposed method outperforms the original algorithm in terms of BER by approximately 1 dB in coding scenarios, while significantly reducing computational complexity. To validate the performance of the model-driven deep learning approach in real-world transmission, a 2×2 MIMO smart communication prototype is developed, and end-to-end air interface transmission tests are conducted. The test results demonstrate that the MCMC detection algorithm enhanced by the model-driven deep learning approach still achieves a significant BER performance advantage with lower computational complexity, thereby confirming the effectiveness and robustness of the proposed solution in practical trans-

mission environments.

Key words: MIMO detection; Markov chain Monte Carlo; model-driven; deep learning; prototyping platform

Foundation Item(s): National Natural Science Foundation of China (No.62261160576, No.62301154, No.623B2019, No.62231019); National Key Laboratory of Advanced Communication Networks Foundation Project (No.SCX23641X011)

1 引言

多输入多输出 (Multiple-Input Multiple-Output, MIMO) 技术在接收端部署大量天线,能够显著提升无线通信系统的吞吐量与频谱效率,是第五代无线通信系统的核心使能技术之一^[1]. 然而,新一代通信系统中的超异构网络、新兴通信场景以及服务需求产生海量数据^[2-4],为应对巨大的传输需求,MIMO 系统中天线数量进一步大幅增长,达到成百乃至上千量级,给接收机信号检测带来全新挑战. 经典的极大似然 (Maximum Likelihood, ML) 检测通过穷举发射符号组合以寻求最大化似然函数的解向量,可达最优性能,然而该算法的计算复杂度随决策变量数目指数级增长,无法在实际场景中大规模应用^[5]. 计算复杂度较低的方案,如线性最小均方误差 (Linear Minimum Mean Square Error, LMMSE) 和迫零 (Zero Forcing, ZF) 等线性检测算法,其误码性能相较于最优检测存在显著差异,难以满足极高可靠性传输的需求.

近年来,人工智能及机器学习的引入为大规模通信信号处理提供了新的实现思路和解决方案. 作为贝叶斯统计机器学习领域中一类重要的随机近似推断方法,马尔可夫链蒙特卡罗 (Markov Chain Monte Carlo, MCMC) 技术通过系统性步骤生成服从目标概率分布的样本用于后续的推断分析,能够以有限的复杂度高效解决信道估计、信号检测、信道译码等接收机信号处理问题,受到广泛关注^[6-9]. MH (Metropolis-Hastings) 是应用最为广泛的 MCMC 算法,该方案利用提议分布实施采样,并按照一定的概率接受或拒绝所得样本以完成对复杂目标分布的近似,其中提议分布是任意易于采样的简单分布. 在实际应用中,MH 算法采样效率较低,需要大量样本才能收敛至最优解附近. 为了减少迭代次数,研究者们引入了梯度下降 (Gradient Descent, GD) 加速算法收敛^[10,11]. 文献[12]基于随机梯度 Langevin 动力学,提出了一种梯度下降辅助的 MCMC 检测算法,在中型的 MIMO 系统中具备同球形译码相近的误码性能. 然而,文献[12]的方案仅针对相移键控 (Phase Shift Keying, PSK) 系统,并不适用于正交幅度调制 (Quadrature Amplitude Modulation, QAM). 文献[13]使用 Nesterov 加速梯度法 (Nesterov's Accelerated Gradient, NAG) 辅助 MCMC 采样,其优点在于 NAG 梯度下降只需使用目标函数一阶导数信息,因此其计算复杂度对变量维度的增加不敏感,有效克服了矩阵求逆计算负担. 然

而,NAG 梯度下降效率较低,算法收敛速度仍有提升空间. 文献[14]提出 MHGD 算法,利用更为高效的牛顿法梯度下降获取最优解所在的方向,指引 MH 随机游走采样,从而加快算法的收敛. 仿真表明,在采样数充足时,梯度下降辅助的 MCMC 检测算法能够取得近 ML 的检测性能. 然而,现有的 MCMC 检测算法依赖于大量采样迭代,引发极高的计算时延、功耗以及算力开销,而在采样数不足时性能不佳. 为适配下一代无线通信系统的低时延需求及计算资源受限特征,亟需改善 MCMC 检测算法在低采样数下的误码性能.

以深度学习为代表的人工智能 (Artificial Intelligence, AI) 技术在移动通信系统设计中大放异彩,实现 AI 与通信的深度融合被视作下一代无线通信系统的关键愿景之一^[15-18]. 在 MIMO 接收机中,深度学习技术已被广泛应用于改进传统检测算法,展现出突破设计瓶颈的巨大潜力. 当前深度学习的应用方案可分为数据驱动和模型驱动两类范式. 数据驱动深度学习需要大量优质数据集,训练开销较大;而模型驱动深度学习基于通信领域专家知识^[19],能够以较小的训练开销显著改善算法性能,受到广泛关注,且具有更强的可解释性和提升系统泛化性的潜力. 文献[20]基于模型驱动深度学习,通过展开已有检测算法并加入可训练参数,提出了一种 DetNet 检测器,误码性能与近似消息传递 (Approximate Message Passing, AMP) 算法相近. 该算法的缺陷在于引入参数过多,导致算法训练开销巨大. 文献[21]的思路与文献[20]类似,利用模型驱动深度学习增强正交近似消息传递 (Orthogonal Approximate Message Passing, OAMP) 算法,提出一种 OAMP-NET 检测器,并在设计过程中额外考虑了信道估计误差. 仿真结果表明,所提 OAMP-NET 算法相比原始的 OAMP 取得了明显的性能增益,同时在全类失配场景下有较强的泛化性与鲁棒性. 然而,OAMP-NET 计算复杂度过高,难以应用于大规模 MIMO 检测. 文献[19]将深度学习技术同软干扰消除 (Soft Interference Cancellation, SIC) 算法^[22]结合,提出一种 DisNet 检测方案,相较于原本的 SIC 算法改善了误码性能,且能较好地适应非线性系统. 然而,文献[19]仅通过代码仿真验证了 DisNet 的性能优势,没有相关的空口传输测试以验证该算法在实际信道场景下的性能增益.

综上所述,模型驱动深度学习增强的检测算法研究还存在诸多的不足. 为解决上述问题,本文提出一种

模型驱动深度学习增强的 MCMC 检测算法,引入并优化可训练参数,改善原有 MCMC 检测算法在低采样数下的误码性能,同时省略部分高开销运算以降低计算复杂度. 主要工作总结如下:

(1) 引入模型驱动深度学习改进已有算法,将 MCMC 检测算法迭代过程展开为多层级联网络结构,向网络合适位置引入可训练参数,提出了模型驱动增强的 MCMC 检测算法 MHGDNNet. 该方案利用深度学习技术离线训练优化网络权重,在信号检测时加载参数,通过预训练的参数代替原算法中部分复杂运算. 仿真结果表明,所提 MHGDNNet 检测算法能够显著优化误码性能,对调制阶数以及信噪比失配场景展现出较强的鲁棒性与泛化能力,且在计算复杂度方面亦能实现明显优势,有效地实现了误码性能和计算复杂度之间的折中.

(2) 考虑到现实信道环境同仿真信道模型差异较大,基于软件无线电(Universal Software Radio Peripheral, USRP)以及 AI 运算单元搭建智能无线传输原型验证平台,并进行 2×2 空口传输实测. 测试结果验证了所提算法在实际信道环境中的性能增益,并证实了 MHGDNNet 在信道失配场景下同样有较好的鲁棒性与泛化性.

