



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

基于稀疏和低秩优化的移动边缘人工智能研究

作者姓名: 杨恺

指导教师: 石远明 研究员 上海科技大学

学位类别: 工学博士

学科专业: 通信与信息系统

培养单位: 中国科学院上海微系统与信息技术研究所

2020 年 6 月

**Sparse and Low-Rank Optimization for Mobile Edge Artificial
Intelligence**

A dissertation submitted to the
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Doctor of Engineering
in **Communication and Information Systems**
By
Yang Kai
Supervisor: Professor Shi Yuanming

**Shanghai Institute of Microsystem and Information Technology,
Chinese Academy of Sciences**

June, 2020

中国科学院大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名: 杨怡
日 期: 2020.05.27

中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关保存和使用学位论文的规定，即中国科学院大学有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延期后适用本声明。

作者签名: 杨怡
日 期: 2020.05.27

导师签名: 石晓明
日 期: 2020.5.27

摘要

人工智能在从语音处理，图像分类到药物研发等众多领域中均取得了重大突破。这是由数据的爆炸性增长，机器学习技术的巨大进步以及计算资源愈来愈强大所造成的。特别地，边缘设备（例如，物联网设备）的大规模部署产生了海量的数据规模，这为获得精准模型并在网络边缘开发各种智能应用提供了机会。但由于变化的信道质量、流量拥塞和数据隐私等问题，致使无法将海量数据全部从终端设备发送到云端进行处理。移动边缘人工智能，即将人工智能模型的推理和训练过程推向移动边缘节点的方法，已经成为了一种极富前景的解决方案。它有望彻底改变未来的 6G 移动网络，实现从“万物互联”到“智慧互联”的范式转变。

移动边缘人工智能需要诸如智能手机和智能车辆之类的边缘设备与无线接入点和基站处的边缘服务器之间的紧密合作，而这给现有的无线通信系统带来了巨大压力。本论文利用多种使能技术在有限资源（如计算、存储、通信带宽、功耗等）的移动边缘实现人工智能模型的训练和推断，具体研究了以下四个方案：首先利用无线多址接入信道的信号叠加性质，提出了一种快速模型聚合方法来解决基于移动设备的分布式模型训练中的通信带宽受限的问题；然后，本文通过重新配置无线信号的传播环境，利用智能反射面技术来进一步提高分布式训练的模型聚合性能；接下来，本文研究了面向移动设备的低延迟分布式模型推理的无线分布式计算框架，为此，利用协作传输和干扰对齐技术设计了一种高通信效率的数据交换策略；最后，本文考虑了基于无线协作传输策略的基于计算任务分流的高能效边缘模型推理方案。

稀疏和低秩建模与优化技术广泛用于信号处理、高维统计、机器学习以及无线通信等应用中。为了设计与优化具有各种类型资源的复杂无线网络，本文研究了稀疏与低秩优化方法来增强移动边缘人工智能系统的性能，以解决由此产生的非凸优化问题、组合优化问题以及不确定性优化问题。本文的详细贡献总结如下：

1. **移动设备分布式训练的快速模型聚合策略：**出于延迟和数据隐私保护的考虑，在网络边缘执行模型训练而不将数据发送到集中式数据中心的方案变得

越来越有吸引力。然而，有限的通信带宽对移动边缘训练的模型聚合问题造成了严重的瓶颈。本文研究了模型聚合的计算结构和无线多址接入信道的信号叠加特性，然后开发了一种基于空中计算的快速模型聚合方法。这是通过设备选择和波束成形的联合设计来实现的，进而将其建模为一个稀疏与低秩优化问题，以支持高效的算法设计。大量的数值仿真结果验证了所提出模型聚合方法的算法优势和优异性能。

2. 智能反射面赋能的移动设备分布式训练快速模型聚合策略：为了克服诸如深度衰落之类的无线链路的不利信号传播条件，本文研究了使用新型智能反射面技术来实现可控的无线电环境，从而提高模型聚合的性能，以实现基于移动设备的分布式模型训练。这依赖于收发机、选择设备与智能反射面相移矩阵的联合设计，导致了稀疏目标函数与更为复杂的非凸双二次约束。为了解决这个问题，本文提出将交替最小化技术和稀疏与低秩优化相结合的方法。使用该方法后能明显从仿真实验结果中看到智能反射面对快速模型聚合的性能提升。

3. 基于移动设备分布式计算的边缘推断的快速数据交换策略：对于无法直接部署在单个设备上的大型模型，本文研究了基于无线分布式计算框架的移动设备端分布式推断系统。该系统的主要瓶颈之一为计算数据交换过程的巨额通信开销。为了解决这个问题，本文提出了一种利用协同传输和干扰对齐技术的快速数据交换方案，并将其建模为一个有仿射约束的低秩优化问题。本文提出了一种计算高效的 DC (difference-of-convex-functions) 算法，从而以较低的计算复杂度得到了满意的数据交换性能。

4. 基于计算任务卸载的边缘推断的高能效协作传输策略：对于具有大尺寸、高计算量的人工智能模型，本文研究了基于移动边缘计算中的计算分流方法的边缘推断解决方案。本文将每个推断任务推送到更多支持边缘计算的无线接入点，通过最小化计算功耗与传输功耗的总和，同时保证针对未知分布的信道状态信息误差的鲁棒性，提出了一种实现高能效处理和鲁棒的无线协作传输的方法。使用基于统计学习的鲁棒优化近似技术，将该问题重新表述为了一个具有非凸二次约束的组稀疏波束成形问题。本文结合迭代最小化加权算法和 DC 正则化方法，设计了一种组稀疏与低秩优化算法，能够在满足服务质量要求的鲁棒性的同时降低总功耗。

关键词：移动边缘人工智能，通信效率，稀疏优化，低秩优化

Abstract

Artificial intelligence (AI) has achieved remarkable breakthroughs in a wide range of fields, ranging from speech processing, image classification to drug discovery. This is driven by the explosive growth of data, advances in machine learning, and easy access to vastly powerful computing resources. Particularly, the wide scale deployment of edge devices (e.g., Internet-of-Things devices) generates an unprecedented scale of data, which provides the opportunity to derive accurate models and develop various intelligent applications at the network edge. However, such enormous data cannot be sent from end devices to the cloud for processing, due to the varying channel quality, traffic congestion and/or privacy concerns. By pushing inference and training processes of AI models to edge nodes, mobile edge AI has emerged as a promising alternative, which is envisioned to revolutionize the future 6G mobile networks and enable the paradigm shift from “connected things” to “connected intelligence”.

AI at mobile edge requires close cooperation among edge devices, such as smart phones and smart vehicles, and edge servers at the wireless access points and base stations, which however imposes enormous pressure on the existing wireless communication system. This dissertation develops a number of enabling technologies to realize AI training and inference at resources (such as computation, storage, communication bandwidth, and power) limited mobile edge, which is realized by the following four proposals. This dissertation first proposes a fast model aggregation approach to address the bottleneck of limited communication bandwidth for on-device distributed model training, by exploring the signal superposition property of a wireless multiple-access channel. This dissertation further leverages the novel intelligent reflecting surface (IRS) technology to further boost the performance of model aggregation, by reconfiguring the wireless propagation environment. Next, this dissertation studies the wireless distributed computing framework for low-latency on-device distributed model inference, for which a communication efficient data shuffling strategy is proposed via cooperative transmission and interference alignment techniques. Last but not the least, this dis-

sertation considers the wireless cooperative transmission strategy for energy-efficient computation offloading based edge model inference.

Sparse and low-rank modeling and optimization have been extensively used in a wide range of applications such as signal processing, high-dimensional statistics, machine learning, as well as wireless communications. To design and operate the complex wireless networks with various types of resources, this dissertation develops sparse and low-rank optimization methods for performance enhancements in mobile edge AI to address the resulting nonconvex, combinatorial, and/or even stochastic optimization problems. Detailed contributions of this dissertation are summarized as follows:

1. Fast model aggregation for on-device distributed model training: It is becoming increasingly attractive to perform model training at network edges without sending data to a centralized data center due to low-latency and privacy concerns. However, the limited communication bandwidth is a key bottleneck of model aggregation for edge training over radio channels. This dissertation studies the computation structure of model aggregation and the signal superposition property of a multi-access channel, followed by developing an over-the-air computation based fast model aggregation approach. This is achieved by joint device selection and beamforming design, which is modeled as a sparse and low-rank optimization problem to support efficient algorithms design. The algorithmic advantages and admirable performance of the proposed methodologies for model training are demonstrated through extensive numerical results.

2. IRS aided fast model aggregation for on-device distributed model training: To avoid the unfavorable signal propagation conditions of wireless links such as deep fading, this dissertation proposes to adopt IRS to develop a smart radio environment, thereby enhancing the performance of model aggregation for on-device distributed model training. It relies on the joint design of transceivers, selected devices and phase shift matrix of IRS, which results in sparse objective and more complicated non-convex bi-quadratic constraints. To address the problem, this dissertation proposes to use sparse and low-rank optimization together with alternating minimization technique, with which the advantages of IRS are demonstrated for fast model aggregation through numerical experiments.

3. Fast data shuffling of wireless distributed computing framework for on-device distributed model inference: For large models that cannot be directly deployed on a single device, this dissertation studies the on-device distributed model inference system based on wireless distributed computing framework. In such system, the communication overhead for information exchange becomes one of the main bottlenecks. To address this issue, this dissertation proposes a fast data shuffling scheme by leveraging cooperative transmission and interference alignment techniques, which is modeled as a low-rank optimization problem with affine constraints. This dissertation proposes a computation efficient DC (difference-of-convex-functions) algorithm to enjoy performance improvement and low computational complexity.

4. Energy-efficient processing and robust wireless cooperative transmission for computation offloading based model inference: For computationally expensive AI models with large model size, this dissertation studies a solution of computation offloading method in mobile edge computing for edge inference. By pushing each inference task to more edge computing enabled wireless access points, an energy-efficient processing and robust wireless cooperative transmission approach is proposed to minimize the total power consumption while guaranteeing the robustness against channel state information errors with unknown distribution. By adopting a statistical learning based robust optimization approximation, the problem is reformulated as a group sparse beamforming problem with nonconvex quadratic constraints. This dissertation provides a group sparse and low-rank optimization algorithm by iteratively reweighted minimization and DC regularization, which is able to achieve the lowest power consumption while meeting the robustness of quality-of-service requirements.

Keywords: Mobile edge artificial intelligence, communication efficiency, sparse optimization, low-rank optimization

目 录

第 1 章 绪论	1
1.1 立题背景	1
1.2 研究挑战与现状	5
1.2.1 移动边缘人工智能面临挑战	5
1.2.2 边缘训练研究现状	6
1.2.3 边缘推断研究现状	7
1.2.4 研究内容与方法	9
1.3 全文架构	12
第 2 章 稀疏优化与低秩优化	15
2.1 稀疏性与低秩性	15
2.2 稀疏优化方法	16
2.2.1 凸松弛方法	17
2.2.2 迭代加权算法	17
2.2.3 组稀疏优化方法	18
2.3 低秩优化方法	19
2.3.1 核范数松弛	19
2.3.2 迭代加权算法	20
2.3.3 其他算法	21
2.4 本章小结	21
第 3 章 移动设备分布式训练的快速模型聚合策略	23
3.1 引言	23
3.2 移动设备分布式训练系统的快速模型聚合	24
3.2.1 移动设备分布式训练系统模型	24
3.2.2 基于空中计算的快速模型聚合方案	27
3.2.3 问题表述	29
3.3 稀疏与低秩优化建模方法	30
3.3.1 问题建模	30
3.3.2 问题分析	32
3.4 稀疏与低秩优化的 DC 算法设计	32
3.4.1 稀疏与低秩函数的 DC 表示	33
3.4.2 统一的 DC 表示框架	34

3.4.3 DC 算法及其复杂度与收敛性分析 ······	35
3.5 仿真结果分析 ······	39
3.5.1 可行性检测 ······	40
3.5.2 所选设备数目与目标均方差大小的关系 ······	41
3.5.3 所提 DC 方法对于联邦学习的性能 ······	42
3.6 本章小结 ······	43
第 4 章 智能反射面赋能的移动设备分布式训练快速模型聚合策略 ······	45
4.1 引言 ······	45
4.2 系统模型与问题表述 ······	46
4.2.1 智能反射面赋能的联邦学习系统 ······	46
4.2.2 问题表述 ······	48
4.3 稀疏与低秩优化算法框架 ······	49
4.3.1 解决稀疏目标函数的两步骤算法框架 ······	49
4.3.2 解决非凸约束的交替低秩优化方法 ······	50
4.3.3 解决低秩约束的 DC 算法 ······	52
4.4 仿真验证与结果分析 ······	55
4.4.1 可行性检测性能评估 ······	55
4.4.2 设备选择性能评估 ······	56
4.4.3 联邦学习性能评估 ······	57
4.5 本章小结 ······	58
第 5 章 基于移动设备分布式计算的边缘推断的快速数据交换策略 ······	59
5.1 引言 ······	59
5.2 基于无线分布式计算的分布式推断系统的数据交换 ······	60
5.2.1 计算模型 ······	60
5.2.2 通信模型 ······	63
5.2.3 可达数据速率与自由度 ······	65
5.3 基于干扰消除技术的低秩优化建模方法 ······	66
5.3.1 干扰对齐条件 ······	66
5.3.2 低秩优化方法 ······	67
5.3.3 问题分析 ······	69
5.4 高效低秩优化 DC 算法提出 ······	71
5.4.1 DC 方法 ······	71
5.4.2 秩函数的一种新的 DC 表示 ······	72

5.4.3 高效 DC 算法	73
5.4.4 计算复杂度与收敛性分析.....	75
5.5 仿真验证与结果分析	76
5.5.1 收敛速度与时间	77
5.5.2 可达自由度与本地存储大小的关系评估	77
5.5.3 可达自由度与天线数目的关系评估	79
5.5.4 可达自由度与移动用户数目的关系评估	79
5.6 本章小结	80
第 6 章 基于计算任务卸载的边缘推断的高能效协作传输策略 .	81
6.1 引言	81
6.2 系统模型与问题表述	84
6.2.1 系统模型	84
6.2.2 功耗模型	86
6.2.3 信道信息不确定性模型.....	87
6.2.4 问题表述	88
6.2.5 问题分析	89
6.3 基于统计学习的联合概率约束的鲁棒优化近似	90
6.3.1 近似联合概率约束的鲁棒优化方法	90
6.3.2 从数据样本中学习高概率区域	91
6.3.3 重新表述鲁棒优化为易处理形式	94
6.3.4 与鲁棒优化近似方法相结合的一种低成本采样策略	95
6.4 用于求解有非凸二次约束的组稀疏波束成形问题的加权功率最小化 方法	96
6.4.1 用于解决非凸二次约束的矩阵升维技术	96
6.4.2 秩为 1 约束的 DC 表示方法.....	97
6.4.3 加权 ℓ_1 最小化方法诱导组稀疏性	98
6.4.4 所提的加权功率最小化方法	98
6.5 仿真结果与分析	100
6.5.1 将信道状态信息不确定性纳入考虑的优势	100
6.5.2 克服想定生成方法的过于保守特性	101
6.5.3 收敛行为	102
6.5.4 总功耗与目标信干噪比的关系	103
6.6 本章小结	105
第 7 章 总结与展望	107
7.1 全文总结	107
7.2 未来研究方向展望	108

附录 A 重要术语中英文及缩写对照表	111
附录 B 公式推导与证明	113
B.1 式 (6.35) 与式 (6.36) 的推导	113
B.2 DC 算法收敛性证明	114
参考文献	117
作者简历及攻读学位期间发表的学术论文与研究成果	127
致谢	131

图形列表

1.1 边缘人工智能应用与系统。	4
1.2 全文组织架构示意图。	12
3.1 移动设备分布式联邦学习模型训练系统。	25
3.2 在选择不同数目的移动设备情况下，使用 FedAvg 算法得到的训练损失函数值与预测准确度。(a) 训练损失函数值，(b) 相对预测准确度。	26
3.3 空中计算原理图。	29
3.4 用于设备选择的两步骤算法框架。	34
3.5 所提出的 DC 算法的收敛。	40
3.6 不同算法下可行性的概率。	41
3.7 所提的 DC 方法的可行性概率与基站端天线数的关系。	42
3.8 不同算法的平均选择设备数目。	43
3.9 (a) FedAvg 算法在不同的设备选择算法下的收敛结果，(b) 通信轮数与所训练的全局模型在测试集的相对准确度的关系。	43
4.1 智能反射面赋能的移动设备分布式联邦学习系统。	47
4.2 所提算法的检测可行性的性能评估。(a) 对比半正定松弛算法 (SDR) 的可行性的概率，(b) 所提的 DC 方法的可行性概率与基站端天线数的关系。	56
4.3 平均选择设备数目与均方误差关系曲线图。	57
4.4 所提智能反射面赋能分布式训练系统模型聚合算法与其他方法性能对比。(a) FedAvg 算法在不同的设备选择算法下的收敛结果，(b) 通信轮数与所训练的全局模型在测试集的相对准确度的关系。	58
5.1 无线分布式计算系统。	61
5.2 用于设备端分布式推断的分布式计算模型。	62
5.3 不同算法的收敛迭代次数与时间。	78
5.4 最大可达对称自由度与每个移动用户本地存储大小 μ 的关系。	78
5.5 最大可达对称自由度与天线数目的关系。移动用户和接入点具有同样数目的天线。	79
5.6 不同算法可达自由度与移动用户数目的关系。	80
6.1 基于边缘服务器推断的系统模型。	84
6.2 AlexNet 能耗分解图。	87
6.3 低成本信道采样方法的时间线示意图。	96

6.4 使用想定生成方法和鲁棒优化近似方法返回可行解的几率与目标信 干噪比 γ 的关系。	103
6.5 所提加权功率最小化方法在不同目标信干噪比 γ 下的收敛行为。 ...	104
6.6 不同目标信干噪比 γ 下在所有边缘服务器处执行推断任务总数的轨 迹。	104
6.7 所提算法的性能评估。(a) 总功耗与目标信干噪比的关系, (b) 边缘服 务器处执行的总任务数目与目标信干噪比的关系。	105

表格列表

6.1 服务质量满足的测试次数	102
-----------------------	-----

符号列表

x	标量
\mathbf{x}	向量
\mathbf{X}	矩阵
$\mathbf{x} \geq 0$	向量每个元素均为非负
$\mathbf{X} \succeq 0$	矩阵为半正定矩阵
$\mathbb{R}(\mathbb{C})^n$	n 维实（复）数向量空间
$\mathbb{R}(\mathbb{C})^{m \times n}$	$m \times n$ 维实（复）数矩阵空间
\mathbb{R}_+^n	n 维非负向量空间
\mathbb{S}_+^n	$n \times n$ 维对称半正定矩阵空间
$(\cdot)^*$	向量、矩阵或者函数的共轭
$(\cdot)^T$	向量或矩阵的转置
$(\cdot)^H$	向量或矩阵的共轭转置
$ x $	标量绝对值或幅值
$ \mathcal{S} $	集合的基数
$\ \mathbf{x}\ _0$	向量的零范数
$\ \mathbf{x}\ _1$	向量的 ℓ_1 范数
$\ \mathbf{x}\ _2$	向量的 ℓ_2 范数
$\ \mathbf{x}\ _\infty$	向量的无穷范数
$\ \mathbf{x}\ _k$	向量的 Ky Fan k 范数
$\ \mathbf{X}\ _*$	矩阵的核范数，又称迹范数
$\ \mathbf{X}\ _2$	矩阵的谱范数
$\ \cdot\ _F$	矩阵的 Frobenius 范数
$\ \cdot\ _{Sp}$	矩阵的 Schatten-p 范数
$\ \mathbf{X}\ _k$	矩阵的 Ky Fan k 范数
$\ \mathbf{X}\ _{2,k}$	矩阵的 Ky Fan $2 - k$ 范数

$\det(\cdot)$	矩阵的行列式
$\text{Tr}(\cdot)$	矩阵的迹
$(\cdot)^{-1}$	取逆
$(\cdot)^+$	伪逆
$\text{rank}(\cdot)$	矩阵的秩
$\text{diag}(\dots)$	以括号中数为对角元素的对角阵
$\mathbb{E}(\cdot)$	求取期望
$\partial(\cdot)$	函数的次微分或次梯度
$\langle \cdot, \cdot \rangle$	内积
$\text{Real}(\cdot)$	取实部操作
$[K]$	集合 $\{1, 2, \dots, K\}$
$\binom{n}{k}$	组合数
\otimes	矩阵克罗内克 (Kronecker) 乘积
\boldsymbol{I}_K	K 维单位阵
$\min\{m, n\}$	取 m 和 n 中的最小值
I_C	指示函数, 条件 C 满足时为 1, 否则为 0
$\mathcal{CN}(\mu, \sigma^2)$	均值为 μ 、方差为 σ^2 的循环对称复高斯分布
$\Pr(\cdot)$	事件发生的概率

第1章 绪论

1.1 立题背景

移动通信网络在过去的几十年里保持了高速的增长趋势，经历了引人注目的发展与改变。移动通信系统经历了数代 (generation, G) 的演进 (Sharma, 2013)，从第一代 (first generation, 1G) 到正在发展的第五代 (fifth generation, 5G) 在速度、技术、数据容量、延迟等多个维度有了巨大的提升。1G 移动无线通信网诞生于 20 世纪 80 年代初，使用的是模拟通信技术来进行语音通话。为给服务更多的用户，第二代 (second generation, 2G) 移动通信网络使用了数字编解码技术，并采用了如 FDMA (frequency division multiple access, 频分多址接入)、TDMA (time division multiple access 时分多址接入)、CDMA (code division multiple access, 码分多址接入) 等多址接入技术，在语音通话之外支持了短信和数据服务，提供了更广泛的通信覆盖。这一代的成员包括了 GSM (global system for mobile communications, 即俗称的 2G)、GPRS (General Packet Radio Service, 俗称 2.5G) 和 EDGE (enhanced data rate for GSM evolution, 俗称 2.75G)。第三代 (third generation, 3G) 移动通信网络开启了移动宽带的体验，提供了更高的传输速率，更大的通信容量，且支持了多媒体。3G 主要包括了三种码分多址技术：CDMA2000、TD-SCDMA (time-division synchronous code division multiple access) 和 W-CDMA (wideband code division multiple access)。第四代 (fourth generation, 4G) 移动通信与固定互联网相集成，支持了无线移动互联网。移动互联网使得大量新兴应用兴起，如移动购物、移动支付、智能家居、移动游戏等，而这也同时成为了促进无线通信技术进步的主要动力。正在到来的 5G 网络将支持增强的移动宽带 (enhanced mobile broadband, eMBB)，超高可靠性和超低延迟 (ultra-reliable and low-latency communications, uRLLC)，以及大规模机器类型通信 (massive machine-type communications, mMTC)。

互联网数据的急剧增加、算力的不断提高以及先进算法的研究开发促使了人工智能 (artificial intelligence, AI) 技术不断取得突破性进展，如语音、文字与图像处理、控制系统等技术，从而在现代社会的各行各业以及日常生活中扮演了越来越重要的角色，其辐射范围涵盖了社交网络、电子商务、远程教育、智慧

医疗、自动驾驶等等。据Bughin 和 Seong (2018) 报告，人工智能预计到 2030 年能给全球国内生产总值（global domestic product, GDP）带来相比于 2018 年的大约 16% 的增长，高达约 13 万亿美元。而随着移动互联网的高速发展，大量的移动设备和移动数据正在产生，无线大数据的时代正在到来 (Bi 等, 2015)。无线通信网络已经不仅仅承载着满足用户通信需求的使命，它既是移动数据传输的纽带，也是产生依赖于大量数据的移动应用的动力。无线通信网络的不断变革也促使了百花齐放的移动设备的出现与交互、新颖的交互方式的产生以及许多新兴商业模式的开创与繁荣。据Cisco (2020) 报告，2018 年全球有 88 亿移动设备与连接，而这一数字到 2023 年预计将增加至 131 亿。而Cisco (2018) 中预测，截止 2021 年所有人、机器与物联网设备产生的数据将高达 85ZB (Zettabyte, 10^{21} bytes)，甚至超过了云计算中心的数据流量的四倍以上。新兴的移动智能应用对于低延迟与数据隐私的要求日益严苛，如自动驾驶车辆、无人机、智能机器人等。传统的基于云计算中心的人工智能需要从大量设备处收集数据至云数据中心 (Sze 等, 2017)，而带宽、时延等网络状况往往处于波动之中，这使得基于云的人工智能面临前所未有的压力。人工智能服务的隐私与安全被提上日程，各国纷纷制定或正在制定更加严格的法律法规来要求服务提供商保护用户的数据隐私。如欧盟颁布了 GDPR (general data protection regulation) 条例 (euG, 2016)，授予用户删除或撤回自己数据的权利。为了解决这些问题，边缘人工智能 (edge AI) (Zhou 等, 2019; Park 等, 2019b; Shi 等, 2020) 应运而生。

边缘人工智能，即将人工智能的训练 (training) 和推断 (inference) 过程从云计算中心推送至更加接近边缘用户的网络边缘。这样可以降低到云端的大量数据传输，减轻网络流量负荷，提高数据隐私保护能力。5G 时代万物互联将推至顶峰，尽管其仍然在初始阶段，学界和业界已经开始展望了下一代无线通信系统 (David 和 Berndt, 2018; Letaief 等, 2019) 的发展蓝图、赋能技术与架构，寄望于将人工智能嵌入无线网络，导致无线通信系统产生从万物互联 (connected things) 到智慧互联 (connected intelligence) 的深刻变革。未来无线网络系统由海量的边缘节点 (edge node) 组成，包括基站、无线接入点等处的边缘服务器 (edge servers) 以及智能手机、车辆、无人机、可穿戴设备等边缘设备 (edge devices)。一般来说，训练人工智能模型需要大量计算资源的支持，特别是对于深度神经网络模型来说。可喜的是，移动边缘计算 (mobile edge computing) (Mao 等, 2017) 的

高速发展使得在移动网络边缘提供云计算能力成为了可能，使得利用边缘节点可以高效地执行人工智能任务。近年来，边缘节点计算能力的持续提升。如华为的麒麟 990 5G 芯片与苹果的 A12 芯片都集成了神经处理单元（neural processing unit, NPU），能够极大程度的加速边缘设备端的人工智能运算。总之，移动边缘计算能力的进步使得移动边缘人工智能成为了一个可行方案，吸引了许多相关研究，成为了智慧互联赋能的下一代无线通信系统的主要研究方向之一。

在移动边缘执行人工智能任务并非易事，绝非仅采用与云计算中心相同的计算与通信技术即可实现。实现移动边缘人工智能最直接的的方案是把人工智能任务直接在移动设备端实现，但是终端设备的计算、存储、电量等资源往往十分有限，使得基于单个移动设备的人工智能方案往往是不可行的。例如，用于计算机视觉的经典卷积神经网络 Alexnet 就具有超过六千万个参数。如图1所示，极具前景的解决方案是设计多个边缘节点的协作机制，整合大量节点的资源来完成需要密集计算与大量存储的人工智能任务。一个典型的移动边缘人工智能训练方案是基于移动设备的联邦学习 (Konečný 等, 2015)，利用多个移动设备的存储、计算与通信资源来共同训练一个人工智能模型。每个移动设备只需根据本地的数据计算出一个本地模型，然后将结果发送给一个融合中心聚合得到全局的模型，然后发送回给每个设备，并不断重复上述过程得到最后训练好的人工智能模型。这其中需要移动设备与融合中心不断交换数据，带来了很大的通信开销，对无线通信系统。此外，一些人工智能推断任务存储占用高、计算任务重，将其推送至具有更强计算能力的边缘服务器是一个好的解决方案，但是同时也会带来较大通信开销。因此，实现移动边缘人工智能还面临着许多挑战，许多问题亟待研究。

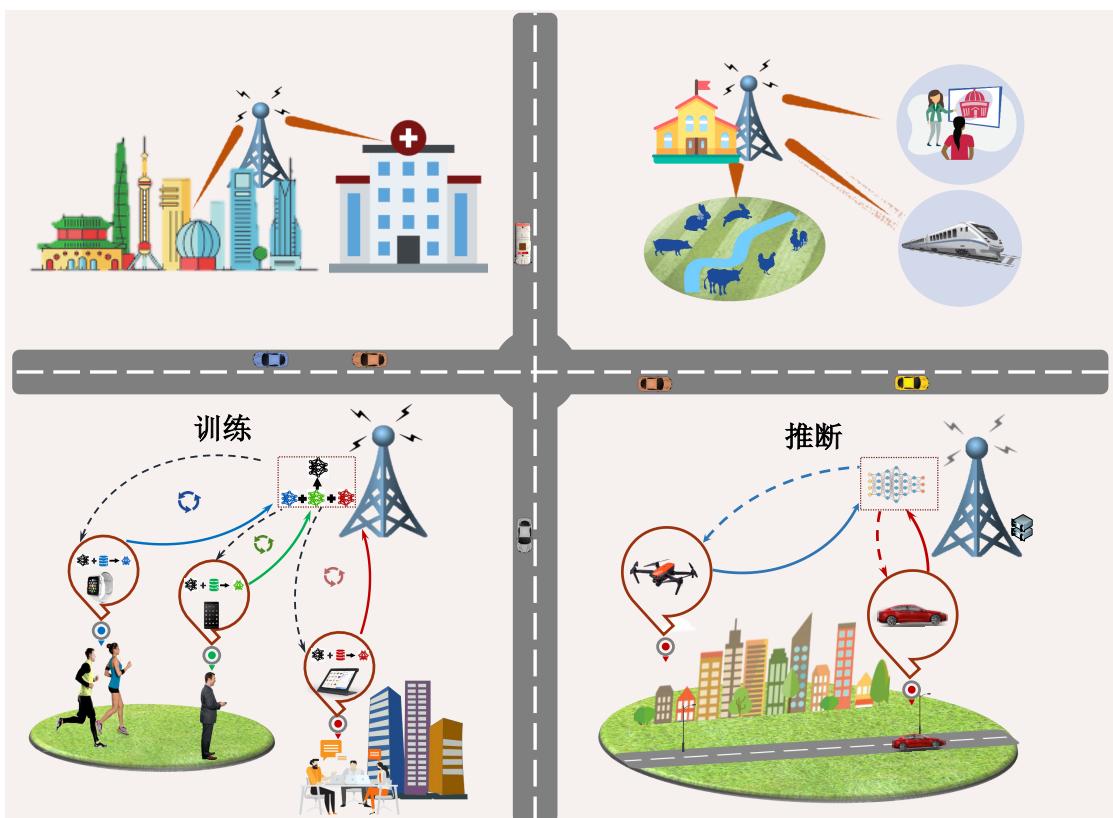


图 1.1 边缘人工智能应用与系统。

Figure 1.1 Applications and systems of mobile edge artificial intelligence.

1.2 研究挑战与现状

1.2.1 移动边缘人工智能面临挑战

一般来说，提供人工智能服务的典型过程包括了从数据中训练出一个机器学习模型，以及根据训练好的模型做推断。机器学习模型的性能可以用模型准确度来衡量，可以通过收集更多的有效数据来提升性能。但是使用海量数据训练机器学习模型十分耗时，一般需要使用分布式的架构，而这会带来额外的通信开销。计算与通信开销随着模型大小急剧增长。在云计算中心中，云计算服务器之间是通过超高带宽网络连接，而且训练数据和模型对于所有节点往往都是可以直接访问的。与此截然不同的是，移动边缘人工智能面临着更为严苛的限制条件：

- **资源受限：**云计算中心使用的是集成了强大的中心处理单元（central processing units, CPU）与图形处理单元（graphic processing units），而移动边缘设备往往只有很有限的计算、存储与电量资源。且大量移动设备与基站或无线接入点处的边缘服务器之间的链路只有有限的频谱资源。例如，用于计算机视觉的经典深度神经网络 AlexNet ([Krizhevsky 等, 2012](#)) 有着超过了六千万个参数。商汤科技的云计算数据中心中使用 512 块 Volta GPU，它们之间通过高达 56Gbps 速率的链路相连，结果破纪录的用了 2.6 分钟训练好了一个 Alexnet 模型 ([Sun 等, 2019](#))。Volta GPUs 是世界上最强大的 GPU 之一，具有 5120 块核心。然而，作为最强大的智能手机之一，华为 Mate 30 Pro 的 Mali-G76 GPU 只有 16 个核。而 5G 的理论最大速率是 10Gbps，平均速率只有 50Mbps。

- **系统异构：**不同移动设备的硬件、网络状况与功耗预算往往差异很大，导致了执行人工智能任务时通信、计算、存储与功耗的异构性，而边缘服务器有着比移动设备更为强大的计算、存储与功耗资源。例如，苹果智能手表 Apple Watch Series 5 的电量仅能支持 10 个小时的音乐播放，这导致移动用户可能只希望在设备处于充电状态时参与执行人工智能训练任务。更糟糕的是，一些连接到计量蜂窝网络的移动设备不愿与其他边缘节点交换信息。

- **隐私与安全限制：**对于一些不断涌现的高风险智能物联网应用来说，人工智能服务的隐私与安全变得非常重要。许多国家与地区纷纷制定了严格的数据隐私保护条例，如欧盟的 GDPR。联邦学习 ([Konečný 等, 2015](#)) 正成为一种极具前景的能够在保护数据隐私的同时合作训练机器学习模型的框架。许多文献

(Chen 等, 2017; Dong 等, 2019) 从鲁棒的算法与系统设计等不同角度设计了分布式的边缘训练方案。

实现高效的移动边缘人工智能是一项极具挑战性的任务，需要协调调度边缘节点来高效地完成训练或者推断任务，且满足各种各样的物理限制与法律法规限制。为了实现这一目标，往往需要联合设计计算、通信与系统方案。基于边缘节点的人工智能方案带来了很大通信开销与功耗，对无线网络系统的通信方式与系统架构提出了挑战。

1.2.2 边缘训练研究现状

在移动网络边缘执行训练任务非常富有挑战性，需要协调大量的边缘节点来共同建立一个人工智能模型。每一个节点往往只有全体训练数据的一部分，而这是移动边缘训练与在云计算中心做训练任务的根本不同点 (Yang 等, 2019b)。边缘节点之间的信息交互带来了巨大的通信开销，特别是由于无线环境下仅有有限的频谱资源，这成为了边缘训练的一大瓶颈。对于模型训练来说，往往需要使用迭代式的算法求解，从而导致需要边缘节点之间进行多轮的信息交互。故为了在资源受限的移动边缘执行人工智能训练任务，需要考虑如何能够降低通信次数、减少每次通信开销。一方面，可以通过提高收敛速度来降低训练模型的通信次数。分布式训练的经典算法是一阶的随机梯度下降法 (stochastic gradient descent)。为了降低其所需的通信次数，文献Yuan 等 (2018); Lee 等 (2017) 使用了缩减方差的梯度算法。另一方面，可以采用梯度的重复使用 (Chen 等, 2018)、量化 (Alistarh 等, 2017) 与稀疏化 (Chen 等, 2018) 等方法来降低每次通信所需要的通信带宽。除此之外，使用（近似的）二阶信息也可以提高算法的收敛速度，如随机的拟牛顿法 (quasi-Newton method) (Byrd 等, 2016)、改良的近似牛顿法 (Wang 等, 2018c) 等。有趣的是，有学者重新回顾了通信理论，发现了能够降低边缘训练通信开销的一些方法。在无线通信系统中，从接收机到发射机的有限反馈 (Love 等, 2008) 对于降低信道自适应传输方法所需要反馈的发射端信道信息比特数十分关键。文献Du 等 (2019) 建立了无线通信中的有限反馈与边缘训练中降低传输开销之间的联系，设计了一种高效的 Grassmannian 量化方法用于压缩训练中所需要的高维梯度信息。

为了保护用户数据的隐私，近年来学者提出了联邦学习 (federated learning) (McMahan 等, 2017; Smith 等, 2017; Bonawitz 等, 2019; Yang 等, 2019b) 的概念，即

在数据留在每个分布式节点本地的同时协作完成训练任务。这十分契合那些移动设备本身就是训练数据的产生地，而且不愿意或者不能将数据发送给数据融合中心的移动智能应用。而且，它也是一种极富前景能够符合 GDPR 等数据保护条例的合作机器学习机制。为了降低训练所需要的通信次数，学者提出了联邦平均（Federated Averaging, FedAvg）算法 (McMahan 等, 2017)，通过在每个节点本地执行更多的计算以提高收敛速度。它是一种经典的联邦学习算法，通过迭代的在每个移动设备执行本地模型更新以及在中心节点处执行全局模型聚合，可以得到训练好的人工智能模型。文献Smith 等 (2017) 研究了联邦多任务学习，提出了通过分布式设备的数据协作训练出多个人工智能模型的方法。

然而，联邦学习技术的应用依然面临着许多挑战。首先，模型训练的完成需要移动设备之间进行频繁的模型更新交互，其通信开销依然很大，限制了联邦学习的规模化 (McMahan 等, 2017; Wang 等, 2018c)；其次，网络中移动设备收集的数据往往是非独立同分布的，这对模型训练带来了很大的统计挑战 (Smith 等, 2017; Zhao 等, 2018)；再次不同移动设备的计算、存储和通信能力各不相同，这种异构性带来了独特的系统挑战，如落后设备可能会导致极大的延迟 (Wang 等, 2018b; Li 等, 2018b)；另外，一些设备可能存在恶意行为 (Blanchard 等, 2017)，这使得大规模分布式学习的安全问题变得也很关键，其可能会导致学习的性能有严重的下降 (Chen 等, 2017)；最后，系统实现也存在一定的问题，比如设备的连接状态可能不够可靠，执行过程有可能中断，而且算法的收敛速度也比集中式的训练要慢 (Bonawitz 等, 2019)。

值得注意的是，尽管联邦学习框架提供了一种可以保护数据隐私的移动设备分布式训练方法，且联邦平均算法通过更多的本地运算降低了所需要的通信次数，但是每次模型聚合需要聚合移动设备的本地模型，其所需要占用很高的通信带宽，这限制了联邦学习的规模化。因此，研究快速模型聚合方法成为了一大挑战。

1.2.3 边缘推断研究现状

在移动边缘执行推断任务的好处是低延迟、隐私性强，这对于无人机、智慧车辆等人工智能应用十分重要，因此吸引了学界和业界的极大注意。举例来说，人工智能在医疗健康 (Reddy 等, 2019) 中展示出了极佳前景，如使用循环神经网络做心力衰竭的检测 (Choi 等, 2016) 和使用强化学习做患者治疗方法的决

策 (Gottesman 等, 2019)。然而大的模型特别是深度神经网络模型需要占用极高的存储与计算资源。因此,有很多关于神经网络模型的压缩算法的研究,如向量量化 (Gong 等, 2014)、二值权重 (Courbariaux 等, 2015)、随机 sketching(Chen 等, 2015)、网络剪枝 (Han 等, 2016) 等方法。另外,执行推断任务的功耗也是一个主要考量,有学者研究了高能效的深度神经网络处理方法,如文献Sze 等 (2017)。

对于低延迟的边缘智能服务来说,如何将训练好的模型部署在离终端用户更近的地方是一个关键问题。然而有限的存储、计算与电量导致了高维的人工智能模型往往是不能够直接部署在移动用户本地的。因此,将人工智能模型推送至邻近的边缘服务器执行所需的大量计算是一种有效的解决方案,这促使了学者们开展基于计算卸载 (offloading) 的边缘推断系统的研究 (Mao 等, 2017)。其中最直接的方法是将整个推断任务推送至邻近边缘服务器处,这种方法特别适合于资源十分有限的物联网设备。有学者研究了这种方案下的数据压缩方案,如文献Mohanarajah 等 (2015) 提出传感器只传输所收集数据的部分帧用于协同三维制图,以减小通信开销。此外,为了保护数据隐私,文献Teerapittayanon 等 (2017) 研究了移动设备与边缘服务器协同计算的方案,即在移动设备端执行一部分的本地处理,然后将结果发送至边缘服务器,从而得到推断结果,返回给移动用户。文献Li 等 (2020) 根据移动设备和边缘服务器不同的计算能力设计了一种将深度学习模型进行分割的方法,将分割后的神经网络分别部署在移动设备和边缘服务器端。

除了基于计算卸载的边缘推断系统以外,还有基于通用计算范式的边缘推断系统,如 MapReduce(Dean 和 Ghemawat, 2008)。这种类型的框架一般考虑的是人工智能模型分布式地部署在多个节点上,适用于模型大小无法直接部署在单个边缘节点上但是可以通过切分在每个节点上部署一部分的推断任务。通过大量边缘节点的分布式计算与数据交互,可以得到每个任务的推断结果。这种系统中,一个关键问题是如何减小边缘节点之间数据传输的通信开销,另外一个关键问题是解决由于不同节点计算能力的异构性带来的某些节点计算过慢的问题。有趣的是,一系列研究发现可以使用编码技术来解决这两个问题。文献Li 等 (2018a) 使用编码技术来降低通信开销,提出了一种可扩展的数据传输方案。文献Li 等 (2017b) 将编码方案扩展到了无线网络中,用于解决基于移动设备的推断任务的分布式计算问题。文献Parrinello 等 (2018) 提出使用编码计算方案,利用

冗余的计算来保证结果能够根据一部分节点的结果即可恢复，从而解决分布式计算中部分节点计算过慢的问题。

尽管这些文献分别针对基于计算卸载与基于通用分布式计算范式的边缘推断方案提出了一系列的使能方法，依然存在着以下问题亟待解决：

1. 将计算任务推送至边缘服务器的计算卸载方案中，没有考虑到物理信道的影响，由于无线网络环境的波动，可能会导致由单个无线接入点提供的通信服务质量不能满足推断任务需求。
2. 基于通用分布式计算范式的边缘推断方案中，现有的数据传输方法并未将无线网络的物理信道纳入考虑，仅以传输所需要的信息量角度来设计传输方案。然而，在无线网络中考虑考虑物理层的无线信道，研究可达传输速率是十分关键的一个问题。

1.2.4 研究内容与方法

解决移动边缘人工智能对于无线通信系统的挑战是一件系统性的工程。从传统的端到端数据传输的角度来看，信息量可以用熵来度量，而香农信源编码理论给出了无损信源编码的极限 (Shannon, 1948)。如果我们只关注于传输方式而不是传输内容的话，香农信息论已经为通信系统中我们能做到的极限提供了完美答案。也就是说，如果固定了系统架构和收发内容的话，端到端的传输问题已经得到了解决。然而，移动边缘人工智能中的通信问题并不是孤立的。从算法的角度来说，传输内容决定了每次信息交互的通信量以及通信次数。这样站在学习算法的角度来看待通信挑战，使得可以通过研发不同的算法来降低每次信息交互的通信量或者降低通信次数。例如，对于一个训练任务来说，可以设计收敛速度更高的算法来降低所需要的通信次数，而有损压缩的方式亦可用来降低每次信息交互的通信量，只需要保证训练得到的模型性能没有明显损失。边缘人工智能系统的设计对于边缘节点之间通信范式的设计也具有极大的影响。例如，基于通用计算范式的分布式推断方案需要设计的大量节点之间进行数据的交互，所设计的传输方案的可扩展性就变得极为重要。而基于移动边缘计算中的计算任务卸载的边缘推断方案中，需要结合节点的计算能力、功耗、通信环境，并结合任务特点设计先进的数据推送策略、边缘计算机制，降低移动设备与边缘服务器之间的通信开销，提高通信效率。

因此，为了实现移动边缘人工智能，实现无线网络从万物互联到智慧互联的

范式转变，需要从边缘人工智能任务的特点与需求出发，针对不同的训练与推断系统架构来设计通信高效、高可靠性以及高能效的移动边缘人工智能方案。对于移动边缘训练来说，本文将就联邦学习中的快速模型聚合问题展开研究，实现移动设备分布式训练。而对于中等大小模型的移动边缘推断问题，本文将考虑基于通用计算范式的方案中的快速数据交换策略，实现基于移动设备分布式计算的边缘推断。对于更高维的人工智能模型推断，如深度神经网络，本文将就基于计算卸载至边缘服务器的方案展开研究，提出高服务质量、高鲁棒性、高能效的边缘推断方案。具体来说，本文研究了下述四个移动边缘人工智能系统与使能方案：

1. 针对移动设备分布式训练的快速数据聚合问题，第3章中建立了通信理论中的网络中计算（in-network computation）问题 ([Giridhar 和 Kumar, 2006](#)) 与模型聚合之间的联系，提出了一种空中计算（over-the-air computation）的快速模型聚合方法，利用无线信道的信号叠加性质提高了联邦学习从移动设备端聚合模型的通信效率。该问题的关键在于最大化参与模型聚合的移动设备的同时保证聚合误差约束，需要同时选择设备与设计接收端波束成形向量，可以表达为一个带有非凸二次约束的稀疏优化问题。非凸的二次约束可以使用矩阵升维技术转化为半正定矩阵的凸约束条件和一个非凸的秩为一的约束。然后设计了基于差分凸函数（difference-of-convex-functions, DC）的稀疏与低秩优化算法，能够更准确的检测非凸二次约束的可行性，进而能够选择更多的移动设备，从而使边缘训练得到更好的性能。
2. 针对移动设备分布式训练的快速数据聚合问题，第4章进一步从改善无线信号传播环境的角度出发，利用新兴的智能反射面技术，进一步加速模型聚合的通信效率。由于智能反射面的存在，模型聚合问题需要接收端波束成形向量、反射面相位矩阵与设备选择的联合设计。该问题可以表达为一个有非凸的双二次约束的稀疏优化问题。为了解决非凸的双二次约束，使用了矩阵升维技术，从而得到了一个关于半正定矩阵的双线性约束和非凸的秩约束。尽管双线性约束依然非凸，可以使用交替优化来解决，再加上差分凸函数技术，得到的算法能够充分发挥智能反射面与空中计算模型聚合技术的优势，提高边缘训练的性能。
3. 针对基于移动设备分布式计算的边缘推断系统的快速数据交换问题，第5章提出了一种高效的传输方案，通过协作传输与干扰消除技术使得每个移动设备

端能够解出所需要的信息。该方案中选择可达速率的一阶描述即自由度为性能度量，将预编码矩阵和解码矩阵嵌入一个低秩矩阵中，可以建立可达自由度与该低秩矩阵的秩的反比关系。从而可以通过求满足干扰消除条件的最小矩阵秩来最大化所提传输方案的可达自由度。该问题为有仿射约束的低秩优化问题，结合问题的结构特点，第5章提出了一种新颖的高效差分凸函数算法，能够在降低现有算法复杂度的基础上保证不带来明显的性能损失。结果显示所提的分布式推断数据交换策略具备可扩展性。

4. 针对基于计算任务卸载的边缘推断系统，第6章提出了一种高能效的协作传输策略。通过将任务推送给多个边缘服务器，使用协作传输能够提高传输的频谱效率，但是会造成边缘服务器计算更多任务，带来更高的功耗。而协作传输需要全局的信道状态信息，而反馈的信道状态信息不可避免地具有不确定性。于是，考虑了任务选择与波束成形联合优化的问题，最小化推断任务计算功率与传输功率之和，以实现边缘推断中高能效的处理与鲁棒性的传输，并将其建模为了一个有概率服务质量约束的组稀疏波束成形问题。为了解决非凸非确定性的概率服务质量约束，采用了基于统计学习的鲁棒优化近似方法，将其转化为了非凸的二次约束。使用矩阵升维技术，可以将该问题转化为一个有非凸秩约束的稀疏优化问题。为了进一步提升稀疏性，提供了一种有差分凸函数正则项的加权功率最小化算法，所返回的结果能够使得在边缘服务器处执行更少的运算任务，从而在保证服务质量的鲁棒性基础上需要更低的功耗。

近年来，稀疏与低秩优化方法在高维统计、信号处理、机器学习等领域得到了广泛的应用，学者们设计了许多高效算法。在无线通信领域中，稀疏与低秩优化也逐渐引起了关注。文献Shi等(2014)提出了基于稀疏优化的波束成形问题，用于最大化云接入网络的能效，而波束成形的组稀疏性恰好对应了基站的开关状态，若一个基站对应的波束成形向量为零，可以让对应的基站进入低功耗的睡眠模式。文献Dai和Yu(2016)进一步利用波束成形向量的组稀疏性研究了云接入网络中高能效的数据分享策略与压缩策略。文献Shi等(2016b)使用稀疏优化来解决云接入网的用户接入控制问题，非负辅助变量每个元素的大小代表了对应的用户可达信噪比与目标信噪比的差距。文献Shi等(2016a)研究了拓扑干扰管理问题，使用低秩矩阵来建模拓扑干扰消除条件中的编解码矩阵，然后将最大化自由度问题重新表述为一个低秩矩阵填充问题求解，矩阵的秩与自由度

成反比。文献 Yang 等 (2019a) 进一步研究了拓扑干扰管理的协作机制，通过解一个低秩优化问题来得到能够最大化自由度的编解码矩阵。使用稀疏优化与低秩优化来建模无线通信中的问题相比传统方法具有算法灵活性的优势，可以利用不同算法的特点结合应用需求选取和设计适合的方法。但也要注意到，在无线通信中的稀疏与低秩优化问题往往有着其独特的问题结构，需要结合问题具体分析，如拓扑干扰管理问题 (Shi 等, 2016a) 所建模的低秩矩阵优化问题，用经典的核范数放缩无法得到低秩的解。本文使用了稀疏优化与低秩优化的方法来建模研究边缘人工智能中的四个关键问题，从中也可以看出所用方法的性能优势。

1.3 全文架构

为了应对移动边缘人工智能对无线通信系统的挑战，本论文研究了如第1.2.4节所述的四个关键议题，组织结构在图1.2中给出了概括表示。具体说来，整个论

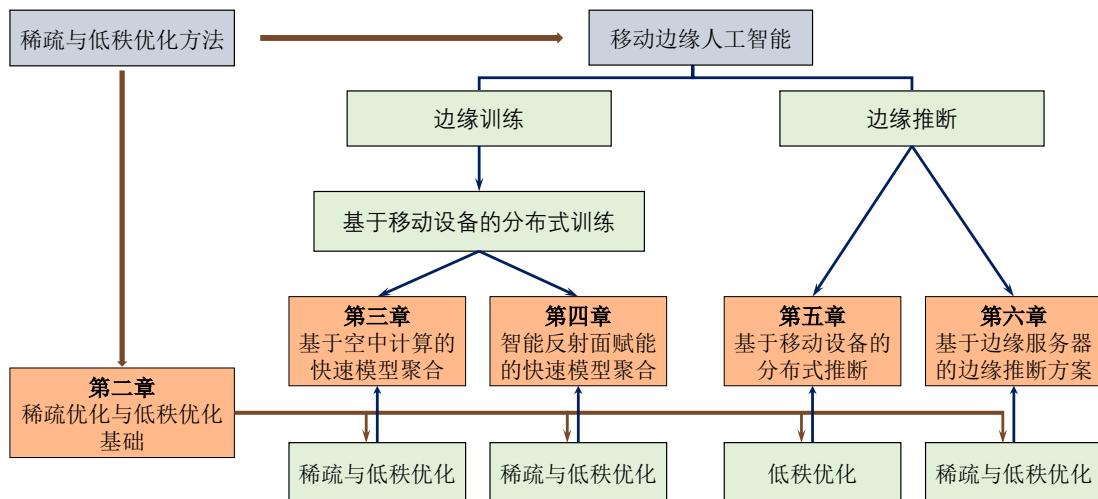


图 1.2 全文组织架构示意图。

Figure 1.2 Illustration of the structure of this thesis.

