



无线通信资源配置中的优化问题与方法

刘亚锋^{1*}, 武哲宇^{1,2}, 陈伟坤³, 戴彧虹¹

1. 中国科学院数学与系统科学研究院计算数学与科学工程计算研究所, 北京 100190;

2. 中国科学院大学数学科学学院, 北京 100049;

3. 北京理工大学数学与统计学院, 北京 100081

E-mail: yafliu@lsec.cc.ac.cn, wuzy@lsec.cc.ac.cn, chenweikun@bit.edu.cn, dyh@lsec.cc.ac.cn

收稿日期: 2021-10-08; 接受日期: 2023-01-11; 网络出版日期: 2023-04-17; * 通信作者

国家优秀青年科学基金 (批准号: 12022116) 和国家自然科学基金 (批准号: 12021001, 11991021, 12288201 和 12101048) 资助项目

摘要 无线通信系统设计中的许多问题归根结底可建模为优化问题. 一方面, 这些优化问题常常具有高度的非线性性, 一般情形下难于求解; 另一方面, 它们又有自身的特殊结构, 如隐凸性和可分性等. 利用优化的方法结合问题的特殊结构求解和处理无线通信系统设计问题是近年来学术界研究的热点. 本文重点介绍和讨论无线通信系统设计中的最优资源配置 (resource allocation) 问题和相关优化方法. 本文以优化方法为主线, 着重介绍这些优化方法在求解无线通信资源配置优化问题中的应用, 主要包括线性锥规划如何揭示非凸问题中的隐凸性, Lagrange 对偶理论如何揭示相关问题最优解的结构, 稀疏优化和整数规划技巧如何帮助建立相关问题的数学模型, 以及半正定松弛、交替最优化和分式规划技巧如何快速求解相关问题. 最后, 本文展望无线通信系统优化设计中的一些研究方向和关键问题.

关键词 半正定松弛 交替最优化 对偶理论 多用户干扰信道 分式规划 无线通信系统优化设计 稀疏优化 隐凸性 整数规划

MSC (2020) 主题分类 90B18, 90C90, 90C25, 90C26

1 引言与背景

1.1 无线通信的背景

最早的无线通信出现在前工业化时期, 这些系统使用狼烟、火炬、信号弹或旗语, 在视距内传输信息. 19 世纪中叶以后, 随着电报、电话的发明以及电磁波的发现, 人类通信领域产生了根本性变革, 实现了利用金属导线来传递信息甚至通过电磁波来进行无线通信, 使神话中的“顺风耳”“千里眼”变成现实. 特别地, 1895 年, 意大利人马可尼 (Guglielmo Marconi) 在英国怀特岛到 30 千米之外的一条拖

英文引用格式: Liu Y-F, Wu Z, Chen W-K, et al. Optimization problems and methods for resource allocation in wireless communications (in Chinese). Sci Sin Math, 2023, 53: 667–696, doi: 10.1360/SSM-2021-0188

船之间成功进行了无线传输, 现代意义下的无线通信从此诞生. 从这一天开始, 无线通信技术迅速发展, 使人们能够在更远的传输距离上实现更好的通信质量、更低的功耗、更小的体积和更便宜的价格, 使无线通信、无线电视和无线网络等成为现实. 1948 年, “信息论之父” Shannon^[133] 发表了题为 “A mathematical theory of communication” 的论文, 建立了通信的数学理论. Shannon 的理论对无线通信技术的发展一直影响至今. 经过几十年的努力, 人们终于可以实现 Shannon 理论预言的传输信道容量 (channel capacity), 也开始思考后 Shannon 时代的传输容量等问题 (参见文献 [174]).

展望未来, 无线通信在几十年内仍将是通信发展的前沿. 未来无线通信的愿景是, 任何人与物可以在任何时间、任何地点进行通信, 信息基础设施可以为每个家庭、每个企业和每个国家提供数字化的智能服务. 为了使未来无线通信的愿景得以实现, 首先必须要攻克许多技术上的难题, 这些难题贯穿于无线通信系统设计的各个方面.

1.2 无线通信对社会和国家未来发展的重要性

无线通信技术^[38, 56, 151] 主要利用无线电磁波来传输信息. 过去几十年, 无线通信技术经历了 4 代快速发展, 极大地改善了人们的生活方式. 特别地, 第 5 代 (fifth-generation, 5G) 移动通信系统是面向 2020 年之后的新一代移动通信系统. 近年来, 5G 和后 5G 已成为世界范围内学术界和工业界的研究热点, 5G 和后 5G 的研发与设计正在如火如荼地进行 (参见文献 [5, 178]), 我国的无线通信技术也由原来的 “一穷二白” 发展到世界相对领先地位. 值得一提的是, 2016 年由我国华为技术有限公司 (以下简称华为) 主推的极化码 (polar code) 战胜了由美国主推的 LDPC (low density parity check code) 方案和法国主推的 Turbo 2.0 方案, 成为了 5G 超高清视频等大流量移动宽带业务场景下控制信道编码的最终方案. 华为主推的编码方案被采纳意味着高通和爱立信等国外公司对通信技术的垄断时代正式结束, 中国通信技术国际地位正在上升. 为落实 “十四五” 期间国家科技创新安排, 2021 年科学技术部发布国家重点研发计划, 支持 13 个 “数学与应用研究” 重点专项, “通信领域若干关键问题的数学理论和方法” 位列其中.

5G 移动通信技术的发展是当前国家战略发展计划的重要组成部分. 2017 年, 李克强总理在《政府工作报告》中专门提及 5G 移动通信技术对于国家未来发展的重要性, 报告指出 “全面实施战略性新兴产业发展规划, 加快新材料、人工智能、集成电路、生物制药、第五代移动通信等技术研发和转化, 做大做强产业集群”. 这是《政府工作报告》首次提及 5G 移动通信技术. 2018 年, 李克强总理在《政府工作报告》中再一次强调发展 5G 移动通信技术的重要性. 相比于 2017 年, 2018 年《政府工作报告》中 5G 移动通信技术的位置更加靠前. 5G 行业如火如荼的 2019 年, 美国政府对华为实施全面制裁和打压, 意在削弱华为和中国在 5G 等尖端领域的发展, 这也从侧面反映出 5G 是当前我国战略发展计划的重要组成部分.

1.3 最优化在通信系统设计中的重要性

最优化^[8, 119, 185] 讨论决策问题的最佳选择, 构造寻求最佳解的计算方法, 研究这些计算方法的理论性质及实际计算表现. 最优化在国防、经济、金融、工程、交通、管理、数据分析、机器学习和人工智能等许多领域有着广泛的应用. 许多其他科学领域的问题都可以归结为最优化问题, 如控制科学中的轨道设计问题、生命科学中的蛋白质折叠问题、信息科学中的模式识别问题和地球科学中的反演问题等.

无线通信系统设计中的很多问题常常可建模为带有特殊结构的最优化问题. 一方面, 这些优化问题常常具有高度的非线性性, 一般情形下难于求解; 另一方面, 它们又有自身的特殊结构, 如隐凸性、稀疏性和可分性等. 利用优化的技术结合问题的特殊结构求解和处理无线通信系统设计问题是近年来的研究热点问题 (参见文献 [10, 54, 102, 162]), 近 20 年, 几乎每年都有这方面的工作获得 IEEE 信号处理学会或者 IEEE 通信学会最佳论文奖. 近几年, 华为在内部成立了多个运筹和优化中心/实验室 (例如在巴黎的 Mathematical and Algorithmic Sciences Laboratory), 集中力量研究无线通信系统优化设计问题. 世界著名优化专家、中国工程院外籍院士、加拿大皇家科学院院士、美国工业与应用数学学会会士、国际顶级信号处理和无线通信期刊 *IEEE Transactions on Signal Processing* 前任主编罗智泉的主要研究方向是优化算法的设计、分析及其在无线通信中的应用. 他在上述方向发表了多篇有影响力的学术论文, 取得了丰硕的研究成果, 从而也获得了 2004、2009、2011 年 IEEE 信号处理学会最佳论文奖、2011 年欧洲信号处理学会最佳论文奖和 2011 年 IEEE 国际通信大会最佳论文奖. 优化在移动通信系统设计中的重要性由此可见一斑.

1.4 无线通信最优资源配置问题

最优资源配置问题 [9, 10, 54, 55, 63, 88, 180, 183] 是无线通信系统设计中的基本问题之一. 最优的分配功率、传输波形和频谱等资源能够极大地提高整个通信系统的传输性能. 从用户的角度, 多用户之间的相互干扰是制约整个通信系统高速稳健传输的最主要因素, 适当地分配系统资源可以有效地消除多用户之间的相互干扰, 从而提高整个通信系统的传输性能. 从系统运营者的角度, 最优地分配系统资源可以有效地提高功率和频谱等昂贵资源的利用率, 从而降低运营费用. 事实上, 系统运营者需要花费几十亿美元才能获取某段频谱的使用权. 所以, 对于系统运营者而言, 功率和频谱的有效利用可以带来更高的回报和更低的运营花费 (参见文献 [56]). 本文重点介绍和讨论无线通信资源配置中的优化问题与方法.

传统的无线通信最优资源配置旨在提高频谱利用率 (spectral efficiency), 常常涉及以下两类优化问题: 第一类优化问题是满足最小传输速率约束下的传输总功率极小化 (power minimization) 问题 [144, 192], 这一类问题主要是从运营商的角度提出的; 第二类优化问题是满足传输功率约束下的系统效用函数极大化 (system utility maximization) 问题 [83, 141, 148], 这一类问题是从用户的角度提出的. 这两类问题相互关联, 一类问题的求解对另一类问题有启发作用. 最小传输速率约束条件下的系统传输总功率极小化模型的缺点是对应的优化问题有时可能不可行, 而第二类优化模型总是可行的. 第二类问题的另一个优点在于可以通过适当选取系统效用函数以平衡系统的整体性能与用户之间的公平性. 当前比较热门的效用函数 [90] 包括和速率 (速率的和)、速率的几何平均、速率的调和平均、最小速率等. 另外一类研究较多的最优资源配置优化模型是能效 (energy efficiency) 最大化问题 [187, 196]. 能效是一个分式形式的指标, 其中分子是系统总的传输速率, 分母是系统总的发射功率. 能效模型是在一定功率约束的条件下极大化能效, 即极大化单位功率支持的传输速率.

无线通信最优资源配置问题具有鲜明的结构和特点. 用户的信干噪比 (signal-to-interference-plus-noise ratio, SINR) 和传输速率是无线通信系统优化设计中最重要的两个指标, 其中信干噪比是信号的功率与干扰和噪声功率之和的比值, 是分式的形式, 而传输速率是信干噪比的一种对数变换. 充分利用这些特殊结构是高效求解相应的最优资源配置问题的关键 (参见文献 [10, 54, 102, 162]). 另外, 许多最优资源配置问题依赖于信道状态, 目的是根据信道状态的变化自适应地分配系统资源以达到某种指标最优. 信道参数一般为复数, 而最优化变量包括波束成形向量等也都是复数, 参数和变量定义在复数

域是无线通信最优资源配置问题的特点之一. 此外, 信道参数变化很快, 这对资源配置问题求解的实时性有一定的要求, 快速实时求解 (允许一定最优性的损失) 是无线通信最优资源配置问题的另一个特点.