2 系统模型与算法简介

2.1 系统模型

考虑一个具有 N_t 根发射天线和 N_r 根接收天线的 MIMO 传输系统. 发送端 N_t 根天线独立发送数据符号,共同组成传输符号 $\mathbf{x} \in \mathcal{Q}^{N_t \times 1}$. 其中, \mathcal{Q} 代表离散的正交振幅调制(Quadrature Amplitude Modulation, QAM)星座集合,调制阶数为 $\log_2 M$ 时星座集合 \mathcal{Q} 中包含 M 个元素. 发送信号通过无线信道和加性噪声影响,接收端信号 $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ 可表示为

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

其中, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ 为信号经历的传输信道, $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ 为信道加性随机噪声. 其中 \mathbb{C} 代表连续的复数空间. MIMO 检测问题即为根据信道信息 \mathbf{H} 以及噪声功率 σ^2 , 从接收向量 \mathbf{y} 中恢复出原始发射信号 \mathbf{x} . 最优检测基于最大后验(Maximum A Posteriori, MAP)准则,即在 $\mathcal{Q}^{N_t \times 1}$ 中寻找最大化后验概率 $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$ 的发射符号向量 \mathbf{x} , 其中 $p(\mathbf{x})$ 代表发送符号 \mathbf{x} 的先验概率,而 $p(\mathbf{y}|\mathbf{x})$ 代表似然函数. 假设 \mathbf{n} 为加性高斯白噪声,即 $\mathbf{n} \sim \text{CN}(\mathbf{0}_{N_r \times 1}, \sigma^2 \mathbf{I}_{N_r})$, 其中 $\mathbf{0}_{N_r \times 1}$ 表示维度为 $N_r \times 1$ 的零向量, \mathbf{I}_{N_r} 表示 $N_r \times N_r$ 的单位阵. 此时接收信号 \mathbf{y} 可看作是以 $\mathbf{H}\mathbf{x}$ 为均值, \mathbf{n} 的能量 σ^2 为方差的高斯随机变量,由此可以得到似然函数 $p(\mathbf{y}|\mathbf{x})$ 表达式为

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right) \quad (2)$$

假设发射符号等概,此时 MAP 准则可转换为 ML 准则,等价于寻找最小化损失函数 $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$ 的传输符号:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{Q}^{N_t \times 1}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \quad (3)$$

式(3)通过穷举发射符号组合实现,其计算复杂度为指数量级,因此难以实际应用. 研究者们提出一系列次优的线性检测算法以降低计算复杂度,如经典的 LMMSE 算法,其估计值 $\hat{\mathbf{x}}$ 可表示为

$$\hat{\mathbf{x}} = (\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I})^{-1} \mathbf{H}^H \mathbf{y} \quad (4)$$

相比于 ML 最优检测, LMMSE 计算复杂度较低,然而该算法误码性能较差,难以满足高可靠性传输需求.

2.2 MCMC 算法基础

MCMC 方法隶属于贝叶斯统计机器学习中的随机性方法,能够获取服从后验分布 $\pi(\cdot)$ 的样本集以便后续推断. 该方法需要构建马尔可夫链 $x^{(0)}, x^{(2)}, \dots, x^{(t)}, \dots$, 且该马尔可夫链的转移概率为 $P(\cdot|\cdot)$, 平稳分布等于目标分布 $\pi(\cdot)$. 平稳分布 $\pi(\cdot)$ 和转移概率 P 满足细致平稳条件,即

$$\pi(i)P(j|i) = \pi(j)P(i|j) \quad (5)$$

将马尔可夫链的每个状态都视作样本,则状态转移程 $x^{(t+1)} \sim P(\cdot|x^{(t)})$ 可视作 MCMC 利用状态转移概率 P 获取新样本. 当 t 足够大时,样本 $x^{(t)}$ 将收敛于平稳分布 $\pi(\cdot)$, 由此获取服从目标分布 $\pi(\cdot)$ 的样本集. 该采样方法要求状态转移概率 P 已知,然而对于一个确定的平稳分布 $\pi(\cdot)$, 满足细致平稳条件的转移概率 P 通常难以直接获得. 作为应用最广泛的 MCMC 算法, MH 基于拒绝采样思想解决该难题,利用式(6)获取候选样本:

$$x^{(t+1)} \sim Q(\cdot|x^{(t)}) \quad (6)$$

其中,提议分布 $Q(\cdot|\cdot)$ 为任意易于采样的分布. 为确保样本集能够收敛到目标分布 $\pi(\cdot)$, MH 算法需要按照一定的概率 α 接受候选样本, α 被称为接受率. 上述方法无需显式获取马尔可夫链对应的转移概率,保证了 MH 算法应用场景的广泛性. 本文 MCMC 检测器的设计将基于 MH 算法进行.

3 模型驱动深度学习增强的 MCMC 检测

3.1 梯度下降辅助的 MCMC 检测算法

传统 MCMC 算法采样效率较低,需要多轮迭代以精确近似目标分布,导致巨大的时间开销. 近年来, GD 和 MCMC 的结合在解决非凸优化问题方面显示出了巨大的潜力^[23], 该方案利用 GD 引导 MCMC 搜索方向, 结

合随机扰动以避免陷入局部极小值. 在梯度下降辅助的 MCMC 检测算法中, 每轮迭代主要分为以下步骤:

(1) 采用梯度下降引导 MCMC 的采样方向. 由于发射符号 \mathbf{x} 被限制在离散集合 $\mathcal{Q}^{N_s \times 1}$ 上, 为使用梯度下降算法, 需要将问题松弛至连续复数空间 $\mathbb{C}^{N_s \times 1}$. 松弛后的 MIMO 检测问题可表达为

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{N_s \times 1}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \quad (7)$$

式(7)将定义在离散 $\mathcal{Q}^{N_s \times 1}$ 上的 NP 难 (Non-deterministic Polynomial-time hard, NP-hard) 问题简化为连续复数空间上的二次凸问题, 极大降低了计算复杂度. 在简化的问题框架下, 算法得以在连续复数空间上使用 GD 方法, 从而揭示状态空间最优解的可能方向. 用 \mathbf{x}_{i-1} 代表前一轮迭代获取的样本, 梯度下降环节可表示为

$$\mathbf{z}_i = \mathbf{x}_{i-1} - \tau \cdot \nabla_{\mathbf{x}} f(\mathbf{x}_{i-1}), \quad i = 1, 2, \dots, N_s \quad (8)$$

其中, \mathbf{z}_i 是本轮迭代获取的梯度下降更新值, \mathbf{x}_{i-1} 表示上一轮的样本, 初始值 \mathbf{x}_0 可在离散星座空间 $\mathcal{Q}^{N_s \times 1}$ 中随机选取, 亦可采用计算复杂度较低的线性检测算法估计值作为初值. τ 代表梯度下降的学习率, N_s 代表最大迭代轮数, 而 $\nabla_{\mathbf{x}} f(\mathbf{x})$ 代表损失函数的梯度值:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = -\mathbf{H}^H (\mathbf{y} - \mathbf{H}\mathbf{x}) \quad (9)$$

式(8)给出了梯度下降的一般形式, 可根据需要选取不同梯度下降算法, 如牛顿法、Nesterov 加速梯度法等.