文后续章节组织结构如下：

1. 在第2章中，讨论了稀疏优化与低秩优化的相关知识，并阐述了求解稀疏与低秩优化问题的困难之处。随后详细介绍了现有的主流方法的思想与原理，为之后章节使用稀疏与低秩优化建模方法奠定了基础。
2. 第3章研究了移动设备分布式训练系统，针对联邦学习的经典算法的模型聚合问题的特点，设计了一种基于空中计算的快速模型聚合策略，在最大化每轮模型聚合选择的设备的同时保证聚合误差满足约束。该问题建模为了一个稀疏

与低秩优化问题，其中稀疏性辅助变量用于设备选择，而低秩模型用于解决非凸二次约束。然后设计了基于差分凸函数的稀疏与低秩优化算法，在快速模型聚合的同时保证了模型训练的性能。

3. 第4章中使用智能反射面来进一步提高移动设备分布式训练系统中的模型聚合的通信效率。该问题也建模为了一个稀疏与低秩优化问题，其中设备选择用稀疏优化来建模，而非凸的双二次约束用低秩优化建模，进一步使用交替优化来求解。该方法改善了无线信号的传播环境，在同样的模型聚合误差下选择了更多的移动设备，提高了模型训练的性能。

4. 第5章考虑了基于移动设备分布式计算的边缘推断系统，利用广泛使用的分布式计算架构来完成每个移动设备的推断任务，提出了一种快速数据交换策略，最大化满足干扰消除条件的自由度。该问题建模为了一个有仿射约束的低秩优化问题，将编解码矩阵嵌入了低秩矩阵变量，矩阵的秩与自由度成反比。为了求解该低秩优化问题，结合问题的结构，提出了一种计算高效的算法。所提的数据交换方案的通信开销具备可扩展性，能够在快速求解出编解码方案的同时保证优异的性能。

5. 第6章中研究了基于计算任务卸载的边缘推断的高能效协作传输策略，将移动用户的推断任务推送至多个边缘服务器，通过在概率服务质量约束下最小化总功耗，为移动设备提供高服务质量、鲁棒性的传输服务。使用基于统计学习的鲁棒优化近似，将该问题重新表述为了一个稀疏与低秩优化问题。其中波束成形向量的组稀疏性对应了每个推断任务是否在每个基站执行的决策，低秩优化用于解决非凸的二次约束。所得的稀疏与低秩优化问题使用了带有差分凸函数正则项的加权最小化算法求解。所提的边缘推断方法功耗低，能够在获得的信道状态信息存在误差的情况下保证鲁棒传输。

6. 第6章总结了全文，并对各章节存在的问题以及最新的研究方向开展了讨论。

第2章 稀疏优化与低秩优化

在绪论中，回顾了无线通信系统的发展史与人工智能所的强大能力与潜力，指出了将智能嵌入移动网络是下一代无线通信系统的发展方向。稀疏优化与低秩优化已经展现出了其在建模无线通信领域相关问题的优势，在许多应用中发挥了重要作用。本章阐述了稀疏性与低秩性的概念与意义，然后对稀疏优化与低秩优化主要的核心算法与思想进行了回顾，为后续章节所需要的基础理论与算法做铺垫。

2.1 稀疏性与低秩性

稀疏性 (sparsity) 是数据的一种极为重要的结构特性，它描述了高维数据所具备的一种低维结构。各种不同形式的稀疏性在高维统计、信号处理、机器学习等领域中起着关键作用。一般来说，稀疏性是指一个信号能够用很少的基本“原子”的叠加来表示。最常见的稀疏的概念指的是一个向量信号 $\mathbf{x} \in \mathbb{R}^n$ 的支集的稀疏，其对应的原子集合为坐标基向量 $\{\mathbf{e}_i\} \subset \mathbb{R}^n$ ，即

$$\mathbf{x} = \sum_{i=1}^k \lambda_i \mathbf{e}_i. \quad (2.1)$$

k 被称作向量 \mathbf{x} 的稀疏度。进一步，将其类比扩充到矩阵频谱的稀疏性，可以得到矩阵 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 低秩 (low-rank) 的概念，其对应的原子集合为秩为 1 的矩阵 $\{\mathbf{u}_i \mathbf{v}_i^\top\} \in \mathbb{R}^{m \times n}$ ，可以从矩阵 \mathbf{X} 的奇异值分解 (singular value decomposition, SVD) 中得到，即

$$\mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^\top = \sum_{i=1}^K \sigma_i \mathbf{u}_i \mathbf{v}_i^\top. \quad (2.2)$$

其中 $K \leq \min\{m, n\}$ 等于矩阵的 \mathbf{X} 的秩。在许多应用中，所涉及的信号或者模型维度很高，但是其能够用很少的基本原子的叠加来表示。例如，压缩感知 (compressed sensing) (Donoho, 2006) 中提出利用信号的稀疏性可以将其从很少的线性测量值中恢复出来，这被广泛应用在磁共振成像 (Lustig 等, 2007)、地震数据恢复 (Herrmann 和 Hennenfent, 2008) 等领域中。而低秩矩阵模型则在推荐系统 (Dong 等, 2020)、遥感图像 (Zhang 等, 2013)、干扰管理 (Yang 等, 2019a) 等领域中发挥了重要作用。

由于稀疏函数（即 ℓ_0 范数）与秩函数均为非凸，不论是稀疏或低秩的目标函数还是约束条件，由此所得的问题是难以直接求解的。这使得学者们提出了一系列的算法来解决非凸的稀疏与低秩函数 (Tropp 和 Wright, 2010; Davenport 和 Romberg, 2016; Udell 等, 2016)。这些算法大体上可以分为两类，凸方法与非凸方法。用一个凸函数来近似 ℓ_0 范数和秩函数，得到一个凸优化问题，进而可以利用凸优化理论 (Boyd 和 Vandenberghe, 2004) 来进行求解。这类方法中最经典的为 ℓ_0 范数的 ℓ_1 范数松弛，与秩函数的核范数松弛方法。非凸方法的一个著名例子为迭代加权最小二乘算法，其基本思想是首先找到一个非凸的近似，然后再利用所得近似问题的结构特性使用迭代加权算法进行求解。对于稀疏优化问题，可以用迭代加权最小二乘算法来最小化 ℓ_0 范数的 ℓ_p 范数近似 ($0 < p < 1$)。对于低秩优化问题，可以用迭代加权最小二乘算法最小化秩函数的 Schatten- p 范数近似 ($0 < p < 1$)。本文考虑的为移动边缘网络来支持人工智能任务的问题，既涵盖了移动设备之间的协作完成训练与推断任务的方案，也研究了在边缘服务器，即基站或无线接入点的强大算力帮助下为移动终端用户提供智能服务的方案。故将分别从应用场景问题的特点出发，来选择和设计稀疏或低秩优化的算法。这依赖于针对问题的稀疏与低秩模型的建立，与针对场景与问题结构的在算法性能和计算复杂度之间的权衡。

2.2 稀疏优化方法

表示稀疏性的 ℓ_0 范数可以用指示函数来表示，即对于向量 $\mathbf{x} \in \mathbb{R}^n$ 有

$$\|\mathbf{x}\|_0 = \sum_{i=1}^n I_{x_i \neq 0}, \quad (2.3)$$

其中指示函数在条件 $x_i \neq 0$ 满足时取 1，不满足时取 0。也就是说， ℓ_0 范数给出了该向量的非零元素数目。稀疏优化的一个典型应用为压缩感知：

例 2.1. 压缩感知 (Donoho, 2006) 问题为从少量线性测量 $\mathbf{y} \in \mathbb{R}^m$ 中恢复稀疏信号 $\mathbf{x} \in \mathbb{R}^n$ ，可以建模为如下优化问题

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}. \quad (2.4)$$

在一些场景中，向量可以分割为许多组 (group) 的变量，需要考虑这些组中的变量是否同时为零，于是可以得到将稀疏性推广了的组稀疏 (group sparse)

的概念。例如对一个由 n 个子向量 $\mathbf{x}_i \in \mathbb{R}^d$ 组成的聚合向量

$$\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T \in \mathbb{R}^{nd}, \quad (2.5)$$

可以定义其组稀疏度为所有子向量 \mathbf{x}_i 作为一个组为非零的数目，即 $\sum_{i=1}^n I_{\mathbf{x}_i \neq 0}$ 。它表示了 \mathbf{x} 的一种组合稀疏性，即每一个子向量 \mathbf{x}_i 作为一个组是否为零向量。稀疏函数的非凸性使得稀疏优化问题的求解困难，本节将展示用于解决稀疏优化问题的不同方法。

2.2.1 凸松弛方法

一个基本的思想是将非凸的 ℓ_0 范数放缩为一个可以诱导稀疏性的凸函数。 ℓ_1 范数是 ℓ_0 范数的凸包络，是最常见的一种凸松弛替代方法。作为其最紧的凸放缩， ℓ_1 范数松弛法得到了广泛的应用。

例 2.2. 在统计学与机器学习领域中，LASSO (least absolute shrinkage and selection operator)(Tibshirani, 1996) 是一种可以同时做变量选择的回归分析方法，即求解

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (2.6)$$

它采用了 ℓ_1 范数正则项来诱导模型变量的稀疏解，从而起到变量选择的作用。

例 2.3. 在优化理论中，检测问题的可行性 (feasibility) 是一个十分关键的问题。其中 ℓ_1 范数被用来解决最小化不可行约束条件个数的问题，这是通过最小化辅助变量的 ℓ_1 范数来实现的，即

$$\underset{\mathbf{x} \in \mathbb{R}_+^n}{\text{minimize}} \quad \|\mathbf{x}\|_1 \quad \text{subject to } g(\mathbf{x}) \leq x_i, i = 1, \dots, m. \quad (2.7)$$

相应的优化问题在优化相关的文献中被称作 sum-of-infeasibilities(Boyd 和 Vandenberghe, 2004)。

2.2.2 迭代加权算法

另外一个著名的诱导稀疏方法是迭代加权最小化 (iteratively reweighted minimization) 算法，如加权 ℓ_1 最小化和加权最小二乘法等。文献Candes 等 (2008) 提出了加权 ℓ_1 最小化算法，通过解一系列的加权 ℓ_1 最小化问题，并根据当前问题解来更新下次迭代的权重的方法，实现比 ℓ_1 松弛法更优异的性能。对于稀疏优化问题

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{x} \in C, \quad (2.8)$$

其中 \mathcal{C} 为约束集合。使用加权 ℓ_1 最小化方法需要在第 t 步解问题

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \sum_{i=1}^n w_i^{[t]} |x_i| \quad \text{subject to } \mathbf{x} \in \mathcal{C}, \quad (2.9)$$

得到解为 $\mathbf{x}^{[t]}$ 。然后根据文献 Candes 等 (2008) 所述，更新的权重 $\mathbf{w}^{[t+1]}$ 取为

$$w_i^{[t+1]} = \frac{1}{|x_i^{[t]}| + \epsilon}, \quad (2.10)$$

其中 $\epsilon > 0$ 为一个光滑参数。这种方法可以看作是使用了非凸的对数和函数来近似 ℓ_0 范数，即 $\sum_{i=1}^n \log(|x_i| + \epsilon)$ ，然后采用 MM (majorization-minimization)(Lange, 2016) 算法来进行求解。

另一种提升诱导稀疏性性能的算法为迭代加权最小二乘法 (iteratively reweighted least squares, IRLS) (Daubechies 等, 2010)，该方法的原理是利用非凸的 ℓ_p 范数 ($0 < p < 1$) 来近似 ℓ_0 范数。向量 \mathbf{x} 的 ℓ_p 范数由下式给出

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n x_i^p \right)^{1/p} \approx \left(\sum_{i=1}^n (x_i^2 + \epsilon^2)^{\frac{p}{2}-1} x_i^2 \right)^{1/p}, \quad (2.11)$$

其中 $\epsilon > 0$ 为一个光滑参数。观察到 ℓ_p 范数的这种近似表达，IRLS 算法的主要迭代步骤是在第 t 步迭代中解一个加权的 ℓ_2 最小化问题，即

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \sum_{i=1}^n w_i^{[t]} x_i^2 \quad \text{subject to } \mathbf{x} \in \mathcal{C}, \quad (2.12)$$

得到解 $\mathbf{x}^{[t]}$ 。然后更新权重 $\mathbf{w}^{[t+1]}$ 为

$$w_i^{[t+1]} = (|x_i^{[t]}|^2 + \epsilon^2)^{-p/2}. \quad (2.13)$$

注意到迭代加权最小化依赖于一个光滑参数 ϵ ，这个参数影响了算法的收敛行为，需要谨慎选取 (Chartrand 和 Yin, 2008; Wang 等, 2018a)。

2.2.3 组稀疏优化方法

组稀疏性往往用来同时选择或者移除一组变量，常用组范数 (group norm) 来诱导组稀疏性。如文献 Bach 等 (2012) 所述，加权的混合 ℓ_1/ℓ_p 范数 ($p > 1$) 可以用于求解组稀疏优化问题，其定义为

$$\|\mathbf{x}\|_{1,p} = \sum_{i=1}^n w_i \|\mathbf{x}_i\|_p = \sum_{i=1}^n w_i \left(\sum_{j=1}^d x_{ij}^p \right)^{\frac{1}{p}}. \quad (2.14)$$

其中 $w_i > 0$ 为权重，一般来说取值为 1。最常见的组范数，加权的混合 ℓ_1/ℓ_2 范数，定义为

$$\|\mathbf{x}\|_{1,2} = \sum_{i=1}^n w_i \|\mathbf{x}_i\|_2. \quad (2.15)$$

通过最小化加权的混合 ℓ_1/ℓ_2 范数可以促使每个 $\|\mathbf{x}_i\|_2$ 趋于零。它是一个凸函数，故而所得的问题可以根据凸优化理论设计算法求解。类似的还有混合 ℓ_1/ℓ_∞ 范数，但是其得到的解可能会有许多元素幅度值相等，这是由其在 0 以外的点处也不可微造成的。同样的，也可以利用迭代加权算法，即通过交替最小化加权的混合 ℓ_1/ℓ_p 范数组范数与更新权重，来进一步增强解的组稀疏性。

2.3 低秩优化方法

近年来有这大量的关于矩阵的低秩结构的相关研究。其中低秩优化的一个典型例子是矩阵补全问题：

例 2.4. 低秩矩阵补全 (Candès 和 Recht, 2009) 在推荐系统、拓扑干扰管理 (Shi 等, 2016a) 等场景得到了广泛的应用。其目标是从观察到的部分元素中恢复出一个低秩矩阵，可以建模为

$$\underset{\mathbf{X}}{\text{minimize}} \text{rank}(\mathbf{X}) \quad \text{subject to } \text{Proj}_{\mathcal{A}}(\mathbf{X}) = \mathbf{Y}, \quad (2.16)$$

其中 $\text{Proj}_{\mathcal{A}}$ 为正交投影算子，使得矩阵 \mathbf{X} 的索引值 (i, j) 在索引集合 \mathcal{A} 中的元素值等于 \mathbf{Y} 对应的元素值，其他值为 0，即

$$Y_{ij} = \begin{cases} X_{ij}, & (i, j) \in \mathcal{A} \\ 0, & \text{其他} \end{cases} \quad (2.17)$$

对于一般的低秩优化问题，学者们研究设计了许多方法来解决非凸秩函数带来的问题，本节就这些算法进行了介绍。

2.3.1 核范数松弛

核范数 (nuclear norm) (Davenport 和 Romberg, 2016)，又被成为迹范数 (trace norm)，是秩函数最经典的一种凸代理，其有效性在许多应用中得到了证明。对于秩最小化问题

$$\underset{\mathbf{X}}{\text{minimize}} \text{rank}(\mathbf{X}) \quad \text{subject to } \mathbf{X} \in \mathcal{C}, \quad (2.18)$$

使用核范数松弛方法得到的优化问题可以表示为

$$\underset{\mathbf{X}}{\text{minimize}} \quad \|\mathbf{X}\|_* \quad \text{subject to } \mathbf{X} \in \mathcal{C}. \quad (2.19)$$

核范数 $\|\mathbf{X}\|_*$ 等于矩阵 \mathbf{X} 所有奇异值的和。它是单位范数的秩为 1 的原子矩阵的集合的凸包，因而是秩函数最紧的凸松弛。核范数最小化问题可以等价转化半正定规划（semidefinite program, SDP）问题来求解，即

$$\begin{aligned} & \underset{\mathbf{X}, \mathbf{W}_1, \mathbf{W}_2}{\text{minimize}} \quad \text{Tr}(\mathbf{W}_1) + \text{Tr}(\mathbf{W}_2) \\ & \text{subject to} \quad \mathbf{X} \in \mathcal{C}, \\ & \quad \begin{bmatrix} \mathbf{W}_1 & \mathbf{X} \\ \mathbf{X}^H & \mathbf{W}_2 \end{bmatrix} \succeq 0 \end{aligned} \quad (2.20)$$

2.3.2 迭代加权算法

类似于促进稀疏性的加权 ℓ_2 最小化算法，有学者研究了基于 Schatten- p 范数 ($0 \leq p \leq 1$) (Mohan 和 Fazel, 2012) 的迭代加权算法。一个矩阵 $\mathbf{X} \in \mathbb{C}^{M \times N}$ 的 Schatten- p 范数的定义是

$$\|\mathbf{X}\|_{Sp} = \left(\sum_{i=1}^K \sigma_i^p(\mathbf{X}) \right)^{1/p}. \quad (2.21)$$

可以观察到，Schatten- p 范数可以看作是其奇异值组成的向量的 p 范数，在 $p \geq 1$ 时是凸的，而在 $0 < p < 1$ 时为非凸的。 $p = 1$ 时即对应了矩阵的核范数。尽管它在 $p < 1$ 时是非凸的，可以用迭代加权最小二乘算法（iterative reweighted least squares algorithm, IRLS），通过交替地最小化加权的 Frobenius 范数和更新权重矩阵 \mathbf{W} 来求解。这是因为观察到

$$\|\mathbf{X}\|_{Sp}^p = \text{Tr}((\mathbf{X}^H \mathbf{X})^{\frac{p}{2}-1} \mathbf{X}^H \mathbf{X}) \quad (2.22)$$

对于非奇异矩阵 \mathbf{X} 都成立。在第 t 步迭代中，可以使用如下方式来更新 \mathbf{X}

$$\underset{\mathbf{X}}{\text{minimize}} \quad \{\text{Tr}(\mathbf{W}^{[t]} \mathbf{X}^H \mathbf{X}) \quad \text{subject to } \mathbf{X} \in \mathcal{C}, \quad (2.23)$$

然后更新权重矩阵 \mathbf{W} 为

$$\mathbf{W}^{[t+1]} = (\mathbf{X}^{[t]H} \mathbf{X}^{[t]} + \gamma^{[k]} \mathbf{I})^{\frac{p}{2}-1}, \quad (2.24)$$

其中 $\gamma^{[k]} \in \mathbb{R}$ 是一个光滑参数，用于保证 $\mathbf{W}^{[t]}$ 有定义。该方法相比核范数松弛法来说，避免了解半正定规划问题带来的高复杂度，求解起来更加高效。

2.3.3 其他算法

除了这些著名方法之外，还有着其他一些不同的方法。文献 (Fazel 等, 2003) 提出了一种基于对数行列式 (log determinant) 的启发式算法，通过将矩阵变量嵌入到一个半正定矩阵中，然后最小化所得矩阵的对数行列式来诱导低秩解。尽管对数行列式也是非凸的，但是由于它是一个光滑函数，可以用一个局部优化算法来求解，即迭代的最小化对数行列式函数在当前点处的一阶泰勒展开。为了降低低秩优化算法内存的占用，学者提出了许多种基于矩阵分解的算法。这是观察到秩为 r 的 $M \times N$ 的矩阵 \mathbf{X} 可以分解为 $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ ，其中 \mathbf{U}, \mathbf{V} 维度分别为 $M \times r$ 和 $N \times r$ 。如文献 (Jain 等, 2013) 提出了用交替最小化 (alternating minimization) 算法，通过交替更新 \mathbf{U} 和 \mathbf{V} 直至算法收敛，来解决低秩矩阵补全问题。另外，对于有秩约束的优化问题，可以通过研究低秩分解因子在黎曼流形上的性质，设计黎曼优化算法。文献 (Vandereycken, 2013) 提出了一种黎曼优化算法来求解低秩矩阵补全问题。但是该方法需要问题的约束具有结构性，并且利用约束条件的流形结构来针对性的设计黎曼优化算法，对具有其他一般约束条件的低秩优化问题不适用。

2.4 本章小结

本章回顾了稀疏与低秩的基本概念，以及主要的稀疏优化与低秩优化的算法的思想与求解方法。这为后续使用稀疏与低秩优化解决移动边缘人工智能奠定了基础，以便针对现有算法的优缺点进行分析，选用合适应用场景的算法，或者提出新思路解决现有方法的缺陷。

第3章 移动设备分布式训练的快速模型聚合策略

本章考虑了移动设备分布式训练系统，即依赖于多个移动设备收集的训练数据与其计算能力，协调合作完成模型的训练过程。许多新涌现的移动智能设备的相关应用对隐私保护有着严格的要求，如无人机，智能车辆等。这催生了一个新生领域联邦学习的诞生，即保护数据隐私的边缘分布式训练。本章就该边缘训练系统中关键性的通信挑战展开研究，提出了一种基于空中计算的快速模型聚合策略。

本章内容安排如下：第3.1节回顾了联邦学习问题与相关研究，指出了联邦学习中快速模型聚合的重要性，并介绍了所提方法的基本思想；第3.2节正式给出了基于移动设备的联邦学习系统模型与所提的快速模型聚合方法的问题表述；第3.3节将其建模为了一个稀疏与低秩优化问题；第3.4节设计了一种基于差分凸函数表示的算法，并进行了算法复杂度与收敛性分析；第3.5对所设计的算法进行了仿真验证与结果分析；最后在第3.6对本章进行了总结。

3.1 引言

联邦学习 (McMahan 等, 2017; Yang 等, 2019b; Wang 等, 2019) 是研究在保护数据隐私下合作共同建立人工智能模型的领域，对隐私敏感、低延迟的边缘智能应用是一种极富潜力的解决方案，能够发挥移动边缘大量分布式智能设备的计算能力。联邦学习在智慧零售、智慧医疗、金融服务、移动内容预测等领域中展示出了极大的应用前景 (Bonawitz 等, 2019; Yang 等, 2019b)。它的典型适用场景是移动设备是数据的产生者或者收集者，并且不愿意将数据传输给第三方的中心服务器。联邦学习使得每个移动设备将它的数据留在本地，而只需要将本地更新的模型进行聚合得到更新的全局模型，从而移动设备可以避免其数据泄漏给其他设备或者中心服务器，保证了数据的隐私性和安全性。然而基于移动设备的联邦学习系统中的无线网络环境带宽有限，频繁的模型聚合会带来巨大的通信开销，特别是当参与的移动设备数目增加时，通过传统的正交传输所需要的通信带宽十分巨大。

联邦平均 (Federated Averaging, FedAvg) 算法 (McMahan 等, 2017) 是一种经

典的联邦学习算法。通过周期性地对每个移动设备的本地更新模型取平均，该算法能够有效地减少中心节点和移动设备的通信次数。本章研究旨在提出快速的模型聚合方案，提高通信效率，降低每次通信所需要的通信带宽。观察到模型聚合过程是由本地更新从移动设备到中心节点传输和计算这些更新的加权平均两部分所组成，本章从而基于空中计算（over-the-air computation, AirComp）的原理，提出了一种计算和通信联合设计的模型聚合方案。该方案利用了无线多址接入信道（multiple-access channel）的信号叠加特性，通过并发传输使得在中心节点处可以直接得到移动设备本地更新的加权平均。应用该方案的关键在于要在选择最多的设备加入到模型聚合中的同时降低聚合误差，以最大化联邦学习算法的性能。这是一个联合设备选择与波束成形问题，本章提出了一种稀疏与低秩优化的建模方法来解决这个问题。基于该问题的独特结构，提供了一种新颖的差分凸函数（difference-of-convex functions, DC）方法，最后通过仿真对该算法的性能增益其性能优势进行了评估。

3.2 移动设备分布式训练系统的快速模型聚合

本节展示了移动设备分布式训练系统模型，并提出了一种计算和通信联合设计的快速模型聚合方案。

3.2.1 移动设备分布式训练系统模型

基于移动设备的联邦学习系统中，每个移动设备的训练数据保持在本地，通过设备之间的合作完成一个全局模型的训练，这带来了低延迟、低功耗和保护数据隐私等大量的好处 (McMahan 等, 2017)。图 1 展示了一个由 M 个单天线的移动设备和一个拥有计算能力的基站 (base station) 组成的联邦学习系统，其中基站有 N 根天线。这个系统的分布式机器学习任务可以表示为

$$\underset{\mathbf{z} \in \mathbb{R}^d}{\text{minimize}} \quad f(\mathbf{z}) = \frac{1}{T} \sum_{j=1}^T f_j(\mathbf{z}), \quad (3.1)$$

其中 \mathbf{z} 是待优化的模型参数，维度为 d ，而 T 是所有训练数据的总个数。这样的模型广泛地涵盖了线性回归、逻辑回归、支撑向量机、深度神经网络等机器学习任务。一般来说，每个函数 f_j 可以表示为 $\ell(\mathbf{z}; \mathbf{x}_j, y_j)$ ，其中 ℓ 是一个损失函数，对应了一个输入输出数据对 (\mathbf{x}_j, y_j) 。这里 $\mathcal{D} = \{(\mathbf{x}_j, y_j) : j = 1, \dots, T\}$ 代表了训练过程涉及到的全体数据集。其中第 i 个移动设备的本地数据集可以用 $\mathcal{D}_i \subseteq \mathcal{D}$

来表示。

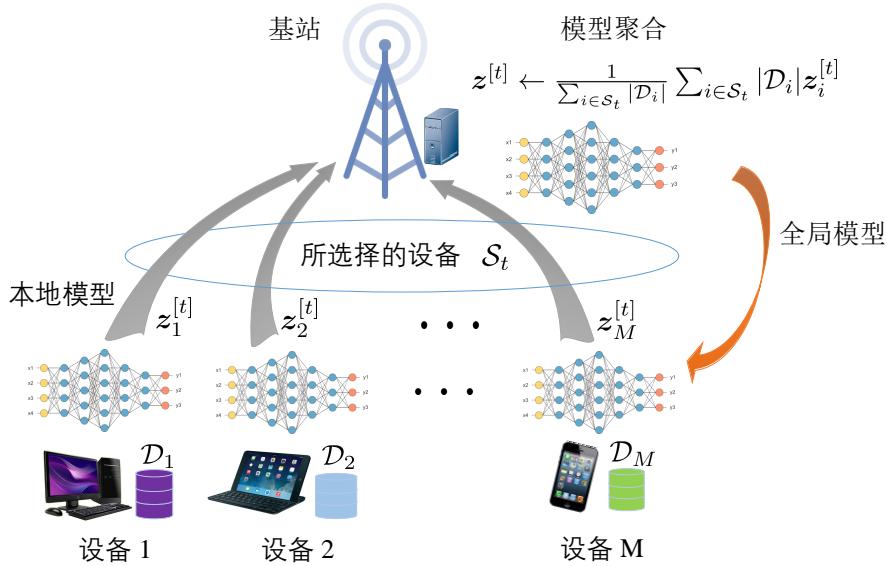


图 3.1 移动设备分布式联邦学习模型训练系统。

Figure 3.1 On-device federated learning system.

在联邦学习中，有限的网络带宽成为了限制全局模型聚合的一个主要因素。FedAvg 算法 (McMahan 等, 2017) 的提出降低了模型聚合所需的通信轮数。该算法的原理可以表示为，在第 t 轮中执行：

- 1) 基站选择一部分的移动设备，其对应的索引集合可以用 $S_t \subseteq \{1, \dots, M\}$ 来表示；
- 2) 基站将上一轮更新了的全局模型 $\mathbf{z}^{[t-1]}$ 发给所有选择的移动设备 S_t ；
- 3) 每一个被选择的移动设备 $i \in S_t$ 基于其拥有的数据集 \mathcal{D}_i 和当前的全局模型参数 $\mathbf{z}^{[t-1]}$ ，在本地运行一个模型更新算法，如随机梯度算法。运行算法得到的输出就是更新了的本地模型 $\mathbf{z}_i^{[t]}$ ；
- 4) 基站聚合所有的本地更新模型 $\mathbf{z}_i^{[t]}, i \in S_t$ ，即计算他们的加权平均作为更新的全局模型 $\mathbf{z}^{[t]}$ 。

FedAvg 算法框架如算法1所示。

本章所关注的是通过设计适用于 FedAvg 算法的快速模型聚合方案来提高联邦学习的通信效率。观察到的 FedAvg 算法的一个核心现象是其统计性能可以通过每一轮选择更多的设备来提高 (McMahan 等, 2017; Wang 和 Joshi, 2018)。图3.2中给出了一个在由十个移动设备组成的联邦学习系统下的示例。基于 CIFAR-

算法 1 FedAvg 算法

```

1: 初始化  $\mathbf{z}_0$ .
2: for  $t = 1, 2, \dots$  do
3:   基站选择一部分移动设备  $\mathcal{S}_t$ 
4:   基站将全局模型  $\mathbf{z}^{[t-1]}$  广播给所有选择的移动设备  $\mathcal{S}_t$ .
5:   for 每个被选择的移动设备  $i, i \in \mathcal{S}_t$  in parallel do
6:     移动设备  $i$  基于  $(\mathcal{D}_i, \mathbf{z}^{[t-1]})$  在本地执行更新算法, 得到  $\mathbf{z}_i^{[t]}$ 
7:   end for
8:   基站端进行模型聚合, 即  $\mathbf{z}^{[t]} \leftarrow \frac{1}{\sum_{i \in \mathcal{S}_t} |\mathcal{D}_i|} \sum_{i \in \mathcal{S}_t} |\mathcal{D}_i| \mathbf{z}_i^{[t]}$ 
9: end for

```

10 数据集 (Krizhevsky 和 Hinton, 2009), 可以用 FedAvg 算法训练一个支撑向量机 (support vector machine, SVM) 分类器, 不同的选择设备数目对应的训练过程中的损失函数值和测试集上的相对预测准确度可以从图上看出。其中相对预测准确度定义为该算法测试集上的预测准确度与随机分类的准确度的比值。该算法测试集上的准确度可以表示为 $\frac{\text{正确预测样本数目}}{\text{测试集样本总数}}$, 而由于每一类样本数目相同, 随机分类的准确度可以表示为 $\frac{1}{\text{总共类别的个数}} = 0.1$. 每一轮中, 移动设备均匀随机的从所有设备中抽取。然而, 每一轮中选择更多的移动设备也使得聚合所有本地更新模型所需的通信开销变得更高。

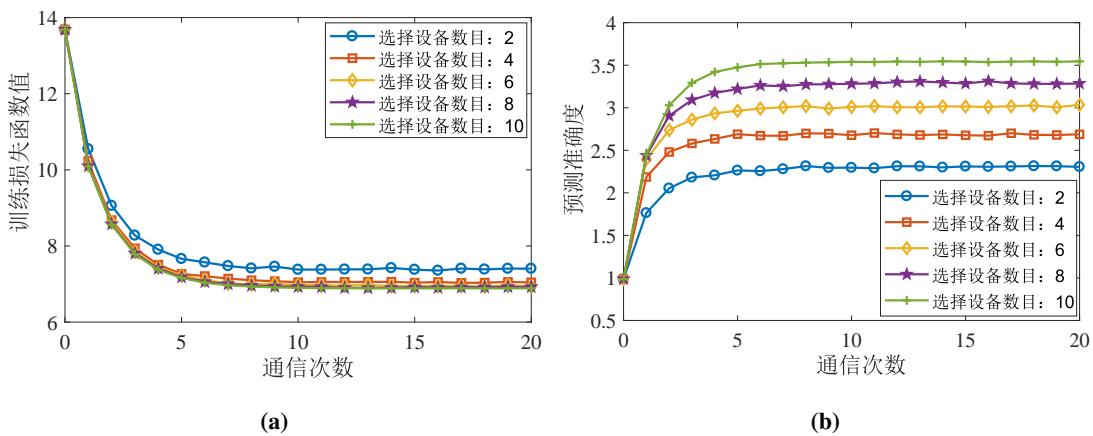


图 3.2 在选择不同数目的移动设备情况下, 使用 FedAvg 算法得到的训练损失函数值与预测准确度。(a) 训练损失函数值, (b) 相对预测准确度。

Figure 3.2 The training loss and prediction accuracy with different number of selected devices using FedAvg algorithm. (a) Training loss. (b) Prediction accuracy.

值得注意的是, 模型聚合的过程需要在基站端得到所选择的移动设备输出

的本地更新，并计算出其加权平均。因此，计算和通信的联合设计成为了快速模型聚合的一个有效的出发点。本章提出的方法是基于空中计算 (Nazer 和 Gastpar, 2007)，利用了无线多址接入信道的信号加和性质。在以前的研究中发现，空中计算相对于传统的通信和计算分离方法在计算分布式移动节点上的信息的一个线性函数时有着更高的通信效率和更低带宽占用的优势 (Nazer 和 Gastpar, 2007; Goldenbaum 等, 2013)，而这个目标是与联邦学习做模型聚合的目标是一致的。进一步可以观察到模型聚合的误差可能会导致预测准确度的严重下降 (Blanchard 等, 2017)。聚合误差可以用均方差来表示，其表达式见公式 (3.7)。为了解决这个问题，将提出一种高效的收发策略来最小化空中计算的模型聚合误差。基于这些关键现象，本章关注于下列两方面去提高联邦学习系统的统计学习性能：

- 最大化每一轮选择的移动设备数以提高训练的收敛速度；
- 最小化模型聚合误差以提高模型的预测准确率。

3.2.2 基于空中计算的快速模型聚合方案

空中计算 (over-the-air computation, AirComp) 是一种极具潜力的快速无线数据聚合方法，其目标是计算多个传输端分布式数据的一个 nomographic 函数，如算术平均值 (Goldenbaum 等, 2013)。空中计算可以利用无线多址接入信道的信号叠加特性，通过并发的传输过程联合计算和通信实现目标函数的计算，相比传统正交传输可以大幅度地提高通信效率。观察到 FedAvg 算法的全局模型聚合过程为计算每个选择的移动设备的本地更新的加权平均值，而其恰恰是空中计算所适用的 nomographic 函数的一种。这激发本节提出了基于空中计算的联邦学习系统的快速数据传输方案。

特别的，FedAvg 算法模型聚合所要计算的目标向量可以表示为

$$\mathbf{z} = \psi \left(\sum_{i \in S} \phi_i(\mathbf{z}_i) \right), \quad (3.2)$$

其中 \mathbf{z}_i 是第 i 个移动设备更新的本地模型， $\phi_i = |\mathcal{D}_i|$ 是设备 i 的预处理标量， $\psi = \frac{1}{\sum_{k \in S} |\mathcal{D}_k|}$ 是基站端的后处理标量，而 S 代表了所选择的移动设备。表示每个设备的预处理前的本地模型的符号向量 $\mathbf{s}_i := \mathbf{z}_i \in \mathbb{C}^d$ 可以假设已经经过归一化处理方差为 1，即 $\mathbb{E}(\mathbf{s}_i \mathbf{s}_i^H) = \mathbf{I}$ 。在每个时隙 $j \in \{1, \dots, d\}$ ，每个移动设备将信

号 $s_i^{(j)} \in \mathbb{C}$ 发送给基站。用

$$g^{(j)} = \sum_{i \in S} \phi_i(s_i^{(j)}) \quad (3.3)$$

来表示在第 j 个时隙下将通过空中计算所计算出的目标函数估计值。

为了简化符号, 忽略时隙索引 j , 直接将 $g^{(j)}$ 和 $s_i^{(j)}$ 分别记为 g 和 s_i . 基站端所收到的信号可以表示为

$$\mathbf{y} = \sum_{i \in S} \mathbf{h}_i b_i s_i + \mathbf{n}, \quad (3.4)$$

其中 $b_i \in \mathbb{C}$ 是发射机标量 (transmitter scalar), $\mathbf{h}_i \in \mathbb{C}^N$ 代表移动设备 i 和基站之间的信道向量, 而 $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ 是噪声向量。发射机存在功率限制, 移动设备 i 的功率限制可以表示为

$$\mathbb{E}(|b_i s_i|^2) = |b_i|^2 \leq P_0, \quad (3.5)$$

其中 $P_0 > 0$ 是最大传输功率。基站端在后处理之前的估计值为

$$\hat{g} = \frac{1}{\sqrt{\eta}} \mathbf{m}^H \mathbf{y} = \frac{1}{\sqrt{\eta}} \mathbf{m}^H \sum_{i \in S} \mathbf{h}_i b_i s_i + \frac{\mathbf{m}^H \mathbf{n}}{\sqrt{\eta}}, \quad (3.6)$$

其中 $\mathbf{m} \in \mathbb{C}^N$ 是接收端波束成形向量, η 是归一化因子。于是目标向量的每个分量可以通过在基站端计算 $\hat{z} = \psi(\hat{g})$ 来获得。

\hat{g} 相对于公式 (3.3) 给出的目标值 g 的误差可以用均方差 (mean-squared-error, MSE) 来表示, 它量化 FedAvg 算法中使用空中计算来做全局模型聚合的性能, 其表达式为

$$\text{MSE}(\hat{g}, g) = \mathbb{E}(|\hat{g} - g|^2) = \sum_{i \in S} \left| \frac{\mathbf{m}^H \mathbf{h}_i b_i}{\sqrt{\eta}} - \phi_i \right|^2 + \sigma^2 \frac{\|\mathbf{m}\|_2^2}{\eta}. \quad (3.7)$$

受文献Chen 等 (2018) 所启发, 可以得到关于发射端波束形成器的如下命题:

命题 3.1. 任意选择一个接收端波束成形向量 \mathbf{m} , 其对应的能够最小化均方差的最优发射机标量由下面的迫零发送器给出:

$$b_i = \sqrt{\eta} \phi_i \frac{(\mathbf{m}^H \mathbf{h}_i)^H}{\|\mathbf{m}^H \mathbf{h}_i\|_2^2}. \quad (3.8)$$

证明. 命题3.1给出的序列 $\{b_i\}$ 有着迫零的结构, 即

$$\sum_{i \in S} \left| \frac{\mathbf{m}^H \mathbf{h}_i b_i}{\sqrt{\eta}} - \phi_i \right|^2 = 0. \quad (3.9)$$

除此之外, 均方误差满足

$$\text{MSE}(\hat{g}, g) \geq \sigma^2 \|\mathbf{m}\|^2. \quad (3.10)$$

因此, 命题3.1中给出的迫零发送端波束成形向量 $\{b_i\}$ 对应了最小的均方误差。

□

由传输功率限制条件 (3.5) 对公式 (3.8) 中的发射机标量 b_i 的限制可以得出

$$\eta = \min_{i \in S} \frac{P_0 \|\mathbf{m}^H \mathbf{h}_i\|_2^2}{\phi_i^2}. \quad (3.11)$$

于是均方差可以表示为

$$\text{MSE}(\hat{g}, g; S, \mathbf{m}) = \frac{\|\mathbf{m}\|_2^2 \sigma^2}{\eta} = \frac{\sigma^2}{P_0} \max_{i \in S} \phi_i^2 \frac{\|\mathbf{m}\|_2^2}{\|\mathbf{m}^H \mathbf{h}_i\|_2^2}. \quad (3.12)$$

为了更清楚的展示空中计算的原理, 将其总结在了图3.3中.