1.5 内容组织和符号约定

本文余下内容组织如下: 无线通信最优资源配置问题都建立在一定的信道模型上, 第 2 节介绍几类常用的无线信道模型. 第 3 节介绍无线通信最优资源配置问题的发展现状, 并详细介绍几类问题中常用的优化方法和案例. 第 4 节给出未来发展方向和展望.

本文符号约定如下: 小写黑斜体字母表示列向量, 大写黑斜体字母表示矩阵. 给定复数 a , $\text{Re}(a)$ 和 $\text{Im}(a)$ 分别表示它的实部和虚部. 给定矩阵 $\mathbf{A}$, $\mathbf{A}^T$ 表示它的转置, $\mathbf{A}^\dagger$ 表示它的共轭转置, $\mathbf{A}^{-1}$ 表示它的逆, A_{ij} 表示它的第 (i, j) 个元素. 相似的符号也适用于向量. 给定向量 $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$, $\|\mathbf{x}\|_p \triangleq (\sum_{k=1}^K |x_k|^p)^{1/p}$ 表示它的 p 范数, 其中 $p \in (0, \infty)$; $\|\mathbf{x}\|$ 表示它的 2 范数; $\|\mathbf{x}\|_0$ 表示向量 $\mathbf{x}$ 中非零元的个数. $\text{Diag}(\mathbf{x})$ 表示由向量 $\mathbf{x}$ 生成的对角矩阵. 给定 (适当维数的) Hermite 矩阵 $\mathbf{A}$ 和 $\mathbf{B}$, $\mathbf{A} \succeq \mathbf{0}$ 表示 $\mathbf{A}$ 是一个半正定矩阵, $\mathbf{A} \succeq \mathbf{B}$ 表示 $\mathbf{A} - \mathbf{B}$ 是一个半正定矩阵, $\text{Trace}(\mathbf{A})$ 表示 $\mathbf{A}$ 的迹, $\text{Rank}(\mathbf{A})$ 表示 $\mathbf{A}$ 的秩. $\mathbf{e}$ 表示适当维数的全 1 向量, $\mathbf{I}$ 表示适当维数的单位矩阵, $\mathbf{0}$ 表示适当维数的全 0 矩阵或向量. $\mathcal{CN}(\mathbf{0}, \Sigma)$ 表示均值为 $\mathbf{0}$ 、方差为 Σ 的复 Gauss 分布.

2 无线信道模型

2.1 多用户干扰信道模型

多用户干扰信道 (multi-user interference channel) 如图 1 所示. 在多用户干扰信道中, 有多个传输端和多个接收端. 每个传输端和接收端装有不同数量的天线. 每个传输端希望给对应的接收端发送信号. 特别地, 当传输端 1 (TX_1) 给接收端 1 (RX_1) 发送信号的同时, TX_1 也给其他的接收端带来了干扰; 反过来, 当 RX_1 接收 TX_1 传输信号的同时, 也接收到其他传输端传输的信号, 这对于 RX_1 来说也是干扰. 这即是多用户干扰信道. 本文主要讨论无线通信中的最优资源分配问题, 这些问题大都基于多用户干扰信道或者它的变种.

在本文中, “用户” 表示系统中的一个传输接收对, 也称为一个 (直接) 链接. 下面给出多用户干扰信道的系统模型. 假设系统中有 K 个用户, 用 $\mathcal{K} = \{1, 2, \dots, K\}$ 表示系统中所有用户的集合. 考虑单载波 (single carrier) 通信系统, 即第 k 个传输端在一个时刻只发送一个信号 $s_k \in \mathbb{C}$ 给对应的接收端 k . 用 $\mathbf{H}_{kj} \in \mathbb{C}^{M_k \times N_j}$ 表示第 j 个传输端到第 k 个接收端的信道矩阵, 其中 M_k 和 N_j 分别表示第 k 个接收端和第 j 个传输端装有的天线个数, 那么接收端 k 接收的信号为

$$\mathbf{y}_k = \mathbf{H}_{kk} \mathbf{v}_k s_k + \sum_{j \neq k} \mathbf{H}_{kj} \mathbf{v}_j s_j + \mathbf{z}_k,$$

其中, $\mathbf{v}_k \in \mathbb{C}^{N_k \times 1}$ 表示传输端 k 使用的传输波束成形向量, $\mathbf{z}_k \in \mathbb{C}^{M_k \times 1}$ 是加性 Gauss 白噪声 (additive white Gaussian noise, AWGN), 服从复 Gauss 分布 $\mathcal{CN}(\mathbf{0}, \sigma_k^2 \mathbf{I})$. 令 $\mathbf{u}_k \in \mathbb{C}^{M_k \times 1}$ 表示接收端 k 使用的接收波束成形向量, 则接收端 k 经过线性处理以后得到的信号为

$$\hat{s}_k = \mathbf{u}_k^\dagger \mathbf{y}_k = \mathbf{u}_k^\dagger \mathbf{H}_{kk} \mathbf{v}_k s_k + \sum_{j \neq k} \mathbf{u}_k^\dagger \mathbf{H}_{kj} \mathbf{v}_j s_j + \mathbf{u}_k^\dagger \mathbf{z}_k.$$

将干扰看成噪声, 那么第 k 个用户的信干噪比可表示为

$$\text{SINR}_k = \frac{|\mathbf{u}_k^\dagger \mathbf{H}_{kk} \mathbf{v}_k|^2}{\sigma_k^2 \|\mathbf{u}_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^\dagger \mathbf{H}_{kj} \mathbf{v}_j|^2}, \quad k \in \mathcal{K}. \quad (2.1)$$

根据 Shannon 信息论的结果 (参见文献 [38]), 第 k 个用户的传输速率可表示为

$$r_k = \log_2(1 + \text{SINR}_k), \quad k \in \mathcal{K}. \quad (2.2)$$

严格地讲, 传输速率的表达式应为 $r_k = \frac{1}{2} \log_2(1 + \text{SINR}_k)$. 但常数 $\frac{1}{2}$ 对后面所考虑的优化问题没有本质影响, 将其简记如 (2.2) 所示.

根据系统中传输端和接收端装有的天线个数, 可将多用户干扰信道做如下分类.

多输入多输出 (multiple-input multiple-output, MIMO) 干扰信道: 当系统中每个传输端和接收端都装有多根天线时, 即当 $M_k \geq 2$ 且 $N_k \geq 2$ ($k \in \mathcal{K}$) 时, 此信道称为 MIMO 干扰信道, 其 SINR 表达式如 (2.1) 所示.

多输入单输出 (multiple-input single-output, MISO) 干扰信道: 当系统中每个传输端装有多根天线且每个接收端装有 1 根天线时, 即当 $M_k = 1$ 且 $N_k \geq 2$ ($k \in \mathcal{K}$) 时, 此信道称为 MISO 干扰信道. 在 MISO 干扰信道中, 信道矩阵 $\mathbf{H}_{kj}$ 退化为行向量 $\mathbf{h}_{kj}^\dagger$ (本文始终用小写黑体表示列向量), 并且不需要接收波束成形向量. 此时, MISO 干扰信道中第 k 个接收端的 SINR 可表示为

$$\text{SINR}_k = \frac{|\mathbf{h}_{kk}^\dagger \mathbf{v}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |\mathbf{h}_{kj}^\dagger \mathbf{v}_j|^2}, \quad k \in \mathcal{K}. \quad (2.3)$$

在 MIMO 和 MISO 干扰信道中, 每个传输端的功率约束可表示为

$$\|\mathbf{v}_k\|^2 \leq \bar{p}_k, \quad k \in \mathcal{K},$$

其中 $\bar{p}_k$ 表示第 k 个传输端的功率上限.

单输入多输出 (single-input multiple-output, SIMO) 干扰信道: 当系统中每个接收端装有多根天线且每个传输端装有 1 根天线时, 即当 $M_k \geq 2$ 且 $N_k = 1$ ($k \in \mathcal{K}$) 时, 此信道称为 SIMO 干扰信道. 在 SIMO 干扰信道中, 信道矩阵 $\mathbf{H}_{kj}$ 退化为列向量 $\mathbf{h}_{kj}$, 传输端 k 所使用的波束成形向量 $\mathbf{v}_k$ 退化为数 v_k . 此时, 第 k 个接收端的 SINR 可表示为

$$\text{SINR}_k = \frac{|\mathbf{u}_k^\dagger \mathbf{h}_{kk}|^2 p_k}{\|\mathbf{u}_k\|^2 \sigma_k^2 + \sum_{j \neq k} |\mathbf{u}_k^\dagger \mathbf{h}_{kj}|^2 p_j}, \quad k \in \mathcal{K},$$

其中 $p_k = |v_k|^2$.

单输入单输出 (single-input single-output, SISO) 干扰信道: 当系统中每个传输端和接收端都装有 1 根天线时, 即当 $M_k = N_k = 1$ ($k \in \mathcal{K}$) 时, 此信道称为 SISO 干扰信道. 在 SISO 干扰信道中, 信道矩阵 $\mathbf{H}_{kj}$ 退化为数 h_{kj} . 此时, 第 k 个接收端的 SINR 可表示为

$$\text{SINR}_k = \frac{g_{kk} p_k}{\eta_k + \sum_{j \neq k} g_{kj} p_j}, \quad k \in \mathcal{K},$$

其中, $g_{kj} = |h_{kj}|^2$ 表示第 j 个传输端到第 k 个接收端的信道增益 (channel gain), $\eta_k = \sigma_k^2$ 表示第 k 个接收端的噪声功率. 在 SIMO 和 SISO 干扰信道中, 每个传输端处的功率约束可表示为

$$0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}.$$

2.2 其他信道模型

下行单波 (unicast) MISO 信道: 下行单波 MISO 信道中有一个装有多根天线的传输端和 K 个装有 1 根天线的接收端. 此信道可以看成是图 1 中的通信系统在传输端允许完全合作, 即 K 个传输端共享所有用户的数据, 所以它们可以看成是一个虚拟 (virtual) 的大传输端. 令 $\mathbf{h}_k = [\mathbf{h}_{k1}^T, \mathbf{h}_{k2}^T, \dots, \mathbf{h}_{kK}^T]^T \in \mathbb{C}^{N \times 1}$ 表示大传输端到第 k 个用户的信道系数, 其中, $\mathbf{h}_{kj} \in \mathbb{C}^{N_j \times 1}$ 表示第 j 个传输端到第 k 个用户的信道系数, $N = \sum_{k \in \mathcal{K}} N_k$. 假设 $\mathbf{w}_k = [\mathbf{w}_{k1}^T, \mathbf{w}_{k2}^T, \dots, \mathbf{w}_{kK}^T]^T \in \mathbb{C}^{N \times 1}$ 为传输端到第 k 个用户的波束成形向量, 其中 $\mathbf{w}_{kj} \in \mathbb{C}^{N_j \times 1}$ 表示第 j 个传输端到第 k 个用户的波束成形向量, 则第 k 个用户接收到的信号为

$$y_k = \mathbf{h}_k^\dagger \mathbf{w}_k s_k + \sum_{j \neq k} \mathbf{h}_k^\dagger \mathbf{w}_j s_j + z_k;$$

第 k 个用户的 SINR 可表示为

$$\text{SINR}_k = \frac{|\mathbf{h}_k^\dagger \mathbf{w}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |\mathbf{h}_k^\dagger \mathbf{w}_j|^2}, \quad k \in \mathcal{K}. \quad (2.4)$$

如果 K 个传输端各自独立传输, 这等价于迫使波束成形向量 $\mathbf{w}_k$ 中除第 k 块外全部置零, 再令 $\mathbf{w}_{kk} = \mathbf{v}_k$, 则 (2.4) 退化为 (2.3).