(2) 在式(8)所导出的更新值基础上, 引入随机扰动项, 并执行离散化处理, 以生成新的样本点:

$$\mathbf{z}_{\text{prop}} = \mathbf{z}_i + \mathbf{d} \quad (10)$$

$$\mathbf{x}_{\text{prop}} = \mathcal{Q}(\mathbf{z}_{\text{prop}}) \quad (11)$$

其中, \mathbf{d} 为高斯随机向量, 用于在梯度值 \mathbf{z}_i 的基础上进行随机游走以获得新的样本值. 随机游走能够避免采样器陷入梯度下降引导的局部最值中, 使得算法能够更为充分地探索整个状态空间. 为保证检测器的最终估计值 $\hat{\mathbf{x}}$ 满足 $\hat{\mathbf{x}} \in \mathcal{Q}^{N_s \times 1}$, 使用 $\mathcal{Q}(\cdot)$ 将 \mathbf{z}_{prop} 映射到离散星座集合 $\mathcal{Q}^{N_s \times 1}$ 上.

(3) 最后, 按一定的接受率 α 接受式(11)所得样本 \mathbf{x}_{prop} , 接受率 α 可表示为

$$\alpha = \min \left\{ 1, \frac{\exp\left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}_{\text{prop}}\|^2\right)}{\exp\left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}_{i-1}\|^2\right)} \right\} \quad (12)$$

其中, \mathbf{x}_{i-1} 代表前一轮所得样本值. 若接受 \mathbf{x}_{prop} , 则有新样本 $\mathbf{x}_i = \mathbf{x}_{\text{prop}}$; 反之若拒绝 \mathbf{x}_{prop} , 则新样本与上一轮所得样本保持一致, 即 $\mathbf{x}_i = \mathbf{x}_{i-1}$.

N_s 轮采样迭代结束后, 所有样本构成样本集 X , 选取 X 中最小化损失函数的元素作为检测估计值 $\hat{\mathbf{x}}$:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in X} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \quad (13)$$

3.2 模型驱动深度学习增强的 MCMC 检测

牛顿法基于目标函数的二阶导数, 展现出下降快、效率高的特性, 能够显著改善 MCMC 检测算法收敛性能. 鉴于此, 本文采用的 MCMC 检测算法将 MH 采样同牛顿梯度下降结合, 称为 MHGD 算法^[14]. MHGD 算法每轮迭代流程可表示为

$$\mathbf{z}_{\text{grad}} = \mathbf{x}_{i-1} + \tau \mathbf{M}_p \mathbf{H}^H \mathbf{r}_{i-1} \quad (14)$$

$$\mathbf{z}_{\text{prop}} = \mathbf{z}_{\text{grad}} + s \mathbf{M}_c \mathbf{v} \quad (15)$$

$$\mathbf{x}_{\text{prop}} = \mathcal{Q}(\mathbf{z}_{\text{prop}}) \quad (16)$$

$$\alpha = \frac{\exp\left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}_{\text{prop}}\|^2\right)}{\exp\left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}_{i-1}\|^2\right)} \quad (17)$$

其中, 式(14)基于牛顿法进行梯度下降, \mathbf{M}_p 为梯度预处理矩阵, 和目标函数海森矩阵的逆有关. 目标函数的海森矩阵可求解如下:

$$\nabla_{\mathbf{x}}^2 f(\mathbf{x}) = \mathbf{H}^H \mathbf{H} \quad (18)$$

基于式(18), 将梯度预处理矩阵定义为 $\mathbf{M}_p = (\mathbf{H}^H \mathbf{H} + \sigma^2/d_{\text{qam}} \mathbf{I})^{-1}$, d_{qam} 代表星座点之间最小距离的一半, $\sigma^2/d_{\text{qam}} \mathbf{I}$ 项的作用在于调整梯度下降的距离. 式(15)对应随机采样步骤, 其中 s 决定随机游走的步长, 取值为 $s = \max(d_{\text{qam}}, \|\mathbf{r}\|/\sqrt{N_r})$, 其中残差范数 $\|\mathbf{r}\| = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|$. 在采样迭代初期, 采样器距离最优解较远, 残差范数 $\|\mathbf{r}\|$ 较大, 使得采样器以较大的步长进行随机游走, 快速搜索样本空间; 当采样点接近最优解时, $\|\mathbf{r}\|/\sqrt{N_r}$ 项较小, 步长相应缩小, 以实现状态空间的精细搜索. d_{qam} 项保证步长不低于星座点间最小距离的一半, 避免算法因步长过小而陷入局部最优解. \mathbf{v} 代表均值为 0、方差为 1 的复高斯随机变量, \mathbf{M}_c 代表协方差矩阵, 定义为 $\mathbf{M}_c = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{T}$, \mathbf{T} 是对角阵, 其主对角线上的元素为 $(\mathbf{H}^H \mathbf{H})^{-1}$ 各列的列范数. 式(16)将随机采样的结果映射到离散星座集合 $\mathcal{Q}^{N_s \times 1}$ 上获得新样本 \mathbf{x}_{prop} , 随后按式(17)中的接受率 α 接受或拒绝 \mathbf{x}_{prop} .

MHGD 可同时部署 N_p 个并行采样器, 每个采样器独立采集 N_s 个样本, 共同构成样本集 X . 并行采样策略充分利用了计算资源, 可修改并行采样数以灵活调整算力分布. 现有研究表明, MHGD 算法在充分采样下可接近 ML 最优性能, 但采样数受限时性能恶化明显. 多轮采样迭代将引发极高的计算时延以及算力开销, 过度依赖采样数以提升性能将严重限制 MHGD 在低时延和资源受限环境中的应用. 此外, 算法参数整定主要依赖启发式策略, 容易导致性能优化存在不确定性和不稳定性, 对误码性能的进一步提升构成了瓶颈. 综上所述, 迫切需要开发一种先进的优化策略, 以增强 MHGD 算法在采样数受限下的误码性能. 此外, 该策略

应当能够系统地调整系统内的超参数,以实现更稳定和高效的性能表现. 因此引入模型驱动深度学习技术,将MHGD检测算法展开为多层级联网络结构,向每层引入可训练参数,可得到模型驱动增强的MHGD检测算法,称为MHGDNet. 图1展示了MHGDNet的网络结构,其中的可训练参数采用深度学习技术调节. 相较于启发式参数整定,深度学习基于大量数据集,减少了人工干预和主观判断,能够获得更优的参数配置. 在采样迭代开始前, N_p 个采样器需要分别确定采样初值. 为避

免LMMSE初始化的矩阵求逆运算,MHGDNet从星座集合 $Q^{N_s \times 1}$ 随机选取符号作为采样起始点. 采样迭代过程中,每一轮经过梯度下降、MH随机游走、离散化以及接受-拒绝环节以获得新的样本,该样本将作为下一轮梯度下降的起始位置. 相比于MHGD算法,MHGDNet在每轮采样引入两个可训练参数 τ_i 和 γ_i , τ_i 表示梯度下降的学习率,而 γ_i 用于调整随机游走采样的步长. 经 τ_i 调整后的梯度下降由式(19)给出:

$$\mathbf{z}_{\text{grad}} = \mathbf{x}_{i-1} + \tau_i \mathbf{M}_p \mathbf{H}^H \mathbf{r}_{i-1} \quad (19)$$