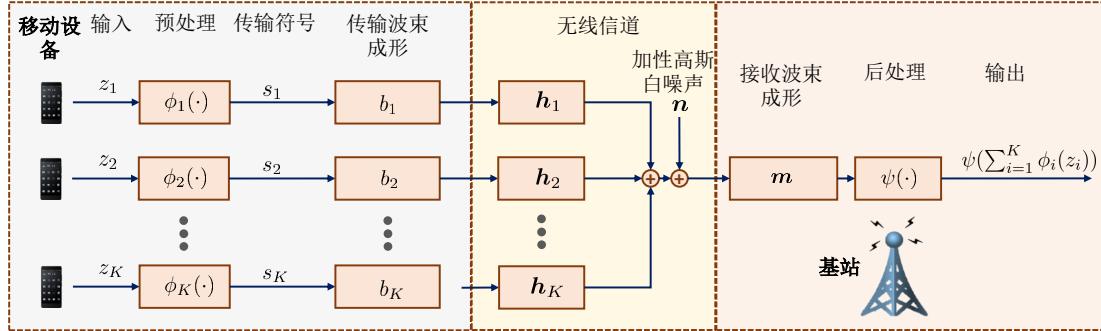


图 3.3 空中计算原理图。

Figure 3.3 Illustration of over-the-air computation.

3.2.3 问题表述

正如第3.2.1节所讨论的, 本章的目标是在最大化选择的移动设备数的同时, 让空中计算所带来的聚合误差尽可能的小。这个问题可以建模为下列混合组合优化问题

$$\underset{S, \mathbf{m} \in \mathbb{C}^N}{\text{maximize}} |S| \quad \text{subject to} \left(\max_{i \in S} \phi_i^2 \frac{\|\mathbf{m}\|_2^2}{\|\mathbf{m}^H \mathbf{h}_i\|_2^2} \right) \leq \gamma, \quad (3.13)$$

其中 $\gamma > 0$ 是全局模型聚合的均方差要求，而 $|\mathcal{S}|$ 表示集合 \mathcal{S} 的基数，即选择的移动设备数。

然而，由于目标函数 $|\mathcal{S}|$ 是一个组合函数，并且约束条件中组合变量 \mathcal{S} 和连续变量 \mathbf{m} 耦合在一起，混合组合问题 (3.13) 是一个很难解决的非凸问题。为了应对非凸的均方差函数，有文献Chen 等 (2018) 建立了非凸均方差约束 (3.13) 和非凸二次约束的联系，用于高效的算法设计。受此启发，优化问题 (3.13) 可以等价地转化为一个最大化可行的非凸二次约束的数目的问题。进一步，可以用稀疏表示的方式来最大化选择的设备数目，而非凸二次约束可以用矩阵升维技术转化为仿射约束和一个额外的秩为 1 的约束。我们将在下一节展示这种稀疏与低秩优化技术以图高效地设计算法求解问题 (3.13)。

3.3 稀疏与低秩优化建模方法

本节提出了一种联邦学习中移动设备选择的稀疏与低秩优化的建模方法。

3.3.1 问题建模

为了设计高效的算法，首先将问题 (3.13) 转化为一个有非凸二次约束的混合组合优化问题，如命题3.2所示。

命题 3.2. 优化问题(3.13)与下面的混合组合优化问题等价：

$$\begin{aligned} & \underset{\mathcal{S}, \mathbf{m} \in \mathbb{C}^N}{\text{maximize}} \quad |\mathcal{S}| \\ & \text{subject to} \quad \|\mathbf{m}\|_2^2 - \gamma_i \|\mathbf{m}^H \mathbf{h}_i\|_2^2 \leq 0, i \in \mathcal{S}, \\ & \quad \|\mathbf{m}\|_2^2 \geq 1, \end{aligned} \tag{3.14}$$

其中 $\gamma_i = \gamma/\phi_i^2$. 于是设计目标变成了在正则性条件 $\|\mathbf{m}\|_2^2 \geq 1$ 下最大化可行的均方差约束 $\|\mathbf{m}\|_2^2 - \gamma_i \|\mathbf{m}^H \mathbf{h}_i\|_2^2 \leq 0$ 个数的问题。

证明. 优化问题(3.13)可以重写为

$$\begin{aligned} & \underset{\mathcal{S}, \mathbf{m} \in \mathbb{C}^N}{\text{maximize}} \quad |\mathcal{S}| \\ & \text{subject to} \quad F_i(\mathbf{m}) = \|\mathbf{m}\|_2^2 - \gamma_i \|\mathbf{m}^H \mathbf{h}_i\|_2^2 \leq 0, i \in \mathcal{S} \\ & \quad \mathbf{m} \neq 0, \end{aligned} \tag{3.15}$$

而它又可以等价转化为

$$\begin{aligned} & \underset{\mathcal{S}, \mathbf{m} \in \mathbb{C}^N}{\text{maximize}} \quad |\mathcal{S}| \\ & \text{subject to} \quad F_i(\mathbf{m})/\tau = \|\mathbf{m}\|_2^2/\tau - \gamma_i \|\mathbf{m}^H \mathbf{h}_i\|_2^2/\tau \leq 0, i \in \mathcal{S} \\ & \quad \|\mathbf{m}\|_2^2 \geq \tau, \tau > 0. \end{aligned} \quad (3.16)$$

通过引入变量 $\tilde{\mathbf{m}} = \mathbf{m}/\sqrt{\tau}$, 优化问题 (3.16) 可以重写为

$$\begin{aligned} & \underset{\mathcal{S}, \tilde{\mathbf{m}} \in \mathbb{C}^N}{\text{maximize}} \quad |\mathcal{S}| \\ & \text{subject to} \quad F_i(\tilde{\mathbf{m}}) = \|\tilde{\mathbf{m}}\|_2^2 - \gamma_i \|\tilde{\mathbf{m}}^H \mathbf{h}_i\|_2^2 \leq 0, \quad i \in \mathcal{S}, \\ & \quad \|\tilde{\mathbf{m}}\|_2^2 \geq 1. \end{aligned} \quad (3.17)$$

因此优化问题 (3.13) 和优化问题 (3.14) 是等价的, 而正则性条件 $\|\mathbf{m}\|_2^2 \geq 1$ 是为了避免奇点 (即 $\mathbf{m} = 0$)。 \square

最大化优化问题(3.14)中的可行的均方差约束个数可通过最小化 x_k 的非零值个数来实现 (Shi 等, 2016b), 即

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}_+^M, \mathbf{m} \in \mathbb{C}^N}{\text{minimize}} \quad \|\mathbf{x}\|_0 \\ & \text{subject to} \quad \|\mathbf{m}\|_2^2 - \gamma_i \|\mathbf{m}^H \mathbf{h}_i\|_2^2 \leq x_i, \forall i, \\ & \quad \|\mathbf{m}\|_2^2 \geq 1. \end{aligned} \quad (3.18)$$

\mathbf{x} 的稀疏结构恰好指示出了选择每个设备的可行性。如果 $x_i = 0$, 选择第 i 个移动设备的同时满足均方差要求就是可行的。

然而问题(3.18)中的均方差约束和和正则性条件都是非凸二次的形式。利用矩阵升维 (matrix lifting) 技术 (Sidiropoulos 等, 2006) 是一种典型的解决方法。具体来说, 通过将向量 \mathbf{m} 升维为一个秩为 1 的半正定 (positive semidefinite, PSD) 矩阵 $\mathbf{M} = \mathbf{m}\mathbf{m}^H$, 优化问题 (3.18) 可以转化为如下的稀疏与低秩优化问题

$$\begin{aligned} \mathcal{P}_{3.1} : & \underset{\mathbf{x} \in \mathbb{R}_+^M, \mathbf{M} \in \mathbb{C}^{N \times N}}{\text{minimize}} \quad \|\mathbf{x}\|_0 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \mathbf{h}_i^H \mathbf{M} \mathbf{h}_i \leq x_i, \forall i, \\ & \quad \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \\ & \quad \text{rank}(\mathbf{M}) = 1. \end{aligned} \quad (3.19)$$

尽管优化问题 $\mathcal{P}_{3.1}$ 依然是非凸的, 有许多有效的算法可以用来解决它。

3.3.2 问题分析

优化问题 $\mathcal{P}_{3.1}$ 是非凸的，它有着稀疏目标函数和低秩约束条件。尽管稀疏函数和秩函数都是非凸的，通过研究不同的问题结构，可以设计出高效的算法求解。

ℓ_1 范数是非凸的稀疏函数，即 ℓ_0 范数，的最常见的一种凸松弛替代方法。相应的用于最大化可行解的优化问题在优化相关的文献中被称作 sum-of-infeasibilities(Boyd 和 Vandenberghe, 2004)。另外一个著名的诱导稀疏的方法是光滑 ℓ_p 最小化方法 (Shi 等, 2016b)，其原理是通过寻找一个 ℓ_0 范数的紧的近似，然后通过迭代地求解加权 ℓ_2 最小化问题求解。然而，这种算法的缺点是其需要谨慎地设置光滑参数，迭代加权算法的收敛性可能会对光滑参数的选择比较敏感 (Chartrand 和 Yin, 2008; Wang 等, 2018a)。

通过简单地去掉优化问题 $\mathcal{P}_{3.1}$ 的秩为 1 的约束条件来求解的方法被称作半正定松弛 (semidefinite relaxation, SDR) 技术 (Luo 等, 2007)。半正定松弛技术被广泛应用在求解非凸二次约束的二次规划问题中。如果求得的解秩不为 1，可以通过著名的高斯随机 (Gaussian randomization) 方法 (Luo 等, 2007) 来得到一个秩为 1 的近似解。然而当天线数增加时，高斯随机方法的性能会由于得到秩为 1 的解的概率下降而变得很差 (Chen 等, 2018; Chen 和 Tao, 2017)。

为了解决现有方法的局限性，本章将提出一种统一的差分凸函数 (difference-of-convex-functions, DC) 规划的方法来解决稀疏与低秩优化问题 $\mathcal{P}_{3.1}$ 。所提的方法能够促进稀疏性，并能准确检测出非凸二次约束的可行性，从而导致相比现有算法有着相当大的性能提升。具体来说，

- 本章将提出一种无参的 DC 方法来增强稀疏性，用于最大化所选的移动设备数。
- 本章将提出一种新颖的 DC 方法来保证得到的解严格满足秩为 1 的约束。注意到所提的 DC 方法能够保证秩为 1 的约束满足，这点对检测设备选择问题中的非凸二次约束是否可行十分关键。

3.4 稀疏与低秩优化的 DC 算法设计

本节将提供一种统一的 DC 优化算法框架用于解决问题 $\mathcal{P}_{3.1}$ 。具体来说，先展示了一种新的 ℓ_0 范数和秩函数的 DC 表示方法，再提供了整个 DC 算法框架，

最后展示了如何求解。

3.4.1 稀疏与低秩函数的 DC 表示

3.4.1.1 稀疏函数的 DC 表示

在介绍 ℓ_0 范数的 DC 表示方法前，这里首先给出其所依赖的向量的 Ky Fan k 范数的定义。

定义 3.1. Ky Fan k 范数 (Fan, 1951): 向量 $\mathbf{x} \in \mathbb{C}^M$ 的 Ky Fan k 范数是 \mathbf{x} 的一个凸函数，它的值是 \mathbf{x} 最大的 k 个绝对值之和，即

$$\|\mathbf{x}\|_k = \sum_{i=1}^k |x_{\pi(i)}|, \quad (3.20)$$

其中 π 是 $\{1, \dots, M\}$ 的一个排列，并且 $|x_{\pi(1)}| \geq \dots \geq |x_{\pi(M)}|$.

如果一个向量的 ℓ_0 范数不大于 k ，则它的 ℓ_1 范数和它的 Ky Fan k 范数相等。基于这个事实， ℓ_0 范数可以用 ℓ_1 范数和 Ky Fan k 范数之差来表示 (Gotoh 等, 2018)：

$$\|\mathbf{x}\|_0 = \min\{k : \|\mathbf{x}\|_1 - \|\mathbf{x}\|_k = 0, 0 \leq k \leq M\}. \quad (3.21)$$

3.4.1.2 低秩约束的 DC 表示

对于半正定矩阵 $\mathbf{M} \in \mathbb{C}^{N \times N}$ ，秩为 1 的约束条件可以等价地表述为

$$\sigma_i(\mathbf{M}) = 0, \forall i = 2, \dots, N, \quad (3.22)$$

其中 $\sigma_i(\mathbf{M})$ 表示矩阵 \mathbf{M} 第 i 大的奇异值。注意到其迹范数 (trace norm) 和谱范数 (spectral norm) 分别为

$$\text{Tr}(\mathbf{M}) = \sum_{i=1}^N \sigma_i(\mathbf{M}) \text{ and } \|\mathbf{M}\|_2 = \sigma_1(\mathbf{M}). \quad (3.23)$$

因此可以得到如下命题：

命题 3.3. 对于半正定矩阵 \mathbf{M} 且有 $\text{Tr}(\mathbf{M}) \geq 1$ 的约束条件存在的情况下，有如下结论

$$\text{rank}(\mathbf{M}) = 1 \Leftrightarrow \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 = 0. \quad (3.24)$$

证明. 如果半正定矩阵 \mathbf{M} 的秩为 1, 由于对所有的 $i \geq 2$ 都有 $\sigma_i(\mathbf{M}) = 0$, \mathbf{M} 的迹范数等于谱范数。等式 $\text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 = 0$ 意味着对于所有的 $i \geq 2$ 都有 $\sigma_i(\mathbf{M}) = 0$, 即 $\text{rank}(\mathbf{M}) \leq 1$. 因此由 $\text{Tr}(\mathbf{M}) \geq 1$ 能得到 $\sigma_1(\mathbf{M}) > 0$. 于是得出若 $\text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 = 0$ 成立, $\text{rank}(\mathbf{M}) = 1$ 也成立。 \square

3.4.2 统一的 DC 表示框架

本节所提出的 DC 表示框架的第一步是先诱导 \mathbf{x} 的稀疏性, 得到的解中可以提取出选择每个设备的优先级。然后通过解一系列的可行性问题来找出最多能够选择多少移动设备的同时满足均方差的要求。这个两步骤的方法如图3.4所示, 每一步需要解一个 DC 规划问题。

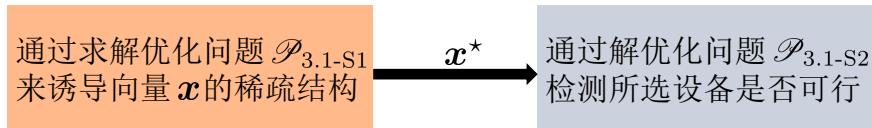


图 3.4 用于设备选择的两步骤算法框架。

Figure 3.4 A two-step framework for device selection.

3.4.2.1 步骤一：诱导稀疏性

为了求解优化问题 $\mathcal{P}_{3.1}$, 需要在第一个步骤解如下的 DC 规划问题:

$$\begin{aligned} \mathcal{P}_{3.1-S1} : & \underset{\mathbf{x}, \mathbf{M}}{\text{minimize}} \quad \|\mathbf{x}\|_1 - \|\mathbf{x}\|_k + \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \mathbf{h}_i^H \mathbf{M} \mathbf{h}_i \leq x_i, \forall i = 1, \dots, M \\ & \quad \mathbf{M} \succeq 0, \quad \text{Tr}(\mathbf{M}) \geq 1, \mathbf{x} \geq 0. \end{aligned} \quad (3.25)$$

对于 k 从 0 到 M , 通过解一系列的问题 $\mathcal{P}_{3.1-S1}$ 可以得到稀疏向量 \mathbf{x}^* 使得目标函数达到零值。注意到当目标函数等于零时, $\text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 = 0$, 秩为 1 的约束会得到满足。

3.4.2.2 步骤二：检测可行性

步骤一所得到的解 \mathbf{x} 刻画了均方差要求与每个设备的可达均方差之间的差距。因此, 第二个步骤将给具有更小 x_k 值的设备 k 赋予更高的被选择优先级。将 \mathbf{x} 的每个元素按降序排列, 即 $x_{\pi(1)} \geq \dots \geq x_{\pi(M)}$. 通过从 1 到 M 增加 k 可以找到最小的 k 使得选择所有在集合 $\mathcal{S}^{[k]}$ 里的移动设备是可行的, 其中集合 $\mathcal{S}^{[k]}$ 是 $\{\pi(k), \pi(k+1), \dots, \pi(M)\}$.

具体来说，如果在集合 $S^{[k]}$ 中的所有设备都可以被选择，如下的优化问题

$$\begin{aligned} & \text{find } \mathbf{m} \\ & \text{subject to } \|\mathbf{m}\|_2^2 - \gamma_i \|\mathbf{m}^H \mathbf{h}_i\|_2^2 \leq 0, \forall i \in S^{[k]} \\ & \quad \|\mathbf{m}\|_2^2 \geq 1 \end{aligned} \tag{3.26}$$

就应是可行的。使用矩阵升维技术它可以等价转化为

$$\begin{aligned} & \text{find } \mathbf{M} \\ & \text{subject to } \text{Tr}(\mathbf{M}) - \gamma_i \mathbf{h}_i^H \mathbf{M} \mathbf{h}_i \leq 0, \forall i \in S^{[k]} \\ & \quad \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \text{rank}(\mathbf{M}) = 1. \end{aligned} \tag{3.27}$$

对于准确检测均方差约束的可行性来说保证秩约束的可行性是十分关键的。它可以通过解如下的 DC 规划问题来解决，即最小化迹范数和谱范数之差：

$$\begin{aligned} \mathcal{P}_{3.1-S2} : & \underset{\mathbf{M}}{\text{minimize}} \quad \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 \\ & \text{subject to } \text{Tr}(\mathbf{M}) - \gamma_i \mathbf{h}_i^H \mathbf{M} \mathbf{h}_i \leq 0, \forall i \in S^{[k]} \\ & \quad \mathbf{M} \succeq 0, \quad \text{Tr}(\mathbf{M}) \geq 1. \end{aligned} \tag{3.28}$$

也就是说，对于集合 $S^{[k]}$ 如果问题 $\mathcal{P}_{3.1-S2}$ 的目标函数值能达到零，可以下结论说所有在 $S^{[k]}$ 中的设备可以被选择的同时满足均方差要求，即优化问题(3.26)对于 $S^{[k]}$ 是可行的。注意得到的解 \mathbf{M}^* 是严格秩为 1 的矩阵，可以通过 Cholesky 分解从中提取出一个可行的接收端波束成形向量 \mathbf{m} ，即 $\mathbf{M}^* = \mathbf{m} \mathbf{m}^H$ 。

本节所提出用来解决联邦学习中稀疏与低秩优化问题的 DC 表示框架总结在算法2中。由于 DC 规划问题依然是非凸的，下一节将会展示如何用 DC 算法求解 DC 规划问题 $\mathcal{P}_{3.1-S1}$ 和 $\mathcal{P}_{3.1-S2}$ 。

3.4.3 DC 算法及其复杂度与收敛性分析

本小节展示了一个具有收敛性保证的 DC 算法，其原理是依次求解原问题和对偶问题的凸放缩直到收敛。

3.4.3.1 DC 表示方法

DC 规划问题 $\mathcal{P}_{3.1-S1}$ 和 $\mathcal{P}_{3.1-S2}$ 是非凸的，有着 DC 目标函数和凸的约束条件。尽管 DC 函数是非凸的，但是其有着很好的问题结构，可以推导出对应的 DC 算法来求解 (Tao 和 An, 1997)。

算法 2 用于解联邦学习中选择设备的问题 $\mathcal{P}_{3.1}$ 的 DC 表示框架

1: **步骤一:** 诱导稀疏性

2: $k \leftarrow 0$

3: **while** 问题 $\mathcal{P}_{3.1-S1}$ 目标函数值非零 **do**

4: 解 DC 规划问题 $\mathcal{P}_{3.1-S1}$ 得到 \mathbf{x}

5: $k \leftarrow k + 1$

6: **end while**

7: **步骤二:** 检测可行性

8: 将 \mathbf{x} 的元素按照降序排列为 $x_{\pi(1)} \geq \dots \geq x_{\pi(M)}$

9: $k \leftarrow 1$

10: **while** 问题 $\mathcal{P}_{3.1-S2}$ 的目标函数值非零 **do**

11: $S^{[k]} \leftarrow \{\pi(k), \pi(k+1), \dots, \pi(M)\}$

12: 解 DC 规划问题 $\mathcal{P}_{3.1-S2}$ 得到 \mathbf{M}

13: $k \leftarrow k + 1$

14: **end while**

15: **输出:** \mathbf{m} (通过做 Cholesky 分解 $\mathbf{M} = \mathbf{m}\mathbf{m}^H$) , 以及选择设备的集合 $S^{[k]} = \{\pi(k), \pi(k+1), \dots, \pi(M)\}$

具体来说，问题 $\mathcal{P}_{3.1-S1}$ 可以等价地重新表述为

$$\underset{\mathbf{x}, \mathbf{M}}{\text{minimize}} \quad f_1 = \|\mathbf{x}\|_1 - \|\mathbf{x}\|_k + \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 + I_{(\mathbf{x}, \mathbf{M}) \in \mathcal{C}_1}, \quad (3.29)$$

而问题 $\mathcal{P}_{3.1-S2}$ 可以等价地重新表述为

$$\underset{\mathbf{M}}{\text{minimize}} \quad f_2 = \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 + I_{\mathbf{M} \in \mathcal{C}_2}. \quad (3.30)$$

这里 $\mathcal{C}_1, \mathcal{C}_2$ 是半正定锥，表示了优化问题 $\mathcal{P}_{3.1-S1}$ 和优化问题 $\mathcal{P}_{3.1-S2}$ 的约束条件。指示函数的定义为

$$I_{(\mathbf{x}, \mathbf{M}) \in \mathcal{C}_1} = \begin{cases} 0, & (\mathbf{x}, \mathbf{M}) \in \mathcal{C}_1 \\ +\infty, & \text{otherwise} \end{cases}. \quad (3.31)$$

DC 函数 f_1, f_2 可以记为 $f_1 = g_1 - h_1, f_2 = g_2 - h_2$. 其中

$$g_1 = \|\mathbf{x}\|_1 + \text{Tr}(\mathbf{M}) + I_{(\mathbf{x}, \mathbf{M}) \in \mathcal{C}_1}, \quad (3.32)$$

$$h_1 = \|\mathbf{x}\|_k + \|\mathbf{M}\|_2, \quad (3.33)$$

$$g_2 = \text{Tr}(\mathbf{M}) + I_{\mathbf{M} \in \mathcal{C}_2}, \quad (3.34)$$

$$h_2 = \|\mathbf{M}\|_2. \quad (3.35)$$

于是优化问题 (3.29) 和问题 (3.30) 可以统一表示为最小化两个凸函数之差的形式

$$\underset{\mathbf{X} \in \mathbb{C}^{m \times n}}{\text{minimize}} \quad f(\mathbf{X}) = g(\mathbf{X}) - h(\mathbf{X}). \quad (3.36)$$

对于复数域的变量 \mathbf{X} ，设计算法需要使用 Wirtinger 积分 (Bouboulis 等, 2012)。DC 算法是通过求解原问题和对偶问题来实现的，但由于原问题 (3.36) 及其对偶问题都是非凸的，需要用到凸松弛的方法。

3.4.3.2 DC 算法

根据 Fenchel 对偶性 (Rockafellar, 2015)，优化问题 (3.36) 的对偶问题是

$$\underset{\mathbf{Y} \in \mathbb{C}^{m \times n}}{\text{minimize}} \quad h^*(\mathbf{Y}) - g^*(\mathbf{Y}), \quad (3.37)$$

其中 g^* 和 h^* 分别是 g 和 h 的共轭函数。共轭函数的定义是

$$g^*(\mathbf{Y}) = \sup_{\mathbf{X} \in \mathbb{C}^{m \times n}} \langle \mathbf{X}, \mathbf{Y} \rangle - g(\mathbf{X}), \quad (3.38)$$

其中, $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Real}(\text{Tr}(\mathbf{X}^H \mathbf{Y}))$ 定义了两个矩阵的内积 (Bouboulis 等, 2012)。经过简化的 DC 算法是, 第 t 次迭代里去求解如下的原问题和对偶问题的凸近似(把凹的部分线性化):

$$\mathbf{Y}^{[t]} = \arg \inf_{\mathbf{Y} \in \mathcal{Y}} h^*(\mathbf{Y}) - [g^*(\mathbf{Y}^{[t-1]}) + \langle \mathbf{Y} - \mathbf{Y}^{[t-1]}, \mathbf{X}^{[t]} \rangle], \quad (3.39)$$

$$\mathbf{X}^{[t+1]} = \arg \inf_{\mathbf{X} \in \mathcal{X}} g(\mathbf{X}) - [h(\mathbf{X}^{[t]}) + \langle \mathbf{X} - \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle]. \quad (3.40)$$

根据 Fenchel biconjugation 定理 (Rockafellar, 2015), 式 (3.39) 可以重写为

$$\mathbf{Y}^{[t]} \in \partial_{\mathbf{X}^{[t]}} h, \quad (3.41)$$

$\partial_{\mathbf{X}^{[t]}} h$ 是 h 对于 \mathbf{X} 变量在 $\mathbf{X}^{[t]}$ 点的次梯度。

因此, 用于解问题 $\mathcal{P}_{3.1-S1}$ 的 DC 算法的迭代值 $\mathbf{x}^{[t]}, \mathbf{M}^{[t]}$ 是如下凸优化问题的解

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{M}}{\text{minimize}} \quad g_1 - \langle \partial_{\mathbf{x}^{[t-1]}} h_1, \mathbf{x} \rangle - \langle \partial_{\mathbf{M}^{[t-1]}} h_1, \mathbf{M} \rangle \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \mathbf{h}_i^H \mathbf{M} \mathbf{h}_i \leq x_i, \forall i = 1, \dots, M, \\ & \quad \mathbf{M} \succeq 0, \quad \text{Tr}(\mathbf{M}) \geq 1, \mathbf{x} \succeq 0. \end{aligned} \quad (3.42)$$

求解优化问题 $\mathcal{P}_{3.1-S2}$ 的迭代值 $\mathbf{M}^{[t]}$ 由下面的优化问题的解给出

$$\begin{aligned} & \underset{\mathbf{M}}{\text{minimize}} \quad g_2 - \langle \partial_{\mathbf{M}^{[t-1]}} h_2, \mathbf{M} \rangle \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \mathbf{h}_i^H \mathbf{M} \mathbf{h}_i \leq 0, \forall i \in S^{[k]}, \\ & \quad \mathbf{M} \succeq 0, \quad \text{Tr}(\mathbf{M}) \geq 1. \end{aligned} \quad (3.43)$$

$h_1 = \|\mathbf{x}\|_k$ 的次梯度可以通过下式来计算 (Gotoh 等, 2018)

$$\partial \|\mathbf{x}\|_k \text{第 } i \text{ 个元素} = \begin{cases} \text{sign}(x_i), & |x_i| \geq |x_{(k)}| \\ 0, & |x_i| < |x_{(k)}| \end{cases}. \quad (3.44)$$

$h_2 = \|\mathbf{M}\|_2$ 的次梯度由如下命题给出。

命题 3.4. $\|\mathbf{M}\|_2$ 的次梯度的值可为 $\mathbf{v}_1 \mathbf{v}_1^H$, 其中 $\mathbf{v}_1 \in \mathbb{C}^N$ 是最大特征值 $\sigma_1(\mathbf{M})$ 对应的特征向量。

证明. 对于半正定矩阵 \mathbf{M} , 其二范数是一个正交不变范数, 它的次微分等于 (Watson, 1992)

$$\partial\|\mathbf{M}\|_2 = \text{conv}\{\mathbf{V}\text{diag}(\mathbf{d})\mathbf{V}^H : \mathbf{d} \in \partial\|\boldsymbol{\sigma}(\mathbf{M})\|_\infty\}, \quad (3.45)$$

其中 conv 表示一个集合的凸包, $\mathbf{M} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^H$ 是矩阵 \mathbf{M} 的奇异值分解, $\boldsymbol{\sigma}(\mathbf{M}) = [\sigma_i(\mathbf{M})] \in \mathbb{C}^N$ 是一个由 \mathbf{M} 所有的奇异值组成的向量。由 $\sigma_1(\mathbf{M}) \geq \dots \geq \sigma_N(\mathbf{M}) \geq 0$ 可得

$$[1, \underbrace{0, \dots, 0}_{N-1}]^H \in \partial\|\boldsymbol{\sigma}(\mathbf{M})\|_\infty. \quad (3.46)$$

因此, $\mathbf{v}_1\mathbf{v}_1^H$ 是 $\|\mathbf{M}\|_2$ 的一个次梯度。 \square

3.4.3.3 计算复杂度和收敛性分析

本章所提的 DC 算法的计算开销包括了在步骤一中解一系列 DC 规划问题 $\mathcal{P}_{3.1-S1}$ 和在步骤二中解 DC 规划问题 $\mathcal{P}_{3.1-S2}$ 。步骤一需要对 k 从 0 到 M 解优化问题 $\mathcal{P}_{3.1-S1}$ 。每一个 DC 规划问题 $\mathcal{P}_{3.1-S1}$ 的求解需要迭代地解半正定规划问题 (3.42)。使用二阶内点法 (Boyd 和 Vandenberghe, 2004) 来求解问题 (3.42) 的每次迭代的计算复杂度是 $\mathcal{O}((N^2 + M)^3)$ 。在步骤二中, 问题 $\mathcal{P}_{3.1-S2}$ 需要通过迭代地求解半正定规划问题 (3.43) 来求解。使用内点法求解问题 (3.43) 的每次迭代的计算复杂度是 $\mathcal{O}(N^6)$ 。注意到在步骤一中 “reweighted+SDR” 方法需要迭代地求解一个半正定规划问题 (即 ℓ_2 最小化问题) 而 “ ℓ_1 +SDR” 方法只需要求解一个单独的半正定规划问题 (即 ℓ_1 最小化问题)。对于这两种算法步骤一中使用内点法解每个半正定规划问题的计算开销都是每次迭代 $\mathcal{O}((N^2 + M)^3)$ 。在步骤二中, “reweighted+SDR” 方法和 “ ℓ_1 +SDR” 方法都需要解一个单独的半正定规划问题, 其用内点法求解的每次迭代复杂度为 $\mathcal{O}(N^6)$ 。因此所提的 DC 算法为了获得更高质量的解, 有着比其他算法更高的计算复杂度。而 “reweighted+SDR” 方法比 “ ℓ_1 +SDR” 方法的复杂度更高。基于文献 (Tao 和 An, 1997) 可知, 求解问题 $\mathcal{P}_{3.1-S1}$ 和 $\mathcal{P}_{3.1-S2}$ 的 DC 算法可以从任意可行的初始点收敛到一个临界点 (critical point), 该收敛性证明详见附录B.2。

3.5 仿真结果分析

本节中通过丰富的数值仿真实验同现有最新的算法做了对比, 来验证所提算法对于联邦学习快速模型聚合方案的性能。每个设备与基站之间的信道系数

向量 \mathbf{h}_i 按独立同分布的复高斯分布生成，即 $\mathbf{h}_i \sim \mathcal{CN}(0, \mathbf{I})$. 平均的传输信噪比 (signal-to-noise-ratio, SNR) P_0/σ^2 设为 20 dB. 假设所有的设备本地都有同样多的数据样本，即 $|\mathcal{D}_1| = \dots = |\mathcal{D}_M|$, 预处理和后处理标量分别取为 $\phi_i = 1, \psi = 1/|\mathcal{S}|$.

3.5.1 可行性检测

考虑一个典型的物联网的设定，由 $M = 20$ 个活跃的移动设备用来做联邦学习，基站有 $N = 6$ 根天线。可行性检测，即检测所选的设备是否可行，其性能对于设备选择来说十分关键。故通过仿真先来评估一下 DC 算法用来检测选择所有设备时的可行性的收敛性。即令问题 $\mathcal{P}_{3.1-S2}$ 中 $\mathcal{S}^{[k]} = \{1, \dots, 20\}$. 在 $\gamma = 5$ dB 和 $\gamma = 3$ dB 的情况下其结果如图3.5所示。它显示出了在 $\gamma = 5$ dB 时目标函数可以下降到零，但是在 $\gamma = 3$ dB 时不能达到零。这表明了所提出的 DC 算法在 $\gamma = 5$ dB 时能够返回一个秩为 1 的解但是在 $\gamma = 3$ dB 时却不能。

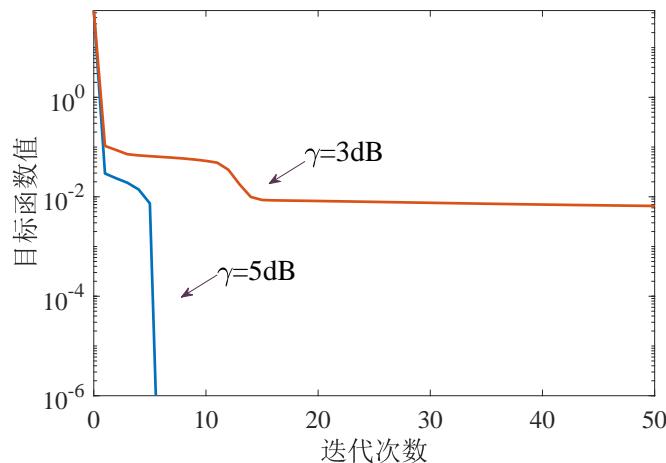


图 3.5 所提出的 DC 算法的收敛。

Figure 3.5 Convergence of the proposed DC algorithm.

接下来把所提出的 DC 方法的可行性检测（即解优化问题 $\mathcal{P}_{3.1-S2}$ ）的性能与下列方法进行比较：

- 半正定松弛法 (SDR) (Luo 等, 2007): 简单地忽略把问题 (3.26) 中的秩为 1 的约束去掉即为半正定松弛法，可用于检测可行性。
- 全局优化算法 (Lu 和 Liu, 2017): 该全局优化方法在最坏情况下有着指数级的时间复杂度 a 。我们将错误的容忍度门限设置为 $\epsilon = 10^{-5}$ 把它的性能作为基准。

平均了 500 次的结果如图3.6所示，其显示出所提的 DC 方法比半正定松弛方法

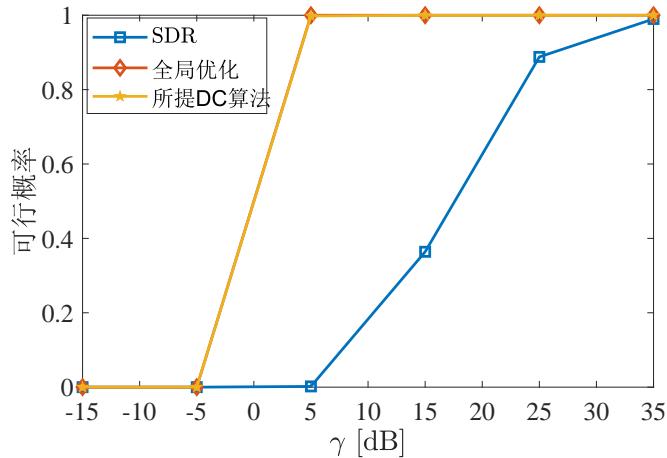


图 3.6 不同算法下可行性的概率。

Figure 3.6 Probability of feasibility with different algorithms.

要好的多，与全局优化方法比几乎达到了最优的性能，从而显示出了所提可行性检测方法的准确性。

接下来评估所提 DC 算法在不同的基站天线数目下的性能。图3.7显示了在不同的目标均方差要求下平均了 500 次信道实现的结果。它表明了通过增加基站端的天线数目可以使得从移动设备出进行快速模型聚合可以满足更严格的均方差要求。

3.5.2 所选设备数目与目标均方差大小的关系

考虑一个由 20 个移动设备和一个 6 根天线的基站组成的网络。依照算法2中所采用的两步骤框架和排序准则，将所提的 DC 算法2与如下最新算法相比较：

- ℓ_1 +SDR (Boyd 和 Vandenberghe, 2004) (Luo 等, 2007): ℓ_1 范数最小化用于在第一个步骤里诱导 \mathbf{x} 的稀疏性，而用半正定松弛来解决步骤一和步骤二中的非凸二次约束条件。

- Reweighted ℓ_2 +SDR (Shi 等, 2016b): 使用光滑化了的 ℓ_p 范数在步骤一里来诱导 \mathbf{x} 的稀疏性，其通过加权 ℓ_2 最小化算法来求解。而用半正定松弛方法来解决步骤一和步骤二里的非凸二次约束。

平均了 500 次信道实现的不同诱导稀疏性和检测可行性算法的平均性能如图3.8所示。它显示出了所提的 DC 算法能够比其他算法选择更多的设备。

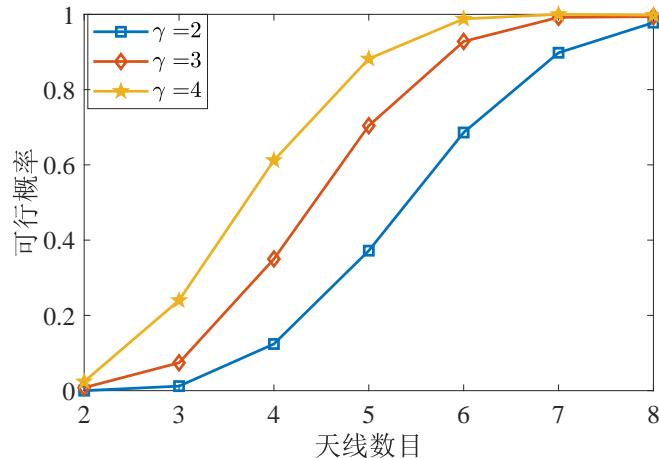


图 3.7 所提的 DC 方法的可行性概率与基站端天线数的关系。

Figure 3.7 Probability of feasibility over the number of BS antennas with the proposed DC approach.

3.5.3 所提 DC 方法对于联邦学习的性能

为了显示所提 DC 方法对于联邦学习中设备选择的性能, 在 CIFAR-10(Krizhevsky 和 Hinton, 2009) 数据集上训练了一个支撑向量机分类器, 在一个由 20 个移动设备和一个 6 天线基站的网络中仿真。CIFAR-10 数据集是一个被广泛用作图像分类的数据集, 它包含了十类不同类别的物体。将所有设备都选择并且模型聚合没有任何误差的情况作为比较的基准。设定 $\gamma = 5\text{dB}$, 运行所有的算法得到平均十次信道实现的结果如图3.9所示。其中训练集和测试集的大小分别选择为 50000 和 10000。仿真结果表明了所提的 DC 方法能够取得更低的训练损失值(如图3.9a), 和更高的预测准确度(如图3.9b)。

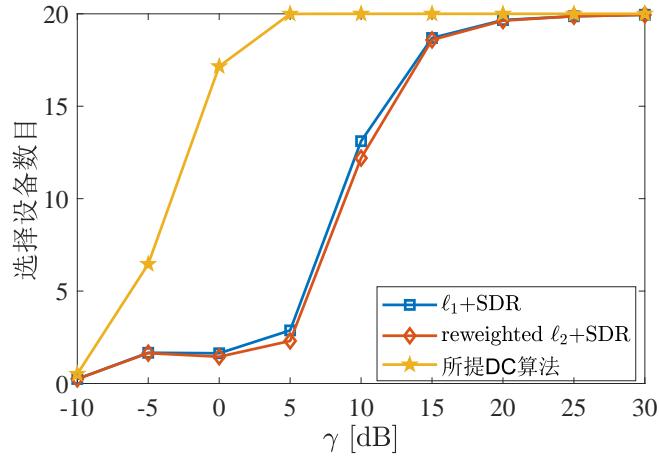


图 3.8 不同算法的平均选择设备数目。

Figure 3.8 Average number of selected devices with different algorithms.

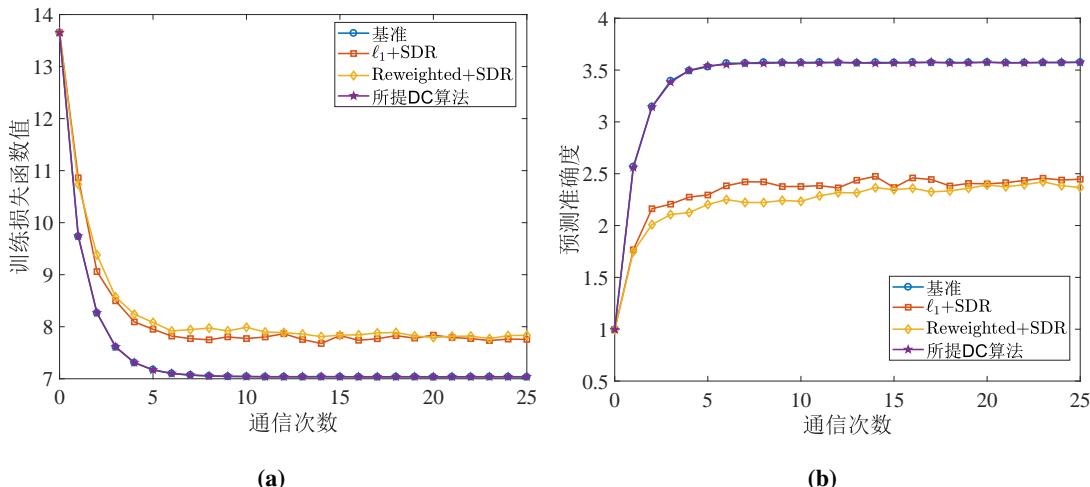


图 3.9 (a) FedAvg 算法在不同的设备选择算法下的收敛结果, (b) 通信轮数与所训练的全局模型在测试集的相对准确度的关系。

Figure 3.9 (a) Convergence of different device selection algorithms for FedAvg. (b) The relationship between communication rounds and test accuracy over random classification of the trained model.

3.6 本章小结

本章针对移动设备分布式联邦训练系统提出了一种新的快速模型聚合方法，核心观察是发现模型聚合恰好符合空中计算适用范围。为了提高模型训练的性能，使用了稀疏与低秩优化的建模方法来最大化选择参与模型聚合的设备数，保证聚合误差要求。然后提供了一个统一的差分凸函数优化框架来诱导稀疏与低秩性，并给出了算法的收敛性分析。仿真结果验证了所提方法相比现有算法的性

能优势。值得注意的是，尽管所提出的模型聚合方法能够降低所需的通信带宽，但是仍然受物理信道状况所限制。在下一章中，将针对这一问题设计有效的系统方案与传输方案，进一步加速模型聚合过程。

第4章 智能反射面赋能的移动设备分布式训练快速模型聚合策略

上一章中研究了移动设备分布式训练系统，根据模型聚合的计算特点设计了基于空中计算的快速模型聚合方法，降低了所需的传输带宽。然而该方案并不能避免无线信道可能的不良信号传播情况，受物理信道状况所限。在上一章的基础上，本章进一步提出使用智能反射面来改善不良信号传播状况，进而提高分布式训练的模型聚合效率。

本章内容安排如下：第4.1节中介绍了智能反射面的原理与作用，回顾了相关的部分研究，并指出了其对分布式训练的模型聚合的重要作用；第4.2节展示了智能反射面赋能的联邦学习系统，以及再加上空中计算技术来做快速模型聚合的问题表述；第4.3节提供了解决该问题的稀疏与低秩优化算法；第4.4节通过数值仿真验证了所提方法的性能，并对结果作出了分析；最后，在第4.5节中总结了本章工作，并提出了一些研究方向。

4.1 引言

尽管空中计算提供了一种利用无线信道中的干扰来降低频谱资源要求的方式，但其性能依然受限于不适宜的无线传播环境。智能反射面（Intelligent Reflecting Surface, IRS）(Yuan 等, 2020) 是一种成本超值的技术，其通过重新配置无线传播环境，极有潜力能更进一步提高频谱效率和能效，降低使用空中计算的模型聚合误差。智能反射面是一个平面的人造超表面，由许多具有可调相移的无源反射元件组成，每个元件都由智能控制器进行软件控制。通过共同控制所有反射元件，智能反射面能够在入射信号上引入所需的相移，可以利用该相移来增强信号功率和（或）减轻同频干扰。此外，事实证明，智能反射面具有能与许多新兴技术集成的巨大潜力，包括机器学习，大规模多输入多输出（MIMO），太赫兹通信和稀疏码多址（sparse code multiple access, SCMA）等(Han 等, 2019)。本章将开发一种由智能反射面赋能的新颖的多路访问方案，以提高第3章所提出移动设备分布式训练中的模型聚合的性能，从而探索由智能反射面增强的空中计算方法为智能应用设计通信效率高的分布式机器学习框架的潜力。

智能反射面中包含了大量的低成本无源反射元件，其能够调节入射信号的

相移，从而改变反射信号的传播。在接收端处原信号与反射信号混合，从而使得智能反射面技术可以显著提高接收信号的功率。文献 (Wu 和 Zhang, 2019) 中提出了一种联合主动波束成形（即发送端的波束成形）和被动波束成形（即智能反射面处的相移）设计的方法以在用户信噪比约束条件下最小化多天线无线接入点处的传输功率。文献 (Huang 等, 2019) 联合设计了下行链路传输功率和智能反射面的相移，以最大化系统的能效。在智能反射面赋能下，本章将设计一个更加高效的基于空中计算的快速模型聚合方法。如第3.2.1节所述，对于分布式训练系统的模型聚合的设计目标是在满足误差要求的情况下最大化选择的设备数目。因此，本章将需要联合设计所选择的移动设备、基站端接收波束成形器以及智能反射面处的相移，为了解决这个问题，提出了一种稀疏与地质优化的建模方法，进而设计了高效算法求解。

4.2 系统模型与问题表述

本节将展示智能反射面赋能的移动设备分布式联邦学习系统，然后给出如何用空中计算来做快速模型聚合。

4.2.1 智能反射面赋能的联邦学习系统

考虑如第3.2.1节所述的联邦学习系统，由一个有 M 根天线的基站和 K 个单天线的移动用户组成（见图3.1）。训练过程的完成需要周期性的收集每个移动设备处的模型更新 \mathbf{z}_i ，即

$$\mathbf{z} = \frac{1}{\sum_{i \in S} |\mathcal{D}_i|} \sum_{i \in S} |\mathcal{D}_i| \mathbf{z}_i, \quad (4.1)$$

\mathcal{D}_i 为第 i 个移动设备的本地数据集。而这个全局模型聚合过程带来了很大的通信开销，第3章中提出了基于空中计算的方案来降低所需的频谱资源，提高通信效率。为了克服信号可能遇到的有害传播环境，以进一步加速模型聚合，本章考虑为该联邦学习系统提供一个具有 N 反射元素的智能反射面，如图4.1所示。特别的，智能反射面利用大量的低成本被动反射元件，从而调整入射信号的相移，以实现与无线传播环境的联合配置。如文献 (Wu 和 Zhang, 2019) 所述，得益于信号反射，智能反射面赋能的系统中接收端信噪比可以得到阶数为 N^2 的增益。

在空中计算的框架下，预处理标量 $\phi_i = |\mathcal{D}_i|$ ，后处理标量 $\psi = \frac{1}{\sum_{i \in S} |\mathcal{D}_i|}$ 。不失一般性，设备 i 处的传输符号 \mathbf{s}_i 假定具有单位功率，即 $\mathbb{E}(\mathbf{s}_i \mathbf{s}_i^H) = \mathbf{I}$ 。每个设

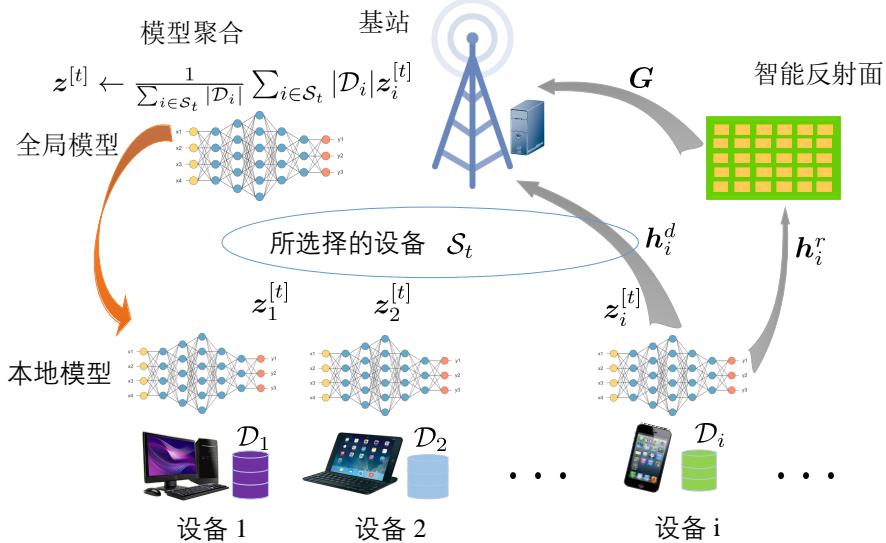


图 4.1 智能反射面赋能的移动设备分布式联邦学习系统。

Figure 4.1 Intelligence reflecting surface empowered on-device federated learning system.