下行多波 (multicast) MISO 信道: 考虑 MISO 下行信道, 其中一个装有多根天线的基站同时给多个单天线用户发送相同的信号. 此系统可以看成是图 1 中的通信系统在传输端允许完全合作且发送的信息 $s_1 = s_2 = \dots = s_K$, 即一个虚拟的大传输端 (基站) 同时给一组用户传输相同的信号. 令 $\mathbf{h}_k$ 表示传输端到第 k 个用户的信道系数, $\mathbf{w} \in \mathbb{C}^{N \times 1}$ 为传输端使用的波束成形向量, 则此系统的 SINR 可表示为

$$\text{SINR} = \min_{k \in \mathcal{K}} \left\{ \frac{|\mathbf{h}_k^\dagger \mathbf{w}|^2}{\sigma_k^2} \right\}.$$

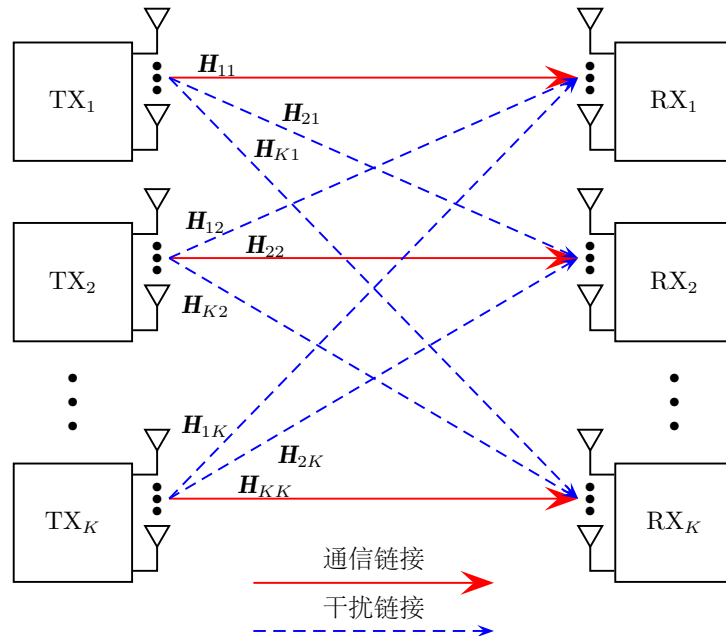


图 1 (网络版彩图) 多用户干扰信道示意图

由此可见, 多波信道中的 SINR (以及传输速率) 由信道最差的用户决定.

多载波 (multicarrier) SISO 干扰信道: 考虑多用户多载波通信系统, 其中 K 个用户共用 M 个载波. 用 g_{kj}^m 表示第 j 个传输端在第 m 个载波上对第 k 个接收端的信道增益, p_j^m 表示第 j 个传输端在第 m 个载波上的发射功率, η_k^m 表示第 k 个接收端在第 m 个载波上的噪声功率, 则第 k 个用户的功率约束和传输速率可分别表示为

$$\begin{aligned} \sum_{m \in \mathcal{M}} p_k^m &\leq \bar{p}_k, \quad k \in \mathcal{K}, \\ R_k &= \sum_{m \in \mathcal{M}} \log_2 \left(1 + \frac{g_{kk}^m p_k^m}{\sum_{j \neq k} g_{kj}^m p_j^m + \eta_k^m} \right), \quad k \in \mathcal{K}, \end{aligned} \quad (2.5)$$

其中 $\mathcal{M} = \{1, 2, \dots, M\}$. 特别地, 当系统中只有一个用户时, 用 α^m 和 η^m 分别表示载波 $m \in \mathcal{M}$ 上的信道增益和噪声功率. 此时用户的传输速率为

$$R = \sum_{m \in \mathcal{M}} \log_2 \left(1 + \frac{\alpha^m p^m}{\eta^m} \right).$$

3 发展现状

在多用户干扰信道或者它的变种中, 用户之间的传输会给其他的用户带来干扰. 目前多用户之间的相互干扰是阻碍无线通信系统高速稳健传输的最主要因素. 最优资源配置通过设计传输波束成形向量 $\{\mathbf{v}_k\}/\{\mathbf{w}_k\}$ 、接收波束成形向量 $\{\mathbf{u}_k\}$ 以及传输功率 $\{p_k\}$ 来减少和消除系统中的干扰, 使得整个系统在最佳状态下进行通信. 本节主要介绍现代优化技术结合问题的特殊结构在无线干扰信道最优资源配置中的几个典型应用案例. 特别地, 主要介绍线性锥规划如何揭示非凸问题中的隐凸性, Lagrange 对偶理论如何揭示相关问题最优解的结构, 稀疏优化和整数规划技巧如何帮助建立相关问题的数学模型, 以及半正定松弛、交替最优化和分式规划技巧如何快速求解相关问题.

3.1 隐凸性和凸等价变形

本小节主要阐述现代优化技术特别是线性锥规划如何揭示非凸资源配置问题中的隐凸性以及将看似非凸的问题转化为凸优化问题 (参见文献 [13]).

SISO 干扰信道功率控制: 考虑多用户 SISO 干扰信道中的功率控制问题, 目标是极小化整个系统的传输功率使得每个用户的 SINR 都大于或者等于预先给定的目标值. 上述功率控制问题可以建模为

$$\begin{aligned} \min_{\{p_k\}} \quad & \sum_{k \in \mathcal{K}} p_k \\ \text{s.t.} \quad & \frac{g_{kk} p_k}{\sum_{j \neq k} g_{kj} p_j + \eta_k} \geq \gamma_k, \quad k \in \mathcal{K}, \\ & 0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}, \end{aligned} \quad (3.1)$$

其中 $\gamma_k > 0$ 为第 k 个用户的 SINR 目标值. 由于上述 SINR 约束等价于

$$g_{kk} p_k \geq \gamma_k \left(\sum_{j \neq k} g_{kj} p_j + \eta_k \right), \quad k \in \mathcal{K},$$

所以问题 (3.1) 实际上是一个线性规划. 不难证明, 问题 (3.1) 的最优解一定在所有 SINR 约束取等号时成立 (参见文献 [1, 186]). 求解功率控制问题 (3.1) 的 Foschini-Miljanic 算法 [50] 正是巧妙地利用了这个特殊结构, 从而将问题 (3.1) 转化为一个线性方程组. Foschini-Miljanic 算法的迭代格式为

$$p_k^{t+1} = \max \left\{ \min \left\{ \frac{\gamma_k}{\text{SINR}_k^t} p_k^t, \bar{p}_k \right\}, 0 \right\}, \quad k \in \mathcal{K}, \quad (3.2)$$

其中, t 代表迭代次数, SINR_k^t 为第 k 个用户在第 t 次迭代的 SINR 值. 上述迭代格式 (3.2) 非常直观, 当第 k 个用户的 SINR 值大于其目标值时, 算法降低第 k 个用户的发射功率; 反之则增加第 k 个用户的发射功率. Foschini-Miljanic 算法本质上即是求解线性方程组的 Jacobi 迭代算法 [57, 126]; 由于原问题有 (功率) 约束, 算法进一步将 Jacobi 迭代得到的结果再进行投影.

Foschini-Miljanic 算法简洁易于实现, 且有如下几个优点. 第一, 只要原问题 (3.1) 可行, 算法总是以线性速度全局收敛 [1, 33, 110]. 第二, Foschini-Miljanic 算法是分布式算法: 每个传输端可以同时更新传输功率而不需要等待其他传输端. 第三, Foschini-Miljanic 算法具有工程实用性. 具体地, 在 Foschini-Miljanic 算法中, 传输端每次更新传输功率时需要相应的接收端反馈当前的 SINR 值, 而这个值在实际通信中容易测量.

下行单波信道波束成形设计: 下行单波信道联合功率和波束成形设计问题可建模为

$$\begin{aligned} \min_{\{\mathbf{w}_k\}} \quad & \sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2 \\ \text{s.t.} \quad & \frac{|\mathbf{h}_k^\dagger \mathbf{w}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |\mathbf{h}_k^\dagger \mathbf{w}_j|^2} \geq \gamma_k, \quad k \in \mathcal{K}. \end{aligned} \quad (3.3)$$

在问题 (3.3) 中, 目标函数 $\sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2$ 表示传输端发射的总功率, $\gamma_k > 0$ 为第 k 个用户的 SINR 目标值. 由于 SINR 约束是非凸的, 问题 (3.3) 看起来是非凸的, 但实际上是一个凸问题. 接下来利用问题的特殊结构, 将其等价地转化为凸优化问题.

半正定松弛 [152] 是求解问题 (3.3) 的一类非常有效的方法. 引入辅助变量

$$\mathbf{H}_k = \mathbf{h}_k \mathbf{h}_k^\dagger \in \mathbb{C}^{N \times N} \quad \text{和} \quad \mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^\dagger \in \mathbb{C}^{N \times N}, \quad k \in \mathcal{K}.$$

问题 (3.3) 关于变量 $\{\mathbf{W}_k\}$ 的半正定松弛为

$$\begin{aligned} \min_{\{\mathbf{W}_k\}} \quad & \sum_{k \in \mathcal{K}} \text{Trace}(\mathbf{W}_k) \\ \text{s.t.} \quad & \text{Trace}(\mathbf{H}_k \mathbf{W}_k) - \gamma_k \sum_{j \neq k} \text{Trace}(\mathbf{H}_k \mathbf{W}_j) \geq \gamma_k \sigma_k^2, \quad k \in \mathcal{K}, \\ & \mathbf{W}_k \succeq \mathbf{0}, \quad k \in \mathcal{K}. \end{aligned} \quad (3.4)$$

相比于原问题 (3.3), 问题 (3.4) 将非凸的秩一约束 $\text{Rank}(\mathbf{W}_k) = 1$ ($k \in \mathcal{K}$) 进行了松弛, 使其变成了一个凸问题. 这也是问题 (3.4) 被称为问题 (3.3) 的半正定松弛的原因.