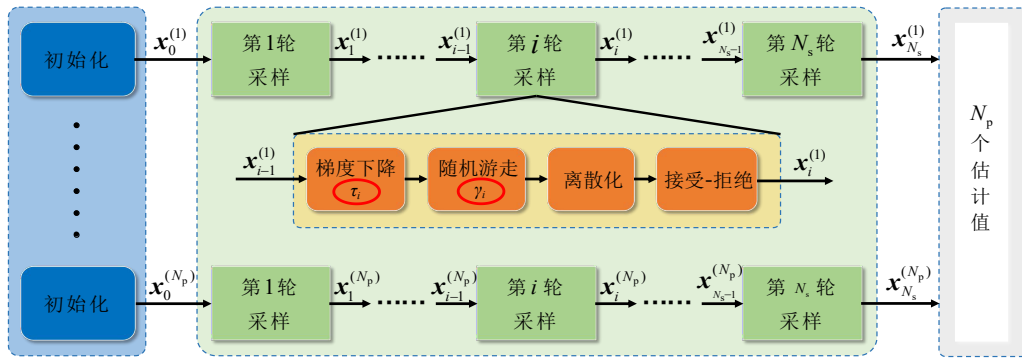


图1 MHGDNet网络结构

文献[14]指出,梯度下降环节的学习率 τ 与前一次采样样本有关,具体可表示为

$$\tau = \frac{\Re(\mathbf{r}_{\text{prev}}^H \mathbf{P} \mathbf{r}_{\text{prev}})}{\|\mathbf{P} \mathbf{r}_{\text{prev}}\|^2} \quad (20)$$

其中, $\Re(\cdot)$ 表示对复数取实部, $\mathbf{r}_{\text{prev}} = \mathbf{y} - \mathbf{H} \mathbf{x}_{\text{prev}}$ 是上一轮样本对应的残差范数,而 $\mathbf{P} = \mathbf{H} \mathbf{M}_p \mathbf{H}^H$. 由式(20)可知,每一轮采样后都需要重新计算学习率 τ ,引入大量矩阵乘法运算,在天线规模增大或采样数较多时将导致计算复杂度的显著提升. 式(19)中的模型驱动实现只需提前训练得到每轮采样对应的学习率 $\{\tau_i\}(i=1,2,\dots,N_s)$,并在算法部署时加载参数即可,避免了计算最优学习率导致的计算开销. 经 γ_i 修正的随机游走步长表示为

$$s_i = \max(\gamma_i d_{\text{qam}}, \|\mathbf{r}\| / \sqrt{N_r}) \quad (21)$$

未经 γ_i 修正的步长取值将 s 限制为始终大于 d_{qam} 以避免算法陷入局部最优,然而在迭代后期该设置易导致算法因步长过大而在最优解区域反复振荡. 通过引入 γ_i 优化,可达成如下效果:在采样初始阶段, γ_i 取值较大,采样器能够以较大的步长快速搜索状态空间,同时避免局部最优解;随着采样轮数增加, γ_i 趋于较小值,进一步限制采样器进行更加细致的搜索. N_p 个采样器使用相同的参数 $\{\tau_i, \gamma_i\}(i=1,2,\dots,N_s)$,独立并行采集 N_s 个样本. MHGDNet引入的参数数量仅与采样数有关,在迭

代轮数为 N_s 的网络中,参数总量为 $2N_s$,与系统规模及调制阶数无关. MHGDNet算法流程由算法1给出. 原始MHGD算法可视为MHGDNet的特殊形式. 当MHGDNet的可训练参数 τ_i 以及 γ_i 均取1时,检测算法每轮迭代的学习率与步长固定,MHGDNet检测退化为原始MHGD算法. 相较于MHGD在每轮迭代中采取固定的学习率与步长,MHGDNet能够动态调整参数 τ_i, γ_i 以适应复杂多变的传输场景.

为进一步提升系统误码性能,加入编解码环节,此时需要检测算法输出每一位比特对应的软信息,即对数似然比(Log Likelihood Ratio, LLR). 使用MCMC检测算法输出软信息,此时传输数据流 b 的第 k 位比特 b_k 对应的LLR计算公式为

$$\begin{aligned} \text{LLR}(b_k) &\approx \log \frac{\exp(-\|\mathbf{y} - \mathbf{H} \mathbf{x}_+\|^2 / \sigma^2)}{\exp(-\|\mathbf{y} - \mathbf{H} \mathbf{x}_-\|^2 / \sigma^2)} \\ &= \frac{1}{\sigma^2} (\|\mathbf{y} - \mathbf{H} \mathbf{x}_-\|^2 - \|\mathbf{y} - \mathbf{H} \mathbf{x}_+\|^2) \end{aligned} \quad (22)$$

其中, $\mathbf{x}_+ = \arg \min_{\mathbf{x} \in X_{k+}} \|\mathbf{y} - \mathbf{H} \mathbf{x}\|^2$, $\mathbf{x}_- = \arg \min_{\mathbf{x} \in X_{k-}} \|\mathbf{y} - \mathbf{H} \mathbf{x}\|^2$. 而 $X_{k+} = X \cap Q_{k+}^{N_s \times 1}$, $X_{k-} = X \cap Q_{k-}^{N_s \times 1}$,其中 $Q_{k+}^{N_s \times 1}$ 代表 $Q^{N_s \times 1}$ 的子集,该子集中所有符号对应的传输比特的第 k 位为1;而 $Q_{k-}^{N_s \times 1}$ 中的符号对应的传输比特的第 k 位为0. 在利用MCMC检测算法输出软信息时,仅有样本集中的元素参与计算,避免了因 $Q^{N_s \times 1}$ 中元素过多而导致高计算复杂度.

算法 1 MHGDNet 检测算法

输入: 梯度预置矩阵 \mathbf{M}_p , 协方差矩阵 \mathbf{M}_c , 检测器的初始估计值 \mathbf{x}_0 , 初始残差范数 $\|\mathbf{r}_0\|^2 = \|\mathbf{y} - \mathbf{H}\mathbf{x}_0\|^2$, 随机游走采样的步长 s , 梯度下降学习率 $\{\tau_i\}(i=0, 1, \dots, N_s)$, 随机游走步长调节参数 $\{\gamma_i\}(i=0, 1, \dots, N_s)$

输出: 检测器估计值 $\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$, \mathcal{X} 为 N_s 次采样获得的样本集 $\{\mathbf{x}_i\}(i=0, 1, \dots, N_s)$

FOR $t=1:N_s$

$\mathbf{w} \sim \text{CN}(\mathbf{0}_{N_r \times 1}, \mathbf{I}_{N_r})$

$\mathbf{z}_{\text{grad}} = \mathbf{x}_{t-1} + \tau_t \mathbf{M}_p \mathbf{H}^H \mathbf{r}_{t-1}$

$\mathbf{z}_{\text{prop}} = \mathbf{z}_{\text{grad}} + s \mathbf{M}_c \mathbf{v}_t$

$\mathbf{x}_{\text{prop}} = Q(\mathbf{z}_{\text{prop}})$

$\alpha = \min \left[\frac{\exp\left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}_{\text{prop}}\|^2\right)}{\exp\left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}_{t-1}\|^2\right)}, 1 \right]$

$u \sim U(0, 1)$

IF $\alpha \geq u$ then

接受 $\mathbf{x}_{\text{prop}}, \mathbf{x}_t = \mathbf{x}_{\text{prop}}$;

ELSE

拒绝 $\mathbf{x}_{\text{prop}}, \mathbf{x}_t = \mathbf{x}_{t-1}$;

END

END

MHGDNet 需要利用深度学习优化所引入参数, 然而迭代过程中需要把随机游走的结果离散化至 QAM 星座点, 这一步骤不可导, 无法直接使用深度学习技术对检测算法进行优化. 因此, 利用连续函数以近似离散化过程. 本文使用双曲正切函数的叠加来近似离散化过程, 近似函数构造如下:

$$g(x) = \frac{-(m-1) + \sum_{i=1}^{m-1} \tanh \left[\eta \cdot \left(x - \frac{2i-m}{\sqrt{E_s}} \right) \right]}{\sqrt{E_s}} + 1 \quad (23)$$

其中, E_s 代表平均符号能量, $m = \log_2 M$ 代表调制阶数. 为了保证近似的准确性, 双曲正切曲线应足够陡峭, 因此系数 η 应取尽可能大的值, 本文取 $\eta = 1000$. 对离散化操作进行近似处理后, 使用深度学习技术更新所引入的参数权重, 从而优化检测算法的误码性能. 模型驱动深度学习方案不改变算法原有的迭代结构, 且省略学习率 τ 的闭式计算, 复杂度相较于 MHGD 算法更低. 将离线训练的参数载入检测算法, 网络将以更合适的步长进行梯度下降与随机采样, 在更低的计算复杂度下获取性能增益.