备在第 $j \in \{1, \dots, d\}$ 个时隙发送信号 $s_i^{(j)} \in \mathbb{C}$ 给基站，简化掉时间索引为 s_i 后，通过空中计算的目标函数可以表示为 $g = \sum_{i \in S} \phi_i \cdot s_i$ 。用 $\mathbf{h}_i^d \in \mathbb{C}^M$, $\mathbf{h}_i^r \in \mathbb{C}^N$ 和 $\mathbf{G} \in \mathbb{C}^{M \times N}$ 来分别表示从设备 i 到基站，从设备 i 到智能反射面，以及从智能反射面到基站的信道系数向量。基站端收到的与反射信号复合的接收信号可以表示为

$$\mathbf{y} = \sum_{i \in S} (\mathbf{G}\Theta \mathbf{h}_i^r + \mathbf{h}_i^d) w_i s_i + \mathbf{n}, \quad (4.2)$$

其中 $\Theta = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_N}) \in \mathbb{C}^{N \times N}$ 是智能反射面的对角相移矩阵, $\theta_n \in [0, 2\pi]$ 。 $w_i \in \mathbb{C}$ 是发射机标量, $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ 是加性复高斯白噪声。

给定接收端波束成形向量 $\mathbf{m} \in \mathbb{C}^M$, 空中计算的目标是直接从接收信号中估计出目标函数, 即

$$\hat{g} = \frac{1}{\sqrt{\eta}} \mathbf{m}^H \mathbf{y} = \frac{1}{\sqrt{\eta}} \mathbf{m}^H \sum_{i \in S} (\mathbf{G}\Theta \mathbf{h}_i^r + \mathbf{h}_i^d) w_i s_i + \frac{1}{\sqrt{\eta}} \mathbf{m}^H \mathbf{n}, \quad (4.3)$$

其中 η 是归一化因子。进而可以对估计出的目标函数值做后处理 $\hat{z} = \psi \cdot \hat{g}$ 从而得到全局的模型聚合结果。所估计处的聚合模型的畸变用均方误差来衡量, 其表达式为

$$\text{MSE}(\hat{g}, g) = \mathbb{E}(|\hat{g} - g|^2) = \sum_{i \in S} \left| \frac{1}{\sqrt{\eta}} \mathbf{m}^H (\mathbf{G}\Theta \mathbf{h}_i^r + \mathbf{h}_i^d) w_i - \phi_i \right|^2 + \frac{\sigma^2 \|\mathbf{m}\|^2}{\eta}. \quad (4.4)$$

发射端标量由如下命题给出:

命题 4.1. 给定解码向量 \mathbf{m} 和相移矩阵 $\boldsymbol{\Theta}$, 最小均方差对应的最优发射端标量由下式给出:

$$w_i = \sqrt{\eta} \phi_i \frac{(\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d))^H}{\|\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2}, \forall i, \quad (4.5)$$

证明. 式 (4.5) 给出的发射端标量 $\{w_i\}$ 具有迫零结构, 即

$$\sum_{i \in \mathcal{S}} |\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d) w_i - \phi_i|^2 = 0. \quad (4.6)$$

另外由均方差的表达式(4.4)可知 $\text{MSE}(\hat{g}, g) \geq \sigma^2 \|\mathbf{m}\|^2 / \eta$.

于是可以得到命题4.5中的迫零发射机标量可以给出最小均方差。 \square

由于每个设备有着最大发送功率限制 $P_0 > 0$, 即 $|w_i|^2 \leq P_0$. 于是归一化因子 η 应为 $\eta = P_0 \min_{i \in \mathcal{S}} \|\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2 / \phi_i^2$, 最小均方差则可以表达为

$$\text{MSE} = \frac{\sigma^2}{P_0} \max_{i \in \mathcal{S}} \phi_i^2 \frac{\|\mathbf{m}\|^2}{\|\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2}. \quad (4.7)$$

4.2.2 问题表述

在联邦学习的模型聚合中, 选择更多的移动设备与降低聚合误差两方面均可以提高其性能。故可以将使用空中计算做模型聚合的问题建模成一个最大化选择设备数的同时满足均方差要求的组合优化问题:

$$\begin{aligned} & \underset{\mathcal{S}, \mathbf{m}, \boldsymbol{\Theta}}{\text{maximize}} \quad |\mathcal{S}| \\ & \text{subject to} \left(\max_{i \in \mathcal{S}} \phi_i^2 \frac{\|\mathbf{m}\|^2}{\|\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2} \right) \leq \gamma, \end{aligned} \quad (4.8)$$

$$0 \leq \theta_n \leq 2\pi, \forall n = 1, \dots, N, \quad (4.9)$$

其中 $\gamma > 0$ 是所要求的均方差, $\boldsymbol{\Theta} = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_N})$. 为助力算法设计, 可以讲均方差约束(4.8)重写为 $F_i(\mathbf{m}) = \|\mathbf{m}\|^2 - \gamma_i \|\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2 \leq 0, i \in \mathcal{S}$, 其中 $\gamma_i = \gamma / \phi_i^2, \mathbf{m} \neq 0$. 进一步将其重写为 $F_i(\mathbf{m}/\sqrt{\tau}) = F_i(\mathbf{m})/\tau \leq 0, i \in \mathcal{S}$, 其中 $\|\mathbf{m}\|^2 \geq \tau, \tau > 0$. 通过引入变量 $\tilde{\mathbf{m}} = \mathbf{m}/\sqrt{\tau}$ 可以将均方差约束(4.8)等价转化为

$$\|\mathbf{m}\|^2 - \gamma_i \|\mathbf{m}^H (\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2 \leq 0, i \in \mathcal{S}, \quad (4.10)$$

$$\|\mathbf{m}\|^2 \geq 1. \quad (4.11)$$

观察到最大化设备数目等价于最大化可行的均方差约束(4.10)的数目，可以将其建模为一个稀疏优化问题

$$\begin{aligned} \mathcal{P}_{4.1} : & \underset{\mathbf{x}, \mathbf{m}, \boldsymbol{\Theta}}{\text{minimize}} \quad \|\mathbf{x}\|_0 \\ & \text{subject to } \|\mathbf{m}\|^2 - \gamma_i \|\mathbf{m}^H(\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r + \mathbf{h}_i^d)\|^2 \leq x_i, \forall i, \end{aligned} \quad (4.12)$$

$$\|\mathbf{m}\|^2 \geq 1, \mathbf{x} \geq 0, \quad (4.13)$$

$$0 \leq \theta_n \leq 2\pi, \forall n = 1, \dots, N. \quad (4.14)$$

由于智能反射面的存在，其与优化问题(3.18)的差异在于其约束条件对 \mathbf{m} 和 \mathbf{h}_i 是非凸的双二次的，即对二者分别为非凸的二次项。再加上稀疏目标函数，优化问题 $\mathcal{P}_{4.1}$ 很难求解。为了解决这个问题，本章将使用一个两步骤的方法来解决稀疏目标函数，用交替低秩优化方法来解决非凸的双二次约束。

4.3 稀疏与低秩优化算法框架

本章提出了一个两步骤的稀疏与低秩优化算法框架来解决优化问题 $\mathcal{P}_{4.1}$ ，其中使用了 ℓ_1 范数松弛来诱导稀疏性，然后用矩阵升维和交替低秩优化方法来解决非凸的双二次约束。

4.3.1 解决稀疏目标函数的两步骤算法框架

为了解决对 \mathbf{m} 和 $\boldsymbol{\Theta}$ 的双二次非凸约束条件，可以使用矩阵升维技术先将其变成双线性约束，然后利用交替优化方法来求解。具体来说，令 $v_n = e^{j\theta_n}$ 可得

$$\mathbf{G}\boldsymbol{\Theta}\mathbf{h}_i^r = \mathbf{G}\text{diag}(\mathbf{h}_i^r)\mathbf{v} = \mathbf{a}_i^H\mathbf{v}, \quad (4.15)$$

其中 $\mathbf{v} = [e^{j\theta_1}, \dots, e^{j\theta_N}]^T$. 问题 $\mathcal{P}_{4.1}$ 中的约束条件(4.12)和(4.14)可以重写为

$$\|\mathbf{m}\|^2 - \gamma_i \|\mathbf{m}^H(\mathbf{a}_i^H\mathbf{v} + \mathbf{h}_i^d)\|^2 \leq x_i, \forall i, \quad (4.16)$$

$$|v_n|^2 = 1, \forall n = 1, \dots, N. \quad (4.17)$$

定义一个向量 $\tilde{\mathbf{v}} = [\mathbf{v}, t]^T$, t 为辅助变量。将 $\tilde{\mathbf{v}}$ 升维为一个秩为一的半正定矩阵 $\mathbf{V} = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H$ 同样可以将向量 \mathbf{m} 也升维为一个秩为一的半正定矩阵 $\mathbf{M} = \mathbf{m}\mathbf{m}^H$. 除此之外，采用著名的 ℓ_1 范数来作为非凸 ℓ_0 范数的凸松弛 (Boyd 和 Vandenberghe, 2004).

用于求解优化问题 $\mathcal{P}_{4.1}$ ，所提的两步骤算法框架的基本思想是：

1. 第一步：用 ℓ_1 范数作为目标函数来诱导稀疏解 \mathbf{x} ，得到选择每个设备的一个优先级；
2. 第二步：求解一系列的可行性检测问题，直至找到在满足均方差要求下能够选择的最大数目的设备。

在第一步中，需要求解如下的低秩矩阵优化问题

$$\begin{aligned} \mathcal{P}_{4.1-S1} : & \underset{\mathbf{x}, \mathbf{M}, \mathbf{V}}{\text{minimize}} \quad \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \text{Tr}(\mathbf{M} \mathbf{A}_i \mathbf{V} \mathbf{A}_i^H) \leq x_i, \forall i, \\ & \quad \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \text{rank}(\mathbf{M}) = 1, \\ & \quad V_{n,n} = 1, \forall n = 1, \dots, N+1, \\ & \quad \mathbf{V} \succeq 0, \text{rank}(\mathbf{V}) = 1, \\ & \quad \mathbf{x} \succeq 0, \end{aligned} \quad (4.18)$$

其中 $\mathbf{A}_i = [\mathbf{a}_i^H, \mathbf{h}_i^d]$. 其解 \mathbf{x} 的每个元素 x_k 刻画了要求的均方差与设备 k 可以达到的均方差之间的差距，故而可以将 x_k 按降序排列后 $x_{\pi(1)} \geq \dots \geq x_{\pi(K)}$ ，较小的 x_k 将被安排更高被选择的优先级。然后通过从 1 到 K 增加 k 来找到最小的 k ，使得选择所有在集合 $S^{[k]} = \{\pi(k), \pi(k+1), \dots, \pi(K)\}$ 内的设备均方差要求可以被满足。检测选择所有在集合 $S^{[k]}$ 里的设备是否可行需要求解如下优化问题

$$\begin{aligned} \mathcal{P}_{4.1-S2} : & \text{find } \mathbf{M}, \mathbf{V} \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \text{Tr}(\mathbf{M} \mathbf{A}_i \mathbf{V} \mathbf{A}_i^H) \leq 0, \forall i \in S^{[k]}, \\ & \quad \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \text{rank}(\mathbf{M}) = 1, \\ & \quad V_{n,n} = 1, \forall n = 1, \dots, N+1, \\ & \quad \mathbf{V} \succeq 0, \text{rank}(\mathbf{V}) = 1. \end{aligned} \quad (4.19)$$

该问题的可行性即对应了选择 $S^{[k]}$ 是否可行。

4.3.2 解决非凸约束的交替低秩优化方法

均方差约束条件(4.18)和(4.19)均为非凸，且变量 \mathbf{M} 和 \mathbf{V} 是高度耦合的。幸运的是，可以观察到其对 \mathbf{M} 和 \mathbf{V} 是凸的仿射约束，故而可以用交替优化方法 (Jiang 和 Shi, 2019) 来解决该非凸约束。即交替的优化基站端的变量 (\mathbf{x}, \mathbf{M}) 与智能反射面端的变量 Θ ，二者中固定其中一个求解另外一个均为一个有秩约束

的半正定规划问题。固定相移矩阵 Θ , 即设备 i 与基站端的等效信道系数向量 $\mathbf{G}\Theta\mathbf{h}_i^r + \mathbf{h}_i^d$ 固定时, 通过求解如下问题来更新变量 (\mathbf{x}, \mathbf{M})

$$\begin{aligned} \mathcal{P}_{4.1-S1.1} : & \underset{\mathbf{x}, \mathbf{M}}{\text{minimize}} \quad \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \text{Tr}(\mathbf{M}\mathbf{H}_i) \leq x_i, \forall i, \\ & \quad \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \text{rank}(\mathbf{M}) = 1, \\ & \quad \mathbf{x} \geq 0, \end{aligned}$$

其中 $\mathbf{H}_i = (\mathbf{G}\Theta\mathbf{h}_i^r + \mathbf{h}_i^d)(\mathbf{G}\Theta\mathbf{h}_i^r + \mathbf{h}_i^d)^H$. 然后固定波束成形向量 \mathbf{M} 与稀疏向量 \mathbf{x} , 通过求解如下问题来更新 Θ

$$\begin{aligned} \mathcal{P}_{4.1-S1.2} : & \text{find } \mathbf{V} \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i (\text{Tr}(\mathbf{R}_i \mathbf{V}) + c_i^2) \leq x_i, \forall i, \\ & \quad V_{n,n} = 1, \forall n = 1, \dots, N+1, \\ & \quad \mathbf{V} \succeq 0, \text{rank}(\mathbf{V}) = 1, \end{aligned}$$

其中

$$c_i = \mathbf{m}^H \mathbf{h}_i^d, \mathbf{b}_i^H = \mathbf{m}^H \mathbf{a}_i^H, \text{and } \mathbf{R}_i = \begin{bmatrix} \mathbf{b}_i \mathbf{b}_i^H, & \mathbf{b}_i c_i \\ c_i^H \mathbf{b}_i^H, & 0 \end{bmatrix}. \quad (4.20)$$

交替执行这两个步骤直到算法收敛即可。同样地, 优化问题 $\mathcal{P}_{4.1-S2}$ 也可以通过交替更新 \mathbf{M} 和 Θ 来求解。固定 Θ , 通过解如下问题来更新 \mathbf{M}

$$\begin{aligned} \mathcal{P}_{4.1-S2.1} : & \text{find } \mathbf{M} \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \text{Tr}(\mathbf{M}\mathbf{H}_i) \leq 0, \forall i \in \mathcal{S}^{[k]}, \\ & \quad \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \text{rank}(\mathbf{M}) = 1, \end{aligned}$$

固定 \mathbf{M} , 通过解如下问题来更新 Θ :

$$\begin{aligned} \mathcal{P}_{4.1-S2.2} : & \text{find } \mathbf{V} \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i (\text{Tr}(\mathbf{R}_i \mathbf{V}) + c_i^2) \leq 0, \forall i \in \mathcal{S}^{[k]}, \\ & \quad V_{n,n} = 1, \forall n = 1, \dots, N+1, \\ & \quad \mathbf{V} \succeq 0, \text{rank}(\mathbf{V}) = 1, \end{aligned}$$

交替低秩优化方法产生的子优化问题 $\mathcal{P}_{4.1-S1.1}$ 、 $\mathcal{P}_{4.1-S1.2}$ 、 $\mathcal{P}_{4.1-S2.1}$ 和 $\mathcal{P}_{4.1-S2.2}$ 有着秩约束, 因此依然是非凸的。虽然可以使用半正定松弛方法直接忽略掉非凸

的秩约束求解该问题，但是其会随着天线数或者反射单元数目的增加而有较低的概率返回满足秩为一的解，从而导致性能损失。为了解决这一问题，本章采用 DC 算法来诱导秩为一的解。

4.3.3 解决低秩约束的 DC 算法

秩为一约束的诱导对于精确检测非凸的双二次约束的可行性起着至关重要的作用，故对所提的两步骤的设备选择算法框架十分关键。因此，这里采用半正定矩阵的秩为一约束的 DC 表示，即

$$\text{rank}(\mathbf{M}) = 1 \Leftrightarrow \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 = 0 \quad (4.21)$$

其中 $\text{Tr}(\mathbf{M}) \geq 1$ 。于是可以添加 DC 正则项来诱导优化问题 $\mathcal{P}_{4.1-S1.1}$ 的秩为一的解，即求解

$$\begin{aligned} \mathcal{P}_{4.1-S1.1'} : & \underset{\mathbf{x}, \mathbf{M}}{\text{minimize}} \quad \|\mathbf{x}\|_1 + \rho (\text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2) \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \text{Tr}(\mathbf{M} \mathbf{H}_i) \leq x_i, \forall i, \\ & \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \mathbf{x} \succeq 0. \end{aligned} \quad (4.22)$$

类似的，通过求解如下问题来检测优化问题 $\mathcal{P}_{4.1-S1.2}$ 的可行性

$$\begin{aligned} \mathcal{P}_{4.1-S1.2'} : & \underset{\mathbf{V}}{\text{minimize}} \quad \text{Tr}(\mathbf{V}) - \|\mathbf{V}\|_2 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i (\text{Tr}(\mathbf{R}_i \mathbf{V}) + c_i^2) \leq x_i, \forall i, \\ & \mathbf{V} \succeq 0, V_{n,n} = 1, \forall n = 1, \dots, N+1. \end{aligned} \quad (4.23)$$

同样地，第二步可行性检测的子优化问题 $\mathcal{P}_{4.1-S2.1}$ 和 $\mathcal{P}_{4.1-S2.2}$ 可以通过求解

$$\begin{aligned} \mathcal{P}_{4.1-S2.1'} : & \underset{\mathbf{M}}{\text{minimize}} \quad \text{Tr}(\mathbf{M}) - \|\mathbf{M}\|_2 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i \text{Tr}(\mathbf{M} \mathbf{H}_i) \leq 0, \forall i \in S^{[k]}, \\ & \mathbf{M} \succeq 0, \text{Tr}(\mathbf{M}) \geq 1, \end{aligned} \quad (4.24)$$

和

$$\begin{aligned} \mathcal{P}_{4.1-S2.2'} : & \underset{\mathbf{V}}{\text{minimize}} \quad \text{Tr}(\mathbf{V}) - \|\mathbf{V}\|_2 \\ & \text{subject to} \quad \text{Tr}(\mathbf{M}) - \gamma_i (\text{Tr}(\mathbf{R}_i \mathbf{V}) + c_i^2) \leq 0, \forall i \in S^{[k]}, \\ & \mathbf{V} \succeq 0, V_{n,n} = 1, \forall n = 1, \dots, N+1, \end{aligned} \quad (4.25)$$

来解决。如果目标函数值能够达到零，则说明可以找到一个秩为一的可行解 \mathbf{M}^* 和 \mathbf{V}^* ，然后通过 Cholesky 分解 $\mathbf{M}^* = \mathbf{m}\mathbf{m}^H$ 提取出接收端波束成形向量 \mathbf{m} ，以及从 $\mathbf{V}^* = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H$ 和 $\tilde{\mathbf{v}} = [\mathbf{v}^0, t^0]^T$ 中提取出智能反射面的相移矩阵 $\Theta = \text{diag}(\mathbf{v}) = \text{diag}(\mathbf{v}^0/t^0)$ 。

非凸的 DC 规划问题具有良好的问题结构，其非凸性是由于两个凸函数之差造成的。DC 算法很好的利用了这种结构，通过连续线性化凹函数部分来求解。具体来说，优化问题 $\mathcal{P}_{4.1-S1.1'}$ 和优化问题 $\mathcal{P}_{4.1-S1.2'}$ 的目标函数可以分别记为 $g_1 - h_1$ 和 $g_2 - h_2$ ，其中

$$g_1 = \|\mathbf{x}\|_1 + \rho \text{Tr}(\mathbf{M}), \quad h_1 = \rho \|\mathbf{M}\|_2 \quad (4.26)$$

$$g_2 = \text{Tr}(\mathbf{V}), \quad h_2 = \|\mathbf{V}\|_2 \quad (4.27)$$

在第 t 步迭代中，问题 $\mathcal{P}_{4.1-S1.1'}$ 的凹函数部分 $-h_1$ 可以线性化为 $-\rho \langle \partial_{\mathbf{M}^{t-1}} \|\mathbf{M}\|_2, \mathbf{M} \rangle$ ，而问题 $\mathcal{P}_{4.1-S1.2'}$ 的凹函数部分 $-h_2$ 可以线性化为 $-\langle \partial_{\mathbf{V}^{t-1}} h_2, \mathbf{V} \rangle$ 。这里 $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Real}(\text{Tr}(\mathbf{X}^H \mathbf{Y}))$ 是两个矩阵的内积。 $\partial_{\mathbf{M}^{t-1}} \|\mathbf{M}\|_2 = \partial \|\mathbf{M}\|_2$ 是 $\|\mathbf{M}\|_2$ 在 \mathbf{M}^{t-1} 点的次梯度， $\partial_{\mathbf{V}^{t-1}} h_2 = \partial \|\mathbf{V}\|_2$ 。而谱范数 $\|\mathbf{V}\|$ 的次梯度可以计算为 $\mathbf{v}_1 \mathbf{v}_1^H$ ，其中 \mathbf{v}_1 是最大的特征值对应的特征向量。

同样的，可以将 DC 规划问题 $\mathcal{P}_{4.1-S2.1'}$ 和优化问题 $\mathcal{P}_{4.1-S2.2'}$ 的目标函数分别记为 $g_3 - h_3$ 和 $g_4 - h_4$ ，其中

$$g_3 = \text{Tr}(\mathbf{M}), \quad h_3 = \|\mathbf{M}\|_2 \quad (4.28)$$

$$g_4 = \text{Tr}(\mathbf{V}), \quad h_4 = \|\mathbf{V}\|_2. \quad (4.29)$$

在第 t 步迭代中，问题 $\mathcal{P}_{4.1-S2.1'}$ 的目标函数的凹函数部分 $-h_3$ 可以线性化为 $-\langle \partial_{\mathbf{M}^{t-1}} h_3, \mathbf{M} \rangle$ ，问题 $\mathcal{P}_{4.1-S2.2'}$ 的凹函数部分可以线性化为 $-\langle \partial_{\mathbf{V}^{t-1}} h_4, \mathbf{V} \rangle$ 。目标函数的凹函数部分 $-h$ 线性化后，所得到的子问题是一个凸优化问题，可以被很高效的解决。该方法可以从任意可行的初始点收敛到临界点。整个所提的用于解决优化问题 $\mathcal{P}_{4.1}$ 的两阶段的稀疏与低秩优化算法可以总结为算法3。

算法 3 所提的用于解决问题 $\mathcal{P}_{4.1}$ 的两步骤的稀疏与低秩优化算法

- 1: **初始化:** $\boldsymbol{\Theta}^1, \epsilon > 0$.
- 2: **步骤一:** 诱导稀疏性
- 3: **for** $t = 1, 2, \dots$ **do**
- 4: 固定 $\boldsymbol{\Theta}^t$, 用 DC 算法求解优化问题 $\mathcal{P}_{4.1-S1.1'}$ 得到解 $\mathbf{M}^t, \mathbf{x}^t$.
- 5: 固定 \mathbf{M}^t 和 \mathbf{x}^t , 用 DC 算法求解优化问题 $\mathcal{P}_{4.1-S1.2'}$ 得到解 $\boldsymbol{\Theta}^{t+1}$.
- 6: **if** 均方差下降值低于 ϵ 或者问题 $\mathcal{P}_{4.1-S1.2}$ 不可行 **then**
- 7: **break for**
- 8: **end if**
- 9: **end for**
- 10: **步骤二:** 检测可行性
- 11: 对 \mathbf{x} 进行降序排序得到 $x_{\pi(1)} \geq \dots \geq x_{\pi(K)}$
- 12: $k \leftarrow 1$
- 13: **while** 问题 $\mathcal{P}_{4.1-S2.1'}$ 或者 $\mathcal{P}_{4.1-S2.2'}$ 不能达到零 **do**
- 14: $S^{[k]} \leftarrow \{\pi(k), \pi(k+1), \dots, \pi(K)\}$
- 15: **for** $t = 1, 2, \dots$ **do**
- 16: 固定 $\boldsymbol{\Theta}^t$, 用 DC 算法求解优化问题 $\mathcal{P}_{4.1-S2.1'}$ 得到解 \mathbf{M}^t .
- 17: 固定 \mathbf{M}^t , 用 DC 算法求解优化问题 $\mathcal{P}_{4.1-S2.2'}$ 得到解 $\boldsymbol{\Theta}^{t+1}$.
- 18: **if** 均方差下降值低于 ϵ , 或者问题 $\mathcal{P}_{4.1-S2.1}$ 或 $\mathcal{P}_{4.1-S2.2}$ 不可行 **then**
- 19: **break for**
- 20: **end if**
- 21: **end for**
- 22: $k \leftarrow k + 1$
- 23: **end while**
- 24: 对得到的解做 Cholesky 分解即, $\mathbf{M}^* = \mathbf{m}\mathbf{m}^H, \mathbf{V}^* = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^H, \tilde{\mathbf{v}} = [\mathbf{v}^0, t^0]^T$.
- 25: **输出:** $\mathbf{m}, \mathbf{v} = \mathbf{v}^0/t^0$, 和选择的设备索引集合 $S^{[k]} = \{\pi(k), \pi(k+1), \dots, \pi(K)\}$

4.4 仿真验证与结果分析

本节使用数值仿真实验来评估所提智能反射面赋能的分布式训练系统的模型聚合问题中所提的两步骤稀疏与低秩优化算法性能。考虑一个三维坐标系统，基站的 $M = 6$ 根天线呈均匀的线性阵列，智能反射面的 $N = 40$ 个反射单元呈均匀的矩形阵列。基站和反射面的位置坐标分别为 $(0, 0, 0)$ 和 $(0, 100, 0)$ ，单位为米。 $K = 20$ 个移动用户分布在 $[-5, 5] \times [95, 105]$ 的区域内（高度坐标为 0），围绕着反射面。路径损耗模型设置为 $L(d) = T_0(d/d_0)^{-\alpha}$ ，其中 T_0 表示参考距离 $d_0 = 1\text{m}$ 处的路损， d 是链路距离，而 α 是路损系数。假设为锐利衰落信道，即

$$\mathbf{h}_i^d = \sqrt{L(d_i^d)} \boldsymbol{\gamma}^d \quad (4.30)$$

$$\mathbf{h}_i^r = \sqrt{L(d_i^r)} \boldsymbol{\gamma}^r \quad (4.31)$$

$$\mathbf{G} = \sqrt{L(d_{IB})} \boldsymbol{\Gamma} \quad (4.32)$$

$$\boldsymbol{\gamma}^d \sim \mathcal{CN}(0, \mathbf{I}), \boldsymbol{\gamma}^r \sim \mathcal{CN}(0, \mathbf{I}), \boldsymbol{\Gamma} \sim \mathcal{CN}(0, \mathbf{I}). \quad (4.33)$$

这里 d_i^d, d_i^r, d_{IB} 分别表示设备 i 到基站的距离、设备 i 到智能反射面的距离以及基站与反射面之间的距离。其他的参数取为 $T_0 = -30\text{dB}$, $P_0 = 20\text{dBm}$, $\sigma^2 = -80\text{dBm}$, $\rho = 20$, $\epsilon = 10^{-3}$.

4.4.1 可行性检测性能评估

因为可行性检测步骤对于设备选择来说至关重要，所以首先比较了解决问题 \mathcal{P}_2 的所提方法与交替半正定松弛方法的性能。所提方法为交替用 DC 正则化方法求解优化问题 $\mathcal{P}_{4.1-S2.1}$ 和 $\mathcal{P}_{4.1-S2.2}$ ，而交替半正定松弛法通过忽略其中的秩约束来交替求解优化问题 $\mathcal{P}_{4.1-S2.1}$ 和 $\mathcal{P}_{4.1-S2.2}$ 。平均了 100 次信道实现，统计返回可行解概率，实验结果如图4.2a所示，结果显示所提的交替 DC 算法性能大大优于交替半正定松弛算法，这意味着能够更加准确的检测出可行性。进一步，评估基站的天线数目对于所提算法性能的影响。在目标均方差要求为 $-8, -9, -10-11\text{dB}$ 四个值下，使用所提的交替 DC 算法平均 100 次实验得到的返回可行性概率如图4.2b所示。结果显示在给定的均方差要求下，增加天线数能够有效的增加设备选择的可行性。

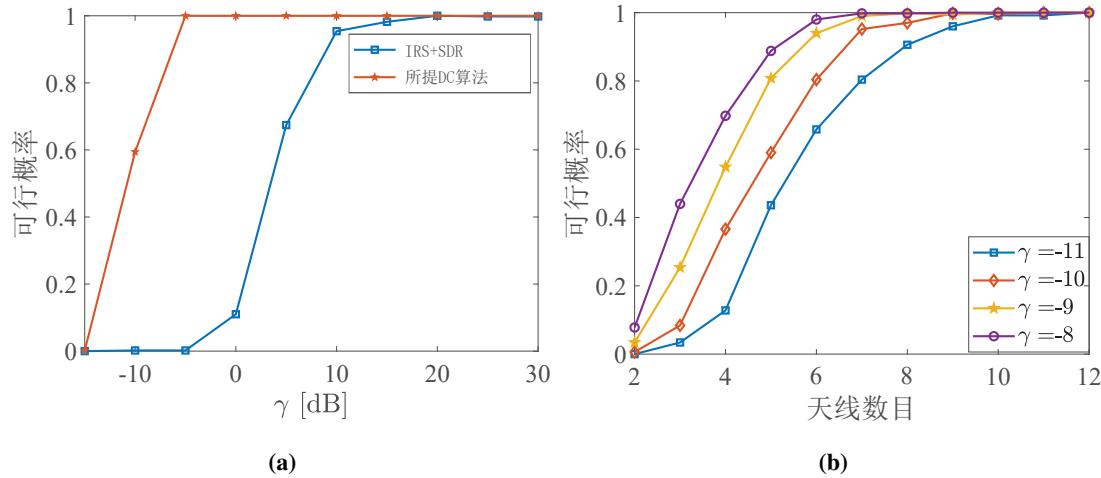


图 4.2 所提算法的检测可行性的性能评估。 (a) 对比半正定松弛算法 (SDR) 的可行性的概率，(b) 所提的 DC 方法的可行性概率与基站端天线数的关系。

Figure 4.2 Performance evaluation in terms of the probability of feasibility. (a) Probability of feasibility of the proposed approach compared with SDR method. (b) Probability of feasibility of the proposed approach over the number of antennas at the base station.

4.4.2 设备选择性能评估

然后，比较所提的两步骤的稀疏与低秩优化算法在给定目标均方差要求下的选择设备数目与使用下列方法的性能差异：

- 智能反射面赋能系统 + 半正定松弛法 (IRS+SDR)：本方法应用在有智能反射面赋能的分布式训练系统中，用半正定松弛法来解决非凸双二次约束；
- 无智能反射面赋能系统 + 半正定松弛法 (SDR without IRS)：相比“半正定松弛法 + 智能反射面”方法，本方法中移除了智能反射面，考虑的是如第3章所述的系统和问题表述，使用半正定松弛法来解决非凸二次约束；
- 无智能反射面赋能系统 +DC 算法 (DC without IRS)：本方法为第3章所提出的没有智能反射面的分布式训练系统，使用及其中所提出的 DC 算法，使用了 DC 算法来解决非凸二次约束。

注意以上所有算法中均使用了 ℓ_1 范数来诱导稀疏性。从图4.3中可以观察到，所提的两步骤的稀疏与低秩优化算法比使用半正定松弛法能够选择更多的移动设备。同时，智能反射面使得系统的性能有了相当大的提升，使得给定目标均方差要求时可以让更多的设备参与进模型聚合中。

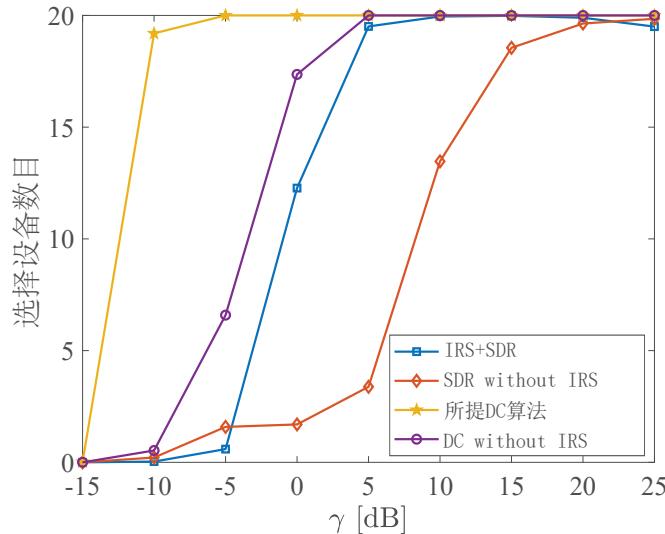


图 4.3 平均选择设备数目与均方误差关系曲线图。

Figure 4.3 Average number selected devices over target MSE.

4.4.3 联邦学习性能评估

使用该 20 个单天线移动设备组成的分布式联邦机器学习系统来完成一个支持向量机分类器的训练任务。采用广泛使用的图像分类数据集 CIFAR-10([Krizhevsky 和 Hinton, 2009](#))，将其均匀随机分割并部署在移动设备上。将所提 DC 方法与“IRS+SDR”、“SDR without IRS”、“DC without IRS”三种系统与算法相对比，得到使用 FedAvg 算法训练支持向量机分类器的性能。同时另外选择了选择所有设备进行无误差模型聚合作为基准性能。仿真结果如图4.4所示，其中4.4a给出了训练集上的损失函数值，而图4.4b显示了所得全局模型在测试集上的预测准确度。结果表明智能反射面赋能的模型聚合可以达到更低的训练损失值与更高的预测准确度，同时所采用的 DC 算法比半正定松弛有着更优异的性能。

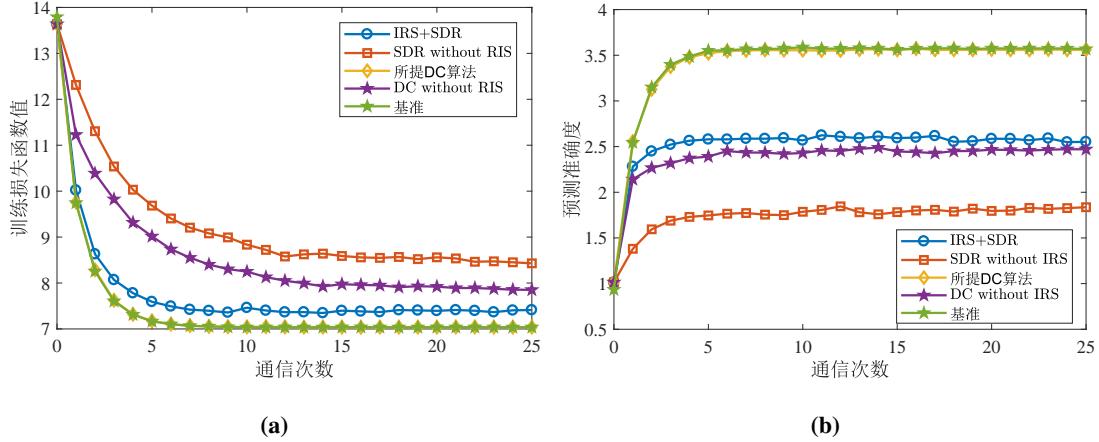


图 4.4 所提智能反射面赋能分布式训练系统模型聚合算法与其他方法性能对比。**(a)** FedAvg 算法在不同的设备选择算法下的收敛结果, **(b)** 通信轮数与所训练的全局模型在测试集的相对准确度的关系。

Figure 4.4 Performance comparisons of different algorithms for fast model aggregation in IRS-empowered on-device federated learning. (a) Convergence comparison for FedAvg between different device selection approaches. (b) Relationship between communication rounds and test accuracy.

4.5 本章小结

本章提出了一种新的智能反射面赋能的空中计算方法, 用于移动设备分布式训练的快速模型聚合。为了确保人工智能模型的性能, 设计了稀疏与低秩优化框架来实现最大化参与模型聚合的设备数目且保证聚合误差要求。其中稀疏性用于设备选择, 而使用交替的低秩优化来解决非凸的双二次约束。仿真结果验证了使用智能反射面技术对于快速模型聚合的巨大性能提升潜力, 也证明了结合智能反射面与空中计算技术, 可以有效降低边缘训练对无线网络带宽的要求。

第5章 基于移动设备分布式计算的边缘推断的快速数据交换策略

前两章就移动边缘分布式训练系统进行了研究，提出了针对性方案以提高模型聚合的通信效率，降低带宽需求。为了支持低延迟的移动边缘人工智能应用，本章将考虑实现边缘推断的使能方案。由于大尺寸的人工智能模型难以直接部署在移动设备上，本章考虑了一种基于移动设备分布式计算的边缘推断方案，基于通用的分布式计算架构，利用大量移动设备的有限的计算、存储、电量等资源池协作完成推断任务。在该方案中，不同移动设备之间需要交换计算中间值，带来了巨大的通信开销，本章就这一问题提出了一种快速数据交换策略。

本章具体内容安排如下：第5.1节中介绍了基于分布式计算架构的推断系统原理与相关研究进展，申明了其中所存在的关键问题是在无线环境下的数据交换带来的通信开销，然后简明介绍了本章所提快速数据交换方法的核心思想与原理；第5.2节系统地阐述了所提方案的系统架构、计算模型与通信模型；第5.3对快速数据交换问题正式给出了表述，并提出了低秩优化方法来建模，并分析了现有方法对该问题的不足之处；第5.4针对现有方法的缺点提出了一种新的低复杂度的DC算法；第5.5章对该算法的性能通过仿真进行了评估，验证了所提算法的性能优势与低计算复杂度的优势，证明了所提快速数据交换方案的有效性；第5.6节对本章进行了小结，并提出了一些研究方向。

5.1 引言

受有限的存储、计算和电量资源的制约，直接将完整的人工智能模型部署在一个单独的终端设备上来执行往往是无法实现的。本章研究了一种基于无线分布式计算的在设备端执行分布式推断（on-device distributed inference）的系统架构，其通过池化了许多分布式设备的计算存储资源来完成每个设备的推断任务。在广泛使用的分布式计算架构如MapReduce等结构中，先将数据集（即用于推断的人工智能模型）进行分割后部署在终端设备上，这个过程被称作数据集放置阶段。每一个终端设备会计算所有任务的的Map函数计算，得到的中间值需要在设备间进行交换，从而使每个设备得到其任务的所有Map函数值。联合这些中间值进行Reduce操作即可以得到所需的推断结果。然而这种系统的其中一个

主要瓶颈在于中间值交换的通信效率 (Li 等, 2017a)。因此, 本章设计了一种高效的数据交换策略。

为了降低分布式计算系统的数据交换通信开销, 有许多研究者设计了不同的编码传输方案。文献 (Li 等, 2018a) 基于编码多播传输方案提出了“Coded MapReduce”的策略, 用于减少有线的分布式计算架构中的数据传输问题。在无线的分布式计算架构下, 文献 (Li 等, 2017b) 提出了一种可伸缩的传输方案, 其通信模型是所有移动设备连接到了一个共同的接入点 (access point, AP) 用于数据交互。每一个移动设备上行通过正交传输将信息发送给接入点, 然后接入点以信道最弱用户的最大可达速率将信息广播传输给所有移动设备。然后其设计了编码方案以降低通信所需的信息比特数。然而, 在无线网络中考虑可达数据传输速率是非常关键的, 而该方案没有考虑物理层的无线信道本身。

本章提出了一种用于高效数据交换的线性编码方案, 采用共信道的上下行传输策略来提高频谱效率。其基本原理是将 Map 阶段每个设备算得的中间值作为辅助信息 (side information), 利用干扰对齐 (interference alignment) 技术进行收发器的设计。该问题可以建模成一个低秩优化问题以最大化可达自由度 (degree-of-freedom, DoF), 即可达速率的一阶表征。由所得低秩优化问题的独特结构出发, 本章提出了一种新的 DC 优化算法, 在大大降低了基于核范数的 DC 算法的复杂度的同时保持了算法的性能优势, 随后通过仿真评估了其性能。

5.2 基于无线分布式计算的分布式推断系统的数据交换

本节将介绍基于无线分布式计算的分布式推断系统架构, 并提出了一种用于数据交换的共信道传输方案。

5.2.1 计算模型

MapReduce 是一种普遍存在的分布式计算框架 (Dean 和 Ghemawat, 2008), 它能够有效地利用多个计算设备来处理有大量数据的计算任务。对于一个具有类 MapReduce 结构的计算任务来说, 待计算的目标函数能够被分解为先计算一系列的 Map 函数值 (该操作可以并行进行), 再对这些值进行 Reduce 操作。因此, 基于 MapReduce 的分布式计算系统能够池化多个设备的计算和存储资源, 从而能够支持实现基于移动设备分布式计算的边缘推断。

考虑一个由 K 个移动用户组成的无线分布式计算系统。如图5.1所示, 这

些移动用户通过一个他们共同连接的接入点进行数据交互。假设每个移动终端具有 L 根天线，而接入点拥有 M 根天线。每个移动设备有一个基于整个数据集（即人工智能模型）的推断任务。整个数据集假设被均匀等分成了 N 个文件 f_1, \dots, f_N ，每一个文件的大小是 F 比特。每个移动用户 k 的目标是基于输入数据 d_k 得到推断计算任务 $\phi_k(d_k; f_1, \dots, f_N)$ 的输出。例如在目标识别中，数据集由许多的物体的特征组成。给定一个图像的特征作为输入，每一个移动用户请求得到图像识别结果。在实际应用中，每个移动用户所拥有的存储十分有限 (Han 等, 2016)，整个数据集无法直接存储在单独一个用户端。于是可以假设每个移动用户的本地存储大小只有 μF 比特，其中 $\mu < N$ ，而整个数据集能够被分布式的存储在 K 个移动用户端，即 $\mu K \geq N$. 令 $\mathcal{F}_k \subseteq [N]$ 表示存储在用户 k 的文件索引值集合，可以得到 $|\mathcal{F}_k| \leq \mu$ 并且 $\cup_{k \in [K]} \mathcal{F}_k = [N]$. 使用 $f_{\mathcal{F}_k} = \{f_n : n \in \mathcal{F}_k\}$ 来表示第 k 个移动用户本地所存储的文件集合。

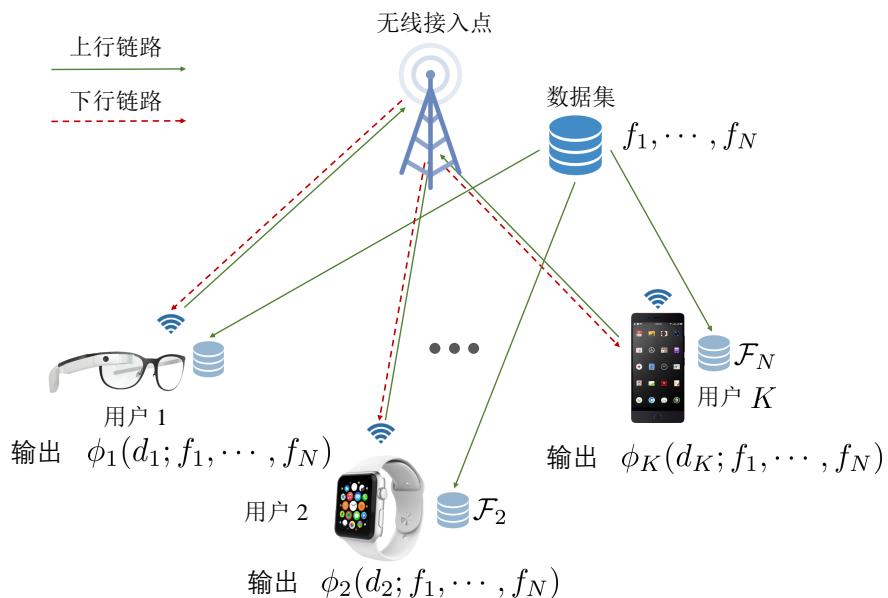


图 5.1 无线分布式计算系统。

Figure 5.1 Wireless distributed computing system.