文献 [7] 证明了上述半正定松弛 (3.4) 与原问题 (3.3) 等价, 即问题 (3.4) 总是存在秩为一的最优解. 文献 [164] 进一步证明了问题 (3.3) 可以等价地转化为一个二阶锥规划问题, 从而揭示了半正定松弛与原问题等价的根本原因. 事实上, 对于 (3.3), 总是可以引入一个角度变化使得所有的 $\mathbf{h}_k^\dagger \mathbf{w}_k$ 都是正的实数, 但不影响 SINR 的值. 因此, 不失一般性, 总是可以假设所有的 $\mathbf{h}_k^\dagger \mathbf{w}_k > 0$. 进而, 问题 (3.3)

等价于如下凸二阶锥规划:

$$\begin{aligned} \min_{\{\mathbf{w}_k, \tau\}} \quad & \tau \\ \text{s.t.} \quad & \sqrt{1 + \frac{1}{\gamma_k} \mathbf{h}_k^\dagger \mathbf{w}_k} \geq \|[\mathbf{h}_k^\dagger \mathbf{W} \sigma_k]\|, \quad k \in \mathcal{K}, \\ & \sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|^2 \leq \tau, \end{aligned} \quad (3.5)$$

其中 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$. 一般而言, 用 CVX 等凸优化软件包求解如问题 (3.5) 所示的二阶锥规划要比求解如问题 (3.4) 所示的半正定规划更高效. 这是因为问题 (3.4) 将原始问题进行了升维, 其变量 $\{\mathbf{W}_k\}$ 的维数为 N^2 ; 而问题 (3.4) 中变量为 $\{\mathbf{w}_k\}$, 与原始问题相同, 维数为 N . 文献 [164] 进一步通过刻画问题 (3.5) 的 KKT (Karush-Kuhn-Tucker) 条件, 提出了一种低复杂度的不动点迭代算法, 可以得到该问题的最优解.

更多的关于无线通信最优资源配置中隐凸性挖掘和凸等价变形的例子, 可参见文献 [90, 95, 108, 135, 146, 181] 等. 文献 [181] 将文献 [7, 164] 中的结果推广到每根天线功率约束的情形下. 文献 [90, 95] 将文献 [7, 164] 中的结果推广到下行 MISO 干扰信道和上行 SIMO 干扰信道中的联合功率和波束成形设计问题. 文献 [108, 135, 146] 将文献 [7, 164] 中的结果推广到信道系数带有误差的情形下.

3.2 Lagrange 对偶理论

本小节主要阐述如何利用 Lagrange 对偶理论 [8] 揭示最优资源分配问题最优解的结构以及设计相应算法.

多载波 SISO 信道功率控制和注水 (waterfilling) 算法 [38]: 单用户多载波 SISO 信道中的功率控制问题考虑如何在多个载波上分配功率以极大化系统的总传输速率. 在数学上, 此问题可建模为

$$\begin{aligned} \max_{\{p^m\}} \quad & \sum_{m \in \mathcal{M}} \log_2 \left(1 + \frac{\alpha^m p^m}{\eta^m} \right) \\ \text{s.t.} \quad & \sum_{m \in \mathcal{M}} p^m \leq P, \quad p^m \geq 0, \quad m \in \mathcal{M}, \end{aligned} \quad (3.6)$$

其中上标 m 表示第 m 个载波. 问题 (3.6) 是一个凸优化问题. 假设 λ 为约束 $\sum_{m \in \mathcal{M}} p^m \leq P$ 对应的 Lagrange 乘子, 则问题 (3.6) 的 (部分) Lagrange 函数为

$$\mathcal{L}(\{p^m\}; \lambda) = - \sum_{m \in \mathcal{M}} \log_2 \left(1 + \frac{\alpha^m p^m}{\eta^m} \right) + \lambda \left(\sum_{m \in \mathcal{M}} p^m - P \right).$$

对 $\mathcal{L}(\{p^m\}; \lambda)$ 关于 $p^m \geq 0$ 求极小可得到问题 (3.6) 的解为

$$p^m(\lambda) = \max \left\{ \frac{1}{\lambda} - \frac{\eta^m}{\alpha^m}, 0 \right\}, \quad m \in \mathcal{M}, \quad (3.7)$$

其中 λ 应满足 $\sum_{m \in \mathcal{M}} p^m(\lambda) = P$. 借助于 Lagrange 乘子, (3.7) 清晰地刻画了问题 (3.6) 解的结构, 即功率分配应当像注水 (参见文献 [38]) 一样, 最优的功率分配应当使得所有分到功率的用户的功率 p^m 与其噪声信道增益比 η^m/α^m 之和在同一水平 $1/\lambda$ 上. 图 2 给出了解 (3.7) 的直观含义, 其中红色表示分配在不同载波上的功率, “水平” 为 1. 上述问题 (3.6) 在每个载波功率约束情形下的闭式解, 可参见文献 [180]. C-RAN (cloud radio access network) 框架下最优无线回程 (backhaul) 分配问题解的结构刻画和 “非线性” 注水算法, 可参见文献 [97].

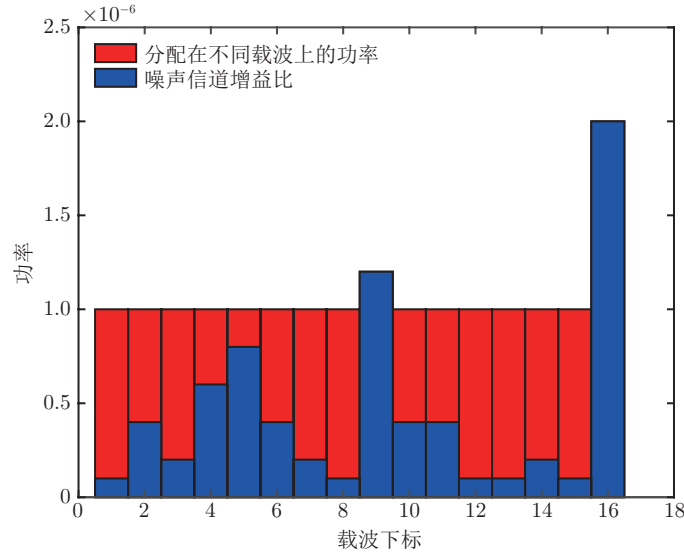


图 2 注水算法示意图

下行单波信道波束成形设计和上下行对偶理论: 考虑单波信道波束成形设计问题 (3.3). 不失一般性, 假设 $\sigma_k = \sigma$, 即每个用户的噪声功率相等. 容易验证, 它的 Lagrange 对偶问题为 (参见文献 [181])

$$\begin{aligned} \max_{\{\lambda_k\}} \quad & \sum_{k \in \mathcal{K}} \lambda_k \sigma^2 \\ \text{s.t.} \quad & I + \sum_{j \in \mathcal{K}} \lambda_j \mathbf{h}_j \mathbf{h}_j^\dagger \succeq \left(1 + \frac{1}{\gamma_k}\right) \lambda_k \mathbf{h}_k \mathbf{h}_k^\dagger, \quad k \in \mathcal{K}, \end{aligned} \quad (3.8)$$

其中 λ_k 是对应于第 k 个 SINR 约束的对偶变量. 考虑如下上行联合功率和波束成形设计问题:

$$\begin{aligned} \min_{\{\hat{\mathbf{w}}_k, \rho_k\}} \quad & \sum_{k \in \mathcal{K}} \rho_k \\ \text{s.t.} \quad & \frac{\rho_k |\mathbf{h}_k^\dagger \hat{\mathbf{w}}_k|^2}{\sigma^2 \|\hat{\mathbf{w}}_k\|^2 + \sum_{j \neq k} \rho_j |\mathbf{h}_j^\dagger \hat{\mathbf{w}}_k|^2} \geq \gamma_k, \quad k \in \mathcal{K}, \end{aligned} \quad (3.9)$$

其中, ρ_k 为上行用户 k 的发射功率, $\hat{\mathbf{w}}_k$ 为接收端为了接收上行用户 k 的发射信号所使用的接收波束成形向量. 容易验证上述问题 (3.9) 关于 $\hat{\mathbf{w}}_k$ 是可分的, 且最优解为线性均方差极小化 (linear minimum mean squared error, LMMSE) 接收向量:

$$\hat{\mathbf{w}}_k = \left(\sigma^2 \mathbf{I} + \sum_{j \in \mathcal{K}} \rho_j \mathbf{h}_j \mathbf{h}_j^\dagger \right)^{-1} \mathbf{h}_k, \quad k \in \mathcal{K}. \quad (3.10)$$

将上述最优解代入问题 (3.9) 可进一步证明问题 (3.9) 等价于

$$\begin{aligned} \min_{\{\rho_k\}} \quad & \sum_{k \in \mathcal{K}} \rho_k \\ \text{s.t.} \quad & \left(1 + \frac{1}{\gamma_k}\right) \rho_k \mathbf{h}_k^\dagger \left(\sigma^2 \mathbf{I} + \sum_{j \in \mathcal{K}} \rho_j \mathbf{h}_j \mathbf{h}_j^\dagger \right)^{-1} \mathbf{h}_k \geq 1, \quad k \in \mathcal{K}. \end{aligned} \quad (3.11)$$

注意到下行问题的最优解总是使得所有的 SINR 取等式. 利用这个事实可以证明问题 (3.8) 与 (3.11) 等价. 所以, 下行问题 (3.3) 与上行问题 (3.9) 等价.

从理论上讲, 非常有趣的是, 上行问题 (3.9) 的最优功率 ρ_k^* 等于下行问题 (3.3) 第 k 个 SINR 约束对应的最优对偶变量 λ_k^* 乘以第 k 个用户的噪声功率 σ^2 , 即 $\rho_k^* = \lambda_k^* \sigma^2$, $k \in \mathcal{K}$. 从算法的角度来讲, 与下行问题 (3.3) 相比, 上行问题 (3.9) 更加容易求解, 所以可以利用上下行对偶理论设计算法求解下行问题. 特别地, 在上行问题 (3.9) 中, 变量 $\hat{w}_k$ 关于 SINR 约束是可分的, 并且有显式解 (3.10); 但下行问题 (3.3) 中变量 w_k 关于所有 SINR 约束耦合在一起, 不易求解. 为方便叙述, 上面假设了每个用户的噪声功率相等. 事实上, 上述结论可推广到噪声功率不等的情形下 (参见文献 [7]).