3.3 复杂度分析

本节分析所提检测器的计算复杂度, 并与基线检测方案进行对比. 由于复数乘法对应的浮点数运算次数在信号检测运算中占据主导, 因而以单次检测复数乘法次数作为计算复杂度衡量指标. 为了简化分析, 假

设收发天线数相等, 即 $N_r = N_t = N$, T 代表期望传播算法 (Expectation Propagation, EP) 的迭代轮数, MHGD-LMMSE 表示采用 LMMSE 检测结果作为采样初值的 MHGD 算法. 表 1 对比了 MHGDNet 以及其他检测算法的计算复杂度. LMMSE 检测包含矩阵乘法 $\mathbf{H}^H \mathbf{H}$ 以及矩阵求逆, 对应计算复杂度分别为 N^3 和 $2N^3$, 故 LMMSE 检测算法总的计算复杂度可表示为 $3N^3 + O(N^2)$. EP 的计算复杂度与迭代轮数 T 正相关, 而 ML 最优检测的复杂度为指数级. MHGD 流程中 $O(N^3)$ 的计算复杂度主要来自以下步骤:

(1) 预处理矩阵 $\mathbf{M}_p = \left(\mathbf{H}^H \mathbf{H} + \sigma^2 / d_{\text{qam}} \mathbf{I} \right)^{-1}$, 包含矩阵乘法 $\mathbf{H}^H \mathbf{H}$ 以及矩阵求逆操作. 若 MHGD 检测算法采用 LMMSE 初始化, 则 $\mathbf{H}^H \mathbf{H}$ 已在 LMMSE 环节获得, \mathbf{M}_p 的计算复杂度只来自矩阵求逆步骤, 为 $2N^3 + O(N^2)$ ^[24]. 若采用随机初始化, 则需要另外计算 $\mathbf{H}^H \mathbf{H}$, 总复杂度为 $3N^3 + O(N^2)$.

(2) 科列斯基分解 (Cholesky decomposition) 计算协方差矩阵, 计算复杂度为 $2/3 N^3 + O(N^2)$.

(3) 式(20)用于确定每轮迭代的学习率, 需要计算矩阵 $\mathbf{P} = \mathbf{H} \mathbf{M}_p \mathbf{H}^H$, 计算复杂度为 $2N^3 + O(N^2)$.

MHGDNet 采用随机初始化, 避免了 LMMSE 初始化中矩阵求逆带来的计算复杂度; 同时通过直接加载离线训练好的参数 τ_i 以确定梯度下降的学习率, 省去了式(20)的运算, 因此 MHGDNet 在 $O(N^3)$ 级别的运算复杂度为 MHGD 算法的 64.71%, 仅为 MHGD-LMMSE 算法的 47.83%. 以上分析表明, 模型驱动深度学习通过离线训练权重, 显著降低了矩阵运算带来的复杂度.

表 1 计算复杂度对比

检测算法	计算复杂度
LMMSE	$3N^3 + O(N^2)$
EP	$(2T+1)N^3 + O(N^2)$
MHGD-LMMSE	$23/3 N^3 + O(N^2)$
MHGD	$17/3 N^3 + O(N^2)$
MHGDNet	$11/3 N^3 + O(N^2)$
ML	$O(M^N)$

3.4 仿真分析

仿真信道采用瑞利衰落信道, 并假定接收端能够获取完美信道状态信息. 编码采用 IEEE 802.11n 标准, 低密度奇偶校验码 (Low-Density Parity-Check codes, LDPC) 码长为 1944, 码率为 3/4, 解码器基于置信传播 (Belief Propagation, BP)^[25] 算法, 迭代轮数为 10 轮. 深度学习训练环节使用小批量梯度下降, 总数据集的大小为 10^5 , 每个轮次 (epoch) 中随机抽取 10^3 数量的样本进行训练, 同时每轮训练后使用验证集进行性能验证以遴选误码性能最优的网络参数, 验证集包含 10^4 个样本. 使用自适应矩

估计(Adaptive moment estimation, Adam)优化器,学习率设置为0.001,总训练轮数设置为1 500,即单次训练所需总样本数为 1.5×10^6 . 针对16 QAM与64 QAM调制分别训练网络参数,16 QAM下的训练信噪比(Signal-to-Noise Ratio, SNR)设置为25 dB,64 QAM下的训练信噪比选取为30 dB,获取参数后用于不同信噪比下的测试. 采取残差范数均值作为损失函数,即

$$\text{Loss} = \frac{1}{L} \sum_{i=0}^{L-1} \| \mathbf{y}_i - \mathbf{H}_i \hat{\mathbf{x}}_i \|^2 \quad (24)$$

其中, L 是一个小批次的样本数, \mathbf{y}_i 与 \mathbf{H}_i 来自训练集, $\hat{\mathbf{x}}_i$ 是本轮网络输出的估计值. 性能测试中,对于无编码系统,以误码率(Bit Error Rate, BER)为评估指标,测试停止条件为错误比特数超过 10^3 或传输总比特数超过 10^7 ;而对于编码系统,以误块率(BLOCK Error Rate, BLER)为评估指标,测试停止条件为错误码块超过10个或传输的总码块超过 10^4 . 除特殊说明外,MHGD检测算法默认从候选向量集 $Q^{N_s \times 1}$ 中随机选取元素作为采样初值. 特别地,采用LMMSE初始化的MHGD检测标记为MHGD-LMMSE. 使用元组 (N_p, N_s) 表示MHGD算法的采样配置,其中 N_p 代表MHGD检测算法并行采样器数, N_s 代表单个采样器的采样数,即MHGD检测算法的迭代轮数. 用于性能对比的基线算法包括LMMSE检测、期望传播^[26]、K-best^[27]、深度学习辅助的GEPNet检测^[17]以及最优的ML检测器,其中EP算法的迭代轮数设置为10.