考虑每个移动用户用于推断的计算任务符合类 MapReduce 的结构，其分解可以用下式来表示 (Li 等, 2017b)

$$\phi_k(d_k; f_1, \dots, f_N) = h_k(g_{k,1}(d_k; f_1), \dots, g_{k,N}(d_k; f_N)). \quad (5.1)$$

其中第 k 个用户基于文件 f_n 可以计算出 Map 函数 $g_{k,n}(d_k; f_n)$ 的值，输出一个大小为 E 比特的中间值 $w_{k,n}$. Reduce 函数 h_k 是一个将所有中间值 $w_{k,1}, \dots, w_{k,N}$

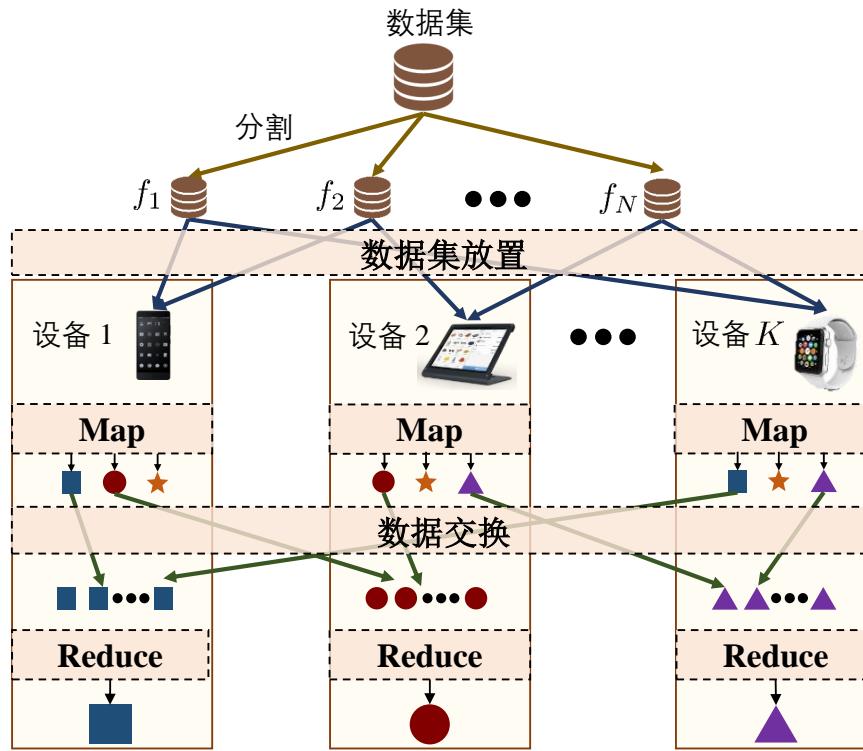


图 5.2 用于设备端分布式推断的分布式计算模型。

Figure 5.2 Distributed computing model for on-device distributed inference.

映射为计算任务 ϕ_k 输出的函数。假设中间值的存储占用足够的小，可以存在每个移动用户处。同时，手机所有的输入数据 d_k 的通信开销可以忽略。如图5.2所示，所有的计算任务可以通过执行如下的四个阶段来完成：

- **数据集放置阶段:** 这个阶段需要决定一个数据集放置策略 \mathcal{F}_k ，然后把文件提前部署在对应的移动用户端。
- **Map 阶段:** 此阶段需要每个移动设备 k 根据本地所有的文件 $f_{\mathcal{F}_k}$ 计算出所有 Map 函数 $g_{k,n}, k \in [K]$ 相应的中间值。
- **数据交换阶段:** 移动用户 k 的计算任务 ϕ_k 的输出也依赖于那些只能被其他用户计算出的中间值，即 $\{w_{k,n} : n \notin \mathcal{F}_k\}$. 因此，移动用户需要在这个阶段通过无线传输交换中间值，以完成计算任务的最终计算。
- **Reduce 阶段:** 每个移动用户 k 通过计算 Reduce 函数，可以将所有需要的中间函数值映射为其计算任务 ϕ_k 的输出，即 $\phi_k(d_k; f_1, \dots, f_N) = h_k(w_{k,1}, \dots, w_{k,N})$. 其中，有限的无线资源使得移动设备之间的数据交换成为了阻碍分布式推断规模化的一个显著瓶颈。

5.2.2 通信模型

如文献 (Li 等, 2017b; Lin 等, 2018) 所述, 通信问题是无线分布式计算架构下完成计算任务的关键瓶颈, 因此本章的研究目标是给定数据集放置策略下提高数据交换阶段的通信效率。所提方案是一个共信道的传输方案, 以高效的交换中间值, 然后把该问题建模成了一个有辅助信息的信息传递问题。具体来说, 将所有的中间值的集合 $\{w_{1,1}, \dots, w_{1,N}, \dots, w_{K,N}\}$ 看作一系列的独立消息 $\{W_1, \dots, W_T\}$, 其中 $T = KN$, 也就是说中间值 $w_{k,n}$ 用消息 $W_{(k-1)N+n}$ 来表示。令 $\mathcal{T}_k \subseteq [T]$ 为移动用户 k 所能算得的中间值的索引集合, 即 $\mathcal{T}_k = \{(j-1)N+n : j \in [K], n \in \mathcal{F}_k\}$. 类似的, 用 $\mathcal{R}_k \subseteq [T]$ 来表示移动用户 k 所需要从其他用户处请求的中间值索引集合, 即 $\mathcal{R}_k = \{(k-1)N+n : n \notin \mathcal{F}_k\}$. 由 MapReduce 分布式计算框架的结构可知有 $\cup_{k \in [K]} \mathcal{T}_k = [T], \mathcal{T}_k \cap \mathcal{R}_k = \emptyset$. 于是, 数据交换问题可以建模成一个有辅助信息的信息传递问题。如图5.1所示, 所提的数据交换的通信模型由上行多址接入 (multiple access) 阶段和下行广播 (broadcasting) 阶段组成。在上行阶段, 接入点需要收集所有移动用户的混合信号, 然后在下行阶段中把这个混合信号广播给每一个移动用户。

每个用户 k 将消息以 r 个信道使用发送出去, 可以表示为

$$\mathbf{x}_k = [\mathbf{x}_k[i]] = \begin{bmatrix} \mathbf{x}_k[1] \\ \vdots \\ \mathbf{x}_k[L] \end{bmatrix} \in \mathbb{C}^{Lr}, \quad (5.2)$$

其中 $\mathbf{x}_k[i] \in \mathbb{C}^r$ 对应了第 i 根天线。令 $H_k^{\text{up}}[s,i]$ 为移动用户 k 的第 i 根天线与接入点的第 s 根天线的信道系数。接入点的第 s 根天线所收到的信号 $\mathbf{y}[s] \in \mathbb{C}^r$ 可以表示为

$$\mathbf{y}[s] = \sum_{k=1}^K \sum_{i=1}^L H_k^{\text{up}}[s,i] \mathbf{x}_k[i] + \mathbf{n}^{\text{up}}[s], \quad (5.3)$$

其中 $\mathbf{n}^{\text{up}}[s] \in \mathbb{C}^r$ 是加性各向同性高斯白噪声。这里考虑了一个准静态衰落信道

(quasi-static fading channel) 模型，在 r 个信道使用期间信道系数保持不变。定义

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}[1] \\ \vdots \\ \mathbf{y}[M] \end{bmatrix} \in \mathbb{C}^{Mr}, \mathbf{n}^{\text{up}} = \begin{bmatrix} \mathbf{n}^{\text{up}}[1] \\ \vdots \\ \mathbf{n}^{\text{up}}[M] \end{bmatrix} \in \mathbb{C}^{Mr}, \quad (5.4)$$

$$\mathbf{H}_k^{\text{up}} = \begin{bmatrix} H_k^{\text{up}}[1, 1] & \cdots & H_k^{\text{up}}[1, L] \\ \vdots & \ddots & \vdots \\ H_k^{\text{up}}[M, 1] & \cdots & H_k^{\text{up}}[M, L] \end{bmatrix} \in \mathbb{C}^{M \times L}. \quad (5.5)$$

接入点处收到的信号可以更紧凑地表示为

$$\mathbf{y} = \sum_{k=1}^K (\mathbf{H}_k^{\text{up}} \otimes \mathbf{I}_r) \mathbf{x}_k + \mathbf{n}^{\text{up}}, \quad (5.6)$$

其中 \otimes 表示 Kronecker 积。在下行广播阶段，接入点将收到的信号 \mathbf{y} 发给每一个移动用户。类似地，第 k 个移动用户收到的信号 $\mathbf{z}_k \in \mathbb{C}^{Lr}$ 表示为

$$\mathbf{z}_k = (\mathbf{H}_k^{\text{down}} \otimes \mathbf{I}_r) \mathbf{y} + \mathbf{n}_k^{\text{down}}, \quad (5.7)$$

其中下行广播阶段的信道系数矩阵 $\mathbf{H}_k^{\text{down}}$ ，以及下行的加性各向同性高斯白噪声 $\mathbf{n}_k^{\text{down}}$ 分别为

$$\mathbf{H}_k^{\text{down}} = \begin{bmatrix} H_k^{\text{down}}[1, 1] & \cdots & H_k^{\text{down}}[1, M] \\ \vdots & \ddots & \vdots \\ H_k^{\text{down}}[L, 1] & \cdots & H_k^{\text{down}}[L, M] \end{bmatrix} \in \mathbb{C}^{L \times M}, \quad (5.8)$$

$$\mathbf{n}_k^{\text{down}} = \begin{bmatrix} \mathbf{n}_k^{\text{down}}[1] \\ \vdots \\ \mathbf{n}_k^{\text{down}}[L] \end{bmatrix} \in \mathbb{C}^{Lr}. \quad (5.9)$$

因此，从移动设备发出信号，经过接入点中转以后再到移动设备 k 收到信号，整体的输入输出关系可以表示为

$$\mathbf{z}_k = \sum_{i=1}^K (\mathbf{H}_k^{\text{down}} \otimes \mathbf{I}_r) (\mathbf{H}_i^{\text{up}} \otimes \mathbf{I}_r) \mathbf{x}_i \quad (5.10)$$

$$\begin{aligned} &+ (\mathbf{H}_k^{\text{down}} \otimes \mathbf{I}_r) \mathbf{n}^{\text{up}} + \mathbf{n}_k^{\text{down}} \\ &= \sum_{i=1}^K (\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{x}_i + \mathbf{n}_k, \end{aligned} \quad (5.11)$$

其中 $\mathbf{H}_{ki} = \mathbf{H}_k^{\text{down}} \mathbf{H}_i^{\text{up}}$ 表示等价的信道状态矩阵, 而 $\mathbf{n}_k = (\mathbf{H}_k^{\text{down}} \otimes \mathbf{I}_r) \mathbf{n}^{\text{up}} + \mathbf{n}_k^{\text{down}}$ 表示有效加性噪声。

从式 (5.11) 可知所提的数据交换通信模型可以等价于一个 K 用户的 MIMO (multiple-input-multiple-output) 干扰信道。这个有辅助信息的通信模型中共有 T 个消息。每一个发送端 k 和接收端 k 都有索引值为 \mathcal{T}_k 的所有信息。第 k 个接收端请求索引值为 \mathcal{R}_k 的所有信息。每一个发送端 k 的平均的功率限制是 $\mathbb{E}[\|\mathbf{x}_k\|_2^2] \leq \rho$, 其中 $\rho > 0$ 是最大的传输功率。本章将提的低秩优化方法仅需要问题能够被表述成一个有辅助信息的干扰消除问题。因此该方法应对其他能够等价成干扰信道的网络模型也适用。

5.2.3 可达数据速率与自由度

令 $R_k(W_l)$ 为移动用户 k 所需的消息 W_l 的可达速率。于是, 存在一种编码方案使得消息 W_l 以 $R_k(W_l)$ 的速率传输的同时, 当码字长度趋于无限长时移动用户 k 解码出 W_l 的错误概率任意小 (Cover 和 Thomas, 2012)。

自由度 (DoF) 是信道容量的一阶表征, 关于自由度的分析与优化被广泛地应用在了干扰信道中 (Cadambe 和 Jafar, 2008; Shi 等, 2016a; Bresler 等, 2014)。文献 (Cadambe 和 Jafar, 2008) 也刻画了完全连接的 K 用户干扰信道的最优自由度。令 $\text{SNR}_{k,l}$ 为第 k 个接收端所需的消息 W_l 的信噪比, 则其自由度定义是 (Cadambe 和 Jafar, 2008)

$$\text{DoF}_{k,l} \triangleq \limsup_{\text{SNR}_{k,l} \rightarrow \infty} \frac{R_k(W_l)}{\log(\text{SNR}_{k,l})}. \quad (5.12)$$

用 $\{\text{DoF}_{k,l} : k \in [K], l \in \mathcal{R}_k\}$ 来表示一个可达的自由度分配集。用 DoF_{sym} 来表示对称自由度, 其定义为对所有的 k, l 都可达的最大自由度。也就是说下列自由度分配是可达的

$$\{\text{DoF}_{k,l} = \text{DoF}_{\text{sym}} : k \in [K], l \in \mathcal{R}_k\}. \quad (5.13)$$

本章选择自由度作为消除数据交换过程中的干扰的性能度量, 不失一般性地, 将最大化无线分布式计算系统中数据交换的可达对称自由度。所用的方法可以很容易的扩展到一般情况。

5.3 基于干扰消除技术的低秩优化建模方法

本节通过建立分布式计算系统的数据交换的干扰对齐条件，从而将问题建模成一个低秩优化问题以找到线性收发器的最大可达自由度。

5.3.1 干扰对齐条件

由于线性编码方案的低复杂性和自由度的最优性，其在收发器设计问题中有着广泛的应用，如干扰对齐 (Cadambe 和 Jafar, 2008) 和索引编码 (index coding) (Maleki 等, 2014)。因此本章将采用线性的编码方案。令 $\mathbf{s}_j \in \mathbb{C}^d$ 为消息 W_j 的表示向量，有 d 个数据流，每个数据流的自由度为 1。则用户 k 的发送信号可以表示为

$$\mathbf{x}_k = \sum_{j \in \mathcal{T}_k} \mathbf{V}_{kj} \mathbf{s}_j. \quad (5.14)$$

其中 \mathbf{V}_{kj} 是用户 k 作用于消息 W_j 的预编码矩阵，它的结构是

$$\mathbf{V}_{kj} = \begin{bmatrix} \mathbf{V}_{kj}[1] \\ \vdots \\ \mathbf{V}_{kj}[L] \end{bmatrix} \in \mathbb{C}^{rL \times d}, \quad (5.15)$$

这里 $\mathbf{V}_{kj}[i] \in \mathbb{C}^{r \times d}$ 对应了用户 k 的第 i 根天线，通过 r 个信道发送的预编码矩阵。同样的，令

$$\mathbf{U}_{kl} = [\mathbf{U}_{kl}[1] \cdots \mathbf{U}_{kl}[L]] \in \mathbb{C}^{d \times Lr} \quad (5.16)$$

为每个消息 W_l 的解码矩阵，其中 $l \in \mathcal{R}_k$ 。则消息 W_l 可以从

$$\tilde{\mathbf{z}}_{kl} = \mathbf{U}_{kl} \mathbf{z}_k = \mathbf{U}_{kl} \sum_{i=1}^K (\mathbf{H}_{ki} \otimes \mathbf{I}_r) \sum_{j \in \mathcal{T}_i} \mathbf{V}_{ij} \mathbf{s}_j + \tilde{\mathbf{n}}_{kl} \quad (5.17)$$

中解码出来，这里 $\tilde{\mathbf{n}}_{kl} = \mathbf{U}_{kl} \mathbf{n}_k$ 。观察到 $\tilde{\mathbf{z}}_{kl}$ 是整个消息集合的线性组合，其可以分解成三个部分：所需消息，干扰，和本地可用消息，即

$$\tilde{\mathbf{z}}_{kl} = \underbrace{\mathcal{I}_1(\mathbf{s}_l)}_{\text{所需消息}} + \underbrace{\mathcal{I}_2(\{\mathbf{s}_j : j \in \mathcal{T}_k\})}_{\text{本地可用消息}} + \underbrace{\mathcal{I}_3(\{\mathbf{s}_j : j \notin \mathcal{T}_k \cup \{l\}\})}_{\text{干扰}} + \tilde{\mathbf{n}}_{kl}. \quad (5.18)$$

线性算子 $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$ 具体表示了

$$\begin{aligned}\mathcal{I}_1(\mathbf{s}_l) &= \sum_{i:l \in \mathcal{T}_i} \mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{V}_{il} \mathbf{s}_l, \\ \mathcal{I}_2(\{\mathbf{s}_j : j \in \mathcal{T}_k\}) &= \sum_{j \in \mathcal{T}_k} \sum_{i:j \in \mathcal{T}_i} \mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{V}_{ij} \mathbf{s}_j, \\ \mathcal{I}_3(\{\mathbf{s}_j : j \notin \mathcal{T}_k \cup \{l\}\}) &= \sum_{j \notin \mathcal{T}_k \cup \{l\}} \sum_{i:j \in \mathcal{T}_i} \mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{V}_{ij} \mathbf{s}_j.\end{aligned}$$

干扰对齐 (Cadambe 和 Jafar, 2008) 是一种处理用户间相互干扰的强大工具。其基本思路是使信号在预期接收器处可分辨，而在其他非预期的接收器处对齐和消除掉。干扰是限制达到高速率传输的关键性因素，消除干扰可以通过建立以下的干扰对齐条件：

$$\det \left(\sum_{i:l \in \mathcal{T}_i} \mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{V}_{il} \right) \neq 0, \quad (5.19)$$

$$\sum_{i:j \in \mathcal{T}_i} \mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{V}_{ij} = 0, \quad j \notin \mathcal{T}_k \cup \{l\}, \quad (5.20)$$

其中 $l \in \mathcal{R}_k, k \in [K]$. 通过设计收发器使得干扰对齐条件 (5.19) 和 (5.20) 满足，可以让每一个用户 k 从信号 $\tilde{\mathbf{s}}_l = \mathcal{I}_1^{-1} (\tilde{\mathbf{z}}_{kl} - \mathcal{I}_2(\{\mathbf{s}_j : j \in \mathcal{T}_k\}))$ 中解出所有的消息 $W_l, l \in \mathcal{R}_k$.

如果式 (5.19) 和式 (5.20) 条件满足，就可以获得无干扰的信道用于在 r 个信道实现上传输 d 维的消息。于是可达的自由度 $\text{DoF}_{k,l}$ 是 d/r . 则整个系统的对称自由度是

$$\text{DoF}_{\text{sym}} = d/r. \quad (5.21)$$

所以可达对称自由度可以通过找满足式 (5.19) 和 (5.20) 的最少的信道实现 r 来最大化。

5.3.2 低秩优化方法

本小节提出一种低秩模型来构建干扰对齐条件 (5.19) and (5.20)，以便于分布式推断中数据交换的高效算法的设计。根据 Kronecker 积的定义可知

$$\mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r) \mathbf{V}_{ij} = \sum_{m=1}^L \sum_{n=1}^L H_{ki}[m, n] \mathbf{U}_{kl}[m] \mathbf{V}_{ij}[n], \quad (5.22)$$

其中 $H_{ki}[m, n]$ 是矩阵 \mathbf{H}_{ki} 第 (m, n) 个元素。定义一系列的矩阵

$$\mathbf{X}_{k,l,i,j} = [\mathbf{X}_{k,l,i,j}[m, n]] = [\mathbf{U}_{kl}[m]\mathbf{V}_{ij}[n]] \quad (5.23)$$

$$= \begin{bmatrix} \mathbf{U}_{kl}[1]\mathbf{V}_{ij}[1] & \cdots & \mathbf{U}_{kl}[1]\mathbf{V}_{ij}[L] \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{kl}[L]\mathbf{V}_{ij}[1] & \cdots & \mathbf{U}_{kl}[L]\mathbf{V}_{ij}[L] \end{bmatrix} \quad (5.24)$$

$$= \begin{bmatrix} \mathbf{U}_{kl}[1] \\ \vdots \\ \mathbf{U}_{kl}[L] \end{bmatrix} \begin{bmatrix} \mathbf{V}_{ij}[1] & \cdots & \mathbf{V}_{ij}[L] \end{bmatrix} \quad (5.25)$$

$$= \tilde{\mathbf{U}}_{kl}\tilde{\mathbf{V}}_{ij}, \quad (5.26)$$

其中 $\tilde{\mathbf{U}}_{kl} \in \mathbb{C}^{Ld \times r}$, $\tilde{\mathbf{V}}_{ij} \in \mathbb{C}^{r \times Ld}$. 进一步记矩阵

$$\mathbf{X} = [\mathbf{X}_{k,l,i,j}] \quad (5.27)$$

$$= \begin{bmatrix} \mathbf{X}_{1,1,1,1} & \cdots & \mathbf{X}_{1,1,1,T} & \cdots & \mathbf{X}_{1,1,K,T} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{X}_{1,T,1,1} & \cdots & \mathbf{X}_{1,T,1,T} & \cdots & \mathbf{X}_{1,T,K,T} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{X}_{K,T,1,1} & \cdots & \mathbf{X}_{K,T,1,T} & \cdots & \mathbf{X}_{K,T,K,T} \end{bmatrix} \quad (5.28)$$

$$= \begin{bmatrix} \tilde{\mathbf{U}}_{11} \\ \vdots \\ \tilde{\mathbf{U}}_{1T} \\ \vdots \\ \tilde{\mathbf{U}}_{KT} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_{11} & \cdots & \tilde{\mathbf{V}}_{1T} & \cdots & \tilde{\mathbf{V}}_{KT} \end{bmatrix} \quad (5.29)$$

$$= \tilde{\mathbf{U}}\tilde{\mathbf{V}}, \quad (5.30)$$

其中 $\tilde{\mathbf{U}} \in \mathbb{C}^{LdKT \times r}$ 和 $\tilde{\mathbf{V}} \in \mathbb{C}^{r \times LdKT}$. 为了设计高效算法, 可以不失一般性的令式 (5.19) 中

$$\sum_{i:l \in \mathcal{T}_i} \mathbf{U}_{kl}(\mathbf{H}_{ki} \otimes \mathbf{I}_r)\mathbf{V}_{il} = \mathbf{I}. \quad (5.31)$$

这样干扰对齐条件 (5.19) 和 (5.20) 就可以重写为

$$\sum_{i:l \in \mathcal{T}_i} \sum_{m=1}^L \sum_{n=1}^L H_{ki}[m, n] \mathbf{X}_{k,l,i,l}[m, n] = \mathbf{I}, \quad (5.32)$$

$$\sum_{i:j \in \mathcal{T}_i} \sum_{m=1}^L \sum_{n=1}^L H_{ki}[m, n] \mathbf{X}_{k,l,i,j}[m, n] = 0, \quad j \notin \mathcal{T}_k \cup \{l\}, \quad (5.33)$$

将等式两边向量化之后可以表示为 $\mathcal{A}(\mathbf{X}) = \mathbf{b}$, 其中线性算子 $\mathcal{A} : \mathbb{C}^{D \times D} \mapsto \mathbb{C}^S$ 是一个 $\{\mathbf{H}_{ki}\}$ 的函数。由 $\mathbf{X} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}$ 可知矩阵 \mathbf{X} 的秩和信道实现数目 r 是相等的, 即

$$\text{rank}(\mathbf{X}) = r. \quad (5.34)$$

因此可以通过解如下的低秩优化问题来最大化可达对称自由度

$$\begin{aligned} \mathcal{P}_{5.1} : & \underset{\mathbf{X} \in \mathbb{C}^{D \times D}}{\text{minimize}} \quad \text{rank}(\mathbf{X}) \\ & \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b}, \end{aligned} \quad (5.35)$$

其中 $\mathbf{D} = \mathbf{LdKT}$. 本章假设了文件放置策略 $\{\mathcal{F}_k\}_{k=1}^K$ 是给定的, 而设计目标是最大化可达自由度。由于每个请求的中间值总能够通过正交传输发送给对应的移动用户, 所以问题 $\mathcal{P}_{5.1}$ 总是可行的。然而由于秩函数的非凸性, 求解优化问题 $\mathcal{P}_{5.1}$ 是计算困难的。

5.3.3 问题分析

低秩优化方法最近在机器学习, 高维统计和推荐系统等领域引起了广泛的关注 (Davenport 和 Romberg, 2016). 然而由于秩函数是非凸性低秩优化问题通常是很难解的。因此, 有许多研究关注于找秩函数可解的表示方法, 进而设计了许多算法求解。

5.3.3.1 核范数松弛

核范数 (Davenport 和 Romberg, 2016) 作为秩函数的凸松弛的有效性已经得到了证明。使用核范数松弛方法得到的优化问题可以表示为

$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b}. \end{aligned} \quad (5.36)$$

其等价的半正定规划形式是

$$\begin{aligned} & \underset{\mathbf{X}, \mathbf{W}_1, \mathbf{W}_2}{\text{minimize}} \quad \text{Tr}(\mathbf{W}_1) + \text{Tr}(\mathbf{W}_2) \\ & \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b} \\ & \quad \begin{bmatrix} \mathbf{W}_1 & \mathbf{X} \\ \mathbf{X}^H & \mathbf{W}_2 \end{bmatrix} \succeq 0, \end{aligned} \tag{5.37}$$

可以通过内点法来求解，以较少的迭代次数得到一个高精度的解。然而作为二阶算法，内点法的计算复杂度高达每次迭代 $\mathcal{O}((S + D^2)^3)$ ，这主要是由牛顿步 (Boyd 和 Vandenberghe, 2004) 造成的。一阶的交替方向乘子法 (alternating direction method of multipliers, ADMM) (Shi 等, 2015c; O'Donoghue 等, 2016) 能够将计算复杂度降低到每次迭代 $\mathcal{O}(SD^2 + D^3)$ (更多细节请参考第5.4.4小节的分析)。给定精度 $\epsilon > 0$ ，它能够在 $\mathcal{O}(1/\epsilon)$ 步迭代内收敛。

然而核范数最小化方法的性能是不能令人满意的，这是由于问题 $\mathcal{P}_{5.1}$ 的仿射约束条件独特的结构造成的。拿一个两用户的情况作为示例， $K = N = 2, \mu = d = L = M = 1$ ，每个移动用户存储了不同的文件，请求另外一个用户所计算的中间值。这种情况下问题 $\mathcal{P}_{5.1}$ 的核范数松弛得到的问题是

$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \mathbf{X} = \begin{bmatrix} \star & \star & 1/H_{12} & 0 \\ 0 & 1/H_{21} & \star & \star \end{bmatrix}, \end{aligned} \tag{5.38}$$

其中 \star 表示对应的值没有约束 (这里去掉了那些全部都没有约束的行和列)。这种情况下，核范数方法返回的是满秩的解，但是原问题的最优的秩为 1. 更进一步来说，第5.5节中使用数值仿真来验证了凸松弛方法的性能劣势。

5.3.3.2 Schatten- p 范数近似与迭代加权最小二乘

使用如第2.3.2节所介绍的迭代加权最小二乘算法可以避免求解半正定规划问题带来的高复杂度，即通过交替地最小化加权的 Frobenius 范数来更新 \mathbf{X} 和更新权重 \mathbf{W} 来求解，即

$$\mathbf{X}^{[t]} = \underset{\mathbf{X}}{\operatorname{argmin}} \{ \text{Tr}(\mathbf{W}^{[k-1]} \mathbf{X}^H \mathbf{X}) : \mathcal{A}(\mathbf{X}) = \mathbf{b} \} \tag{5.39}$$

$$\mathbf{W}^{[t]} = (\mathbf{X}^{[t]H} \mathbf{X}^{[t]} + \gamma^{[k]} \mathbf{I})^{\frac{p}{2}-1}, \tag{5.40}$$

其中 $\gamma^{[k]} > 0$ 是光滑参数。然而，将其应用于求解问题 $\mathcal{P}_{5.1}$ 时仿射约束的较差结构使得它的性能仍然不足。本章将基于秩函数的 DC 表示提出一种新的 DC 算法，使性能得到可观的提升。

5.4 高效低秩优化 DC 算法提出

本节针对数据交换所建模的低秩优化问题提出了秩函数的一个新的 DC 表示，进而设计了一个高效的 DC 算法。

5.4.1 DC 方法

文献 (Gotoh 等, 2018) 提出了一种秩函数的 DC 表示并设计了 DC 算法来求解问题 $\mathcal{P}_{5.1}$ 。这里先介绍矩阵的 Ky Fan 范数的定义。

定义 5.1. Ky Fan k 范数 (Watson, 1993): 矩阵 \mathbf{X} 的 Ky Fan k 范数是矩阵 \mathbf{X} 的一个凸函数，等于它的最大的 k 个奇异值之和，即

$$\|\mathbf{X}\|_k = \sum_{i=1}^k \sigma_i(\mathbf{X}), \quad (5.41)$$

其中 $\sigma_i(\mathbf{X})$ 是 \mathbf{X} 的第 i 大的奇异值。

由定义 5.1 可知，如果一个矩阵的秩为 r ，它的 Ky Fan r 范数等于其核范数。于是可以得到秩函数的一个 DC 表示。对任意的矩阵 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 来说有下式成立 (Gotoh 等, 2018):

$$\text{rank}(\mathbf{X}) = \min\{k : \|\mathbf{X}\|_* - \|\mathbf{X}\|_k = 0, k \leq \min\{m, n\}\}. \quad (5.42)$$

借助 Ky Fan k 范数来表示秩函数使得问题 $\mathcal{P}_{5.1}$ 可以通过找使得如下问题最优目标函数值为 0 的最小的 k 来求解：

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{C}^{D \times D}}{\text{minimize}} \quad \|\mathbf{X}\|_* - \|\mathbf{X}\|_k \\ & \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b}, \end{aligned} \quad (5.43)$$

目标函数是两个凸函数 $\|\mathbf{X}\|_*$ 和 $\|\mathbf{X}\|_k$ 之差。由于非凸的 DC 目标函数的存在，可以采用 MM (majorization-minimization) 算法 (Tao 和 An, 1997; Le Thi 和 Dinh, 2018) 来求解，通过将问题中的 $\|\mathbf{X}\|_k$ 线性化为 $\text{Tr}(\partial \|\mathbf{X}_t\|_k^H \mathbf{X})$ ，在第 $(t+1)$ 次迭

代中求解得到的凸的子问题

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{C}^{D \times D}}{\text{minimize}} \quad \|\mathbf{X}\|_* - \text{Tr}(\partial \|\mathbf{X}_t\|_k^H \mathbf{X}) \\ & \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b} \end{aligned} \quad (5.44)$$

来解决。这里 \mathbf{X}_t 是第 t 次迭代中问题 (5.44) 的解。 $\partial \|\mathbf{X}_t\|_k$ 表示 $\|\mathbf{X}\|_k$ 在 \mathbf{X}_t 处的次微分，根据文献 (Watson, 1993) 可以将其选为

$$\partial \|\mathbf{X}_t\|_k = \{\mathbf{U} \text{diag}(\mathbf{q}) \mathbf{V}^H, \mathbf{q} = [\underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{D-k}]\}, \quad (5.45)$$

其中 $\mathbf{X}_t = \mathbf{U} \Sigma \mathbf{V}^H$ 是 \mathbf{X}_t 的奇异值分解。

不幸的是，这种 DC 方法有个主要缺点，其在每次迭代过程中需要求解一个核范数最小化问题 (5.44)。而核范数最小化问题的计算复杂度，即使用一阶 ADMM 算法也需要 $\mathcal{O}(\frac{1}{\epsilon}(SD^2 + D^3))$ 的计算复杂度以达到 ϵ 的精度，这样的高复杂度对于数据交换问题来说是不适的。特别是对于由大量移动用户组成的分布式推断系统来说，需要提出高效的算法求解。故接下来将会展示一种新颖、高效的 DC 方法来解决问题 $\mathcal{P}_{5.1}$ ，这基于所要提出的新颖的秩函数的 DC 表示方法。

5.4.2 秩函数的一种新的 DC 表示

可以观察到，问题 (5.43) 的目标函数中的核范数函数导致了高复杂度的计算。为了克服这个缺陷，本节提出了秩函数的一种新的 DC 表示方法。首先介绍如下定义：

定义 5.2. 对任意整数 $1 \leq k \leq \min\{m, n\}$ ，矩阵 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 的 Ky Fan $2-k$ 范数 (Doan 和 Vavasis, 2016) 定义为一个由 \mathbf{X} 最大的 k 个奇异值组成的子向量的 ℓ_2 范数，即

$$\|\mathbf{X}\|_{k,2} = \left(\sum_{i=1}^k \sigma_i^2(\mathbf{X}) \right)^{1/2}. \quad (5.46)$$

其中 $\sigma_i(\mathbf{X})$ 是矩阵 \mathbf{X} 的第 i 大的奇异值。

Ky Fan $2-k$ 范数是一个酉不变 (unitarily invariant) 范数，如文献 (Doan 和 Vavasis, 2016) 所述可以通过如下的半正定规划问题计算出来

$$\begin{aligned} \|\mathbf{X}\|_{k,2}^2 &= \underset{z, \mathbf{U}}{\text{minimize}} \quad kz + \text{Tr}(\mathbf{U}) \\ &\text{subject to} \quad z\mathbf{I} + \mathbf{U} \succeq \mathbf{X}^H \mathbf{X}, \\ & \quad \mathbf{U} \succeq 0. \end{aligned} \quad (5.47)$$

注意到 $\text{rank}(\mathbf{X}) = r$ 意味着矩阵 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 最小的 $\min\{m, n\} - r$ 个奇异值均为 0. 基于这个事实可以得到如下命题

命题 5.1. 对于矩阵 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 有下式成立

$$\text{rank}(\mathbf{X}) \leq k \Leftrightarrow \|\mathbf{X}\|_F = \|\mathbf{X}\|_{k,2}. \quad (5.48)$$

另外有

$$\text{rank}(\mathbf{X}) = \min\{k : \|\mathbf{X}\|_F^2 - \|\mathbf{X}\|_{k,2}^2 = 0, k \leq \min\{m, n\}\}. \quad (5.49)$$

证明. 给定已知条件 $\text{rank}(\mathbf{X}) \leq k$, 可以得到 $\sigma_i(\mathbf{X}) = 0 \forall i > k$, 进而有 $\|\mathbf{X}\|_F = \|\mathbf{X}\|_{k,2}$. 反过来可以从 $\|\mathbf{X}\|_F = \|\mathbf{X}\|_{k,2}$ 推出 $\sigma_i(\mathbf{X}) = 0 \forall i > k$, 于是可得矩阵 \mathbf{X} 的秩不超过 k .

令矩阵 \mathbf{X} 的秩为 r . 则有 $\sigma_i(\mathbf{X}) = 0 \forall i > r$, $\sigma_i(\mathbf{X}) > 0 \forall i \leq r$. 由于 $\|\mathbf{X}\|_F = \|\mathbf{X}\|_{k,2}$ 当且仅当 $\text{rank}(\mathbf{X}) \leq k$ 时成立, 使得 $\|\mathbf{X}\|_F^2 - \|\mathbf{X}\|_{k,2}^2 = 0$ 成立的最小的 k 等于 r . 反过来, 由 $r = \min\{k : \|\mathbf{X}\|_F^2 - \|\mathbf{X}\|_{k,2}^2 = 0\}$ 可知 $\sigma_i(\mathbf{X}) = 0 \forall i > r$, 同时 $\sigma_i(\mathbf{X}) > 0 \forall i \leq r$. 故有 $\text{rank}(\mathbf{X}) = r$. \square

5.4.3 高效 DC 算法

利用所提的秩函数的 DC 表示方法, 可以通过将 k 从 1 增加到 $\min\{m, n\}$ 连续求解如下问题

$$\begin{aligned} \mathcal{P}_{5.1-\text{DC}} : & \underset{\mathbf{X} \in \mathbb{C}^{D \times D}}{\text{minimize}} \quad \|\mathbf{X}\|_F^2 - \|\mathbf{X}\|_{k,2}^2 \\ & \text{subject to } \mathcal{A}(\mathbf{X}) = \mathbf{b} \end{aligned} \quad (5.50)$$

直到优化问题 $\mathcal{P}_{5.1-\text{DC}}$ 的目标函数达到 0, 来得到最小的秩 r . 由于目标函数是两个凸函数之差, 优化问题 $\mathcal{P}_{5.1-\text{DC}}$ 是一个 DC 规划问题。

为了得到简化形式的 DC 算法 (Tao 和 An, 1997), 先将优化问题 $\mathcal{P}_{5.1-\text{DC}}$ 等价地写为

$$\underset{\mathbf{X} \in \mathbb{C}^{m \times n}}{\text{minimize}} \quad \|\mathbf{X}\|_F^2 + I_{(\mathcal{A}(\mathbf{X})=\mathbf{b})}(\mathbf{X}) - \|\mathbf{X}\|_{k,2}^2 \quad (5.51)$$

其中指示函数 I 是

$$I_{(\mathcal{A}(\mathbf{X})=\mathbf{b})}(\mathbf{X}) = \begin{cases} 0, & \mathcal{A}(\mathbf{X}) = \mathbf{b} \\ +\infty, & \text{otherwise} \end{cases}. \quad (5.52)$$

可以用 Wirtinger 积分 (Bouboulis 等, 2012) 处理复数域的问题。令 $g(\mathbf{X}) = \|\mathbf{X}\|_F^2 + I_{\{\mathcal{A}(\mathbf{X})=\mathbf{b}\}}(\mathbf{X})$, $h(\mathbf{X}) = \|\mathbf{X}\|_{k,2}^2$. 由于 $\{\mathbf{X} : \mathcal{A}(\mathbf{X}) = \mathbf{b}\}$ 是一个仿射空间, 所以函数 g 和 h 都是凸的。

根据第3.4.3.2节所述, 简化的 DC 算法原理是用连续凸近似 (successive convex approximation, SCA) 来更新原始变量和对偶变量, 即用于解优化问题 $\mathcal{P}_{5.1-DC}$ 的具体迭代过程可以表示为

$$\mathbf{Y}^{[t]} = \arg \inf_{\mathbf{Y} \in \mathcal{Y}} h^*(\mathbf{Y}) - [g^*(\mathbf{Y}^{[t-1]}) + \langle \mathbf{Y} - \mathbf{Y}^{[t-1]}, \mathbf{X}^{[t]} \rangle], \quad (5.53)$$

$$\mathbf{X}^{[t+1]} = \arg \inf_{\mathbf{X} \in \mathcal{X}} g(\mathbf{X}) - [h(\mathbf{X}^{[t]}) + \langle \mathbf{X} - \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle]. \quad (5.54)$$

利用 Fenchel biconjugation 定理 (Rockafellar, 2015) 可知, 式 (5.53) 可以概括为

$$\mathbf{Y}^{[t]} \in \partial h(\mathbf{X}^{[t]}). \quad (5.55)$$

因此优化问题 $\mathcal{P}_{5.1-DC}$ 可以通过更新原始变量和对偶变量 $\mathbf{X}^{[t+1]}, \mathbf{Y}^{[t]}$ 来求解, 更新公式为

$$\mathbf{Y}^{[t]} \in \partial \|\mathbf{X}^{[t]}\|_{k,2}^2 \quad (5.56)$$

$$\mathbf{X}^{[t+1]} = \arg \inf_{\mathbf{X} \in \mathcal{X}} \{ \|\mathbf{X}\|_F^2 - \langle \mathbf{X}, \mathbf{Y}^{[t]} \rangle : \mathcal{A}(\mathbf{X}) = \mathbf{b} \}. \quad (5.57)$$

命题 5.2. $\|\mathbf{X}\|_{k,2}^2$ 的一个次梯度可以表示为

$$\partial \|\mathbf{X}\|_{k,2}^2 := 2\mathbf{U}\boldsymbol{\Sigma}_k\mathbf{V}^H, \quad (5.58)$$

其中 $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ 是矩阵 $\mathbf{X} \in \mathbb{C}^{D \times D}$ 的奇异值分解, $\boldsymbol{\Sigma}_k$ 保留了矩阵 $\boldsymbol{\Sigma}$ 最大的 k 个对角元素。

证明. 首先矩阵 \mathbf{X} 的 Ky Fan $2-k$ 范数是酉不变的 (在对 \mathbf{X} 作酉变换下保持不变)。这可以从奇异值的酉不变性质以及下式中得到

$$\|\mathbf{X}\|_{k,2}^2 = \|\boldsymbol{\sigma}(\mathbf{X})\|_{k,2}^2 = \sum_{i=1}^k \sigma_i^2(\mathbf{X}). \quad (5.59)$$

这里 $\boldsymbol{\sigma} = [\sigma_i(\mathbf{X})] \in \mathbb{R}^D$ 表示由矩阵 \mathbf{X} 的所有奇异值组成的向量。 $\|\boldsymbol{\sigma}(\mathbf{X})\|_{k,2}$ 表示向量 $\boldsymbol{\sigma}(\mathbf{X})$ 的 Ky Fan $2-k$ 范数。 $\|\boldsymbol{\sigma}(\mathbf{X})\|_{k,2}^2$ 关于 $\boldsymbol{\sigma}(\mathbf{X})$ 的次梯度是

$$\mathbf{c} \in \mathbb{R}^D : c_i = \begin{cases} 2\sigma_i(\mathbf{X}), & i \leq k \\ 0, & i > k \end{cases}. \quad (5.60)$$

由文献 (Watson, 1992) 中提供的正交不变范数的次微分定理可知

$$\{\mathbf{U} \text{diag}(\mathbf{d}) \mathbf{V}^H : \mathbf{X} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H, \mathbf{d} \in \partial \|\boldsymbol{\sigma}(\mathbf{X})\|_{k,2}\} \subseteq \partial \|\mathbf{X}\|_{k,2}. \quad (5.61)$$

于是可以得到

$$2\mathbf{U} \boldsymbol{\Sigma}_k \mathbf{V}^H \in \partial \|\mathbf{X}\|_{k,2}^2, \quad (5.62)$$

其中 $\boldsymbol{\Sigma}_k$ 为

$$\boldsymbol{\Sigma}_k \text{的第}(i,j)\text{个元素} := \begin{cases} \sigma_i(\mathbf{X}), & i = j \text{ 且 } i \leq k \\ 0, & \text{其他} \end{cases}. \quad (5.63)$$

□

由于式 (5.57) 是一个简单的二次规划 (quadratic programming, QP) 问题, 所提 DC 算法需要如式 (5.56) 和式 (5.57) 的迭代过程, 比起来求解核范数最小化问题 (5.44) 其计算更加的高效。具体来说, 根据式 (5.56) 和式 (5.57) 可知, $\mathbf{X}^{[t+1]}$ 可以写成以下二次规划问题的解:

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{C}^{D \times D}}{\text{minimize}} \quad \|\mathbf{X} - \frac{1}{2} \partial \|\mathbf{X}^{[t]}\|_{k,2}^2\|_F^2 \\ & \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b}. \end{aligned} \quad (5.64)$$

这是一个有仿射约束条件的最小二乘问题, 它的解是到仿射空间的正交投影, 闭式解可以写为

$$\mathbf{X}^{[t+1]} = (\mathbf{I} - \mathcal{A}^+ \mathcal{A})(\frac{1}{2} \partial \|\mathbf{X}^{[t]}\|_{k,2}^2) + \mathcal{A}^+(\mathbf{b}), \quad (5.65)$$

其中 $\mathcal{A}^+ = \mathcal{A}^H (\mathcal{A} \mathcal{A}^H)^{-1}$. 因此所提出的 DC 算法的整体过程可以总结为算法4。

5.4.4 计算复杂度与收敛性分析

正如在第5.3.3.1节所分析, 用二阶内点法解一个核范数松弛问题的计算复杂度是每次迭代 $\mathcal{O}((S + D^2)^3)$, 可以用 CVX 工具箱 (Grant 和 Boyd, 2014) 来实现。而所提的 DC 算法是给定秩 r 计算一系列式 (5.65) 的过程。由于 \mathcal{A} 和 \mathcal{A}^+ 都可以提前算出来并存储下来, 每次迭代的计算复杂度来自于矩阵向量乘积和计算次梯度两个部分。由线性算子 \mathcal{A} 的维度是 $\mathbb{C}^{D \times D} \mapsto \mathbb{C}^S$ 可知矩阵向量乘积部分的计算复杂度是 $\mathcal{O}(SD^2)$. 而使用式 (5.58) 来计算次梯度的计算复杂度主要来自于计算奇异值分解, 其复杂度为 $\mathcal{O}(D^3)$. 因此所提 DC 算法的复杂度是每次迭代 $\mathcal{O}(SD^2 + D^3)$. 而用于解核范数最小化问题 (5.44) 的一阶交替方向乘子法需要通

算法 4 用于求解优化问题 $\mathcal{P}_{5.1}$ 的 DC 算法

```

1: 输入:  $\mathcal{A}, \mathbf{b}$ 
2: for  $r = 1, \dots, \min\{m, n\}$  do
3:   初始化  $\mathbf{X}_r^{[0]} \in \mathbb{C}_*^{m \times n}$ 
4:   while 不收敛 do
5:      $\mathbf{X}_r^{[t+1]} = (\mathbf{I} - \mathcal{A}^+ \mathcal{A})(\frac{1}{2}\partial \|\mathbf{X}_r^{[t]}\|_{k,2}^2) + \mathcal{A}^+(\mathbf{b})$ 
6:   end while
7:   if  $\text{rank}(\mathbf{X}_r) \leq r$  then
8:     返回  $\mathbf{X}_r$ 
9:   end if
10:  end for
11: 输出:  $\mathbf{X}_r$  和  $\text{rank}(\mathbf{X}_r)$ .