文献 [36, 125, 131, 155] 利用上下行对偶理论设计有效算法求解下行问题; 文献 [181] 从凸优化的角度将上下行对偶理论统一到 Lagrange 对偶框架下, 并进一步将上述结果推广到每根天线都有功率约束的情形; 文献 [19] 将上述上下行对偶理论推广到带有自干扰的全双工 (full duplex) 通信情形; 文献 [47, 86] 进一步将上下行对偶结果推广到带有 (压缩) 中继 (relay) 的框架下, 并设计了有效的算法能够求得问题的全局最优解. 注意到在带有中继的框架下, 系统有两跳, 第一跳是从基站到中继, 第二跳是从中继到用户; 但是现有的上下行对偶结果都只考虑一跳, 即从基站直接到用户这一跳.

动态谱管理 (dynamic spectrum management) 问题的零间隙对偶理论: 考虑如下多载波 SISO 干扰信道中的动态谱管理问题:

$$\begin{aligned} \max_{\{p_k^m\}} \quad & U(R_1, R_2, \dots, R_K) \\ \text{s.t.} \quad & \sum_{m \in \mathcal{M}} p_k^m \leq \bar{p}_k, \quad k \in \mathcal{K}, \\ & p_k^m \geq 0, \quad k \in \mathcal{K}, \quad m \in \mathcal{M}, \end{aligned} \quad (3.12)$$

其中, $U(\cdot)$ 为适当选择的效用函数, R_k 的表达式见 (2.5). 问题 (3.12) 的对偶问题为

$$\begin{aligned} \min_{\{\lambda_k\}} \quad & d(\{\lambda_k\}) \\ \text{s.t.} \quad & \lambda_k \geq 0, \quad k \in \mathcal{K}, \end{aligned} \quad (3.13)$$

其中

$$d(\{\lambda_k\}) = \max_{\{p_k^m \geq 0\}} \left\{ U(R_1, R_2, \dots, R_K) - \sum_{k \in \mathcal{K}} \lambda_k \left(\sum_{m \in \mathcal{M}} p_k^m - \bar{p}_k \right) \right\}.$$

用 p_M^* 和 d_M^* 分别表示原始问题 (3.12) 和对偶问题 (3.13) 的最优值. 从弱对偶定理可知 $d_M^* - p_M^* \geq 0$. 一般情形下, 当原问题非凸时, 存在严格对偶间隙, 即 $d_M^* - p_M^* > 0$. 事实上, 问题 (3.12) 是 NP (non-deterministic polynomial time)-困难的 (参见文献 [107]).

当载波数 M 趋于无穷大时, 文献 [182] 观察到一个有趣的现象: 当目标函数取为和速率时, 上述问题的对偶间隙趋于 0. 文献 [182] 进一步给出了这个现象的工程解释: 当载波数目趋于无穷时, 动态谱管理问题 (3.12) 满足所谓的时分 (time-sharing) 性质; 此时多个用户可以时分复用 (time-division multiplexing) “几乎相同” 的载波, 此时用户的速率区域为凸的. 文献 [107] 进一步利用 Lyapunov 定理严格证明了, 若问题 (3.12) 中的目标函数关于所有 R_k 联合凹且对每一个 R_k 单调递增, 则上述零对偶间隙结论成立. 关于动态谱管理问题取各种不同效用函数的复杂性分析结果, 可参见文献 [107]. 文献 [89] 进一步回答了文献 [107] 中遗留的动态谱管理问题复杂性分析的一个公开问题. 关于动态谱管理问题 (3.12) 的其他算法设计与分析, 可参见文献 [17, 105, 179, 182].

3.3 半正定松弛

本小节给出半正定松弛^[152]技巧应用于求解两类特殊非凸二次约束二次优化资源分配问题的例子.

MIMO 检测 (detection): MIMO 检测问题是现代数字通信中的基本问题之一 (参见文献 [175]). 在数学上, MIMO 信道的输入输出关系可表示为

$$\mathbf{r} = \mathbf{H}\mathbf{x}^* + \mathbf{v},$$

其中, $\mathbf{r} \in \mathbb{C}^m$ 为接收端接收到的信号, $\mathbf{H} \in \mathbb{C}^{m \times n}$ 为信道矩阵, $\mathbf{x}^* \in \mathbb{C}^n$ 表示传输的真实信号, $\mathbf{v} \in \mathbb{C}^m$ 为加性噪声 (m 和 n 分别为接收端和发送端的天线个数). 如果使用 M -PSK (M -ary phase shift keying) 的调制 (modulation) 方式, 那么 $\mathbf{x}^*$ 的每一个分量 x_i^* 应属于一个有限集合, 即 $x_i^* \in \{e^{j\theta} \mid \theta = 2j\pi/M, j = 0, 1, \dots, M-1\}$, $i = 1, 2, \dots, n$, 其中 j 为虚数单位. 假设信道信息 $\mathbf{H}$ 已知, MIMO 检测问题希望基于收到的信号 $\mathbf{r}$ 恢复传输的信号 $\mathbf{x}^*$. 数学上, MIMO 检测问题可建模为

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{C}^n} \quad & \|\mathbf{H}\mathbf{x} - \mathbf{r}\|^2 \\ \text{s.t.} \quad & |x_i|^2 = 1, \quad \arg(x_i) \in \mathcal{A}, \quad i = 1, 2, \dots, n, \end{aligned} \quad (3.14)$$

其中, $\arg(\cdot)$ 表示复数的辐角, $\mathcal{A} = \{0, 2\pi/M, \dots, 2(M-1)\pi/M\}$.

基于半正定松弛的方法是求解问题 (3.14) 最有效的方法之一. 此方法的基本思想是将问题 (3.14) 先松弛为一个 (复的或者实的) 半正定松弛, 然后通过圆整 (rounding) 的方法得到原问题的一个可行解. 特别地, 问题 (3.14) 的传统复半正定松弛是

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{X}} \quad & \text{Trace}(\mathbf{Q}\mathbf{X}) + 2\text{Re}(\mathbf{c}^\dagger \mathbf{x}) \\ \text{s.t.} \quad & X_{ii} = 1, \quad i = 1, 2, \dots, n, \\ & \mathbf{X} \succeq \mathbf{x}\mathbf{x}^\dagger, \end{aligned} \quad (3.15)$$

其中 $\mathbf{Q} = \mathbf{H}^\dagger \mathbf{H}$, $\mathbf{c} = -\mathbf{H}^\dagger \mathbf{r}$, $\text{Re}(\cdot)$ 表示取实部. 由于它的对角特殊结构, 内点算法可在 $\mathcal{O}(n^{3.5})$ 时间内求得 (3.15) 的最优解 (参见文献 [74]). 数值实验表明, 与球解码 (sphere decoding) 算法^[49, 156]相比, 基于半正定松弛的方法在误码率 (symbol error rate, SER) 性能和时间复杂性上做到了非常好的权衡 (参见文献 [74]).

学者们广泛关注的另一个问题是问题 (3.14) 的半正定松弛在什么条件下是紧的, 即什么条件下半正定松弛可以求解得到原问题的最优解. 关于 $M = 2$ 时半正定松弛紧性的相关结果, 可参见文献 [66, 67, 74, 75, 145]. 最近, 文献 [100] 利用辐角信息, 在传统的半正定松弛基础上添加了有效辐角割 (参见文献 [98, 99]), 提出了一个新的半正定松弛, 并在 $M > 2$ 的情形下给出了保证该半正定松弛为紧的充分条件:

$$\lambda_{\min}(\mathbf{H}^\dagger \mathbf{H}) \sin \frac{\pi}{M} > \|\mathbf{H}^\dagger \mathbf{v}\|_\infty,$$

其中, $\lambda_{\min}(\mathbf{H}^\dagger \mathbf{H})$ 表征了信道的好坏, $\sin \frac{\pi}{M}$ 刻画了 PSK 调制阶数 M 对问题难度的影响, $\|\mathbf{H}^\dagger \mathbf{v}\|_\infty$ 刻画了噪声的影响. 简而言之, 调制的阶数越大、信道越差、噪声 (对信道矩阵列空间的影响) 越大, 上述条件越难满足. 更多关于 $M > 2$ 时相关半正定松弛何时是紧的以及 MIMO 检测问题可以精确求解的最新结果, 可参见文献 [71, 82, 100]. 关于 MIMO 检测问题的其他有效算法, 可参见文献 [14, 68, 82, 101, 103, 120, 194].

下行多波 MISO 信道波束成形: 下行多波 MISO 信道波束成形设计问题希望极小化系统总传输功率, 同时满足所有用户的 SINR 约束. 在数学上, 此问题可建模为

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & \frac{|\mathbf{h}_k^\dagger \mathbf{w}|^2}{\sigma_k^2} \geq \gamma_k, \quad k \in \mathcal{K}. \end{aligned} \quad (3.16)$$

不难得到问题 (3.16) 的半正定松弛为

$$\begin{aligned} \min_{\mathbf{W}} \quad & \text{Trace}(\mathbf{W}) \\ \text{s.t.} \quad & \text{Trace}(\mathbf{H}_k \mathbf{W}) \geq \gamma_k \sigma_k^2, \quad k \in \mathcal{K}, \\ & \mathbf{W} \succeq \mathbf{0}, \end{aligned} \quad (3.17)$$

其中, $\mathbf{H}_k = \mathbf{h}_k \mathbf{h}_k^\dagger \in \mathbb{C}^{N \times N}$, $\mathbf{W} = \mathbf{w} \mathbf{w}^\dagger \in \mathbb{C}^{N \times N}$.

当用户数 $K \leq 4$ 时, 半正定松弛 (3.17) 有秩一解^[3], 此时半正定松弛 (3.17) 对于原问题 (3.16) 是紧的. 因为问题 (3.16) 是 NP- 困难的^[144], 所以半正定松弛 (3.17) 对于原问题 (3.16) 一般而言不是紧的. 文献 [144] 首次提出利用 Gauss 随机化的技巧得到原问题的一个可行解: 首先利用 Gauss 随机化技巧在半正定松弛解的基础上产生一些备选点, 然后对这些备选点作适当的尺度 (scale) 变换使其满足所有的约束, 最后从这些可行点中找出使得原问题目标函数最小的点. 特别地, 对于任意解 $\mathbf{w}$, 如下解 $\hat{\mathbf{w}}$ 满足所有的约束:

$$\hat{\mathbf{w}} = \frac{\sqrt{\gamma_k} \sigma_k \mathbf{w}}{\min\{|\mathbf{h}_1^\dagger \mathbf{w}|, |\mathbf{h}_2^\dagger \mathbf{w}|, \dots, |\mathbf{h}_K^\dagger \mathbf{w}|\}}.$$

通过上述的尺度变换得到问题 (3.16) 的可行解充分利用了其约束的特性, 即约束中没有线性项. 因此, 在半正定松弛解的基础上得到一个可行解的复杂度非常低, 与求解半正定松弛 (3.17) 相比, 它的计算复杂度可以忽略. 所以, 半正定松弛加 Gauss 随机化技巧求解单组多波波束成形设计问题 (3.16) 的计算复杂度是多项式的. 当问题的规模 (即用户的个数) 比较小时, 上述方法可以得到质量很好的解. 文献 [106, 144] 从理论上分析了半正定松弛算法的性能, 证明其性能随着用户数 K 的增加而线性衰减. 求解单组多波波束成形设计问题 (3.16) 的其他算法可参见文献 [58, 99, 150]. 特别地, 文献 [99] 给出了现有的关于问题 (3.16) 的几个有效算法的数值比较.