无编码场景下,由检测算法直接输出估计值 $\hat{\mathbf{x}}$,映射至候选向量集 $Q^{N_s \times 1}$ 后直接获得解调比特. 图2展示了 8×8 MIMO、16 QAM调制系统中,在不同采样数下MHGDNet相较于MHGD的性能增强效果. 仿真结果表明,在低采样数下,如 $N_s = 2$ 和 $N_s = 4$,模型驱动深度学习能够显著提升MHGD检测算法的误码性能,取得接近一个数量级的误码性能增益. 随着采样数增加,性能增强效果逐渐减弱,在 $N_s = 32$ 时MHGDNet相较于原始的MHGD检测算法仅

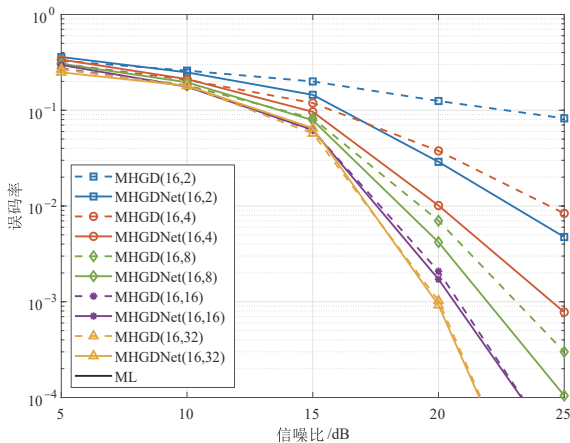
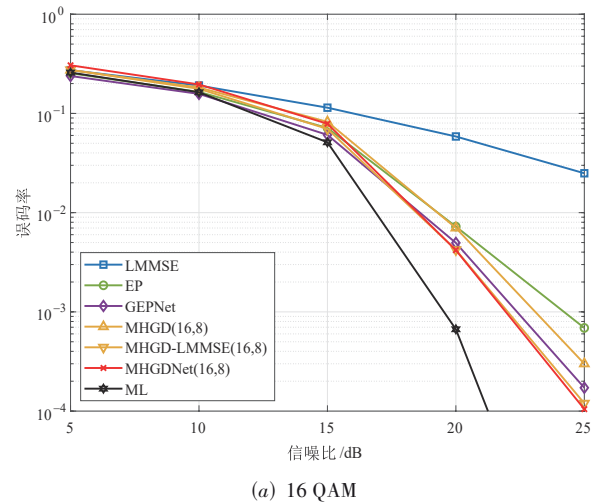


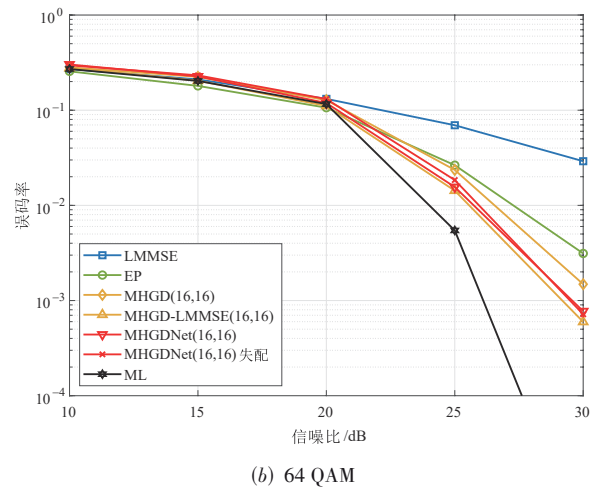
图2 8×8 MIMO、16 QAM调制系统,模型驱动在不同采样数下对MHGD检测算法的增强效果

有微小的性能增益. 这可解释为高采样数下MHGD已能较为充分地搜索状态空间,误码性能提升潜力相对有限. 注意到 $N_s = 32$ 下MHGD的误码性能相较于ML最优基线仅有不到0.2 dB的差距,这表明随着采样数的增加,MHGDNet能够收敛至ML最优性能.

图3展示了 8×8 MIMO系统中,不同调制阶数下MHGDNet与其他检测算法的误码性能对比. 在16 QAM系统中,所用的并行采样器数量 $N_p = 16$,采样数 $N_s = 8$,而64 QAM下采用 $N_p = N_s = 16$. 不同调制阶数下的仿真结果均表明,传统LMMSE检测方案误码性能较差,相较于其余检测算法有较大差距. 这是因为LMMSE检测算法在线性运算的约束下最小化均方误差,降低了运算复杂度,但是不可避免地牺牲了一定的误码性能. 在收发天线数较多的MIMO系统中,LMMSE性能损失明显^[1,17]. 图3(a)结果表明,在16 QAM系统中,MHGDNet的误码性能相较于



(a) 16 QAM



(b) 64 QAM

图3 8×8 MIMO、不同调制阶数下,MHGDNet与基线检测算法误码性能对比

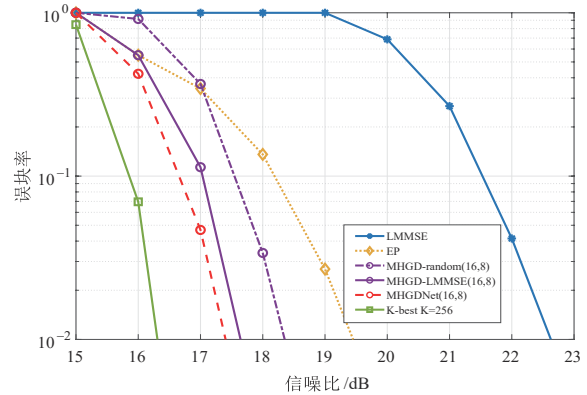
EP算法取得近一个数量级的优势,与MHGD-LMMSE相当.相近的误码性能下,MHGDNet省略了LMMSE初始化中的矩阵求逆步骤,能够更好地实现误码性能和计算复杂度的折中.和典型的深度学习辅助的MIMO检测算法GEPNet相比,MHGDNet同样能够取得性能优势.对比训练开销,训练GEPNet所需的总样本数为 6×10^7 ^[17],而训练MHGDNet所需总样本数仅为 1.5×10^6 .由此可见,MHGDNet在检测性能优于GEPNet的同时,在训练开销方面表现出显著优势.图3(b)表明,在64 QAM系统中MHGDNet相较于MHGD检测依然能取得显著性能增益,与MHGD-LMMSE相比仅有低于0.2 dB的性能差距.在64 QAM调制系统中,相较于16 QAM系统,模型驱动深度学习的性能增强效果有所减弱.这一现象可以归因于64 QAM系统中待搜索的样本空间更加庞大,算法遍历解空间所需的搜索量指数级增加.为考察MHGDNet对调制阶数的鲁棒性,将16 QAM下训练所得参数用于64 QAM系统仿真验证,对应误码曲线标记为“失配”.仿真结果表明,调制阶数失配造成的性能恶化微乎其微,特定调制阶数下所得参数能够较好地泛化至不同调制阶数的系统中.此外,尽管权重参数均在固定信噪比下训练获得,MHGDNet在各个信噪比上均有一定的性能增益,这表明所提模型驱动深度学习方案对调制阶数以及信噪比失配场景均有较好的鲁棒性.

图4展示了在 8×8 MIMO编码系统中所提MHGDNet的性能增益效果.由图4(a)可知,16 QAM下MHGDNet误码性能较随机初始化的MHGD提升近1 dB,较LMMSE初始化的MHGD提升近0.2 dB,性能增益显著.与LMMSE初始化优化采样出发点不同,所提模型驱动深度学习方案通过优化每一次梯度下降的距离与每一次采样的步长,能够提升样本集 X 的整体质量,从而在计算LLR时提供更加准确的置信信息.图4(b)展示了64 QAM下所提方案的增益效果.此时MHGDNet的误码性能与LMMSE初始化下的MHGD基本持平,优于随机初始化下的MHGD算法约0.6 dB.与图3(b)的结果类似,64 QAM调制方式下待搜索的样本空间更加庞大,模型驱动深度学习的优化效果相比于16 QAM的情况有所减弱,但依然能使MHGDNet在随机初始化下取得与LMMSE初始化相当的误码性能.