```

过奇异值分解解一系列的半正定锥投影问题，其给定目标精度 ϵ 的计算复杂度是 $\mathcal{O}(\frac{1}{\epsilon}(SD^2 + D^3))$. 因此，所提的带有闭式解的方法来解优化问题 DC 规划问题 (5.50) 的 DC 算法十分高效，而解 DC 规划 (5.43) 需要解一系列的核范数最小化问题，计算复杂度高的多。对于迭代加权最小二乘算法，用投影梯度下降法 (Mohan 和 Fazel, 2012) 来求解迭代过程式 (5.39) 和式 (5.40) 的计算复杂度是 $\mathcal{O}((SD^2 + D^3)\log \frac{1}{\epsilon})$.

利用算子 \mathcal{A} 的稀疏结构可以很高效的实现所提的 DC 算法。矩阵向量成绩的计算复杂度往往是比较小的，特别是当 L 和 d 比移动用户数小的多的时候。具体来说，线性算子 \mathcal{A} 的稀疏度是

$$\sum_{k=1}^K \sum_{l \in \mathcal{R}_k} \sum_{j \neq \mathcal{T}_k} |\{i : j \in \mathcal{T}_i\}| L^2 d^2. \quad (5.66)$$

举例来说，对于一个由五个移动用户和十个文件的数据集组成的单天线分布式推断系统，如果每个用户本地存了六个文件且每个中间值用单数据流传输的话，可知 $D = 250$ 而线性算子 \mathcal{A} 的稀疏度仅有 920。

根据文献 (Tao 和 An, 1997) 所述，给定秩参数 k ，所提的解优化问题 $\mathcal{P}_{5.1-\text{DC}}$ 的 DC 算法4可以从任意初始点收敛到临界点，该收敛性证明详见附录B.2。

5.5 仿真验证与结果分析

本节通过数值实验来比较所提 DC 算法（即算法4）与下列算法的性能：

- **核范数松弛:** 为了评估核范数松弛方法 (5.37) 的性能, 本节用 CVX 工具箱实现了第5.3.3.1节介绍的内点法来求解。
- **迭代加权最小二乘:** 文献 (Mohan 和 Fazel, 2012) 采用了光滑化的 Schatten- p 范数来近似秩函数。如第5.3.3.2节所述, 可用迭代加权算法来解决相应的非凸问题。通过交叉验证将其中的 p 值选为了 0.5。
- **核范数 DC:** 如第5.4.1节所述, 该 DC 算法基于核范数和 Ky Fan k 范数的差, 由文献 (Gotoh 等, 2018) 所提出, 我们用“核范数 DC”来表示该算法。

在所有仿真中均考虑的是对称的情形, 即所有的移动设备和接入点都有 $L = M$ 根天线, 并选择最大的可达对称自由度 (5.21) 作为性能度量。信道系数依照独立同分布的复高斯分布随机生成, 即 $\mathbf{H}_{ki} \sim \mathcal{CN}(0, \mathbf{I})$. 每个算法的秩定为奇异值大于 10^{-5} 的个数。给定 r , 所提 DC 算法会在第 $(r + 1)$ 个奇异值小于 10^{-5} 的时候停止, 即 $\sigma_{r+1}(\mathbf{X}) < 10^{-5}$. 由于“核范数 DC”算法的计算复杂度在问题规模较大的时候十分巨大, 仿真中只在问题规模较小时对该算法进行了评估, 即第5.5.1节和第5.5.2节中。

5.5.1 收敛速度与时间

本小节考虑了一个 5 用户的单天线无限分布式计算系统用于分布式推断, 比较了迭代加权算法, “核范数 DC”算法和所提的 DC 算法达到收敛所需的迭代步数和计算时间。假设每个移动用户本地存储了总数据集的 10 个文件中的 5 个文件。设定 $r = 13$ 的情况下运行每一个算法, 将矩阵 \mathbf{X} 的第 $(r + 1)$ 个奇异值取为代价函数, 观察每个算法的收敛行为, 结果如图5.3所示。该结果显示出“核范数 DC”算法在很少几步迭代之后就达到收敛, 而所提的 DC 算法需要最多的迭代步数。然而, 所提的 DC 算法的整体计算复杂度最低、收敛用时最少。这是因为它的每次迭代过程的计算复杂度很低。

5.5.2 可达自由度与本地存储大小的关系评估

考虑由五个单天线移动用户和一个单天线接入点组成的无线分布式计算平台用于分布式的推断服务。整个数据集由 10 个文件组成, 而每个移动用户本地存储了 5 至 9 个文件, 每个消息使用单数据流传输。通过仿真实验来评估每个算法的最大可达对称自由度, 每个算法平均了 100 次观察得到的自由度和本地存储大小的关系。如图5.4所示, 当每个移动设备有能力存储更多文件时, 使用每

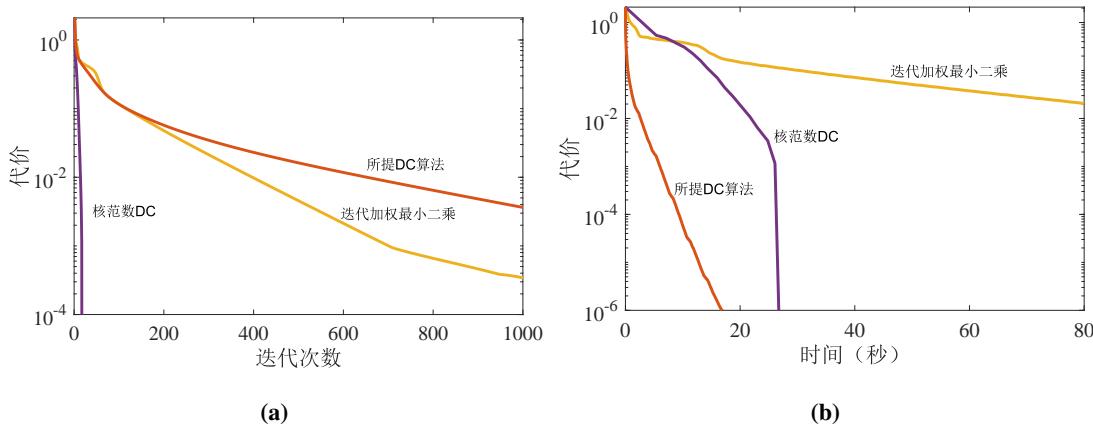
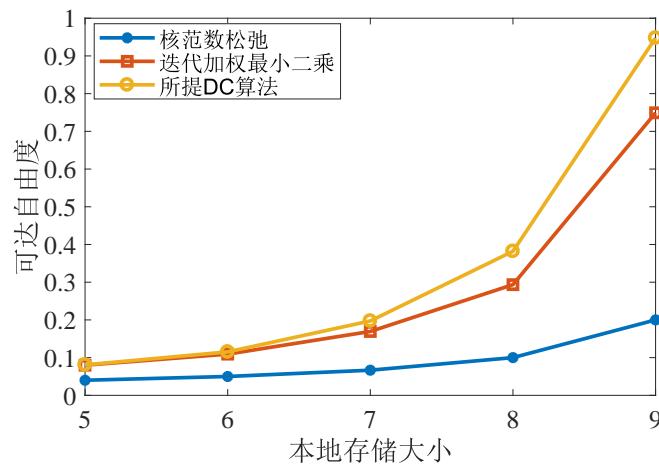


图 5.3 不同算法的收敛迭代次数与时间。

Figure 5.3 Convergence rate and convergence time with difference algorithms.

个算法得到的可达对称自由度有了显著增长。这是因为设备之间有了更多的协作机会，且每个用户获得了数据集中更多文件使得需要更少的中间值交换了。所提 DC 算法的性能和“核范数 DC”算法的性能基本一致，二者都比迭代加权最小二乘算法和核范数松弛法的性能优异许多。同时所提的 DC 算法计算复杂度在所有四个算法中最低。这个实验结果也证实了相比 Schatten- p 范数近似方法，所提的秩函数的 DC 表示方法具有优势，而核范数松弛的性能差于二者。

图 5.4 最大可达对称自由度与每个移动用户本地存储大小 μ 的关系。**Figure 5.4 The maximum achievable symmetric DoF over local storage size μ of each mobile user.**

5.5.3 可达自由度与天线数目的关系评估

这里考虑由 8 个移动用户和一个接入点组成的无线分布式计算系统用于分布式推断，每个移动用户本地存储了整个数据集 4 个文件中的一个。假设每个移动用户和接入点均配备了同样数目的天线。使用不同的天线数目值来评估该系统的复用增益。每个点平均 100 次的仿真结果如图5.5所示。

可以看出，使用所提的 DC 算法和迭代加权最小二乘算法得到的可达对称自由度随着天线数目增加而线性提升。但是使用核范数松弛算法的可达自由度随着天线增加是不变的，这是由于问题的特殊结构所造成的。这个测试说明了所提的收发机设计框架能够在增大天线数目时达到线性增益。它也说明了核范数松弛方法用于解决数据交换问题时的内在缺陷。所提的 DC 方法性能优于迭代加权最小二乘算法和核范数松弛方法。

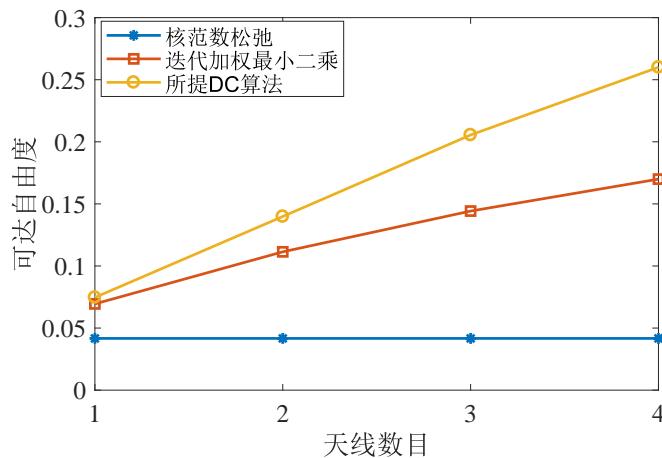


图 5.5 最大可达对称自由度与天线数目的关系。移动用户和接入点具有同样数目的天线。

Figure 5.5 The maximum achievable symmetric DoF over the number of antennas when the mobile users and the AP are equipped with same number of antennas.

5.5.4 可达自由度与移动用户数目的关系评估

正如文献 (Li 等, 2017b) 所指出的那样，由于计算任务随着网络规模的增长而线性增长，有限的通信带宽是一大瓶颈。因此，基于无线分布式计算的分布式推断系统的可伸缩性变得至关重要。本实验中评估了当用户数目增加时可达自由度的变化情况。考虑一个单天线的无线分布式计算系统，数据集可以分成 5 个文件，每个移动用户本地最多仅能存 2 个文件。假设均匀的放置策略，即每个用户存 $\mu = 2$ 个文件，数据集中的每个文件被 $\mu K/N = 2K/5$ 个用户所存储。考虑

单数据流的情形，即 $d = 1$ 。对每个算法都平均了 100 次实验，其平均可达对称自由度如图5.6所示。结果显示，所提的 DC 算法的性能能在网络规模增大时几乎保持不变，具备可伸缩性。对比之下，迭代加权最小二乘算法和核范数松弛算法的可达自由度都有了显著的下降。尽管当用户数增加时整个系统包含的需要传输的消息增加，由于每个文件存在了更多的移动用户处，移动用户协作的机会也在增长。所提出的算法可以利用这种协作的机会，而其他算法却不能。但是从理论上证明所提 DC 算法的可伸缩性仍然是一个有趣同时富有挑战的问题。

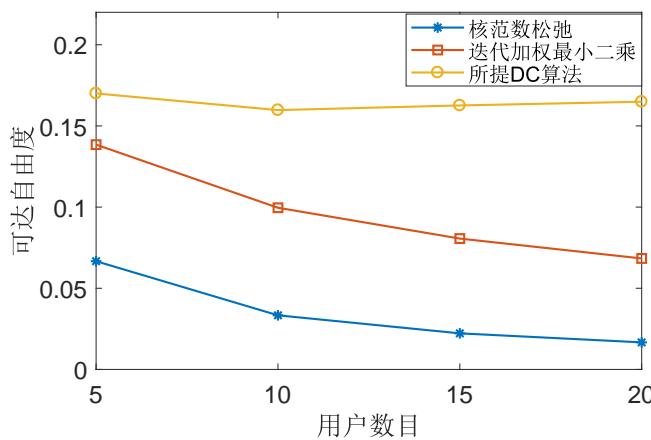


图 5.6 不同算法可达自由度与移动用户数目的关系。

Figure 5.6 The achievable DoF with different algorithms over the number of mobile users.

概括起来，所提的 DC 算法相比其他基准方法能够达到更高的自由度，其利用了数据交换问题的特殊结构。同时，所提 DC 算法的可达自由度在移动用户数目增加时能几乎保持不变。

5.6 本章小结

本章研究了基于移动设备分布式计算的边缘推断系统中的快速数据交换问题，提出了一种共信道的传输方案，利用干扰消除技术将编解码矩阵设计问题建模为一个低秩优化问题。为了高效求解该低秩优化问题并克服现有算法性能较差的缺点，提出了一种新的矩阵秩的 DC 表示方法，并结合问题的结构设计了一个高效的 DC 算法，使得每步迭代都具有闭式解。数值仿真显示除了所提方法能够比核范数松弛与迭代加权算法有更高的可达自由度，而比最新提出的低秩优化 DC 算法计算更加的高效。另外，在均匀数据集放置的情况下，所提方案的可达自由度几乎不随着参与的设备数目的增加而改变。

第6章 基于计算任务卸载的边缘推断的高能效协作传输策略

上一章考虑了基于通用分布式计算架构的边缘推断解决方案，使用多个有限资源的移动设备来协作完成推断任务。而对于计算更为密集的推断任务来说，将推断任务的输入上传给邻近的边缘服务器（如无线接入点）来帮助执行推断任务，然后将推断结果返回给终端设备是一种有效的解决办法。这一类系统架构可以统称为基于计算任务卸载的边缘推断系统，通过协作传输可以提高传输的频谱效率，但是所需要在边缘服务器处进行更多计算。而能效也是这种边缘处理系统的一大关键问题。本章提出了一种高能效的边缘推断方案，同时能够应对下行传输时信道估计是不准确性，保证鲁棒性传输。

本章的具体内容安排如下：第6.1节对现有的基于计算任务卸载的边缘推断方案相关工作进行了总结，指出了所考虑的协作方案所具备的潜力与意义，并概括了所提方案的基本原理；第6.2节详细给出了系统模型与问题表述，建模了一个由概率服务质量约束的组稀疏优化问题；第6.3节提供了一种基于统计学习的鲁棒优化近似方法来解决概率服务质量约束，将问题重新表述为一个具有非凸二次约束的组稀疏优化问题；第6.4节提供了稀疏与低秩优化算法框架，用低秩优化解决非凸二次约束，用迭代加权算法解决目标函数中的组稀疏函数；第6.5节通过仿真来评估与验证所提方法的有效性，以及相对其他方法的优势；第6.6节对本章进行了小结，并给出了本工作的不足之处与潜在的研究方向。

6.1 引言

将人工智能模型部署在网络边缘可以降低人工智能应用的延迟并提高其数据安全性，但是直接将模型部署在移动设备上存在许多限制。对于计算量较大的推断任务来说，一个可行的解决方案是将任务传输给邻近移动用户的边缘服务器来执行。网络边缘服务器具有更强大的计算能力和存储能力，能够更加高效的完成计算任务。近年来涌现了一些针对这种基于计算任务卸载的边缘推断方案的研究。其中一种策略是将所有的任务都卸载到边缘服务器上(Mao等, 2017)，可以称其为基于服务器的边缘推断，它对于资源严重受限的物联网设备的推断任务特别适合。这种情况下，整个人工智能模型部署在边缘服务器上，而移动设

备将原始输入数据上传至服务器用于计算得到推断结果。为了缓解原始数据传输对带宽受限的边缘人工智能系统带来的通信开销，文献Mohanarajah 等 (2015) 提出仅发送关键帧数据的方法来降低带宽需求；文献Liu 等 (2018b) 观察到传统压缩方法（如JPEG）往往是从人类视觉出发的，提出了一种从深度神经网络出发的数据编码压缩方案，实现了更高的压缩率的同时不带来图像识别精度的损失。出于对数据隐私保护的考虑，另外一种方案是考虑任务的层级分布式结构，即将推断任务进行分割，仅将部分任务推送至边缘服务器，让其基于移动设备计算得到的中间值从而得到模型的推断结果。文献Hauswald 等 (2014) 考虑了将图像分类的流程进行分割，发现把特征提取过程部署在移动设备，而将剩下的交给边缘服务器能够使得运行时间最少。文献Teerapittayananon 等 (2017) 提出将超大规模的深度神经网络模型分割后部署在移动设备、边缘服务器和云端，移动设备端先执行前面几层网络，剩下的计算任务依次在边缘服务器和云端执行，当结果的精度满足了应用的需求时就可以提前退出，避免后续的计算冗余。最近，文献Li 等 (2020) 提出了一种联合设计模型分割策略和提前退出策略的方法，根据移动设备和边缘服务器的不同计算能力以及复杂的网络环境，实现根据应用需求的低延迟推断。

现有这些方法的研究基础是给定了边缘服务器的计算能力，而注意到协作传输 (Gesbert 等, 2010) 是一种广为人知的降低同信道干扰和提高传输频谱效率的方式。这启发了本章提出计算任务卸载的边缘推断的高能效协作传输策略，通过将每个任务在多个边缘服务器同时计算，得到的结果使用协作传输来提高推断结果传输的服务质量。如图6.1所示，本章所考虑的边缘推断系统中，每个移动用户的输入数据（如一张原始涂鸦画）上传至具备边缘计算能力的无线接入点，利用提前训练好的深度学习模型（如英伟达的人工智能系统 GauGAN (Gau)）将每个任务在多个边缘服务器处执行，然后使用多个无线接入点协同波束成形技术将推断输出结果（如风景图）传回给移动用户。值得注意的是，该方案对于能够保护数据隐私的层级分布式结构边缘推断方案也是适用的，具体可见第6.2.4节的注解讨论。在这样的系统中，考虑上下行链路的无线传输问题都十分重要。除了低延迟之外，由于处理深度神经网络的高计算复杂度，提高能效 (Sze 等, 2017) 也是一个很关键的问题，有许多工作研究了模型压缩方法 (Han 等, 2016; Cheng 等, 2018) 来降低处理深度神经网络的能耗。

对于所考虑的方案，存在着通信与计算的权衡问题。具体来说，如果一个任务在更多的边缘服务器处执行，通过协作传输可以给下行传输带来更高的通信服务质量，然而这会导致执行深度学习模型带来更高的计算功耗，而能效也是边缘处理系统中一个至关重要的因素 (Shi 等, 2014)。因此，本章提出了一种计算任务分配与下行波束成形联合设计方案，通过保证服务质量的基础上最小化传输功耗和计算功耗之和，以实现低延迟与高能效的边缘处理。执行深度神经网络模型推断任务的功耗可以通过估计能量 (Yang 等, 2017) 与计算时间来确定。观察到下行聚合波束成形向量的组稀疏结构 (Shi 等, 2014; Tao 等, 2016) 与在边缘服务器执行的任务集合这一组合变量可以联系起来。另外，协作传输策略需要全局信道状态信息 (channel state information, CSI)，而获取的信道状态信息中不可避免的存在着不确定性，这可能是由于基于训练的信道估计方法 (Yang 等, 2018)、有限反馈 (Mo 和 Heath, 2018)、仅获取部分信道状态信息 (Shi 等, 2015b) 或者信道状态信息获取的延迟 (Maddah-Ali 和 Tse, 2012)。因此本章考虑了边缘推断系统中的任务选择与下行波束成形联合设计问题，以实现高效的处理与对于信道状态信息误差的鲁棒性传输。该问题被表述为一个有概率服务质量约束条件 (Shi 等, 2015b) 的组合稀疏波束成形设计问题。

为了解决非凸非确定性的概率服务质量约束，本章采用了一种基于统计学习的方法 (Hong 等, 2017) 将其近似为一个鲁棒优化约束，它可以克服想定生成方法 (scenario generation, SG) (Nemirovski 和 Shapiro, 2006) 的过于保守性，也避免了随机优化方法 (Shi 等, 2015b) 的高计算复杂度。进而采用了矩阵升维技术与低秩优化技术来解决鲁棒优化近似带来的非凸二次约束。但是该方法会导致组稀疏优化最常用的方法，混合 ℓ_1/ℓ_2 范数最小化法不可用，尽管文献 Shi 等 (2015a) 展示了混合 ℓ_1/ℓ_2 范数的一种可用的二次变体，但是其性能仍有欠缺。于是本章使用了迭代加权算法来增强诱导稀疏性，同时采用了 DC 正则项来促使秩为一的约束满足。仿真结果验证了所提的鲁棒优化近似方法来建模信道状态信息不确定性对于传输的鲁棒性，证实了所提协作传输方案具备通过将一个任务在更多边缘服务器计算所具备的提高边缘推断服务质量的潜力。

6.2 系统模型与问题表述

本节介绍了所考虑的边缘推断的系统模型和功率消耗模型，然后展示了高能效的边缘处理方案，以及在信道状态信息有误差情况下的概率服务质量约束。

6.2.1 系统模型

考虑一个由 N 个有 L 根天线的无线接入点和 K 个单天线的移动用户组成的边缘处理网络，如图6.1所示，每个无线接入点作为边缘计算节点具备一定的计算能力。每个移动用户 k 有一个深度学习推断任务 $\phi_k(d_k)$ ，其输入是 d_k 。深度学习任务的执行不依赖于云数据中心，而直接在无线接入点的边缘服务器处，从而给高风险的应用提供低延迟和保护数据隐私的服务，比如无人机和智能车辆等 (Park 等, 2019a)。本章考虑的系统中，所有神经网络模型 ϕ_k 提前下载到无线接入点处以供执行。在第一个阶段中，每一个无线接入点从移动用户处收集了所有任务的输入数据 $\{d_k\}_{k=1}^K$ 。在第二个阶段中，每一个无线接入点会选择性的执行一部分的推断任务，并通过协作式的下行传输将输出推断结果发送给对应的移动用户，从而为其提供低延迟的智能服务。本章集中考虑第二个阶段的任务选择和下行传输波束成形联合设计问题。

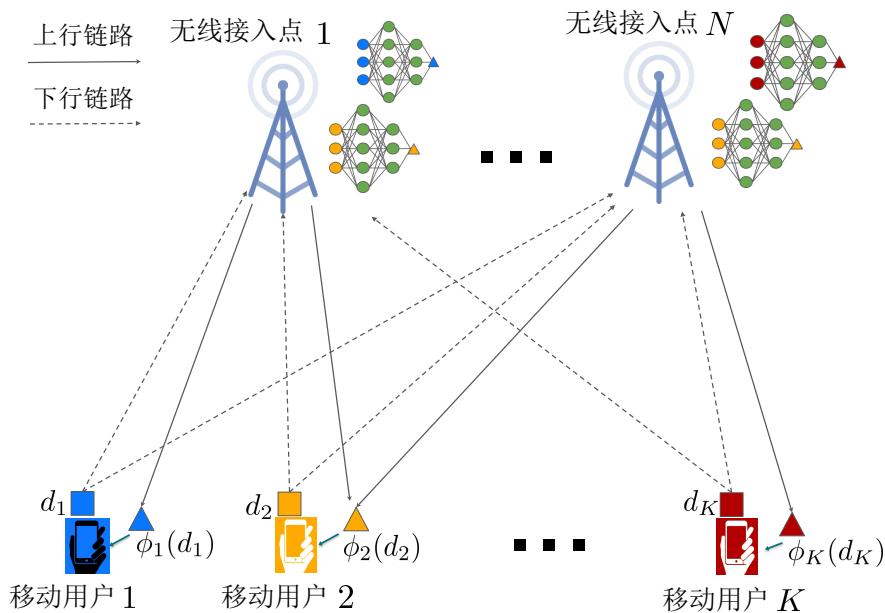


图 6.1 基于边缘服务器推断的系统模型。

Figure 6.1 System model of in-edge cooperative inference.

令 $\phi_k(d_k)$ 为移动用户 k 所需要的输出结果， $s_k \in \mathbb{C}$ 为待发送的经过编码的

数据符号标量, $\boldsymbol{v}_{nk} \in \mathbb{C}^L$ 为第 n 个无线接入点处消息 $\phi_k(d_k)$ 对应的波束成形向量。考虑这样的下行传输场景中所有的输入数据 d_k 已经被所有的无线接入点所收集。则第 l 个移动用户所收到的信号可以表示为

$$y_k = \sum_{n=1}^N \sum_{l=1}^K \mathbf{h}_{kn}^H \boldsymbol{v}_{nl} s_l + z_k, \quad (6.1)$$

其中 $\mathbf{h}_{kn} \in \mathbb{C}^L$ 是第 n 个无线接入点与第 k 个移动用户之间的信道系数向量, $z_k \sim \mathcal{CN}(0, \sigma_k^2)$ 是加性各向同性高斯白噪声。假设所有的数据符号 s_k 是相互独立的且和噪声独立, 并具有单位功率, 即 $\mathbb{E}[|s_k|^2] = 1$ 。用 $[K]$ 来表示集合 $\{1, \dots, K\}$ 。用 $\mathcal{A} \subseteq \{(n, k) : n \in [N], k \in [K]\}$ 来表示推断任务在无线接入点处的一个可行的分配方案, 即对于所有的 $(n, k) \in \mathcal{A}$, 计算任务 ϕ_k 将在第 n 个无线接入点处执行。考虑到聚合波束成形向量的组稀疏结构

$$\boldsymbol{v} = [\boldsymbol{v}_{11}^H, \dots, \boldsymbol{v}_{N1}^H, \dots, \boldsymbol{v}_{NK}^H]^H \in \mathbb{C}^{NKL}, \quad (6.2)$$

可知如果推断任务 k 不在无线接入点 n 处执行, 即 $(n, k) \notin \mathcal{A}$ 时, 波束成形向量 \boldsymbol{v}_{nk} 可以设置为全零向量。用 $\mathcal{T}(\boldsymbol{v})$ 来表示 \boldsymbol{v} 的组稀疏模式, 即

$$\mathcal{T}(\boldsymbol{v}) = \{(n, k) | \boldsymbol{v}_{nk} \neq 0\}. \quad (6.3)$$

移动用户 k 所收信号的信干噪比 (signal-to-interference-plus-noise-ratio, SINR) 等于

$$\text{SINR}_k(\boldsymbol{v}; \mathbf{h}_k) = \frac{|\mathbf{h}_k^H \boldsymbol{v}_k|^2}{\sum_{l \neq k} |\mathbf{h}_k^H \boldsymbol{v}_l|^2 + \sigma_k^2}, \quad (6.4)$$

其中 \mathbf{h}_k 和 \boldsymbol{v}_k 分别为

$$\mathbf{h}_k = [\mathbf{h}_{k1}^H, \dots, \mathbf{h}_{kN}^H]^H \in \mathbb{C}^{NL}, \quad (6.5)$$

$$\boldsymbol{v}_k = [\boldsymbol{v}_{1k}^H \dots \boldsymbol{v}_{Nk}^H]^H \in \mathbb{C}^{NL}. \quad (6.6)$$

聚合信道系数向量表示为

$$\mathbf{h} = [\mathbf{h}_1^H, \dots, \mathbf{h}_K^H]^H \in \mathbb{C}^{NKL}. \quad (6.7)$$

第 n 个无线接入点的发送功率约束是

$$\mathbb{E} \left[\sum_{l=1}^K \|\boldsymbol{v}_{nl} s_l\|_2^2 \right] = \sum_{l=1}^K \|\boldsymbol{v}_{nl}\|_2^2 \leq P_n^{\text{Tx}}, n \in [N], \quad (6.8)$$

其中 P_n^{Tx} 是最大发送功率。

6.2.2 功耗模型

尽管深度学习的广泛应用给智能系统带来了大量机遇，能耗是其最关键的考虑因素之一 (Xu 等, 2018)。事实上，深度神经网络推断任务的能耗主要来自于访问存储器。正如 (Han 等, 2016) 所指出的，32 位动态随机存取存储器 (dynamic random access memory, DRAM) 的一个存储器访问操作消耗的能量是 640pJ，而 32 位静态随机存取存储器 (static random access memory, SRAM) 的一个缓存访问操作仅消耗 5pJ 能量，一个 32 位的浮点加法操作消耗的能量是 0.9pJ。大型的深度神经网络模型往往无法容纳于移动设备，因为其需要更多的开销很大的 DRAM 存储器访问。因此小模型可以直接部署在移动设备，而对于大模型更好的解决方案是上传至边缘服务器处执行。令在第 n 个无线接入点处执行计算任务 ϕ_k 的功耗是 P_{nk}^c ，则所有边缘计算节点的总功率消耗可以表示为

$$P^c = \sum_{n,k} P_{nk}^c I_{(n,k) \in \mathcal{T}(\boldsymbol{v})}, \quad (6.9)$$

其中指示函数 I 的值在 $(n, k) \in \mathcal{T}(\boldsymbol{v})$ 时为 1，其他情况下为 0。因此，包括传输功耗和计算功耗在内的总功耗为

$$P = \sum_{n,k} \frac{1}{\eta_n} \|\boldsymbol{v}_{nk}\|_2^2 + \sum_{n,k} P_{nk}^c I_{(n,k) \in \mathcal{T}(\boldsymbol{v})}, \quad (6.10)$$

其中 η_n 是功率放大器效率。

深度神经网络，特别是深度卷积神经网络 (convolutional neural networks, CNNs)，变成了实际智能服务中一个不可或缺的最新范式。它的高能耗问题吸引了许多研究者进行高能效的神经网络结构设计 (Han 等, 2016)。于是，估计神经网络的能耗变成了边缘推断的一个关键问题，文献 (Ene) 中开发了一个估计工具。执行一个推断任务的能耗包括了计算部分和数据移动部分 (Yang 等, 2017)。计算部分的能耗可以通过对每一层中的乘加 (multiply-and-accumulate, MAC) 操作计数，然后将其与计算核心中每个 MAC 运算的能耗进行加权来计算。数据移动的能耗是通过计算相应硬件中存储器层次结构每个级别的访问存储器的数量，并用相应级别的访问存储器的能耗进行加权来计算的。在这里，我们说明了如何在 Eyeriss 芯片上使用经典的卷积神经网络 (即由 5 个卷积层和 3 个全连接层组成的 AlexNet(Krizhevsky 等, 2012)) 估算执行图像分类任务的计算功耗。能量估算工具以网络配置为输入，并根据三种数据类型 (权重，输入特征映射，输出特

征映射) 的计算部分和数据移动部分, 输出每层估计出的能量分解。图6.2显示了估计出的在 Eyeriss 芯片上运行的每层消耗的能量, 总能耗为四部分之和。单位能量是指每个 MAC 操作的能量。基于整体的能耗, 计算的功耗可以用能耗除以计算时间来获得。

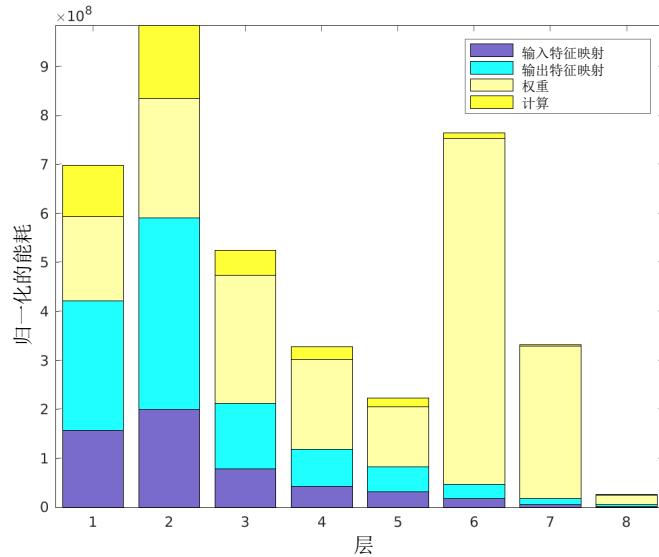


图 6.2 AlexNet 能耗分解图。

Figure 6.2 Energy consumption breakdown of the AlexNet.

6.2.3 信道信息不确定性模型

对于高风险的应用来说, 传输的鲁棒性是一个关键要求。实际上获得的信道状态信息 \mathbf{h} 中不可避免的存在不确定性, 本章中考虑了这一不确定性, 从而提供鲁棒的信息传输。这种不确定性可能来自于基于训练的信道估计方法 (Yang 等, 2018)、有限精度的信道反馈 (Mo 和 Heath, 2018)、仅获得了部分信道状态信息 (Shi 等, 2015b) 以及信道状态信息获取存在延迟 (Maddah-Ali 和 Tse, 2012)。这里参考文献 (Liu 等, 2018a; Fang 等, 2017) 采用了信道状态信息的加性误差模型, 即

$$\mathbf{h} = \hat{\mathbf{h}} + \mathbf{e}, \quad (6.11)$$

其中 $\hat{\mathbf{h}} \in \mathbb{C}^{NKL}$ 是所估计的聚合信道向量, $\mathbf{e} \in \mathbb{C}^{NKL}$ 是信道状态信息中存在的未知分布的随机误差, 其期望值为零。利用概率服务质量 (probabilistic quality-of-service) 约束 (Shi 等, 2015b) 来刻画给移动用户发送推断结果的鲁棒性

$$\Pr(\text{SINR}_k(\mathbf{v}; \mathbf{h}_k) \geq \gamma_k) \geq 1 - \zeta, \forall k \in [K]. \quad (6.12)$$

这里 ζ 是宽容度 (tolerance level), “ $\text{SINR}_k \geq \gamma_k$ ” 被称作安全条件 (safe condition)

6.2.4 问题表述

在所提出的用于深度学习推断的边缘处理框架中, 计算和通信存在一个基础的权衡关系。具体来说, 在边缘节点处执行更多的推理任务将产生更高的计算功耗, 而由于协作传输带来增益, 下行链路的传输功耗将降低。本章提出了一种高能效和鲁棒传输的方法, 以在满足概率服务质量约束和发射功率约束的同时, 将总网络功耗降至最低。这个问题可以表示为下列概率组稀疏波束成形问题:

$$\begin{aligned} \mathcal{P}_{6.1-\text{CCP}} : \min_{\boldsymbol{v} \in \mathbb{C}^{NKL}} & \sum_{n,k} \frac{1}{\eta_n} \|\boldsymbol{v}_{nk}\|_2^2 + \sum_{n,k} P_{nk}^c I_{(n,k) \in \mathcal{T}(\boldsymbol{v})} \\ \text{s.t. } & \Pr(\text{SINR}_k(\boldsymbol{v}; \boldsymbol{h}_k) \geq \gamma_k) \geq 1 - \zeta, k \in [K] \end{aligned} \quad (6.13)$$

$$\sum_{k=1}^K \|\boldsymbol{v}_{nk}\|_2^2 \leq P_n^{\text{Tx}}, n \in [N]. \quad (6.14)$$

在边缘推断中, 数据隐私是另外一个关键的考虑因素。高风险应用中的移动用户往往是不愿意将他们的原始数据暴露给无线接入点。为了避免原始数据暴露, 许多文献研究了层级分布式结构。如文献 (Li 等, 2020) 通过决定一个深度神经网络模型的分割点, 并把分割后的两部分分别部署在移动设备和能够进行边缘计算的无线接入点上。由于仅上传了分割点之前层的输出, 这种方法可以保护数据隐私。而所提的框架也可以应用到这种保护数据隐私的层级分布式架构中。这种情况下, 输入 d_k 为在移动设备 k 本地计算出的分割点之前深度神经网络的输出。计算任务 ϕ_k 变成了使用分割点之后层来计算推断结果的任务。

为了实现服务质量对信道状态信息误差的鲁棒性, 可以通过收集 D 个独立同分布的不完美的信道状态信息样本作为数据集 $\mathcal{D} = \{\tilde{\boldsymbol{h}}^{(1)}, \dots, \tilde{\boldsymbol{h}}^{(D)}\}$, 用于学习出一个信道状态信息的不确定性模型。基于数据集 \mathcal{D} , 目标是设计一个波束成形向量 \boldsymbol{v} 使得安全条件以不低于 $1 - \zeta$ 的概率满足。然而由于不知道随机误差的先验分布, 某一种解决方法的统计保证往往是用对某一容忍度 $1 - \epsilon$ 的置信水平 $1 - \delta$ 来表述的, 如想定生成方法 (Nemirovski 和 Shapiro, 2006)。也就是说, 对于给定的 \boldsymbol{v} 和 \mathcal{D} ,

$$\Pr(\text{SINR}_k(\boldsymbol{v}; \boldsymbol{h}_k) \geq \gamma_k) \geq 1 - \epsilon \quad (6.15)$$

的置信水平不低于 $1 - \delta$, 其中 $0 < \epsilon < 1, 0 < \delta < 1$. 因此违反安全条件的概率上

限为

$$\Pr(\text{SINR}_k(\boldsymbol{v}; \boldsymbol{h}_k) < \gamma_k) < \delta + \epsilon(1 - \delta). \quad (6.16)$$

通过取值 ϵ 和 δ 使得 $\zeta > \delta + \epsilon(1 - \delta)$ 成立，安全条件 (6.12) 能够保证成立。

本章考虑块衰落信道 (block fading channel)，假设其信道分布在 T_s 个时间块内保持不变 (Liu 等, 2019)，信道系数向量在每个块内保持不变。注意到每个时间块内收集 D 个信道样本来训练误差模型会导致很高的信令开销，故将于第6.3.4节中介绍一种可以与所提的解决概率服务质量约束的方法相结合的低成本信道采样策略。

6.2.5 问题分析

直接求解联合概率约束 (6.13) 通常是一个极度困难的任务 (Nemirovski 和 Shapiro, 2006)，特别当对于不确定性本身没有确切的信息时。这里将展示一种不需要知道随机误差的先验分布的通用框架，以支持边缘推断。一个自然的想法是找概率服务质量约束 (6.13) 的可计算的近似。

6.2.5.1 想定生成方法

想定生成方法 (scenario generation, SG) (Nemirovski 和 Shapiro, 2006) 是一个著名的解决概率约束问题的方法，通过获得 D 个随机信道系数向量 \boldsymbol{h} 的独立样本，并施加每个样本都要满足的目标服务质量约束 $\text{SINR}_k \geq \gamma_k, k \in [K]$ 。然而，当增加采样样本数目的时候这种方法会变得更加的保守，这是因为可行区域的体积会下降，这可能会导致问题变得不可行。另外，所需的样本数目 D 须满足 $\sum_{i=1}^{NKL-1} \binom{D}{i} \epsilon^i (1 - \epsilon)^{D-i} \leq \delta$ ，其中 $1 - \delta$ 是式 (6.12) 定义的概率服务质量约束的置信水平。由此可知，所需的最小样本数目 D 在 ϵ 较小时大概随着 $1/\epsilon$ 线性增加，也随着 NKL 线性增加，这导致了想定生成方法存在伸缩性的问题。

6.2.5.2 随机规划方法

为了解决想定生成方法存在的过于保守的问题，(Shi 等, 2015b) 提供了一种随机规划的方法。其原理是找了一个概率约束的 DC 近似，得到的具有 DC 约束的问题可以在每次迭代中使用蒙特卡洛方法结合连续凸近似方法来求解。然而，这种方法的计算开销随着样本树木 D 而线性增加，这严重的影响了获得高鲁棒性的解。同时，随机规划方法在有限样本数目下的统计保证仍然未知。

为了解决现有方法的局限性，将于第6.3节中展示一种通过统计学习来得到概率约束的一个鲁棒优化近似的方法(Hong 等, 2017)。这种方法的主要有点是其所需最小样本数目仅有 $\log \delta / \log(1 - \epsilon)$ ，而计算开销于样本数目是无关的。

6.3 基于统计学习的联合概率约束的鲁棒优化近似

本节用鲁棒优化来近似优化问题 $\mathcal{P}_{6.1\text{-CCP}}$ 中的联合概率约束，用一种统计学习的方法来学习鲁棒优化中的高概率区域的形状和大小。

6.3.1 近似联合概率约束的鲁棒优化方法

鲁棒优化 (robust optimization) (Hong 等, 2017) 使用了安全近似方法，施加了一个让安全条件在随机变量落在某个几何集时总是满足的约束。具体来说，联合概率约束 (6.13) 的鲁棒优化近似可以表示为

$$\text{SINR}_k(\mathbf{v}; \mathbf{h}_k) \geq \gamma_k, \mathbf{h}_k \in \mathcal{U}_k, \forall k \in [K] \quad (6.17)$$

其中 \mathcal{U}_k 即为 \mathbf{h}_k 所在的高概率区域 (high probability region)。作为联合概率约束的近似，鲁棒优化方法返回的解应使得概率服务质量约束以高置信水平满足。鲁棒优化近似的方法是通过用数据集 \mathcal{D} 构建一个高概率区域 \mathcal{U}_k 来实现的，其中高概率区域 \mathcal{U}_k 包含了 \mathbf{h}_k 的 $1 - \epsilon$ 大小的容量，即

$$\Pr(\mathbf{h}_k \in \mathcal{U}_k) \geq 1 - \epsilon, \quad (6.18)$$

以至少 $1 - \delta$ 的置信度成立。通过对高概率区域内的元素施加服务质量约束，即式 (6.17)，概率服务质量约束 (6.15) 的置信水平就至少为 $1 - \delta$ 。于是可以得到优化问题 $\mathcal{P}_{6.1\text{-CCP}}$ 的鲁棒优化近似为

$$\begin{aligned} \mathcal{P}_{6.1\text{-RO}} : & \underset{\mathbf{v}, \mathbf{h}}{\text{minimize}} \quad \sum_{n,k} \frac{1}{\eta_n} \|\mathbf{v}_{nk}\|_2^2 + \sum_{n,l} P_{nl}^c I_{(n,k) \in \mathcal{T}(\mathbf{v})} \\ & \text{subject to} \quad \text{SINR}_k(\mathbf{v}; \mathbf{h}_k) \geq \gamma_k, \mathbf{h}_k \in \mathcal{U}_k, k \in [K] \\ & \quad \sum_{k=1}^K \|\mathbf{v}_{nk}\|_2^2 \leq P_n^{\text{Tx}}, n \in [N]. \end{aligned} \quad (6.19)$$

不确定性集合 \mathcal{U}_k 的几何形状的选择对于鲁棒优化近似方法的可解性和性能来说十分关键。受鲁棒优化的可解性启发，椭圆体和多面体是最常见的基础不确定集合的选择。不确定性集合进一步可以增广为这些基础集合的交集或者

并集。本章选取了椭圆体的不确定性集合来建模每一组的信道系数向量 \mathbf{h}_k 的不确定性，这是考虑到了椭圆体集合在建模信道状态信息不确定性的广泛使用 (Shi 等, 2015a; Hanif 等, 2013) 以及它的可解性（将于第6.3.3节展示）。椭圆体的高概率区域 \mathcal{U}_k 可以参数化为

$$\mathcal{U}_k = \{\mathbf{h}_k : \mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{B}_k \mathbf{u}_k, \mathbf{u}_k^H \mathbf{u}_k \leq 1\}. \quad (6.20)$$