当前关于单组多波波束成形设计问题 (3.16) 的相关延伸和推广主要包括如下两个方面. 第一, 单组多波波束成形设计问题 (3.16) 发展新的物理层 (physical layer) 传输策略是一个非常有趣的研究方向. 特别地, 文献 [128, 163, 169, 170] 独立地提出了一个基于 Alamouti 的编码技术, 这个编码技术数学上可以看成是从问题 (3.16) 的秩一传输技术推广到秩二传输. 文献 [169] 证明了, 相比于秩一的半正定松弛, 秩二半正定松弛的性能随着用户数的增加成平方根衰减. 第二, 单组多波波束成形设计问题 (3.16) 可以推广到多组多波波束成形设计问题, 即系统中的用户根据需要的信息可以分为多组, 基站给每个组内的用户发送相同的信息, 不同组的用户发送不同的信息. 关于这方面的工作, 可参见文献 [12, 21, 35, 72].

3.4 交替最优化方法

本小节举例阐述交替最优化方法^[167] 结合问题特殊结构在求解最优资源分配问题中的应用.

MIMO 干扰信道最小 SINR 最大化: 考虑 MIMO 干扰信道中的最小 SINR 最大化 (等价于最小速率最大化) 问题:

$$\begin{aligned} & \max_{\{\mathbf{u}, \mathbf{v}\}} \min_{k \in \mathcal{K}} \{\text{SINR}_k(\mathbf{u}_k, \mathbf{v})\} \\ & \text{s.t. } \|\mathbf{u}_k\|^2 = 1, \quad \|\mathbf{v}_k\|^2 \leq \bar{p}_k, \quad k \in \mathcal{K}, \end{aligned} \quad (3.18)$$

其中 $\text{SINR}_k(\mathbf{u}_k, \mathbf{v})$ 由 (2.1) 给出. 注意到第 k 个用户的 SINR 表达式 (2.1) 不依赖于 $\{\mathbf{u}_j\}_{j \neq k}$, 所以用 $\text{SINR}_k(\mathbf{u}_k, \mathbf{v})$ 表示第 k 个用户的 SINR. 文献 [91] 将问题 (3.18) 的变量分为两块, 即接收端波束成形向量 $\mathbf{u}$ 和传输端波束成形向量 (包含功率分配) $\mathbf{v}$, 进而交替地优化 $\mathbf{u}$ 和 $\mathbf{v}$. 特别地, 当变量 $\mathbf{v}$ 固定时, 问题 (3.18) 关于变量 $\mathbf{u}$ 有显式解, 即线性方差极小化 (LMMSE) 接收向量; 当变量 $\mathbf{u}$ 固定时, 问题 (3.18) 关于变量 $\mathbf{v}$ 的问题为 MISO 干扰信道中最小 SINR 最大化问题, 可在多项式时间内求解 (参见文献 [90, 164]).

一般情形下, 即使每个子问题都精确求解, 我们仍然不能保证交替最优化算法的收敛性 (参见文献 [123]). 目标函数非光滑部分以及约束中变量的可分性对于交替最优化算法的收敛性非常重要. 容易构造反例证明, 当目标函数非光滑部分或者约束中的变量不可分时, 交替最优化算法产生的点列可能收敛到一个无意义的点 (参见文献 [64]). 尽管问题 (3.18) 的目标函数是非光滑的且目标函数中变量 $\mathbf{u}$ 和 $\mathbf{v}$ 耦合在一起, 但是文献 [91] 利用变量 $\mathbf{u}_k$ 只出现在 SINR_k 中的特殊结构证明了上述交替最优化算法产生的点列收敛到问题 (3.18) 的稳定点. 为进一步提高上述算法的计算有效性, 文献 [93] 提出了一个非精确的交替最优化方法求解问题 (3.18), 其每次迭代非精确地求解关于变量 $\mathbf{v}$ 的优化问题. 更多的关于交替最优化方法求解传输接收波束成形设计和功率控制的工作, 可参见文献 [18, 19, 36, 77, 95, 122, 140, 148].

MIMO 干扰信道和速率最大化: 考虑 MIMO 干扰信道中的和速率最大化问题:

$$\begin{aligned} & \max_{\{\mathbf{u}_k, \mathbf{v}_k\}} \sum_{k \in \mathcal{K}} R_k(\mathbf{u}_k, \{\mathbf{v}_k\}) \\ & \text{s.t. } \|\mathbf{v}_k\|^2 \leq \bar{p}_k, \quad k \in \mathcal{K}, \end{aligned} \quad (3.19)$$

其中

$$R_k(\mathbf{u}_k, \{\mathbf{v}_k\}) = \log_2 \left(1 + \frac{|\mathbf{u}_k^\dagger \mathbf{H}_{kk} \mathbf{v}_k|^2}{\sigma_k^2 \|\mathbf{u}_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^\dagger \mathbf{H}_{kj} \mathbf{v}_j|^2} \right), \quad k \in \mathcal{K}.$$

令

$$e_k = |1 - \mathbf{u}_k^\dagger \mathbf{H}_{kk} \mathbf{v}_k|^2 + \mathbf{u}_k^\dagger \left(\sum_{j \neq k} \mathbf{H}_{kj} \mathbf{v}_j \mathbf{v}_j^\dagger \mathbf{H}_{kj}^\dagger + \sigma_k^2 \mathbf{I} \right) \mathbf{u}_k, \quad k \in \mathcal{K}.$$

上述定义的 e_k 实际为用户 k 对应的均方差 (mean square error, MSE). 容易验证

$$\max_{\mathbf{u}_k} R_k(\mathbf{u}_k, \{\mathbf{v}_k\}) = \max_{\mathbf{u}_k} \log_2(e_k^{-1}), \quad k \in \mathcal{K}.$$

上述问题的最优解为 LMMSE 接收向量. 进一步地, 引入辅助变量 w_k , 可以证明原问题 (3.19) 等价于如下问题:

$$\begin{aligned} & \min_{\{\mathbf{u}_k, \mathbf{v}_k, w_k\}} \sum_{k \in \mathcal{K}} (w_k e_k - \ln(w_k)) \\ & \text{s.t. } \|\mathbf{v}_k\|^2 \leq \bar{p}_k, \quad k \in \mathcal{K}. \end{aligned} \quad (3.20)$$

尽管原问题 (3.19) 关于变量 $\mathbf{v}_k$ 是高度非线性的, 但是它的等价问题 (3.20) 关于每一个变量的优化问题都是“简单”问题. 特别地, 问题 (3.20) 关于变量 $\{w_k\}$ 有显式解 $w_k = 1/e_k$; 问题 (3.20) 关于 $\{\mathbf{u}_k\}$ 有显式解 (LMMSE); 问题 (3.20) 关于 $\{\mathbf{v}_k\}$ 为带有球约束的凸二次规划, 也非常容易求解. 基于上述观察, 求解问题 (3.19) 的 WMMSE (weighted MMSE) 算法^[34, 130, 141] 交替地优化每一组变量, 所以 WMMSE 算法本质上是一个交替最优化方法, 其巧妙之处在于利用原问题 (3.19) 的特殊结构, 将其升维为一个更高维空间中的非线性程度较低的等价优化问题 (3.20).

上面以 MIMO 干扰信道中的和速率最大化问题为例简要介绍了 WMMSE 算法. 实际上, WMMSE 算法可以用来求解 MIMO 干扰广播信道中更一般的效用函数的最大化问题 (参见文献 [141]). 进一步地, 文献 [62] 证明了 WMMSE 算法实际上是一种特殊的连续凸逼近算法, 其中的辅助变量 $\{w_k\}$ 是凸逼近过程中对应的系数. WMMSE 算法进一步的推广和应用, 可参见文献 [40, 65, 80, 127]; 交替最优化方法及其在各个领域中的应用, 可参见文献 [8, 64, 167].

3.5 稀疏优化

本小节举例说明稀疏优化^[195] 技巧在最优资源分配问题的模型建立和算法设计方面的应用.

MIMO 干扰信道基站部分合作: 考虑多用户干扰信道, 如图 1 所示: 第一种 (极端) 情形是各个基站之间服务用户的信息独立, 即第 k 个基站服务第 k 个用户, 所有的基站合作设计各自的波束成形向量 (见问题 (3.19)); 第二种 (极端) 情形是所有基站完全共享所有用户的信息, 这样所有的基站可以看成是一个大的虚拟基站 (见问题 (3.3)). 在上述第一种情形下, 基站不需要共享用户的信息, 其优点在于基站与基站之间的通信负载很低; 其缺点在于整个系统的传输性能可能很差. 在上述第二种情形下, 所有基站共享所有用户的信息, 所有基站完全合作, 所以可以看成是一个虚拟的大基站, 其优点在于整个系统的传输性能会比第一种好很多; 其缺点在于, 当系统中的基站数量较多时基站之间的通信负载过高.

在异构网络 (heterogeneous network) 中, 存在大量的微型 (micro/pico) 基站, 对于某一个用户, 这些基站中的一部分可以合作形成一个虚拟基站服务该用户, 这种方式介于上述两种极端情形之间, 称为部分合作 (partial cooperation) 或者自适应合作 (adaptive cooperation)^[117, 180, 189]. 特别地, 文献 [65] 首先假设所有的基站完全合作, 然后在此基础上强迫其波束成形向量中的某些块为零, 形成部分合作. 数学上, 假设 $\mathbf{w}_k = [\mathbf{w}_{k1}^T, \mathbf{w}_{k2}^T, \dots, \mathbf{w}_{kK}^T]^T \in \mathbb{C}^{N \times 1}$ 为所有基站对第 k 个用户使用的波束成形向量, 其中第 j 块 $\mathbf{w}_{kj} = 0$ 等价于第 j 个基站没有参与第 k 个用户的传输, 所以希望向量 $\mathbf{w}_k$ 是块稀疏的. 数学上, 诱导块稀疏的一种有效的方法是添加正则项 $\|\mathbf{w}_{kj}\|_2$ ^[184]. 基于上述观察, 文献 [65] 利用稀疏优化的思想建立了 MIMO 干扰信道基站部分合作波束成形设计问题的稀疏优化模型:

$$\begin{aligned} & \max_{\{\mathbf{u}_k, \mathbf{w}_k\}} \sum_{k \in \mathcal{K}} R_k(\mathbf{u}_k, \{\mathbf{w}_k\}) - \lambda \sum_k \sum_j \|\mathbf{w}_{kj}\|_2 \\ & \text{s.t.} \quad \sum_{k \in \mathcal{K}} \|\mathbf{w}_{kj}\|^2 \leq \bar{p}_j, \quad j \in \mathcal{K}, \end{aligned} \quad (3.21)$$

其中,

$$\begin{aligned} R_k(\mathbf{u}_k, \{\mathbf{w}_k\}) &= \log_2 \left(1 + \frac{|\mathbf{u}_k^\dagger \mathbf{H}_k \mathbf{w}_k|^2}{\sigma_k^2 \|\mathbf{u}_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^\dagger \mathbf{H}_k \mathbf{w}_j|^2} \right), \quad k \in \mathcal{K}, \\ \mathbf{H}_k &= [\mathbf{H}_{k1}, \mathbf{H}_{k2}, \dots, \mathbf{H}_{kK}], \quad k \in \mathcal{K}; \end{aligned}$$

$\lambda > 0$ 为控制稀疏度的正则化参数, 即基站的合作程度; $\sum_{k \in \mathcal{K}} \|\mathbf{w}_{kj}\|^2 \leq \bar{p}_j$ 为第 j 个基站的传输功率约束, $j \in \mathcal{K}$. 文献 [65] 进一步发展了 S-WMMSE (sparse WMMSE) 算法求解问题 (3.21). 关于基站部分合作的更多工作, 可参见文献 [16, 32, 40, 46, 73, 136, 143, 149].