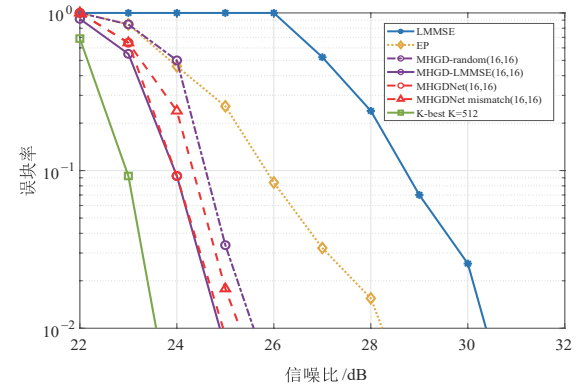
4 原型验证平台及结果分析

4.1 平台介绍

为了验证所提模型驱动深度学习对MCMC检测算法在实际信道环境中的增强效果,将算法部署至智能通信原型验证平台上进行空口传输测试.原型验证平台如图5所示,采用高性能AI平台进行AI模型训练推理以完成基带信号处理,采用软件无线电实现射频信号收发.高性能AI



(a) 16QAM



(b) 64QAM

图4 8×8 MIMO系统,1 944码长,3/4码率LDPC编码,不同调制阶数下MHGDNet与基线检测算法性能比较

平台采用NVIDIA Jetson AGX Orin,简称为Jetson.本文所用Jetson搭载2048核NVIDIA Ampere架构GPU,GPU主频最高1.3 GHz,每秒最多可进行275万亿次操作.软件无线电采用NI公司的USRP-2974,该设备基于LabView可重配置I/O架构,单机集成两个射频通道,最多可支持 2×2 MIMO收发功能.Jetson和USRP经路由器连接在同一以太网内,通过用户数据报协议(User Datagram Protocol,UDP)交换数据.图6展示了原型验证平台的工作流程.发射端Jetson平台进行LDPC编码、调制等预处理,产生频域传输符号,执行网络编码,打包为UDP报文并循环发送至路由器,同时在显示器上实时显示发送内容;USRP-2974循环侦听UDP包,解包后获得OFDM基带信号,经过I/Q双路调制,产生射频信号从天线发射,经由无线信道到达接收端;USRP-2974同时执行循环接收程序,将经过射频放大和基带信号预处理后的接收信号打包成UDP报文传送至路由器;接收端Jetson循环接收解包,执行网络解码,对解码所得数据报进行信道估计、信号检测、信道译码等操作,解调比特结果与原始发送比特进行比对,评估误码性能,实现对传输可行性和有效性的验证.

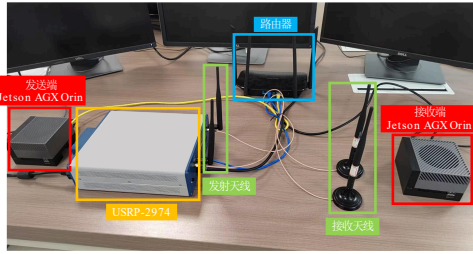


图5 2×2 MIMO智能通信原型验证平台

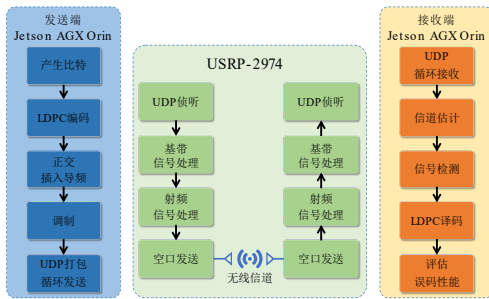


图6 信号传输流程图

图7展示了所传空口数据对应的帧结构。一帧完整的传输数据包括帧头的同步序列以及20个时隙。同步序列长度为600点,用于定时同步,接收端从空口接收数据后,利用序列相关获得发送数据的起始位置。每个时隙包含长度256点的OFDM符号以及64点的循环前缀(Cyclic Prefix, CP)。在每个OFDM符号中,150个子载波用于传输导频和有效数据符号,105个子载波用作保护带,1个子载波为直流偏移,而150个有效子载波由25个导频符号和125个数据符号组成。系统使用梳状导频,每个导频符号后跟随5个数据符号。将频域上的OFDM符号进行逆傅里叶变换得到时域数据,将时域数据的后64位复制并拼接到头部作为CP,即得到完整的单个时隙数据。准备完整的一帧传输数据后,USRP-2974将信号上变频至2 GHz载频并通过两路射频通道发送,所用传输带宽为350 kHz。

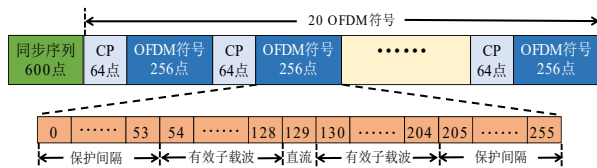


图7 传输数据帧结构

4.2 空口测试结果

进行空口传输测试时,系统采用16 QAM调制,EP迭代数 $T=10$,MHGD算法并行采样数 $N_p=16$,样本数 $N_s=8$,采用与仿真中相同的编解码配置。设置收发天线距离分别为0.5 m、0.7 m和1 m并进行信号传输测试,记录BLER性能,测试结果如表2所示。实测结果表明,实际信道环

境中,MHGDNet方案相较于LMMSE以及EP基线算法而言,表现出显著的性能优势,与仿真结果符合较好。同MHGD检测相比,MHGDNet在不同传输距离下均能够取得一定的性能增益,在13.5 dB处的BLER相较于MHGD降低约20%。这是因为MHGDNet能够基于空口传输数据,利用深度学习技术调节参数,从而在全新的环境下保持性能优势。由于实测系统的收发天线之间存在直直径,真实传输环境同系统模型所假设的信道模型并不匹配,因此模型驱动深度学习的增强效果相较于仿真场景有所减弱,其优势主要表现为计算复杂度的降低。MHGDNet利用可训练参数避免了每轮迭代中的学习率更新,在相近误码性能下,计算复杂度仅为MHGD算法的64.71%。这表明MHGDNet在实际传输中依然能够相较于原始算法更好地实现误码性能与计算复杂度的折中。

表2 原型验证平台空口测试结果

距离/m	检测算法	SNR=12.5 dB	SNR=13 dB	SNR=13.5 dB
0.5	LMMSE	1.0	1.0	0.846 1
	EP	0.846 1	0.407 4	4.602×10^{-2}
	MHGD	0.647 1	0.379 3	1.315×10^{-2}
	MHGDNet	0.647 1	0.366 7	1.08×10^{-2}
0.7	LMMSE	1.0	1.0	1.0
	EP	1.0	0.733 3	0.392 8
	MHGD	0.785 7	0.255 8	4.214×10^{-2}
	MHGDNet	0.785 7	0.234 0	3.416×10^{-2}
1	LMMSE	1.0	1.0	0.954 5
	EP	1.0	0.533 3	3.537×10^{-2}
	MHGD	0.687 5	0.323 5	1.047×10^{-2}
	MHGDNet	0.687 5	0.275 0	8.545×10^{-3}

5 结论

为提升现有MCMC检测算法在低采样数下的误码性能,本文提出了MHGDNet,该方案将原始MHGD检测算法的迭代过程展开为多层级联网络结构,向网络中引入可训练参数以代替部分复杂运算,并通过深度学习技术优化权重,从而以更低的复杂度提升算法误码性能。仿真结果表明,在不同调制阶数下,MHGDNet均能够明显改善MHGD检测算法在低采样数下的误码率。进一步将仿真测试扩展至编码场景,MHGDNet相较于MHGD检测算法依然取得约0.5 dB至1 dB的性能增益,且无需针对编码场景进行额外训练。此外,测试结果显示MHGDNet算法在各个信噪比上均能获得性能增益,表明模型驱动深度学习增强方案对信噪比不匹配的环境有较好的鲁棒性。将MHGDNet算法部署至智能通信原型验证平台进行实际测试,空口实测结果表明MHGDNet算法对真实信道环境具有优异的鲁棒性。