这里参数 $\mathbf{B}_k \in \mathbb{C}^{NL \times NL}$ 和 $\hat{\mathbf{h}}_k \in \mathbb{C}^{NL}$ 将从数据集 \mathcal{D} 中习得，在第6.3.2节中将展示如何学习。在这之后将会于第6.3.3节中展示鲁棒优化近似问题 $\mathcal{P}_{6.1\text{-RO}}$ 的一个可解的重新表述。

6.3.2 从数据样本中学习高概率区域

注意到 (6.17) 仅仅给出了可行性的保证，即保证了联合概率约束以置信水平不低于 $1 - \delta$ 成立，但是它的保守程度仍然是一个具有挑战性的问题。一般来说，如果优化问题 $\mathcal{P}_{6.1\text{-RO}}$ 的可行区域更大的话，它就是优化问题 $\mathcal{P}_{6.1\text{-CCP}}$ 的一个更不保守的近似。因此，具有更小体积的高概率区域 \mathcal{U} 因其提供了更大的可行区域而是更好的选择。进一步，可以通过减小高概率区域的体积使得概率服务质量约束的置信水平离 $1 - \delta$ 更加接近，而不是仅仅比它大。本小节展示了一种统计学习的方法 (Hong 等, 2017) 来确定高概率区域 \mathcal{U} 的参数，包含了形状学习步骤和通过分位点估计的大小校准步骤两步。首先要把数据集 \mathcal{D} 中的样本分成两部分，即 $\mathcal{D}^1 = \{\tilde{\mathbf{h}}^{(1)}, \dots, \tilde{\mathbf{h}}^{(D_1)}\}$ 和 $\mathcal{D}^2 = \{\tilde{\mathbf{h}}^{(D_1+1)}, \dots, \tilde{\mathbf{h}}^{(D)}\}$ ，每一个部分用于一个步骤。

6.3.2.1 形状学习

每一个椭圆体集合 \mathcal{U}_k 可以重新参数化为

$$\mathcal{U}_k = \{\mathbf{h}_k : (\mathbf{h}_k - \hat{\mathbf{h}}_k)^T \Sigma_k^{-1} (\mathbf{h}_k - \hat{\mathbf{h}}_k) \leq s_k\}, \quad (6.21)$$

其中 $\hat{\mathbf{h}}_k$ 和 Σ_k 是椭圆体 \mathcal{U}_k 的形状参数，而 $s_k > 0$ 决定了它的大小， $\Sigma_k/s_k = \mathbf{B}_k \mathbf{B}_k^H$. 令所有观察到的 \mathbf{h}_k 的样本表示为 $\mathcal{D}_k = \mathcal{D}_k^1 \cup \mathcal{D}_k^2 = \{\tilde{\mathbf{h}}_k^{(j)}\}_{j=1}^D$. 形状参数 $\hat{\mathbf{h}}_k$ 可以选为样本平均值，即

$$\hat{\mathbf{h}}_k = \frac{1}{D_1} \sum_{j=1}^{D_1} \tilde{\mathbf{h}}_k^{(j)}, \quad (6.22)$$

为了降低椭圆体的复杂度，这里忽略每个 $\{\mathbf{h}_{kn}\}$ 之间的关联性，令 $\boldsymbol{\Sigma}_k$ 为一个块对角矩阵，每个对角块是 \mathbf{h}_{kn} 的第一部分数据集的样本方差，

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{k1} & & \\ & \ddots & \\ & & \boldsymbol{\Sigma}_{kN} \end{bmatrix}, \text{ 其中}$$

$$\boldsymbol{\Sigma}_{kn} = \frac{1}{D_1 - 1} \sum_{j=1}^{D_1} (\tilde{\mathbf{h}}_{kn}^{(j)} - \hat{\mathbf{h}}_{kn})(\tilde{\mathbf{h}}_{kn}^{(j)} - \hat{\mathbf{h}}_{kn})^H. \quad (6.23)$$

6.3.2.2 使用分位点估计的大小校准

使用第二部分的数据集 \mathcal{D}_k^2 来对椭圆体的大小参数 s_k 做校准。其核心思想是对 \mathcal{D}_k^2 中的数据样本做一个变换，然后估计出置信水平不低于 $1 - \delta$ 的一个 $(1 - \epsilon)$ -分位点。令

$$\mathcal{G}(\xi) = (\xi - \hat{\mathbf{h}}_k)^T \boldsymbol{\Sigma}_k^{-1} (\xi - \hat{\mathbf{h}}_k) \quad (6.24)$$

为从 \mathbf{h}_k 所在的随机空间到 \mathbb{R} 的一个映射。大小参数 s_k 将基于所有在 \mathcal{D}_{nk}^2 中的样本值，选为 $\mathcal{G}(\xi)$ 函数值的潜在分布的 $(1 - \epsilon)$ -分为点的估计值。其中 $(1 - \epsilon)$ -分为点用 $q_{1-\epsilon}$ 来表示，其定义是

$$\Pr(\mathcal{G}(\xi) \leq q_{1-\epsilon}) = 1 - \epsilon. \quad (6.25)$$

具体来说，计算 \mathcal{G} 对数据集 \mathcal{D}_k^2 中每一个样本的函数值，得到了一系列的观察值 G_1, \dots, G_{D-D_1} ，其中 $G_j = \mathcal{G}(\mathbf{h}_k^{(D_1+j)})$ 。然后将观察值按降序进行排序得到 $G_{(1)} \leq \dots \leq G_{(D-D_1)}$ ，根据如下的命题可知，第 t^* 个值 $G_{(j^*)}$ 即为 $\mathcal{G}(\xi)$ 潜在分布的 $(1 - \epsilon)$ 分位点的一个上界。

命题 6.1. 如果 s_k 设为

$$s_k = G_{(j^*)}, \text{ 其中 } j^* \text{ 为}$$

$$\min_{1 \leq j \leq D-D_1} \left\{ j : \sum_{k=0}^{j-1} \binom{D-D_1}{k} (1-\epsilon)^k e^{D-D_1-k} \geq 1 - \delta \right\}, \quad (6.26)$$

则命题 s_k 是潜在分布的 $(1 - \epsilon)$ -分位点的一个上界成立的置信水平是 $1 - \delta$ ，即

$$\Pr(s_k \geq q_{1-\epsilon}) \geq 1 - \delta. \quad (6.27)$$

证明. 根据分位点 $q_{1-\epsilon}$ 的定义有

$$\begin{aligned} & \Pr(G_{(j)} \geq q_{1-\epsilon}) \\ &= \Pr(G_{(k)} < q_{1-\epsilon}, k = 0, \dots, j-1) \\ &= \sum_{k=0}^{j-1} \binom{D - D_1}{k} (1-\epsilon)^k \epsilon^{D-D_1-k}. \end{aligned} \quad (6.28)$$

因此 $G_{(j^\star)}$ 是所有具有置信水平 $1 - \delta$ 的潜在分布的 $(1 - \epsilon)$ -分位点的上界中最小的一个。 \square

使用所展示的两个步骤可以习得随机的信道状态信息 h_k 的一个高概率区域 \mathcal{U} 。这种基于统计学习的鲁棒优化近似方法的统计保证由如下命题给出：

命题 6.2. 假设每个数据集 D_k 中的样本数据是独立同分布的，且从一个连续分布中生成。数据集分成了两个独立部分 D_k^1 和 D_k^2 。每一个不确定性集合选作 $\mathcal{U}_k = \{h_k : (h_k - \hat{h}_k)^T \Sigma_k^{-1} (h_k - \hat{h}_k) \leq s_k\}$ 。他们的参数 \hat{h}_k , Σ_k , 和 s_k 分别通过式 (6.22)、式 (6.23) 以及式 (6.26) 确定。如此优化问题 $\mathcal{P}_{6.1-RO}$ 的任意可行解都能保证概率服务质量约束 (6.15) 以至少 $1 - \delta$ 的置信度满足。

证明. 由于 \mathcal{G} 仅依赖于 D_k^1 , 可知

$$\Pr_{D_k^1}(\mathbf{v} \in \mathcal{V}) = \Pr_{D_k^1}(G_{(t^\star)} \geq q_{1-\epsilon}) \geq 1 - \delta. \quad (6.29)$$

因此可得 $\Pr(\text{SINR}_k \geq \gamma_k) \geq 1 - \epsilon$ 以至少 $1 - \delta$ 的置信水平满足。 \square

注意到仅当

$$\sum_{k=0}^{D-D_1-1} \binom{D - D_1}{k} (1-\epsilon)^k \epsilon^{D-D_1-k} \geq 1 - \delta, \quad (6.30)$$

时 j^\star 才存在, 也就是说需要满足 $1 - (1 - \epsilon)^{D-D_1} \geq 1 - \delta$. 换句话说, 为了达到概率服务质量约束 (6.13) 的 $1 - \delta$ 置信水平, 所需的最小样本数目是 $D > D - D_1 \geq \log \delta / \log(1 - \epsilon)$. 矩阵 \mathbf{B}_k 可以通过下式来计算

$$\mathbf{B}_k = \sqrt{s_k} \boldsymbol{\Delta}_k, \quad (6.31)$$

其中 $\boldsymbol{\Delta}_k$ 是 $\boldsymbol{\Sigma}_k$ 的 Cholesky 分解, 即 $\boldsymbol{\Sigma}_k = \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^H$. 整个从数据集 \mathcal{D} 中学习高概率区域 \mathcal{U} 的过程可以总结为算法5。

算法 5 高概率区域 \mathcal{U}_k 的统计学习方法

- 1: **输入:** 数据集 $\mathcal{D} = \{\tilde{\mathbf{h}}^{(1)}, \dots, \tilde{\mathbf{h}}^{(D)}\}$.
 - 2: **for** $k = 1, \dots, K$ **do**
 - 3: **数据分割:** 将 \mathbf{h}_k 的数据样本 \mathcal{D}_k 随机分割为两部分 \mathcal{D}_k^1 和 \mathcal{D}_k^2 .
 - 4: **形状学习:** 基于数据集 \mathcal{D}_k^1 , 将形状参数 $\hat{\mathbf{h}}_k$ 和 Σ_k 如式 (6.22) 和式 (6.23) 选取。
 - 5: **大小校准:** 计算 \mathcal{G} 对于 \mathcal{D}_k^2 中的样本值的函数值, 将大小参数 s_k 设为 $G_{(j^\star)}$, 其中 \mathcal{G} 由式 (6.24) 给出, j^\star 按式 (6.26) 来计算。
 - 6: 用 Cholesky 分解 $\Sigma_k = \mathbf{A}_k \mathbf{A}_k^H$ 来计算出 $\mathbf{B}_k = \sqrt{s_k} \mathbf{A}_k$.
 - 7: **end for**
 - 8: **输出:** 所有的 $\hat{\mathbf{h}}_k$ 和 \mathbf{B}_k .
-

6.3.3 重新表述鲁棒优化为易处理形式

根据式 (6.20) 中的椭圆体不确定性模型, 鲁棒优化近似问题 (6.17) 可以重写为

$$\mathbf{h}_k^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{h}_k \geq \sigma_k^2 \quad (6.32)$$

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{B}_k \mathbf{u}_k, \mathbf{u}_k^H \mathbf{u}_k \leq 1, \quad (6.33)$$

其中 $\mathbf{u}_{nk} \in \mathbb{C}^L$. 通过定义矩阵

$$\mathbf{H}_k = \begin{bmatrix} \hat{\mathbf{h}}_k & \mathbf{B}_k \end{bmatrix} \in \mathbb{C}^{NL \times (NL+1)}, \quad (6.34)$$

并使用 S-过程 (Boyd 和 Vandenberghe, 2004), 可以得到式 (6.32) 和 (6.33) 如下的易于处理的重新表述形式:

$$\mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{H}_k \succeq \mathbf{Q}_k \quad (6.35)$$

$$\lambda_k \geq 0. \quad (6.36)$$

其中 $\lambda = [\lambda_k] \in \mathbb{R}_+^K$, \mathbf{Q}_k 为

$$\mathbf{Q}_k = \begin{bmatrix} \lambda_k + \sigma_k^2 & 0 \\ 0 & -\lambda_k \mathbf{I}_{NL} \end{bmatrix} \in \mathbb{C}^{(NL+1) \times (NL+1)}. \quad (6.37)$$

根据式 (6.32) 和 (6.33) 推导出式 (6.35) 和式 (6.36) 的详细过程留于附录B.1中。

因此所提的对于优化问题 $\mathcal{P}_{6.1-\text{CCP}}$ 的鲁棒优化近似可以表示为一个如下的有非凸二次约束的组稀疏波束成形问题：

$$\begin{aligned} \mathcal{P}_{6.1-\text{RGS}} : & \underset{\boldsymbol{v} \in \mathbb{C}^{NKL}, \lambda \in \mathbb{R}^K}{\text{minimize}} \sum_{n,l} \frac{1}{\eta_n} \|\boldsymbol{v}_{nl}\|_2^2 + \sum_{n,l} P_{nl}^c I_{(n,l) \in \mathcal{T}(\boldsymbol{v})} \\ & \text{subject to} \quad \mathbf{H}_k^H \left(\frac{1}{\gamma_k} \boldsymbol{v}_k \boldsymbol{v}_k^H - \sum_{l \neq k} \boldsymbol{v}_l \boldsymbol{v}_l^H \right) \mathbf{H}_k \succeq \mathbf{Q}_k, \lambda_k \geq 0, \forall k \in [K] \\ & \quad \sum_{l=1}^K \|\boldsymbol{v}_{nl}\|_2^2 \leq P_n^{\text{Tx}}, \forall n \in [N]. \end{aligned} \quad (6.39)$$

解决优化问题 $\mathcal{P}_{6.1-\text{RGS}}$ 的计算复杂度是与样本数目 D 无关的。一种有效的得到非凸二次约束的二次规划问题的近似解的方式是将问题变量，即聚合波束成形向量升维为一个秩为 1 的半正定矩阵 $\mathbf{V} = \boldsymbol{v} \boldsymbol{v}^H$ ，然后简单的忽略秩为 1 的非凸约束，这种方法被成为半正定松弛法 (Luo 等, 2007)。然而，由于简单的忽略了秩为 1 的约束，这种方法所得的解可能对于原来的非凸二次约束是不可行的。为了诱导具有非凸二次约束的组稀疏性，文献 (Shi 等, 2015a) 采用了加权的混合 ℓ_1/ℓ_2 范数的一种二次变形。本章将用一种迭代加权最小化方法来进一步增强聚合波束成形向量的组稀疏性，这种方法已经在云无线接入网中显示了其有效性 (Dai 和 Yu, 2016; Shi 等, 2016b)。除此之外，为了提高迭代加权方法的每个子问题中的非凸二次约束的可行性，本章提供了一种 DC 方法来诱导秩为 1 的解。值得一提的是，由于信道状态信息的不确定性，上下行链路的对偶性质不能用于高效的解决式 (6.35) 中的鲁棒服务质量约束。

6.3.4 与鲁棒优化近似方法相结合的一种低成本采样策略

考虑块衰落信道，信道分布在一段时间内保持不变，这个时间段称为信道统计信息的相干间隔 (Liu 等, 2019)。信道统计信息的相干间隔是由 T_s 个块组成每个块被称作一个信道状态信息的相干间隔，在这段间隔内信道系数向量保持不变。然而，在每个块内手机 D 个信道样本会导致很高的信令开销。为了解决这个问题，本部分提供了一个低成本的采样策略用于鲁棒传输，其时间线如图6.3所示。

在信道统计信息的相干间隔的开始处，收集 D 个独立同分布的信道样本，记为 \mathcal{D} 。基于数据集 \mathcal{D} 可以利用式 (6.22) 得到所估计的信道系数向量 $\hat{\mathbf{h}}_k$ ，利用式 (6.31) 得到估计处的误差 \boldsymbol{e}_k 所在的高概率区域进而得到参数 \mathbf{B}_k 。在第一个

块中的传输需要利用式(6.34)结合这两部分得到 $\{\mathbf{H}_k\}$, 然后解相应的优化问题 $\mathcal{P}_{6.1\text{-RGS}}$. 对于其他第 t 块来说 ($t > 1$), 可以收集少至 1 个信道系数向量的样本, 用其样本均值来作为估计的信道系数 $\hat{\mathbf{h}}[t]$ 。然后将其替换所估计的信道系数 $\hat{\mathbf{h}}$ 并保持误差信息参数 $\{\mathbf{B}_k : k \in [K]\}$, 可以构建出第 t 个块的参数 $\{\mathbf{H}_k[t]\}$ 为

$$\mathbf{H}_k[t] = [\hat{\mathbf{h}}_k[t] \ \mathbf{B}_k], \forall k \in [K], \quad (6.40)$$

然后带入解优化问题 $\mathcal{P}_{6.1\text{-RGS}}(\{\mathbf{H}_k[t]\})$ 来得到发送波束成形向量。这种方法可以极大程度的减少采样带来的信令开销, 其有效性将于第6.5.1节中通过数值仿真来展示。

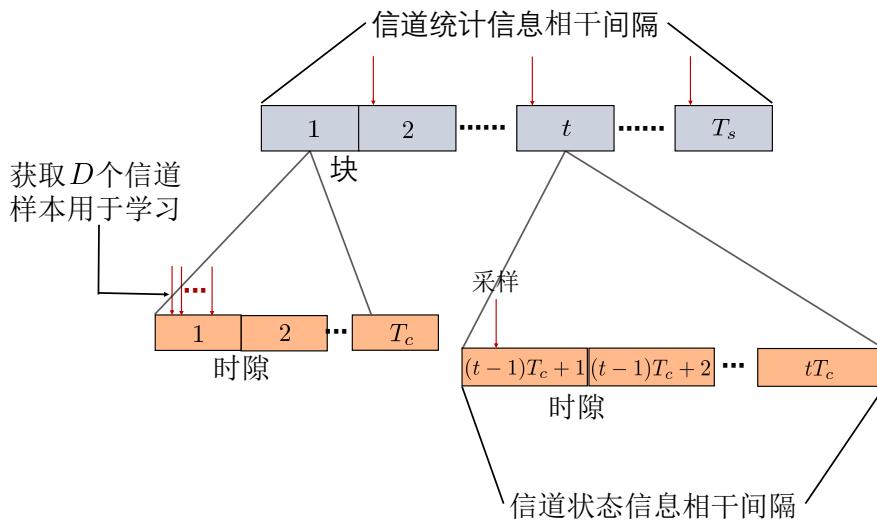


图 6.3 低成本信道采样方法的时间线示意图。

Figure 6.3 Timeline of a cost-effective channel sampling strategy.

6.4 用于求解有非凸二次约束的组稀疏波束成形问题的加权功率最小化方法

本节提供了一种加权功率最小化的方法来诱导问题 $\mathcal{P}_{6.1\text{-RGS}}$ 中的组稀疏结构。该问题中的非凸二次约束可以用矩阵升维技术将其重新表述为关于一个秩为 1 的半正定矩阵的凸约束, 进而其可以通过 DC 方法来诱导秩为 1 的解。

6.4.1 用于解决非凸二次约束的矩阵升维技术

可以观察到约束条件(6.35)关于 $\mathbf{v}\mathbf{v}^H$ 是凸的, 但是关于变量 \mathbf{v} 本身是非凸的约束。这启发了利用矩阵升维技术(Luo 等, 2007)来解决问题 $\mathcal{P}_{6.1\text{-RGS}}$ 中的非

凸二次约束。记

$$\mathbf{V}_{ij}[s, t] = \mathbf{v}_{si}\mathbf{v}_{tj}^H \in \mathbb{C}^{L \times L} \quad (6.41)$$

$$\mathbf{V}_{ij} = \begin{bmatrix} \mathbf{V}_{ij}[1, 1] & \cdots & \mathbf{V}_{ij}[1, N] \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{ij}[N, 1] & \cdots & \mathbf{V}_{ij}[N, N] \end{bmatrix} = \mathbf{v}_i\mathbf{v}_j^H \in \mathbb{C}^{NL \times NL} \quad (6.42)$$

$$\mathbf{V} = \mathbf{v}\mathbf{v}^H = \begin{bmatrix} \mathbf{V}_{11} & \cdots & \mathbf{V}_{1K} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{K1} & \cdots & \mathbf{V}_{KK} \end{bmatrix} \in \mathbb{S}_+^{NKL}, \quad (6.43)$$

其中 \mathbb{S}_+^{NKL} 表示厄米特 (Hermitian) 的半正定矩阵的集合。如此聚合波束成形向量 \mathbf{v} 被提升为了一个秩为 1 的半正定矩阵 \mathbf{V} 。优化问题 $\mathcal{P}_{6.1\text{-RGS}}$ 的约束条件 C_k , 即式 (6.35), 可以等价重写为如下半正定约束

$$\mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{H}_k \succeq \mathbf{Q}_k, \quad (6.44)$$

而发送功率约束 (6.39) 可以等价重写为

$$\sum_{l=1}^K \|\mathbf{v}_{nl}\|_2^2 = \sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n, n]) \leq P_n^{\text{Tx}}, \forall n = 1, \dots, N. \quad (6.45)$$

因此, 使用矩阵升维技术可以得到优化问题 $\mathcal{P}_{6.1\text{-RGS}}$ 的一个等价表述

$$\begin{aligned} \mathcal{P}_{6.1} : \underset{\mathbf{V}, \lambda}{\text{minimize}} \quad & \sum_{n,l} \left(\frac{1}{\eta_n} \text{Tr}(\mathbf{V}_{ll}[n, n]) + P_{nl}^c I_{\text{Tr}(\mathbf{V}_{ll}[n, n]) \neq 0} \right) \\ \text{subject to} \quad & \mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{H}_k \succeq \mathbf{Q}_k, \lambda_k \geq 0, \forall k \in [K] \end{aligned} \quad (6.46)$$

$$\sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n, n]) \leq P_n^{\text{Tx}}, \forall n \in [N] \quad (6.47)$$

$$\mathbf{V} \succeq 0, \text{rank}(\mathbf{V}) = 1. \quad (6.48)$$

注意秩为 1 的约束条件导致了这个问题的约束条件仍然是非凸的。

6.4.2 秩为 1 约束的 DC 表示方法

对于一个非零的半正定矩阵 $\mathbf{V} \in \mathbb{S}_+^{NKL}$ 来说, 当且仅当只有其最大奇异值非零时它的秩为 1, 即

$$\sigma_i(\mathbf{V}) = 0, i = 2, \dots, NKL, \quad (6.49)$$

其中 $\sigma_i(\mathbf{V})$ 是矩阵 \mathbf{V} 第 i 大的奇异值。半正定矩阵 \mathbf{V} 的迹范数和谱范数分别为

$$\text{Tr}(\mathbf{V}) = \sum_{i=1}^{NKL} \sigma_i(\mathbf{V}), \|\mathbf{V}\| = \sigma_1(\mathbf{V}). \quad (6.50)$$

因此可以得到矩阵 \mathbf{V} 的秩为 1 约束的如下 DC 表示：

$$\mathcal{R}(\mathbf{V}) = \text{Tr}(\mathbf{V}) - \|\mathbf{V}\| = 0. \quad (6.51)$$

由迹范数和谱范数都是凸函数可知 \mathcal{R} 是关于 \mathbf{V} 的一个 DC 函数。

6.4.3 加权 ℓ_1 最小化方法诱导组稀疏性

在提高云无线接入网络的能效问题 (Dai 和 Yu, 2016; Shi 等, 2016b) 中, 加权 ℓ_1 最小化方法显示出了它在增强组稀疏性方面的优势。 ℓ_1 范数是非凸的 ℓ_0 范数的一个著名的凸近似。加权 ℓ_1 最小化方法的提出是为了进一步增强稀疏性, 通过迭代地执行最小化加权的 ℓ_1 范数与更新权重步骤。在优化问题 $\mathcal{P}_{6.1}$ 的目标函数中, 指示函数 $I_{\text{Tr}(\mathbf{V}_{ll}[n,n]) \neq 0}$ 可以解释为 $\text{Tr}(\mathbf{V}_{ll}[n,n])$ 的 ℓ_0 范数。因此可以用 $w_{nl} \text{Tr}(\mathbf{V}_{ll}[n,n])$ 来近似 $I_{\text{Tr}(\mathbf{V}_{ll}[n,n]) \neq 0}$, 引用加权的 ℓ_1 最小化技术, 交替的最小化目标函数与按照下式更新权重

$$w_{nl} = \frac{c}{\text{Tr}(\mathbf{V}_{ll}[n,n]) + \tau}, \quad (6.52)$$

其中 $\tau > 0$ 是一个正则化因子常量, $c > 0$ 是一个常数。如果 $\text{Tr}(\mathbf{V}_{ll}[n,n])$ 较小, 加权 ℓ_1 最小化方法会给对应的收发对 (n, l) 施加更大的权重, 这样会促使推断任务 l 不在第 n 个边缘服务器处执行。

6.4.4 所提的加权功率最小化方法

本部分提供了一个加权功率最小化方法, 其结合了矩阵升维技术、DC 表示与加权 ℓ_1 最小化技术。在第 j 步中通过解如下优化问题来更新 $\mathbf{V}^{[j+1]}$

$$\begin{aligned} & \underset{\mathbf{V}, \lambda}{\text{minimize}} \quad \sum_{n,l} \left(\frac{1}{\eta_n} + w_{nl}^{[j]} P_{nl}^c \right) \text{Tr}(\mathbf{V}_{ll}[n,n]) \\ & \text{subject to} \quad \mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{H}_k \succeq \mathbf{Q}_k, \lambda_k \geq 0, \forall k \in [K] \\ & \quad \sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n,n]) \leq P_n^{\text{Tx}}, \forall n \in [N] \\ & \quad \mathbf{V} \succeq 0, \text{rank}(\mathbf{V}) = 1, \end{aligned} \quad (6.53)$$

而权重 $\{w_{nl}^{[j]}\}$ 的更新是通过式(6.52), 权重在算法的开始初始化为1。

为了解决有非凸秩为1约束的优化问题(6.53), 可以使用式(6.51)给出的DC表示, 即求解如下的DC规划问题

$$\begin{aligned} \mathcal{P}_{6.1-\text{DC}} : \underset{\boldsymbol{V}, \lambda}{\text{minimize}} \quad & \sum_{n,l} \left(\frac{1}{\eta_n} + w_{nl}^{[j]} P_{nl}^c \right) \text{Tr}(\boldsymbol{V}_{ll}[n,n]) + \mu \mathcal{R}(\boldsymbol{V}) \\ \text{subject to} \quad & \boldsymbol{H}_k^H \left(\frac{1}{\gamma_k} \boldsymbol{V}_{kk} - \sum_{l \neq k} \boldsymbol{V}_{ll} \right) \boldsymbol{H}_k \succeq \boldsymbol{Q}_k, \lambda_k \geq 0, \forall k \in [K] \\ & \sum_{l=1}^K \text{Tr}(\boldsymbol{V}_{ll}[n,n]) \leq P_n^{\text{Tx}}, \forall n \in [N] \\ & \boldsymbol{V} \succeq 0, \end{aligned} \quad (6.54)$$

其中 $\mu > 0$ 是正则化参数。尽管DC问题也是非凸的, 优化问题 $\mathcal{P}_{6.1-\text{DC}}$ 可以用简化的DC算法高效求解, 即迭代地对凹部分线性化(Tao和An, 1997)。第 t 步迭代中需要解

$$\begin{aligned} \underset{\boldsymbol{V}, \lambda}{\text{minimize}} \quad & \sum_{n,l} \left(\frac{1}{\eta_n} + w_{nl}^{[j]} P_{nl}^c \right) \text{Tr}(\boldsymbol{V}_{ll}[n,n]) + \mu (\text{Tr}(\boldsymbol{V}) - \text{Tr}(\boldsymbol{G}^{(t)} \boldsymbol{V})) \\ \text{subject to} \quad & \boldsymbol{H}_k^H \left(\frac{1}{\gamma_k} \boldsymbol{V}_{kk} - \sum_{l \neq k} \boldsymbol{V}_{ll} \right) \boldsymbol{H}_k \succeq \boldsymbol{Q}_k, \lambda_k \geq 0, \forall k \in [K] \\ & \sum_{l=1}^K \text{Tr}(\boldsymbol{V}_{ll}[n,n]) \leq P_n^{\text{Tx}}, \forall n \in [N] \\ & \boldsymbol{V} \succeq 0, \end{aligned} \quad (6.55)$$

其中 $\boldsymbol{G}^{(t)}$ 是谱范数在 $\boldsymbol{V}^{(t)}$ 处的一个次梯度。它可以计算为 $\partial \|\boldsymbol{V}\|_2 = \boldsymbol{u}_1 \boldsymbol{u}_1^H$, 其中 \boldsymbol{u}_1 是矩阵 \boldsymbol{V} 的最大特征值对应的特征向量。这样的DC算法能够保证从任意的可行初始点收敛到优化问题 $\mathcal{P}_{6.1-\text{DC}}$ 的一个临界点(Tao和An, 1997)。

当加权 ℓ_1 最小化算法收敛到一个秩为1的解 $\boldsymbol{V}^{[j]}$ 时, 可以对其做Choleskey分解 $\boldsymbol{V}^{[j]} = \boldsymbol{v}^\star \boldsymbol{v}^{\star H}$ 从而提取出聚合波束成形向量 \boldsymbol{v}^\star 。所提的加权功率最小化方法的整个过程总结在算法6中。

算法 6 求解优化问题 $\mathcal{P}_{6.1}$ 的加权功率最小化方法

```

1: 初始化:  $\mathbf{V}^{[0]}, w_{nl}$ .
2: while 不收敛 do
3:    $\mathbf{V}^{(0)} \leftarrow \mathbf{V}^{[j]}$ .
4:   while 不收敛 do
5:     更新  $\mathbf{V}^{(t)}$  为优化问题 (6.55) 的解。
6:   end while
7:    $\mathbf{V}^{[j+1]} \leftarrow \mathbf{V}^{(t)}$ .
8:   根据式 (6.52) 更新权重  $\{w_{nl}^{[j+1]}\}$ .
9: end while
10: 通过 Choleskey 分解  $\mathbf{V}^{[j]} = \mathbf{v}^* \mathbf{v}^{*H}$  得到  $\mathbf{v}^*$ .
11: 输出:  $\mathbf{v}^*$ .

```

6.5 仿真结果与分析

本节提供了丰富的数值实验来比较所提出的方法与其他方法的性能。该边缘推断系统中，随机产生了 4 个无线接入点，其位于 $(\pm 400, \pm 400)$ 米的位置，而 $K = 4$ 个移动用户随机坐落于 $[-800, 800] \times [-800, 800]$ 米的方形区域中。每一个无线接入点配备了 $L = 2$ 根天线。有误差的信道系数向量模型按如下方式设置：第 n 个无线接入点和第 k 个移动用户的信道系数为 $\mathbf{h}_{kn} = 10^{-L(d_{kn})/20}(\mathbf{c}_{kn} + \mathbf{e}_{kn})$ 。路径损耗模型设置为 $L(d_{kn}) = 128.1 + 37.6 \log_{10} d_{kn}$ ，瑞利分布的小尺度衰落系数为 $\mathbf{c}_{kn} \sim \mathcal{CN}(0, \mathbf{I})$ ，加性误差是 $\mathbf{e}_{kn} \sim \mathcal{CN}(0, 10^{-4} \mathbf{I})$ 。如第 6.3.2 节中所述， D_1 的大小决定了所学的不确定性集合的形状的准确度，而 D_2 的大小决定了不确定性集合的大小校准的准确度。为了平衡这两点，假设收集到的 \mathbf{h}_{kn} 的 D 个独立样本平均分为两半，分别用于学习不确定椭圆体集合的形状和大小，即 $D_1 = D_2 = D/2$ 。每个无线接入点的功放效率取为 $\eta_1 = \dots = \eta_N = 1/4$ 。平均最大发送功率设置为 $P_1^{\text{Tx}} = \dots = P_N^{\text{Tx}} = 1W$ ，而在每个无线接入点处计算一个任务的计算功耗设置为 $P_{nk}^c = 0.60W$ 。将目标信干噪比设置为 $\gamma_1 = \dots = \gamma_K = \gamma$ ，容忍度取为 $\epsilon = 0.05$ ，置信水平取为 $\delta = 0.05$ 。正则化参数 τ 设为 10^{-6} ， μ 取为 10。

6.5.1 将信道状态信息不确定性纳入考虑的优势

本章中考虑了信道状态信息的不确定性，提出了一种基于统计学习的鲁棒优化近似方法来求解。为了进一步降低信道采样的开销，第 6.3.4 节中提供了一种

低成本的采样策略。现通过仿真来观察这种方法相比于不考虑信道状态信息误差的方法的性能，假设每一个任务都被所有的无线接入点计算。具体来说，在训练阶段的一个信道状态信息相干间隔内收集 $D = 200$ 个独立同分布的信道样本。在测试阶段，仅收集一个信道样本 $\mathbf{h}^{(1)}$ ，然后根据式 (6.40) 来构建 \mathbf{H}_k ，求解如下问题

$$\begin{aligned} & \underset{\mathbf{V}, \lambda}{\text{minimize}} \quad \sum_{n,l} \left(\frac{1}{\eta_n} \text{Tr}(\mathbf{V}_{ll}[n,n]) + P_{nl}^c \right) \\ & \text{subject to} \quad \mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{H}_k \succeq \mathbf{Q}_k, \lambda_k \geq 0, \forall k \in [K] \\ & \quad \sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n,n]) \leq P_n^{\text{Tx}}, \forall n \in [N] \\ & \quad \mathbf{V} \succeq 0. \end{aligned} \tag{6.56}$$

不考虑不确定性的波束成形设计方案作为对比，即求解优化问题

$$\begin{aligned} & \underset{\mathbf{V}, \lambda}{\text{minimize}} \quad \sum_{n,l} \left(\frac{1}{\eta_n} \text{Tr}(\mathbf{V}_{ll}[n,n]) + P_{nl}^c \right) \\ & \text{subject to} \quad \mathbf{h}_k^{(1)H} \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{h}_k^{(1)} \geq \sigma_k^2, \forall k \\ & \quad \sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n,n]) \leq P_n^{\text{Tx}}, \forall n, \\ & \quad \mathbf{V} \succeq 0. \end{aligned} \tag{6.57}$$

值得注意的是为了对比公平二者均使用了半正定松弛方法。这里产生了 40000 次信道采样实现用于测试，对于所提的方法，每 200 次实现重新产生一次训练数据集。计算出每个方法得到的解对应每个移动用户达到的信干噪比，即 $\text{SINR}_k(\mathbf{v}; \tilde{\mathbf{h}})$ ，其中 $\tilde{\mathbf{h}}$ 是真实的信道系数向量，然后统计出有多少次信道实现中目标每个移动用户的服务质量约束是满足的，即 $\text{SINR}_k \geq \gamma_k$ 。仿真的结果展示于表 6.1 中，它说明了所提的鲁棒优化近似方法能够很大程度上提高服务质量对于信道状态信息误差的鲁棒性，且仅使用了一种低成本的采样策略。

6.5.2 克服想定生成方法的过于保守特性

正如第 6.2.5 节所指出的那样，想定生成方法过于保守，因为它强制目标服务质量约束要对所有样本都成立，从而导致了更小的可行区域。这里用数值仿真实验来展示所提鲁棒优化近似方法在克服保守性上的优势。考虑鲁棒优化近似方

表 6.1 服务质量满足的测试次数

Table 6.1 Number of tests that QoS is met.

用户索引	1	2	3	4
所提方案	39946	39946	39946	39946
不考虑不确定性方案	15205	15123	15197	15214

法的可行性问题

$$\begin{aligned}
 & \text{find } \mathbf{V}, \lambda \\
 \text{subject to } & \mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{H}_k \geq \mathbf{Q}_k, \lambda_k \geq 0, \forall k \in [K], \\
 & \sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n, n]) \leq P_n^{\text{Tx}}, \forall n \in [N], \\
 & \mathbf{V} \succeq 0,
 \end{aligned} \tag{6.58}$$

而想定生成方法的可行性问题是

$$\begin{aligned}
 & \text{find } \mathbf{V} \\
 \text{subject to } & \mathbf{h}_k^{(i)H} \left(\frac{1}{\gamma_k} \mathbf{V}_{kk} - \sum_{l \neq k} \mathbf{V}_{ll} \right) \mathbf{h}_k^{(i)} \geq \sigma_k^2, \forall k, i \\
 & \sum_{l=1}^K \text{Tr}(\mathbf{V}_{ll}[n, n]) \leq P_n^{\text{Tx}}, \forall n, \\
 & \mathbf{V} \succeq 0.
 \end{aligned} \tag{6.59}$$

为了比较的公平性，二者均采用了半正定松弛的方法。每次实现中收集 $D = 200$ 个独立同分布的信道样本，每个算法运行 25 次随机实现，比较想定生成方法和鲁棒优化近似方法的返回可行解的几率。如图6.4所示，仿真结果显示基于统计学习的鲁棒优化近似方法能够相当程度上提高返回可行解的几率。

6.5.3 收敛行为

通过选定加权参数为 $c = 1/\ln(1 + \tau^{-1})$ ，由

$$I_{x \neq 0} = \|x\|_0 = \lim_{\tau \rightarrow 0} \ln(1 + x\tau^{-1})/\ln(1 + \tau^{-1}) \tag{6.60}$$

和文献(Dai 和 Yu, 2016)可知，所提的加权功率最小化方法算法6本质上即为近似了 ℓ_0 范数，并在约束条件(6.46)和(6.47)下，使用 MM(majorization-minimization)

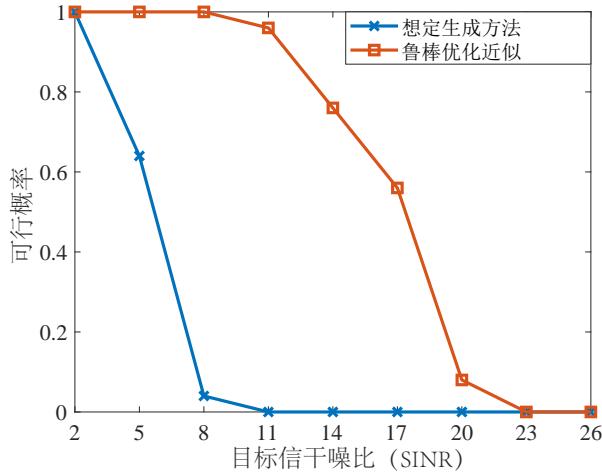


图 6.4 使用想定生成方法和鲁棒优化近似方法返回可行解的几率与目标信噪比 γ 的关系。

Figure 6.4 Probability of feasibility using scenario generation and the robust optimization approximation approach over the target SINR γ .

技术来最小化近似的目地函数

$$f(\mathbf{V}) = \sum_{n,l} \left(\frac{1}{\eta_n} \text{Tr}(\mathbf{V}_{ll}[n,n]) + P_{nl}^c \frac{\ln(1 + \tau^{-1} \text{Tr}(\mathbf{V}_{ll}[n,n]))}{\ln(1 + \tau^{-1})} \right) + \mu \mathcal{R}(\mathbf{V}). \quad (6.61)$$

图6.5表明了所提的加权功率最小化方法在收集了 $D = 200$ 个信道样本情况的收敛行为（用目标函数值 f 来表示）。相应的，聚合波束成形向量 \mathbf{v} 的组稀疏度的轨迹，即在所有无线接入点处所执行的推断任务数目，也展示于了图6.6。结果表明在边缘服务器处执行任务的数目随着目标服务质量 γ 的增加而增长，进而导致了边缘推断系统需要更高的总功耗。

6.5.4 总功耗与目标信噪比的关系

然后，选取了 $D = 200$ 个独立同分布的信道样本，通过数值仿真实验来比较求解问题 $\mathcal{P}_{6.1}$ 的不同算法的性能，包括所提的加权功率最小化方法与下列最新算法：

- 混合 $\ell_1/\ell_2 +$ 半正定松弛 (mixed $\ell_1/\ell_2 +$ SDR): 文献 (Shi 等, 2015a) 使用了这种算法，利用了加权混合 ℓ_1/ℓ_2 范数的一种二阶变形来诱导组稀疏性，以及半正定松弛 (Luo 等, 2007) 来解决非凸二次约束。
- 迭代加权算法 + 半正定松弛 (reweighted+SDR): 为了提高云无线接入网络下行传输的能效，(Dai 和 Yu, 2016) 采用了迭代加权最小化方法来诱导组稀疏性。本算法中采用了这种组稀疏性诱导方法，并用半正定松弛 (Luo 等, 2007) 来

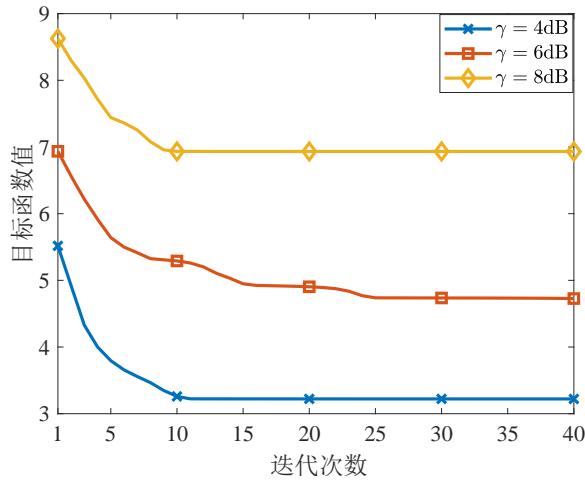


图 6.5 所提加权功率最小化方法在不同目标信噪比 γ 下的收敛行为。

Figure 6.5 Convergence behavior of the proposed reweighted power minimization approach with different target SINR γ .

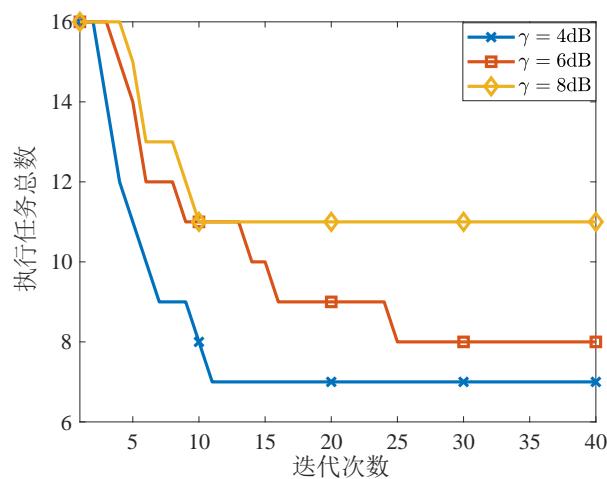


图 6.6 不同目标信噪比 γ 下在所有边缘服务器处执行推断任务总数的轨迹。

Figure 6.6 Trajectories of the total number of inference tasks performed at all edge computing nodes with different target SINR γ .

解决非凸二次约束。

- 协同波束成形 + 半正定松弛 (CB+SDR): 本方法中假设所有任务都在每一个无线接入点处执行, 然后用协同波束成形 (coordinated beamforming, CB) 的方法在概率服务质量约束下最小化传输功率。

同样按照第6.5.3节中也设置参数 $c = 1/\ln(1 + \tau^{-1})$ 。所有算法平均 100 次信道实现的性能结果如图6.7a和图 Fig. 6.7b所示。图6.7a给出了每个算法的平均总功耗, 结果显示所提的方法能够比其他方法有更低的总功耗, 这得益于其能够有更强的诱导出组稀疏解的能力 (如图6.7b所示)。注意 “协同波束成形 + 半正定松弛” 方法的总任务数目一直是 $KN = 16$.

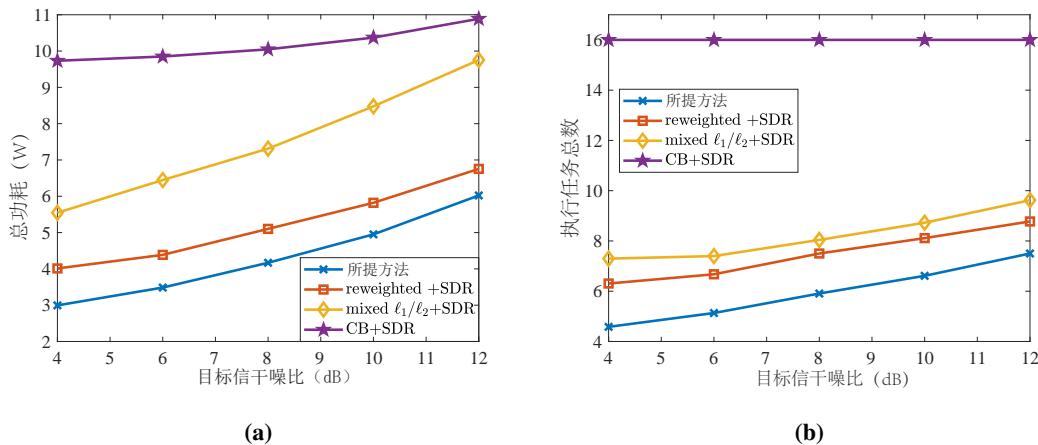


图 6.7 所提算法的性能评估。(a) 总功耗与目标信干噪比的关系, (b) 边缘服务器处执行的总任务数目与目标信干噪比的关系。

Figure 6.7 Performance of the proposed approach. (a) Total power consumption over target SINR. (b) Total # of tasks performed at APs over target SINR.