SISO 干扰信道联合功率/接入控制: 考虑多用户 SISO 干扰信道中的联合功率/接入控制问题, 即当问题 (3.1) 不可行时, 如何从用户集合 $\mathcal{K}$ 中选出尽可能多的用户, 使得系统能够同时支持这些用户 (系统支持一个用户指相应用户的 SINR 大于或等于其 SINR 目标), 同时最小化总传输功率. 为了叙述方便, 引入归一化信道: $\mathbf{q} = (q_1, q_2, \dots, q_K)^T$ 表示归一化以后的功率分配向量; $\mathbf{c} = (c_1, c_2, \dots, c_K)^T$ 表示归一化后的噪声向量, 其中

$$c_k = \frac{\gamma_k \eta_k}{g_{kk} \bar{p}_k} > 0, \quad k \in \mathcal{K};$$

$\mathbf{A} \in \mathbb{R}^{K \times K}$ 表示归一化以后的信道矩阵, 其中 $\mathbf{A}$ 的第 (k, j) 个元素为

$$a_{kj} = \begin{cases} 1, & \text{如果 } k = j, \\ -\frac{\gamma_k g_{kj} \bar{p}_j}{g_{kk} \bar{p}_k}, & \text{如果 } k \neq j. \end{cases}$$

$|a_{kj}|$ 可以看作归一化以后的信道增益. 矩阵 $\mathbf{A}$ 的对角元素都是 1, 非对角元素都是非正的实数, 这个结构特征对于设计算法求解联合功率/接入控制问题非常重要. 基于归一化信道和矩阵 $\mathbf{A}$ 的特殊结构, 文献 [92] 证明了联合功率/接入控制问题等价于如下稀疏优化问题:

$$\begin{aligned} \min_{\{\mathbf{q}\}} \quad & \|\mathbf{A}\mathbf{q} - \mathbf{c}\|_0 + \alpha \bar{\mathbf{p}}^T \mathbf{q} \\ \text{s.t.} \quad & \mathbf{0} \leq \mathbf{q} \leq \mathbf{e}, \end{aligned} \quad (3.22)$$

其中参数 α 满足 $0 < \alpha < \alpha_1 \triangleq 1/\bar{\mathbf{p}}^T \mathbf{e}$. 文献 [92] 进一步证明了问题 (3.22) 的最优解 $\mathbf{q}^*$ 一定满足 $\mathbf{c} - \mathbf{A}\mathbf{q}^* \leq \mathbf{0}$, 且 $(\mathbf{c} - \mathbf{A}\mathbf{q}^*)_k = 0$ 当且仅当第 k 个用户的 SINR 等于其 SINR 目标 γ_k . 文献 [92] 提出用 ℓ_1 凸模型逼近 ℓ_0 模型 (3.22), 并发展了一个有效的用户逐步删除算法求解联合功率/接入控制问题. 文献 [94] 提出用如下 ℓ_q ($0 < q < 1$) 非凸模型逼近 ℓ_0 模型 (3.22):

$$\begin{aligned} \min_{\{\mathbf{x}, \mathbf{q}\}} \quad & \|\mathbf{A}\mathbf{q} - \mathbf{c}\|_q^q + \alpha \bar{\mathbf{p}}^T \mathbf{q} \\ \text{s.t.} \quad & \mathbf{0} \leq \mathbf{q} \leq \mathbf{e}. \end{aligned} \quad (3.23)$$

文献 [94] 推广了文献 [52, 69, 177] 中的算法, 发展了一个多项式时间的势函数下降算法来求解得到非凸逼近模型 (3.23) 的稳定点, 进而发展了一个基于非凸逼近的用户逐步删除算法求解联合功率/接入控制问题. 数值实验表明基于非凸模型的算法性能确实好于基于凸模型的算法性能. 文献 [94] 也证明只要参数 q 选取得充分小, 那么 ℓ_q 逼近模型 (3.23) 具有精确恢复性质, 即 ℓ_q 模型 (3.23) 与 ℓ_0 模型 (3.22) 有相同的最优解, 同时举例说明凸 ℓ_1 逼近模型不具有精确恢复性质. 关于 SISO 干扰信道联合功率/接入控制的其他算法, 可参见文献 [4, 45, 110, 116]; 关于联合波束成形和接入控制的工作, 可参见文献 [112, 113, 142, 157]; 关于形如 (3.23) 非凸非光滑问题的理论分析和算法设计, 可参见文献 [96].

值得注意的是, 这种实际应用中的不可行问题也驱动了相关优化理论与方法的发展. 关于这方面的进展, 可参见文献 [15, 41, 42]. 其中, 文献 [42] 考虑了一般形式的锥规划问题, 试图在不可行程度最小的点中寻找一个使得目标函数最优的点. 本小节考虑的问题正是这种一般问题的特殊情形.

活跃用户检测 (user activity detection): 海量机器类通信的最大挑战在于基站能从大量潜在连接的用户中快速准确地确定出活跃用户, 然后忽略非活跃用户, 只对活跃用户进行数据收发, 这样可以节省大量的系统控制开销 (参见文献 [27, 84]). 此问题的数学模型为

$$\mathbf{Y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k \sqrt{g_k} \mathbf{h}_k^T + \mathbf{Z}, \quad (3.24)$$

其中 $\mathbf{Y} \in \mathbb{C}^{L \times N}$ 为基站收到的观测信号; α_k 表示第 k 个用户是否活跃, $\alpha_k = 1$ 表示第 k 个用户活跃, $\alpha_k = 0$ 表示第 k 个用户不活跃; $\mathbf{s}_k \in \mathbb{C}^L$ 为第 k 个用户的导频序列, 每个用户的导频序列不同, 但两两不一定正交; $\sqrt{g_k} \mathbf{h}_k \in \mathbb{C}^{N \times 1}$ 表示基站和第 k 个用户间的信道向量, g_k 代表大尺度衰落 (large-scale fading), $\mathbf{h}_k$ 代表瑞利衰落 (Rayleigh fading), 服从标准正态分布; $\mathbf{Z} \in \mathbb{C}^{L \times N}$ 为 Gauss 噪声矩阵, 每个分量均值为 0、方差为 σ^2 . 在上述问题中, K 为所有用户的个数, L 为导频序列的长度, N 为基站端天线的个数. 活跃用户检测问题希望基于观测信号 $\mathbf{Y}$ 和导频序列矩阵 $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K] \in \mathbb{C}^{L \times K}$ 快速且准确地检测出活跃用户, 即哪些用户对应 $\alpha_k = 1$, 哪些用户对应 $\alpha_k = 0$. 在上述问题中, 总用户数 K 一般较大, 但在每个时刻活跃用户的个数相对于 K 较小.

利用向量 $[\alpha_1, \alpha_2, \dots, \alpha_K]^T$ 的稀疏性, 文献 [30, 87, 132] 将上述活跃用户检测问题建模为基于压缩感知的行稀疏优化问题, 提出了利用近似信息传递 (approximate message passing, AMP) 算法求解上述模型, 并分析了算法的理论检测性能. 求解上述模型可以同时检测出活跃用户并恢复活跃用户的信道信息. 活跃用户检测问题的另一种建模方式是将 $\{\gamma_k \triangleq g_k \alpha_k\}$ 看成确定的未知参数, 利用信道和噪声的统计特性将相关问题看作参数估计问题 (参见文献 [61]). 具体地, 因为 $\{\mathbf{h}_k\}$ 和噪声 $\mathbf{Z}$ 都服从正态分布, 所以 $\mathbf{Y}$ 的第 n 列 $\mathbf{y}_n \in \mathbb{C}^{L \times 1}$ 服从 $\mathbf{y}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{S} \mathbf{T} \mathbf{S}^\dagger + \sigma_w^2 \mathbf{I})$. 进而估计参数 $\boldsymbol{\gamma}$ 的最大似然估计模型可以等价写为

$$\begin{aligned} \min_{\boldsymbol{\gamma}} \quad & \log |\mathbf{S} \mathbf{T} \mathbf{S}^\dagger + \sigma_w^2 \mathbf{I}| + \text{Trace}((\mathbf{S} \mathbf{T} \mathbf{S}^\dagger + \sigma_w^2 \mathbf{I})^{-1} \hat{\boldsymbol{\Sigma}}) \\ \text{s. t. } \quad & \boldsymbol{\gamma} \geq \mathbf{0}, \end{aligned} \quad (3.25)$$

其中, $\hat{\boldsymbol{\Sigma}} = \mathbf{Y} \mathbf{Y}^\dagger / N$ 为采样协方差矩阵, $\mathbf{T} = \text{Diag}\{\boldsymbol{\gamma}\} \in \mathbb{R}^{K \times K}$, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_K]^T \in \mathbb{R}^{K \times 1}$. 由于上述模型仅依赖于观测矩阵 $\mathbf{Y}$ 的采样协方差矩阵, 因此又称为基于协方差矩阵的模型. 数值实验和理论分析都表明基于协方差矩阵模型的检测效果好于基于压缩感知稀疏优化模型的检测效果 (参见文献 [29, 48]). 特别地, 文献 [29] 刻画了基于协方差矩阵模型 (3.25) 的解的渐近一致性和估计误差分布, 并证明了当天线数 N 趋于无穷时基于协方差矩阵的模型 (3.25) 可以无误地检测 $\mathcal{O}(L^2)$ 个活跃用户.