参考文献

- [1] YANG S S, HANZO L. Fifty years of MIMO detection: The Road to large-scale MIMOs[J]. *IEEE Communications Surveys & Tutorials*, 2015, 17(4): 1941-1988.
- [2] YOU X H, WANG C X, HUANG J, et al. Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts[J]. *Science China Information Sciences*, 2020, 64(1): 110301.
- [3] WANG C X, YOU X H, GAO X Q, et al. On the road to 6G: Visions, requirements, key technologies, and testbeds[J]. *IEEE Communications Surveys & Tutorials*, 2023, 25(2): 905-974.
- [4] ALBREEM M A, JUNTTI M, SHAHABUDDIN S. Massive MIMO detection techniques: A survey[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(4): 3109-3132.
- [5] HOCHWALD B M, BRINK S TEN. Achieving near-capacity on a multiple-antenna channel[J]. *IEEE Transactions on Communications*, 2003, 51(3): 389-399.
- [6] HE Y F, ZHANG J, JIN S, et al. Model-driven DNN decoder for turbo codes: Design, simulation, and experimental results[J]. *IEEE Transactions on Communications*, 2020, 68(10): 6127-6140.
- [7] JIN J J, LIANG X, XU Y H, et al. LDPC decoder based on Markov chain Monte Carlo method[C]//2018 IEEE Asia Pacific Conference on Circuits and Systems. Piscataway: IEEE, 2018: 219-222.
- [8] HEDSTROM J C, YUEN C H, CHEN R R, et al. Achieving near MAP performance with an excited Markov chain Monte Carlo MIMO detector[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(12): 7718-7732.
- [9] HUANG J T, KIM Y H. MCMC decoding of LDPC codes with BP preprocessing[C]//GLOBECOM 2020 - 2020 IEEE Global Communications Conference. Piscataway: IEEE, 2020: 1-5.
- [10] WELLING M, TEH Y W. Bayesian learning via stochastic gradient Langevin dynamics[C]//Proceedings of the 28th International Conference on Machine Learning (ICML-11). New York: ACM, 2011: 681-688.
- [11] ZHOU X Y, LIANG L, ZHANG J, et al. Decentralized massive MIMO detection using mini-batch gradient-based MCMC[C]//2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications. Piscataway: IEEE, 2024: 456-460.
- [12] WU Z W, LI H. Stochastic gradient Langevin dynamics for massive MIMO detection[J]. *IEEE Communications Letters*, 2022, 26(5): 1062-1065.
- [13] ZHOU X Y, LIANG L, ZHANG J, et al. Gradient-based Markov chain Monte Carlo for MIMO detection[J]. *IEEE Transactions on Wireless Communications*, 2024, 23(7): 7566-7581.
- [14] GOWDA N M, KRISHNAMURTHY S, BELOGOLOVY A. Metropolis-Hastings random walk along the gradient descent direction for MIMO detection[C]//ICC 2021 - IEEE International Conference on Communications. Piscataway: IEEE, 2021: 1-7.
- [15] QIN Z J, YE H, LI G Y, et al. Deep learning in physical layer communications[J]. *IEEE Wireless Communications*, 2019, 26(2): 93-99.
- [16] YE H, LIANG L, LI G Y, et al. Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(5): 3133-3143.
- [17] KOSASIH A, ONASIS V, MILOSLAVSKAYA V, et al. Graph neural network aided MU-MIMO detectors[J]. *IEEE Journal on Selected Areas in Communications*, 2022, 40(9): 2540-2555.
- [18] ZHOU X Y, ZHANG J, SYU C W, et al. Model-driven deep learning-based MIMO-OFDM detector: Design, simulation, and experimental results[J]. *IEEE Transactions on Communications*, 2022, 70(8): 5193-5207.
- [19] HU Q, GAO F F, ZHANG H, et al. Understanding deep MIMO detection[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(12): 9626-9639.
- [20] SAMUEL N, DISKIN T, WIESEL A. Deep MIMO detection[C]//2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications. Piscataway: IEEE, 2017: 1-5.
- [21] HE H T, WEN C K, JIN S, et al. Model-driven deep learning for MIMO detection[J]. *IEEE Transactions on Signal Processing*, 2020, 68: 1702-1715.
- [22] WANG X D, POOR H V. Iterative (turbo) soft interference cancellation and decoding for coded CDMA[J]. *IEEE Transactions on Communications*, 1999, 47(7): 1046-1061.
- [23] MA Y A, CHEN Y S, JIN C, et al. Sampling can be faster than optimization[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(42): 20881-20885.
- [24] MA L, DICKSON K, MCALLISTER J, et al. QR decomposition-based matrix inversion for high performance embedded MIMO receivers[J]. *IEEE Transactions on Signal Processing*, 2011, 59(4): 1858-1867.
- [25] GALLAGER R. Low-density parity-check codes[J]. *IRE Transactions on Information Theory*, 1962, 8(1): 21-28.

[26] CÉSPEDES J, OLMOS P M, SÁNCHEZ-FERNÁNDEZ M, et al. Expectation propagation detection for high-order high-dimensional MIMO systems[J]. IEEE Transactions on Communications, 2014, 62(8): 2840-2849.

[27] GUO Z, NILSSON P. Algorithm and implementation of the K-best sphere decoding for MIMO detection[J]. IEEE Journal on Selected Areas in Communications, 2006, 24(3): 491-503.

作者简介



曹益泉 男, 2024 年获得东南大学通信工程专业学士学位. 目前正在东南大学信息科学与工程学院攻读硕士学位. 主要研究方向为物理层通信、机器学习以及人工智能.

E-mail: 220240883@seu.edu.cn



周星宇 男, 2021 年获得南京邮电大学通信工程专业学士学位. 目前正在东南大学信息科学与工程学院攻读博士学位. 主要研究方向为基于机器学习的通信信号处理和收发器设计.

E-mail: xy_zhou@seu.edu.cn



张静 女, 2015 年获得中国矿业大学信息与通信工程专业学士学位, 并于 2018 年和 2022 年分别获得东南大学信息与通信工程专业硕士和博士学位. 目前在东南大学信息科学与工程学院担任讲师职务. 主要研究方向为物理层通信、人工智能以及收发器设计.

E-mail: jingzhang@seu.edu.cn



梁乐 男, 2012 年获得东南大学信息工程学士学位, 2015 年获得维多利亚大学电气工程硕士学位, 2018 年获得乔治亚理工学院电气与计算机工程博士学位. 目前在东南大学国家移动通信研究所工作. 主要研究方向为无线通信与机器学习. 中国电子学会会员编号: E190034275M.

E-mail: lliang@seu.edu.cn



李勇 男, 分别于 2009 年、2011 年和 2014 年在哈尔滨工业大学获得学士、硕士和博士学位. 目前在中国电子科技集团公司第 54 研究所先进通信网络国家重点实验室担任教授级高级工程师. 主要研究方向为无线人工智能通信、物理层安全和调制理论.

E-mail: young_li_54@126.com



金石 男, 于 1996 年在桂林电子科技大学获得通信工程学士学位, 2003 年在南京邮电大学获得硕士学位, 2007 年在东南大学获得信息与通信工程博士学位. 目前在东南大学国家移动通信研究所任教. 主要研究方向为无线通信、随机矩阵理论和信息论.