通过所有这些数值仿真结果可知, 所提供的基于统计学习的鲁棒优化近似方法与加权功率最小化方法相结合, 能够为边缘推断提供高能效的边缘处理以及高鲁棒性的传输服务。

6.6 本章小结

本章研究了基于计算任务卸载的边缘推断系统, 结合无线信道传输的特性提出了一种高能效处理以及鲁棒性传输的方案, 在保证服务质量以高概率满足的条件下最小化计算与传输的总功率。基于统计学习的方法, 该问题可以使用组稀疏与低秩优化模型来解决, 根据问题的特殊结构, 采用了迭代加权算法来诱导

组稀疏性，同时采用 DC 的正则项来诱导秩为 1 的约束，以满足非凸二次约束的可行性。数值仿真实验显示出了相比其他方法所提方法可以达到最低的总功耗，且避免了其他解决信道状态信息误差不确定性的缺点。

第7章 总结与展望

本章对全文的工作内容进行了总结，并指出了未来的研究方向。

7.1 全文总结

本文着力于在移动网络边缘实现人工智能，克服其对于无线通信系统的严峻挑战，以实现下一代无线通信系统从万物互联到智慧互联的转变。对于边缘训练与边缘推断两方面，共研究了四个使能方案，分别为移动设备分布式训练系统中的快速模型聚合，智能反射面赋能的移动设备分布式训练快速模型聚合策略，基于移动设备分布式计算的边缘推断的快速数据交换策略，以及基于计算任务卸载的边缘推断的高能效协同传输策略。为了解决所建模的非凸的、组合的、非确定性的问题，提出了使用稀疏与低秩优化方法来重新表述，以便于高效、高性能的灵活算法设计，并得出了一些有意义的结果与结论。

第2章中概括地介绍了稀疏优化与低秩优化的基本概念，并详细介绍了解决非凸的稀疏与低秩优化问题的主流思想与实现方法，为后续章节使用稀疏与低秩优化建模分析移动边缘人工智能中的相关问题，设计相关算法，奠定了基础。

第3章与第4章研究了移动设备分布式训练系统，其优势在于可以利用大量移动设备的分布式资源与数据来建立一个人工智能模型，同时可以避免原始数据的上传从而保护数据隐私。为了解决该系统中全局模型聚合对通信带宽的严苛要求，第3章结合了模型聚合计算的特殊结构与无线信道特有的信号叠加性质，提出了基于空中计算的快速模型聚合方案。空中计算应用于边缘训练的模型聚合中会带来两方面的影响，参与模型聚合的设备增多、模型聚合的误差降低均能提升模型训练的性能。于是，设计目标为最大化选择设备的同时保证聚合误差满足要求。本文提出了一种稀疏与低秩优化方法来解决该问题，稀疏性用于设备选择，而低秩优化用于解决非凸的二次约束。所采用的DC优化方法对于检测设备选择的可行性有着相当优势，从而使得所采用的DC算法能够大大改善现有方法的性能，使得在快速模型聚合的基础上模型训练的效果优于其他方法。

第4章在第3章的基础上，提出使用近年来提出的智能反射面技术来改善无线信号的传播性质，结合空中计算技术，从根本上提升模型聚合的通信效率。智能

反射面赋能的分布式训练系统中的快速模型聚合策略需要联合设计收发机、设备选择以及智能反射面的相移矩阵，这使得约束条件变为了更为复杂的非凸双二次约束。为了解决该问题，在用稀疏性来解决设备选择问题之外，本文采用了交替优化与低秩优化相结合的策略，交替地使用 DC 算法来求解两个稀疏与低秩优化问题。仿真结果也证明了使用智能反射面技术对于降低带宽需求与提升训练性能的巨大潜力。

第5章研究了基于移动设备分布式计算的边缘推断系统，其对于符合分布式计算架构的推断任务可以利用大量移动设备的存储与计算等资源协作实现，完成单个移动设备所无法直接部署并执行的推断任务。对于分布式计算系统来说，数据交换的通信开销是一个主要瓶颈，为了解决该问题，本文提出了一种基于协同传输和干扰对齐的快速数据交换策略，进而将编解码矩阵设计问题建模为一个有仿射约束的秩最小化问题。尽管最近提出的 DC 算法能够达到优异的性能，但是对于该问题每次迭代需要解一个半定规划问题计算复杂度很高，针对这一问题，本文提出了一种新颖的、计算高效的 DC 算法，在大大降低复杂的同时保证了没有明显的性能损失。另外，在均匀数据放置策略下，所提方案的可达自由度几乎不随着移动设备数目的增加而改变，具备可扩展性。

第6章研究了基于计算任务卸载的边缘推断系统，该方案适合于模型巨大、对计算能力要求很高的推断任务，将密集的计算分流给边缘服务器执行，再将结果返回给移动用户。为了提高该方案的通信服务质量，可以通过多个边缘服务器的协作传输实现，但是这需要每个任务被更多边缘服务器执行，需要更高的功耗，本章提出了一种高能效处理与鲁棒性下行传输的方案，最小化计算与传输功耗的同时保证传输对于所得信道状态信息误差的鲁棒性。利用一种基于统计学习的鲁棒优化近似方法，该问题转化为一个有非凸二次约束的组稀疏优化问题。利用低秩条件的 DC 表示能更精确的检测非凸二次约束的可行性，而该问题的结构使得用迭代加权算法能够更好的诱导组稀疏性，算法能够在保证通信质量鲁棒性基础上以更低功耗将推断结果返回给移动用户。

7.2 未来研究方向展望

本文对移动边缘人工智能进行了深入研究，将无线通信系统、信号传播特性与边缘人工智能的训练与推断任务特点相结合，设计了四个使能方案，取得了一

定的成果。然而，在本文的研究基础之上仍有着进一步研究的空间：

1. 本文先后提出的基于空中计算技术与智能反射面技术的快速模型聚合方法，能够有效的提高移动设备分布式模型训练的通信效率。值得注意的是，所提方法是结合了联邦学习模型聚合的计算特点与无线信道的性质而设计的。波束成形的设计依赖于完美信道状态信息的假设，而研究信道状态信息的有限反馈和不确定性对模型聚合的影响是一个有意义的研究方向。此外，模拟的空中计算方案难以直接在当前的数字通信系统中实现，研究空中计算的数字调制方案也是一个有趣的方向。另外，由于本方案是针对每轮通信的数据聚合过程降低带宽要求，对于带宽和延迟要求更为苛刻的边缘训练任务来说，可以从结合训练算法设计的角度出发进一步降低通信需求，如利用二阶信息加快收敛速度从而降低总通信开销。除此之外，对于人工智能训练任务来说，模型聚合的安全问题也十分关键。一个重要的研究方向是设计对恶意攻击具有鲁棒性的聚合方法。

2. 本文针对基于移动设备分布式计算的推断系统提出了一种快速数据交换策略。然而该设计仅考虑了给定文件放置策略下的数据传输问题，而没有对文件放置策略的设计进行研究。一个有意思的研究方向是考虑如何设计最优的文件放置策略以进一步提高通信效率。另外，本文使用自由度作为性能度量，在有限信噪比下的收发机设计问题也值得进一步研究。除此之外，尽管仿真结果显示除了所提方法对于移动用户的增长具备可扩展性，从理论上证明该扩展性仍然有待进一步研究。

3. 本文针对基于计算任务卸载的边缘推断系统设计了冗余计算与协同传输的方案，该方法假设了每个推断任务在多个边缘服务器处单独执行。可以更进一步的考虑将推断任务的计算图进行分割以后，自由地部署在多个边缘服务器，然后同时考虑计算与传输的协作。另外，本文所提方法考虑的是推断结果大小远高于输入数据大小的情况，适合于生成模型等应用，可以进一步研究对于一般情况同时考虑上下行链路的通信开销与能耗。除此之外，由于文献 Dai 和 Yu (2016); Wang 等 (2018a) 中提供的收敛性保证的前提条件不能满足，本文结合迭代加权算法与 DC 算法的收敛性的理论保证尚不明确，依然是一个开放性问题。

移动边缘人工智能是下一代无线通信系统的演进方向，也将成为人工智能必不可少、举足轻重的组成部分。实现移动边缘人工智能是一件系统性的工程，本文提供了边缘训练与推断的几种典型系统与使能方案，未来针对不同的应用

场景挖掘需求，联合设计合理的人工智能算法、通信系统架构与传输方案是本文未来的研究重点。

附录 A 重要术语中英文及缩写对照表

缩写词	英文全称	中文释义
AI	artificial intelligence	人工智能
AirComp	over-the-air computation	空中计算
CNN	convolutional neural network	卷积神经网络
DC	Difference-of-Convex-functions	差分凸函数
DNN	deep neural network	深度神经网络
DoF	degree-of-freedom	自由度
FedAvg	federated averaging	联邦平均算法
IRLS	iterative reweighted least squares	迭代加权最小二乘算法
IRS	intelligent reflecting surface	智能反射面
MIMO	multiple-input-multiple-output	多输入多输出
MMSE	minimum mean-square-error	最小均方差
MSE	mean-square-error	均方差
PSD	positive semidefinite	半正定
QoS	quality-of-service	服务质量
QP	quadratic programming	二次规划
SCA	successive convex approximation	连续凸近似
SDP	Semidefinite Programming	半定规划
SINR	signal-to-interference-plus-noise-ratio	信干噪比
SNR	signal-to-noise-ratio	信噪比
SVD	singular value decomposition	奇异值分解
SVM	support vector machine	支持向量机

附录 B 公式推导与证明

B.1 式 (6.35) 与式 (6.36) 的推导

首先将式 (6.33) 写为

$$\mathbf{h}_k \tau_k = \hat{\mathbf{h}}_k \tau_k + \mathbf{B}_k \tilde{\mathbf{u}}_k, \tilde{\mathbf{u}}_k^H \tilde{\mathbf{u}}_k \leq \tau_k^2, \quad (\text{B.1})$$

其中 $\mathbf{u}_k = \tilde{\mathbf{u}}_k / \tau_k \in \mathbb{C}^L$, $\tau_k > 0$. 令

$$\mathbf{x}_k = \left[\tau_k^H \quad \tilde{\mathbf{u}}_k^H \right]^H \in \mathbb{C}^{NL+1}, \quad (\text{B.2})$$

可以得到

$$\mathbf{h}_k \tau_k = \mathbf{H}_k \mathbf{x}_k. \quad (\text{B.3})$$

然后可知

$$\mathbf{h}_k^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{h}_k - \sigma_k^2 \geq 0 \quad (\text{B.4})$$

$$\Leftrightarrow (\mathbf{h}_k \tau_k)^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{h}_k \tau_k - \sigma_k^2 \tau_k^2 \geq 0 \quad (\text{B.5})$$

$$\Leftrightarrow (\mathbf{H}_k \mathbf{x}_k)^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{H}_k \mathbf{x}_k - \sigma_k^2 \tau_k^2 \geq 0 \quad (\text{B.6})$$

$$\Leftrightarrow \mathbf{x}_k^H \mathbf{P}_k^0 \mathbf{x}_k \geq 0, \quad (\text{B.7})$$

其中 $\mathbf{P}_k^0 \in \mathbb{S}^{NL+1}$ 由下式给出

$$\mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{H}_k - \begin{bmatrix} \sigma_k^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (\text{B.8})$$

同样地, $\tilde{\mathbf{u}}_k^H \tilde{\mathbf{u}}_k \leq \tau_k^2$, 可以重写为

$$\mathbf{x}_k^H \mathbf{P}_k^1 \mathbf{x}_k \geq 0, \quad (\text{B.9})$$

其中 $\mathbf{P}_k^1 \in \mathbb{S}^{NL+1}$ 为

$$\mathbf{P}_k^1 = \begin{bmatrix} 1 \\ & -\mathbf{I}_N \end{bmatrix} \quad (\text{B.10})$$

可以用 S-过程

$$\mathbf{x}_k^H \mathbf{P}_k^1 \mathbf{x}_k \geq 0 \implies \mathbf{x}_k^H \mathbf{P}_k^0 \mathbf{x}_k \geq 0, \quad (\text{B.11})$$

来得到其等价条件，即

$$\mathbf{P}_k^0 \geq \lambda_k \mathbf{P}_k^1, \lambda_k \geq 0. \quad (\text{B.12})$$

如此便得到了联合概率约束 (6.13) 的一个易于处理的重新表述，即

$$\mathbf{H}_k^H \left(\frac{1}{\gamma_k} \mathbf{v}_k \mathbf{v}_k^H - \sum_{l \neq k} \mathbf{v}_l \mathbf{v}_l^H \right) \mathbf{H}_k \succeq \mathbf{Q}_k, \quad (\text{B.13})$$

其中 $\lambda = [\lambda_1, \dots, \lambda_K] = [\lambda_{nk}] \in \mathbb{R}_+^{N \times K}$, \mathbf{Q}_k 由下式给出

$$\mathbf{Q}_k = \begin{bmatrix} \lambda_k + \sigma_k^2 & \\ & -\lambda_k \mathbf{I}_{NL} \end{bmatrix} \in \mathbb{C}^{(NL+1) \times (NL+1)}. \quad (\text{B.14})$$

B.2 DC 算法收敛性证明

第3.4.3.2节、第4.3.3节、第5.4.3节、第6.4.4节中使用的 DC 算法框架的迭代过程可以统一表示为式3.39与式3.40，根据 Fenchel biconjugation 定理可以重写为

$$\mathbf{Y}^{[t]} \in \partial_{\mathbf{X}^{[t]}} h, \mathbf{X}^{[t]} \rangle, \quad (\text{B.15})$$

$$\mathbf{X}^{[t+1]} \in \partial_{\mathbf{Y}^{[t]}} g^*. \quad (\text{B.16})$$

下面证明该算法从任意一个可行的初始点收敛到临界点。

由 $\mathbf{Y}^{[t]} \in \partial_{\mathbf{X}^{[t]}} h$ 可知

$$h(\mathbf{X}^{[t+1]}) \geq h(\mathbf{X}^{[t]}) + \langle \mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle. \quad (\text{B.17})$$

不等式两边都加上 $-g(\mathbf{X}^{[t+1]})$ 可得

$$(g - h)(\mathbf{X}^{[t+1]}) \leq g(\mathbf{X}^{[t+1]}) - \langle \mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle - h(\mathbf{X}^{[t]}). \quad (\text{B.18})$$

类似地，由 $\mathbf{X}^{[t+1]} \in \partial g^*(\mathbf{Y}^{[t]})$ 可得

$$g(\mathbf{X}^{[t]}) \geq g(\mathbf{X}^{[t+1]}) + \langle \mathbf{X}^{[t]} - \mathbf{X}^{[t+1]}, \mathbf{Y}^{[t]} \rangle + \|\mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}\|_F^2. \quad (\text{B.19})$$

不等式两边都减去 $h(\mathbf{X}^{[t]})$ 可得

$$g(\mathbf{X}^{[t+1]}) - \langle \mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle - h(\mathbf{X}^{[t]}) \leq (g - h)(\mathbf{X}^{[t]}) - \|\mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}\|_F^2. \quad (\text{B.20})$$

另外，

$$\mathbf{X}^{[t+1]} \in \partial g^*(\mathbf{Y}^{[t]}) \Leftrightarrow \langle \mathbf{X}^{[t+1]}, \mathbf{Y}^{[t]} \rangle = g(\mathbf{X}^{[t+1]}) + g^*(\mathbf{Y}^{[t]}) \quad (\text{B.21})$$

$$\mathbf{Y}^{[t]} \in \partial h(\mathbf{X}^{[t]}) \Leftrightarrow \langle \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle = h(\mathbf{X}^{[t]}) + h^*(\mathbf{Y}^{[t]}). \quad (\text{B.22})$$

于是可知

$$g(\mathbf{X}^{[t+1]}) - \langle \mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}, \mathbf{Y}^{[t]} \rangle - h(\mathbf{X}^{[t]}) = h^*(\mathbf{Y}^{[t]}) - g^*(\mathbf{Y}^{[t]}). \quad (\text{B.23})$$

根据式 (B.18) 与式 (B.20) 可得

$$\begin{aligned} (g - h)(\mathbf{X}^{[t+1]}) &\leq h^*(\mathbf{Y}^{[t]}) - g^*(\mathbf{Y}^{[t]}) \\ &\leq (g - h)(\mathbf{X}^{[t]}) - \|\mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}\|_F^2. \end{aligned} \quad (\text{B.24})$$

再加上

$$(g - h)(\mathbf{X}) \geq 0, \quad (\text{B.25})$$

可得该算法使得目标函数值收敛，且有

$$\lim_{t \rightarrow \infty} \|\mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}\|_F^2 = 0. \quad (\text{B.26})$$

在每个极限点处均有

$$(g - h)(\mathbf{X}^{[t+1]}) = (g - h)(\mathbf{X}^{[t]}) \quad (\text{B.27})$$

与

$$\|\mathbf{X}^{[t+1]} - \mathbf{X}^{[t]}\|_F^2 = 0 \quad (\text{B.28})$$

成立。因此可知

$$(g - h)(\mathbf{X}^{[t+1]}) = h^*(\mathbf{Y}^{[t]}) - g^*(\mathbf{Y}^{[t]}) = (g - h)(\mathbf{X}^{[t]}). \quad (\text{B.29})$$

从式 (B.22) 可得

$$h(\mathbf{X}^{[t+1]}) + h^*(\mathbf{Y}^{[t]}) = g(\mathbf{X}^{[t+1]}) + g^*(\mathbf{Y}^{[t]}) = \langle \mathbf{X}^{[t+1]}, \mathbf{Y}^{[t]} \rangle, \quad (\text{B.30})$$

即有

$$\mathbf{Y}^{[t]} \in \partial h(\mathbf{X}^{[t+1]}). \quad (\text{B.31})$$

于是得到了 $\mathbf{Y}^{[t]} \in \partial g(\mathbf{X}^{[t+1]}) \cap \partial h(\mathbf{X}^{[t+1]})$ ，这意味着极限点 $\mathbf{X}^{[t+1]}$ 是 $g - h$ 的一个临界点。故给定任意可行初始点，该 DC 算法能够收敛到临界点。

参考文献

- ALISTARH D, GRUBIC D, LI J, et al. QSGD: Communication-efficient SGD via gradient quantization and encoding[C]//Advances in Neural Information Processing Systems (NeurIPS). 2017: 1709-1720.
- Anon. CNN Energy Estimation Website[EB/OL]. <http://energyestimation.mit.edu>.
- Anon. Stroke of Genius: GauGAN Turns Doodles into Stunning, Photorealistic Landscapes[EB/OL]. <https://blogs.nvidia.com/blog/2019/03/18/gaugan-photorealistic-landscapes-nvidia-research/>.
- Anon. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [J/OL]. OJ L 119, 2016:1-88. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:2016:119:FULL&from=EN>.
- BACH F, JENATTON R, MAIRAL J, et al. Optimization with sparsity-inducing penalties[J]. Foundations and Trends® in Machine Learning, 2012, 4(1):1-106.
- BI S, ZHANG R, DING Z, et al. Wireless communications in the era of big data[J]. IEEE Communications Magazine, 2015, 53(10):190-199.
- BLANCHARD P, GUERRAOUI R, STAINER J, et al. Machine learning with adversaries: Byzantine tolerant gradient descent[C]//Advances in Neural Information Processing Systems (NeurIPS). 2017: 119-129.
- BONAWITZ K, EICHNER H, GRIESKAMP W, et al. Towards federated learning at scale: System design[J]. arXiv preprint arXiv:1902.01046, 2019.
- BOUBOULIS P, SLAVAKIS K, THEODORIDIS S. Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients[J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(3):425-438.
- BOYD S, VANDENBERGHE L. Convex optimization[M]. Cambridge University Press, 2004.
- BRESLER G, CARTWRIGHT D, TSE D. Feasibility of interference alignment for the MIMO interference channel[J]. IEEE Transactions on Information Theory, 2014, 60(9):5573-5586.
- BUGHIN J, SEONG J. Assessing the economic impact of artificial intelligence[J]. ITUTrends Issue Paper No. 1, 2018.
- BYRD R H, HANSEN S L, NOCEDAL J, et al. A stochastic quasi-newton method for large-scale optimization[J]. SIAM Journal on Optimization, 2016, 26(2):1008-1031.

- CADAMBE V R, JAFAR S A. Interference alignment and degrees of freedom of the K-user interference channel[J]. IEEE Transactions on Information Theory, 2008, 54(8):3425-3441.
- CANDÈS E J, RECHT B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6):717.
- CANDES E J, WAKIN M B, BOYD S P. Enhancing sparsity by reweighted ℓ_1 minimization[J]. Journal of Fourier Analysis and Applications, 2008, 14(5-6):877-905.
- CHARTRAND R, YIN W. Iteratively reweighted algorithms for compressive sensing[C]//2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2008: 3869-3872.
- CHEN C Y, CHOI J, BRAND D, et al. Adacom: Adaptive residual gradient compression for data-parallel distributed training[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- CHEN E, TAO M. ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems[J]. IEEE Transactions on Communications, 2017, 65(6):2685-2698.
- Chen L, Qin X, Wei G. A uniform-forcing transceiver design for over-the-air function computation [J]. IEEE Wireless Communications Letters, 2018, 7(6):942-945.
- CHEN T, GIANNAKIS G, SUN T, et al. Lag: Lazily aggregated gradient for communication-efficient distributed learning[C]//Advances in Neural Information Processing Systems (NeurIPS). 2018: 5050-5060.
- CHEN W, WILSON J, TYREE S, et al. Compressing neural networks with the hashing trick[C]// International Conference on Machine Learning (ICML). 2015: 2285-2294.
- CHEN Y, SU L, XU J. Distributed statistical machine learning in adversarial settings: Byzantine gradient descent[C]//Proceedings of the ACM on Measurement and Analysis of Computing Systems: volume 1. New York, NY, USA: ACM, 2017: 44:1-44:25.
- CHENG Y, WANG D, ZHOU P, et al. Model compression and acceleration for deep neural networks: The principles, progress, and challenges[J]. IEEE Signal Processing Magazine, 2018, 35(1):126-136.
- CHOI E, SCHUETZ A, STEWART W F, et al. Using recurrent neural network models for early detection of heart failure onset[J]. Journal of the American Medical Informatics Association, 2016, 24(2):361-370.
- CISCO. Cisco global cloud index: Forecast and methodology, 2016-2021 white paper[J/OL]. 2018. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html>.
- CISCO. Cisco annual internet report (2018-2023) white paper[J/OL]. 2020. [https:](https://)

<http://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>.

- COURBARIAUX M, BENGIO Y, DAVID J P. Binaryconnect: Training deep neural networks with binary weights during propagations[C]//Advances in Neural Information Processing Systems (NeurIPS). 2015: 3123-3131.
- COVER T M, THOMAS J A. Elements of information theory[M]. John Wiley & Sons, 2012.
- Dai B, Yu W. Energy efficiency of downlink transmission strategies for cloud radio access networks [J]. IEEE Journal on selected areas in Communications, 2016, 34(4):1037-1050.
- DAUBECHIES I, DEVORE R, FORNASIER M, et al. Iteratively reweighted least squares minimization for sparse recovery[J]. Communications on Pure and Applied Mathematics, 2010, 63 (1):1-38.
- DAVENPORT M A, ROMBERG J. An overview of low-rank matrix recovery from incomplete observations[J]. IEEE Journal of Selected Topics in Signal Processing, 2016, 10(4):608-622.
- DAVID K, BERNDT H. 6G vision and requirements: Is there any need for beyond 5G?[J]. IEEE Veh. Technol. Mag., 2018, 13(3):72-80.
- DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1):107-113.
- DOAN X V, VAVASIS S. Finding the largest low-rank clusters with Ky Fan $2-k$ -norm and ℓ_1 -norm [J]. SIAM Journal on Optimization, 2016, 26(1):274-312.
- DONG J, YANG K, SHI Y. Ranking from crowdsourced pairwise comparisons via smoothed riemannian optimization[J]. ACM Transactions on Knowledge Discovery from Data, 2020, 14(2): 1-26.
- DONG Y, CHENG J, HOSSAIN M, et al. Secure distributed on-device learning networks with byzantine adversaries[J]. IEEE Network, 2019, 33(6):180-187.
- DONOHO D L. Compressed sensing[J]. IEEE Transactions on information theory, 2006, 52(4): 1289-1306.
- DU Y, YANG S, HUANG K. High-dimensional stochastic gradient quantization for communication-efficient edge learning[C]//2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). 2019.
- FAN K. Maximum properties and inequalities for the eigenvalues of completely continuous operators [J]. Proceedings of the National Academy of Sciences of the United States of America, 1951, 37 (11):760-766.
- FANG F, ZHANG H, CHENG J, et al. Joint user scheduling and power allocation optimization for energy efficient NOMA systems with imperfect CSI[J]. IEEE Journal on selected areas in Communications, 2017, 35(12):2874-2885.

- FAZEL M, HINDI H, BOYD S P. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices[C]//Proceedings of the 2003 American Control Conference: volume 3. IEEE, 2003: 2156-2162.
- GESBERT D, HANLY S, HUANG H, et al. Multi-cell MIMO cooperative networks: A new look at interference[J]. IEEE Journal on Selected Areas in Communications, 2010, 28(9):1380-1408.
- GIRIDHAR A, KUMAR P. Toward a theory of in-network computation in wireless sensor networks [J]. IEEE Communications Magazine, 2006, 44(4):98-107.
- GOLDENBAUM M, BOCHE H, STAŃCZAK S. Harnessing interference for analog function computation in wireless sensor networks[J]. IEEE Transactions on Signal Processing, 2013, 61(20): 4893-4906.
- GONG Y, LIU L, YANG M, et al. Compressing deep convolutional networks using vector quantization[J]. arXiv preprint arXiv:1412.6115, 2014.
- GOTOH J Y, TAKEDA A, TONO K. DC formulations and algorithms for sparse optimization problems[J]. Mathematical Programming, 2018, 169(1):141-176.
- GOTTESMAN O, JOHANSSON F, KOMOROWSKI M, et al. Guidelines for reinforcement learning in healthcare[J]. Nat Med, 2019, 25(1):16-18.
- GRANT M, BOYD S. CVX: Matlab software for disciplined convex programming, version 2.1 [EB/OL]. 2014. <http://cvxr.com/cvx>.
- HAN S, HUANG Y, MENG W, et al. Optimal power allocation for SCMA downlink systems based on maximum capacity[J]. IEEE Transactions on Communications, 2019, 67(2):1480-1489.
- HAN S, MAO H, DALLY W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding[C]//2016.
- HANIF M F, TRAN L N, TÖLLI A, et al. Efficient solutions for weighted sum rate maximization in multicellular networks with channel uncertainties[J]. IEEE Transactions on Signal Processing, 2013, 61(22):5659-5674.
- HAUSWALD J, MANVILLE T, ZHENG Q, et al. A hybrid approach to offloading mobile image classification[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2014: 8375-8379.
- HERRMANN F J, HENNENFENT G. Non-parametric seismic data recovery with curvelet frames [J]. Geophysical Journal International, 2008, 173(1):233-248.
- HONG L J, HUANG Z, LAM H. Learning-based robust optimization: Procedures and statistical guarantees[J]. arXiv preprint arXiv:1704.04342, 2017.
- Huang C, Zappone A, Alexandropoulos G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication[J]. IEEE Transactions on Wireless Communications, 2019, 18(8):4157-4170.

- JAIN P, NETRAPALLI P, SANGHAVI S. Low-rank matrix completion using alternating minimization[C]//Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing. 2013: 665-674.
- JIANG T, SHI Y. Over-the-air computation via intelligent reflecting surfaces[C]//2019 IEEE Global Communications Conference (GLOBECOM). 2019.
- KONEčNÝ J, MCMAHAN H B, RAMAGE D. Federated optimization: Distributed optimization beyond the datacenter[C]//NIPS Optimization for Machine Learning Workshop. 2015.
- KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images: 4[R]. University of Toronto, 2009: 7.
- KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[M]//Advances in Neural Information Processing Systems (NeurIPS). 2012: 1097-1105.
- LANGE K. MM optimization algorithms: volume 147[M]. SIAM, 2016.
- LE THI H A, DINH T P. DC programming and DCA: thirty years of developments[J]. Mathematical Programming, 2018:1-64.
- LEE J D, LIN Q, MA T, et al. Distributed stochastic variance reduced gradient methods by sampling extra data with replacement[J]. The Journal of Machine Learning Research, 2017, 18(1):4404-4446.
- Letaief K B, Chen W, Shi Y, et al. The roadmap to 6G: AI empowered wireless networks[J]. IEEE Communications Magazine, 2019, 57(8):84-90.
- LI E, ZENG L, ZHOU Z, et al. Edge AI: On-demand accelerating deep neural network inference via edge computing[J]. IEEE Transactions on Wireless Communications, 2020, 19(1):447-457.
- LI S, MADDAH-ALI M A, AVESTIMEHR A S. Coding for distributed fog computing[J]. IEEE Communications Magazine, 2017, 55(4):34-40.
- LI S, YU Q, MADDAH-ALI M A, et al. A scalable framework for wireless distributed computing [J]. IEEE/ACM Transactions on Networking, 2017, 25(5):2643-2654.
- LI S, MADDAH-ALI M A, YU Q, et al. A fundamental tradeoff between computation and communication in distributed computing[J]. IEEE Transactions on Information Theory, 2018, 64(1): 109-128.
- LI S, KALAN S M M, AVESTIMEHR A S, et al. Near-optimal straggler mitigation for distributed gradient methods[C]//2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). 2018b: 857-866.
- LIN Y, HAN S, MAO H, et al. Deep gradient compression: Reducing the communication bandwidth for distributed training[J]. International Conference on Learning Representations (ICLR), 2018.

- Liu A, Chen X, Yu W, et al. Two-timescale hybrid compression and forward for massive MIMO aided C-RAN[J]. IEEE Transactions on Signal Processing, 2019, 67(9):2484-2498.
- LIU B, ZHOU F, LU G, et al. Energy efficient and robust beamforming for MISO cognitive small cell networks[J]. IEEE Internet of Things Journal, 2018.
- LIU Z, LIU T, WEN W, et al. DeepN-JPEG: A deep neural network favorable JPEG-based image compression framework[C]//Proceedings of the 55th Annual Design Automation Conference. ACM, 2018b: 1-6.
- LOVE D J, HEATH R W, LAU V K, et al. An overview of limited feedback in wireless communication systems[J]. IEEE Journal on selected areas in Communications, 2008, 26(8):1341-1365.
- LU C, LIU Y F. An efficient global algorithm for single-group multicast beamforming[J]. IEEE Transactions on Signal Processing, 2017, 65(14):3761-3774.
- LUO Z Q, SIDIROPOULOS N D, TSENG P, et al. Approximation bounds for quadratic optimization with homogeneous quadratic constraints[J]. SIAM Journal on Optimization, 2007, 18(1):1-28.
- LUSTIG M, DONOHO D, PAULY J M. Sparse MRI: The application of compressed sensing for rapid mr imaging[J]. Magnetic Resonance in Medicine, 2007, 58(6):1182-1195.
- MADDAH-ALI M A, TSE D. Completely stale transmitter channel state information is still very useful[J]. IEEE Transactions on Information Theory, 2012, 58(7):4418-4431.
- MALEKI H, CADAMBE V R, JAFAR S A. Index coding – an interference alignment perspective [J]. IEEE Transactions on Information Theory, 2014, 60(9):5402-5432.
- MAO Y, YOU C, ZHANG J, et al. A survey on mobile edge computing: The communication perspective[J]. IEEE Communications Surveys Tutorials, 2017, 19(4):2322-2358.
- MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics: volume 54. 2017: 1273-1282.
- MO J, HEATH R W. Limited feedback in single and multi-user mimo systems with finite-bit ADCs [J]. IEEE Transactions on Wireless Communications, 2018, 17(5):3284-3297.
- MOHAN K, FAZEL M. Iterative reweighted algorithms for matrix rank minimization[J]. The Journal of Machine Learning Research, 2012, 13:3441-3473.
- MOHANARAJAH G, USENKO V, SINGH M, et al. Cloud-based collaborative 3D mapping in real-time with low-cost robots[J]. IEEE Transactions on Automation Science and Engineering, 2015, 12(2):423-431.
- NAZER B, GASTPAR M. Computation over multiple-access channels[J]. IEEE Transactions on Information Theory, 2007, 53(10):3498-3516.
- NEMIROVSKI A, SHAPIRO A. Convex approximations of chance constrained programs[J]. SIAM Journal on Optimization, 2006, 17(4):969-996.

- O'DONOGHUE B, CHU E, PARIKH N, et al. Conic optimization via operator splitting and homogeneous self-dual embedding[J]. *Journal of Optimization Theory and Applications*, 2016, 169(3):1042-1068.
- PARK J, SAMARAKOON S, BENNIS M, et al. Wireless network intelligence at the edge[J]. *Proceedings of the IEEE*, 2019, 107(11):2204-2239.
- PARK J, SAMARAKOON S, BENNIS M, et al. Wireless network intelligence at the edge[J]. *Proceedings of the IEEE*, 2019, 107(11):2204-2239.
- PARRINELLO E, LAMPIRIS E, ELIA P. Coded distributed computing with node cooperation substantially increases speedup factors[C]//2018 IEEE International Symposium on Information Theory (ISIT). 2018: 1291-1295.
- REDDY S, FOX J, PUROHIT M P. Artificial intelligence-enabled healthcare delivery[J]. *Journal of the Royal Society of Medicine*, 2019, 112(1):22-28.
- ROCKAFELLAR R T. Convex analysis[M]. Princeton university press, 2015.
- SHANNON C E. A mathematical theory of communication[J]. *Bell System Technical Journal*, 1948, 27(3):379-423.
- SHARMA P. Evolution of mobile wireless communication networks-1G to 5G as well as future prospective of next generation communication network[J]. *International Journal of Computer Science and Mobile Computing*, 2013, 2(8):47-53.
- SHI Y, ZHANG J, LETAIEF K B. Group sparse beamforming for green Cloud-RAN[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(5):2809-2823.
- SHI Y, ZHANG J, LETAIEF K B. Robust group sparse beamforming for multicast green Cloud-RAN with imperfect CSI[J]. *IEEE Transactions on Signal Processing*, 2015, 63(17):4647-4659.
- SHI Y, ZHANG J, LETAIEF K B. Low-rank matrix completion for topological interference management by Riemannian pursuit[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(7):4703-4717.
- SHI Y, ZHANG J, LETAIEF K B. Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty[J]. *IEEE Transactions on Signal Processing*, 2015, 63(4):960-973.
- SHI Y, ZHANG J, O'DONOGHUE B, et al. Large-scale convex optimization for dense wireless cooperative networks[J]. *IEEE Transactions on Signal Processing*, 2015, 63(18):4729-4743.
- SHI Y, CHENG J, ZHANG J, et al. Smoothed L_p -minimization for green Cloud-RAN with user admission control[J]. *IEEE Journal on selected areas in Communications*, 2016, 34(4):1022-1036.
- SHI Y, YANG K, JIANG T, et al. Communication-efficient edge AI: Algorithms and systems[J]. arXiv preprint arXiv:2002.09668, 2020.

- SIDIROPOULOS N D, DAVIDSON T N, LUO Z Q. Transmit beamforming for physical-layer multicasting[J]. *IEEE Transactions on Signal Processing*, 2006, 54(6):2239-2251.
- SMITH V, CHIANG C K, SANJABI M, et al. Federated multi-task learning[C]//*Advances in Neural Information Processing Systems (NeurIPS)*. 2017: 4424-4434.
- SUN P, FENG W, HAN R, et al. Optimizing network performance for distributed dnn training on gpu clusters: Imagenet/alexnet training in 1.5 minutes[J]. *arXiv preprint arXiv:1902.06855*, 2019.
- SZE V, CHEN Y H, YANG T J, et al. Efficient processing of deep neural networks: A tutorial and survey[J]. *Proceedings of the IEEE*, 2017, 105(12):2295-2329.
- Tao M, Chen E, Zhou H, et al. Content-centric sparse multicast beamforming for cache-enabled cloud ran[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(9):6118-6131.
- TAO P D, AN L T H. Convex analysis approach to DC programming: Theory, algorithms and applications[J]. *Acta Mathematica Vietnamica*, 1997, 22(1):289-355.
- TEERAPITTAYANON S, MCDANEL B, KUNG H T. Distributed deep neural networks over the cloud, the edge and end devices[C]//2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). 2017: 328-339.
- TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 58(1):267-288.
- Tropp J A, Wright S J. Computational methods for sparse solution of linear inverse problems [J]. *Proceedings of the IEEE*, 2010, 98(6):948-958.
- UDELL M, HORN C, ZADEH R, et al. Generalized low rank models[J]. *Foundations and Trends® in Machine Learning*, 2016, 9(1):1-118.
- VANDEREYCKEN B. Low-rank matrix completion by riemannian optimization[J]. *SIAM Journal on Optimization*, 2013, 23(2):1214-1236.
- WANG H, ZHANG F, WU Q, et al. Nonconvex and nonsmooth sparse optimization via adaptively iterative reweighted methods[J]. *arXiv:1810.10167*, 2018.
- WANG J, JOSHI G. Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms[J]. *arXiv preprint arXiv:1808.07576*, 2018.
- WANG S, TUOR T, SALONIDIS T, et al. When edge meets learning: Adaptive control for resource-constrained distributed machine learning[C]//*IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. 2018b.
- WANG S, ROOSTA-KHORASANI F, XU P, et al. GIANT: globally improved approximate newton method for distributed optimization[C]//*Advances in Neural Information Processing Systems (NeurIPS)*. 2018c: 2332-2342.
- Wang X, Han Y, Wang C, et al. In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning[J]. *IEEE Network*, 2019, 33(5):156-165.

- WATSON G A. Characterization of the subdifferential of some matrix norms[J]. Linear Algebra and its Applications, 1992, 170:33-45.
- WATSON G. On matrix approximation problems with Ky Fan k norms[J]. Numerical Algorithms, 1993, 5(5):263-272.
- WU Q, ZHANG R. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming[J]. IEEE Transactions on Wireless Communications, 2019, 18(11):5394-5409.
- XU X, DING Y, HU S X, et al. Scaling for edge inference of deep neural networks[J]. Nature Electronics, 2018, 1(4):216.
- YANG F, CAI P, QIAN H, et al. Pilot contamination in massive MIMO induced by timing and frequency errors[J]. IEEE Transactions on Wireless Communications, 2018, 17(7):4477-4492.
- YANG K, SHI Y, DING Z. Generalized low-rank optimization for topological cooperation in ultra-dense networks[J]. IEEE Transactions on Wireless Communications, 2019, 18(5):2539-2552.
- YANG Q, LIU Y, CHEN T, et al. Federated machine learning: Concept and applications[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2):12.
- YANG T J, CHEN Y H, SZE V. Designing energy-efficient convolutional neural networks using energy-aware pruning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 5687-5695.
- YUAN K, YING B, LIU J, et al. Variance-reduced stochastic learning by networked agents under random reshuffling[J]. IEEE Transactions on Signal Processing, 2018, 67(2):351-366.
- YUAN X, ZHANG Y J, SHI Y, et al. Reconfigurable-intelligent-surface empowered 6G wireless communications: Challenges and opportunities[J]. arXiv preprint arXiv:2001.00364, 2020.
- ZHANG H, HE W, ZHANG L, et al. Hyperspectral image restoration using low-rank matrix recovery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(8):4729-4743.
- ZHAO Y, LI M, LAI L, et al. Federated learning with non-iid data[J]. arXiv preprint arXiv:1806.00582, 2018.
- ZHOU Z, CHEN X, LI E, et al. Edge intelligence: Paving the last mile of artificial intelligence with edge computing[J]. Proceedings of the IEEE, 2019, 107(8):1738-1762.

作者简历及攻读学位期间发表的学术论文与研究成果

作者简历

杨恺，男，河南省洛阳市人，1993年出生，中国科学院上海微系统与信息技术研究所博士研究生。

通讯地址：上海市浦东新区华夏中路393号，上海科技大学

邮编：201210

电子邮件：yangkai@shanghaitech.edu.cn

教育经历

2011年09月——2015年07月，在大连理工大学信息与通信工程学院电子信息工程专业获得学士学位。

2015年09月——2017年07月，在中国科学院上海微系统与信息技术研究所进行硕士阶段学习。

2017年09月——2020年07月，在中国科学院上海微系统与信息技术研究所攻读博士学位。

已发表(或正式接受)的学术论文：

期刊论文

1. **Yang K**, Shi Y, Zhou Y, et al. Federated machine learning for intelligent IoT via reconfigurable intelligent surface[M]. IEEE Network (2020年04月正式接受)
2. **Yang K**, Shi Y, Yu W, et al. Energy-efficient processing and robust wireless cooperative transmission for edge inference[J]. IEEE Internet of Things Journal, doi:10.1109/JIOT.2020.2979523 (2020年03月正式接受)
3. **Yang K**, Jiang T, Shi Y, et al. Federated learning via over-the-air computation[J]. IEEE Transactions on Wireless Communications, 2020, 19(3):2022-2035.
4. **Yang K**, SHI Y, DING Z. Data shuffling in wireless distributed computing via low-rank optimization[J]. IEEE Transactions on Signal Processing, 2019, 67(12):3087-3099.

5. **Yang K, SHI Y, DING Z.** Generalized low-rank optimization for topological co-operation in ultradense networks[J]. IEEE Transactions on Wireless Communications, 2019, 18(5):2539-2552.
6. DONG J, **Yang K, SHI Y.** Ranking from crowdsourced pairwise comparisons via smoothed riemannian optimization[J]. ACM Transactions on Knowledge Discovery from Data, 2020, 14(2):1-26.
7. Dong J, **Yang K, Shi Y.** Blind demixing for low-latency communication[J]. IEEE Transactions on Wireless Communications, 2019, 18(2):897-911.

会议论文

1. **Yang K, Jiang T, Shi Y, et al.** Federated learning based on over-the-air computation[C]//2019 IEEE International Conference on Communications (ICC). 2019: 1-6.
2. **Yang K, Shi Y, Ding Z.** Low-rank optimization for data shuffling in wireless distributed computing[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018: 6343–6347.
3. **Yang K, Shi Y, Ding Z.** Generalized matrix completion for low complexity transceiver processing in cache-aided fog-ran via the Burer-Monteiro approach[C]//2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP). 2017: 863-867.
4. **Yang K, Shi Y, Zhang J, et al.** A low-rank approach for interference management in dense wireless networks[C]//2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP). 2016: 708-712.
5. **Yang K, Shi Y, Ding Z.** Low-rank matrix completion for mobile edge caching in Fog-RAN via Riemannian optimization[C]//2016 IEEE Global Communications Conference (GLOBECOM). 2016: 1-6.
6. Jiang T, **Yang K, Shi Y.** Pliable data shuffling for on-device distributed learning[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019: 7460-7464.
7. Hua S, **Yang K, Shi Y.** On-device federated learning via second-order optimiza-

- tion with over-the-air computation[C]//2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall). 2019a: 1-5.
8. Dong J, **Yang K**, Shi Y. Blind demixing for low-latency communication[C]//2018 IEEE Wireless Communications and Networking Conference (WCNC). 2018: 1-6.
 9. Dong J, **Yang K**, Shi Y. Ranking from crowdsourced pairwise comparisons via smoothed matrix manifold optimization[C]//2017 IEEE International Conference on Data Mining Workshops (ICDMW). 2017: 949-956.
 10. Hua S, Yang X, **Yang K**, et al. Deep learning tasks processing in fog-ran[C]//2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall). 2019b: 1-5.

参加的研究项目及获奖情况:

研究项目

触感网络中移动边界计算的研究, 国家自然科学基金青年科学基金项目, 编号 61601290, 2017 年 01 月至 2019 年 12 月。

密集移动雾计算接入网络中大规模优化算法, 上海市青年科技英才杨帆计划, 编号 16YF1407700, 2016 年 06 月至 2019 年 05 月。

获奖情况

1. 2017 年中国科学院大学研究生国家奖学金
2. 2017 年中国科学院大学优秀学生
3. 2016 年上海科技大学三好学生

致 谢

五年在上海的硕博求学生涯转瞬即逝，回想以前所遇到的困难艰辛与求知求学的快乐，感谢家人对我求学生涯的支持，感谢老师们对我的不吝赐教，感谢同学朋友师弟师妹对我的帮助与陪伴，感谢学校对我提供的优秀平台，这一切要在 2020 年画上一个暂时的句号了。

本文建立在博士期间的取得的科研成果之上，这要感谢期间耐心、悉心指导我从事科研的上海科技大学的石远明老师和加州大学戴维斯分校的丁峙老师，让我掌握了基本的科研方法与应持的端正科研态度。感谢多伦多大学的郁炜老师，让我在科研合作与学术交流等方面的能力得到了提升。感谢香港理工大学的张军老师，还有上海科技大学的许多老师，都曾给予过我有益的指导。感谢共度五年硕博时光的室友蔡朋浩，还有薛志鹏、杨付乾、张霄宇、蒋涛、董佳琳等其他一众学长同学师弟师妹学弟学妹们，通过共同讨论科研问题给了我许多启发，也在生活上给了我许多帮助。感谢在多伦多大学的陈晞涵、陈致霖、郑曦、沈闡明等人，让我在加拿大度过了一年收获良多的时光。特别感谢杨展鹏师弟，帮助我得到了第 4 章的一部分仿真结果。

此外，感谢我的家人们在我读研期间提供的无私帮助。感谢我的父母及其他亲人们，他们对我生活关怀备至，对我的求学坚定支持，这是我能够完成学业的重要动力。感谢我在上海的弟弟妹妹，在上海与你们经历了一些轻松愉快的闲暇时刻，舒缓了学业与科研的压力。感谢我的妻子与我共同分享快乐、排解忧愁，与我共度了珍贵的五年时光，并将继续携手前行。