现有求解问题 (3.25) 的算法包括坐标下降算法 [28, 61, 70]、期望最小化 (expectation minimization) 算法 [165] 和 SPICE (sparse iterative covariance-based estimation) 算法 [176], 其中坐标下降算法在这 3 类方法中求解效率最高. 坐标下降算法利用问题的特殊结构每次只更新一个变量 (固定其他变量), 每次更新都有闭式解, 因此计算量很小. 文献 [158] 利用问题真解的稀疏性设计了积极集算法求解问题 (3.25). 积极集算法的基本思想是, 每次迭代挑选出所有用户的一个子集 (称为积极集), 然后求解定义在积极集上的小规模问题. 由于问题真解的稀疏性, 如果每次积极集选取得恰当, 则定义在积极集上的优化问题的维数相比于原问题的维数将会小很多, 从而极大地提高计算效率. 数值实验表明积极集算法的计算效率高于没有利用问题真解稀疏性的坐标下降算法 [158]. 关于活跃用户检测问题的最新进展, 如用户和数据联合检测、测量不完全同步、多小区协作检测等, 可参见文献 [31, 51, 79, 85, 134, 159–161].

3.6 分式规划

本小节阐述分式规划^[147]技巧在能效模型的求解和最优资源分配模型的简化方面的应用.

SISO 干扰信道能效极大化: 近年来,“能效”作为绿色通信的重要指标受到广泛关注. 这里,能效是指单位功率支持的传输速率. 多用户 SISO 干扰信道中的能效极大化问题可以建模为

$$\begin{aligned} \max_{\{p_k\}} \quad & \frac{\sum_{k \in \mathcal{K}} \log_2(1 + \frac{g_{kk}p_k}{\sum_{j \neq k} g_{kj}p_j + \eta_k})}{\sum_{k \in \mathcal{K}} p_k + P_{c,k}} \\ \text{s.t.} \quad & 0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}. \end{aligned} \quad (3.26)$$

上述目标函数表示整个系统的总能效, 其中分子和分母分别表示总传输速率和总发射功率. 能效这一指标具有分式形式, 因此, 分式规划技巧在能效模型的求解中发挥着重要作用. 下面考虑问题 (3.26) 的求解.

经典的分式规划算法, 如 Dinkelbach 算法^[44] 和 Charnes-Cooper 算法^[20,129] 均要求分子是非负凹函数, 分母是恒正凸函数, 而问题 (3.26) 的分子关于变量 $\{p_k\}$ 是非凹的. 一种经典的做法是, 利用序列分式规划技巧, 即通过求解一系列简单的凹 - 凸分式问题, 逐步逼近得到原问题的稳定点. 这种方法实际上是将凸逼近思想^[111] 和分式规划技巧结合起来. 具体地, 这是一个双层迭代算法. 外层迭代基于当前迭代点 $\{p_k^t\}$ 更新分子的下界 $g(\mathbf{p}, \mathbf{p}^t)$, 内层迭代利用 Dinkelbach 算法求解当前的凹 - 凸问题. 这里, $g(\mathbf{p}, \mathbf{p}^t)$ 是速率函数在当前点的凹近似 (参见文献 [111]). 对于上述 SISO 能效模型 (3.26), 类似文献 [121] 中的做法, 可以选取

$$g(\mathbf{p}, \mathbf{p}^t) = \sum_{k \in \mathcal{K}} \left(a_k^t \log_2 \left(\frac{g_{kk}p_k}{\sum_{j \neq k} g_{kj}p_j + \eta_k} \right) + b_k^t \right),$$

其中,

$$a_k^t = \frac{\text{SINR}_k^t}{1 + \text{SINR}_k^t}, \quad b_k^t = \log_2(1 + \text{SINR}_k^t) - \frac{\text{SINR}_k^t}{1 + \text{SINR}_k^t} \log_2(\text{SINR}_k^t),$$

SINR_k^t 为用户 k 在第 t 次迭代的 SINR 值. 再对 $\{p_k\}$ 进行变量代换 $p_k = e^{\tilde{p}_k}$, 可以得到相应的凹 - 凸模型.

文献 [154, 188] 将上述结果分别推广到多载波 SISO 干扰信道和 MIMO 中继干扰信道中. 文献 [118] 利用序列分式规划技巧求解了多用户 MIMO 系统的能效模型. 利用分式规划技巧求解能效模型的更多文献, 可参见文献 [11, 53]. 利用分式规划技巧求解能效模型的综述文章, 可参见文献 [187].

SISO 干扰信道上行用户调度 (user scheduling) 与功率分配: 信干噪比和传输速率是衡量无线通信性能的重要指标. 信干噪比, 如 (2.1)、(2.3) 和 (2.4), 天然具有分式形式, 而速率函数 (2.2) 是对数函数与分式函数的复合. 这种复杂的函数形式使得相应的无线通信问题, 特别是和速率极大化问题很难求解. 针对此类问题的特殊结构, 文献 [137, 138] 提出了二次变换和 Lagrange 对偶变换, 不仅大大提升了此类问题的求解效率和质量, 也为分式规划理论注入了新鲜血液.

下面以 SISO 信道上行用户调度和功率分配问题为例, 阐述二次变换和 Lagrange 对偶变换的有效性. 考虑蜂窝网络 (cellular network), 每个基站及与它关联的用户构成一个小区 (cell). 在 SISO 系统中, 每个基站只有一根天线, 因此即使一个小区中有多个用户, 在每个时隙 (time slot) 也只能有一个用户向基站发送数据. 如何选取工作的用户以及为这些用户分配多少功率是我们解决的问题. 在

数学上, 这个问题可以被建模为

$$\begin{aligned} \max_{\{\mathbf{p}, \mathbf{s}\}} \quad & \sum_{i \in \mathcal{B}} \ln \left(1 + \frac{g_{is_i} p_{s_i}}{\sum_{j \neq i} g_{is_j} p_{s_j} + \eta_i} \right) \\ \text{s.t.} \quad & 0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}, \\ & s_i \in \mathcal{K}_i \cup \{\emptyset\}, \end{aligned} \quad (3.27)$$

其中, $\mathcal{B}$ 是所有小区的集合, $\mathcal{K}_i$ 是第 i 个小区中用户的集合, $\mathbf{s}$ 是调度变量, $s_i = k$ 指在第 i 个小区中用户 k 给基站 i 传送数据, 而 $s_i = \emptyset$ 指第 i 个小区中没有用户传输数据. 这里用 $\ln(\cdot)$ 函数代替 $\log_2(\cdot)$ 函数, 不影响问题的最优解. 上述问题是一个复杂的离散优化问题, 不仅目标函数是非凸的, 而且离散的调度变量 $\{s_j\}$ 交织在一起, 这给问题的求解带来了很大的困难. 文献 [138] 考虑利用二次变换和 Lagrange 对偶变换对上述模型进行简化. 首先, 利用 Lagrange 对偶变换 [138], 可以将分式函数从对数函数中分离出来, 得到如下等价问题:

$$\begin{aligned} \max_{\{\mathbf{p}, \mathbf{s}, \boldsymbol{\gamma}\}} \quad & \sum_{i \in \mathcal{B}} \left(\ln(1 + \gamma_i) - \gamma_i + \frac{(1 + \gamma_i) g_{is_i} p_{s_i}}{\sum_{j \in \mathcal{B}} g_{is_j} p_{s_j} + \eta_i} \right) \\ \text{s.t.} \quad & 0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}, \\ & s_i \in \mathcal{K}_i \cup \{\emptyset\}. \end{aligned} \quad (3.28)$$

进一步应用二次变换 [137], 可以将上述目标函数中分式部分的分子与分母分离开来, 得到

$$\begin{aligned} \max_{\{\mathbf{p}, \mathbf{s}, \boldsymbol{\gamma}, \mathbf{y}\}} \quad & \sum_{i \in \mathcal{B}} \left(\ln(1 + \gamma_i) - \gamma_i - y_i^2 \eta_i + 2y_i \sqrt{(1 + \gamma_i) g_{is_i} p_{s_i}} - \sum_{j \in \mathcal{B}} y_j^2 g_{is_j} p_{s_j} \right) \\ \text{s.t.} \quad & 0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}, \\ & s_i \in \mathcal{K}_i \cup \{\emptyset\}. \end{aligned} \quad (3.29)$$

容易看出, 当 $(\mathbf{s}, \mathbf{p})$ 固定时, 问题 (3.28) 关于 $\boldsymbol{\gamma}$ 有显式解. 当 $(\mathbf{s}, \mathbf{p})$ 和 $\boldsymbol{\gamma}$ 固定时, 问题 (3.29) 关于 $\mathbf{y}$ 有显式解. 当 $\boldsymbol{\gamma}$ 和 $\mathbf{y}$ 固定时, s_i 只出现在目标函数的第 i 项中, 与 $\{s_j\}_{j \neq i}$ 无关. 因此, $(\mathbf{s}, \mathbf{p})$ 也容易确定. 求解问题 (3.27) 的分式规划算法 [138] 基于如上观察, 交替优化每一组变量进行求解. 此外, 文献 [139] 证明了上述交替更新变量的方法本质上是一种 MM (minorization-maximization) 算法, 并将二次变换和 Lagrange 对偶变换推广到矩阵情形.

目前正被广泛研究的混合波束成形 (hybrid beamforming) 模型 [6, 78] 和智能反射面 (intelligent reflecting surface) 模型 [168, 173] 均含有变量的乘积项. 这种复杂的结构给问题的求解带来了更大的挑战. 二次变换和 Lagrange 对偶变换可对此类复杂的模型进行简化, 关于这方面的研究可参见文献 [59, 76, 109]. 新的通信技术和设备也会产生更多的带有特殊结构的优化问题, 如通信感知一体化中的波束成形设计问题 [81]、带有一比特或离散恒包络约束的波束成形设计问题 [171, 172] 等. 这些优化问题也会驱动更多的优化方法的研究.

3.7 整数规划

本小节给出线性整数规划 [37] 技巧在网络切片 (network slicing) 问题的模型建立中的应用.

网络切片问题: 考虑网络切片问题, 该问题试图将多个虚拟网络请求映射到共享的物理网络. 这里, 多个虚拟网络请求被称为功能服务链 (service function chain, SFC); 物理网络可表示为 $\mathcal{G} = (\mathcal{I}, \mathcal{L})$, 其

中 $\mathcal{I}$ 和 $\mathcal{L}$ 分别为网络中节点的集合和边的集合. 借助网络虚拟化技术 (network function virtualization), 可在网络中的一些节点 ($\mathcal{V} \subseteq \mathcal{I}$) 提供必要的功能处理, 如防火墙、负载均衡器和深层报文检测等 (参见文献 [114]). 称能提供功能处理的节点为功能节点. 图 3 给出了物理网络的一个简单例子, 其中蓝色节点 (C 和 E) 表示功能节点. 每一条功能服务链 $k \in \mathcal{K}$ 是从初始节点 $S(k) \in \mathcal{I}$ 到目的节点 $D(k) \in \mathcal{I}$ 的数据流, 该数据流要求有序处理特定的功能单元 $f_1^k \rightarrow f_2^k \rightarrow \cdots \rightarrow f_{\ell_k}^k$ (其中 $\ell_k \in \mathbb{Z}_+$) [115, 193]. 图 4 给出了一条要求有序处理功能单元 $f_1^k \rightarrow f_2^k$ 的功能服务链. 将功能服务链 k 映射到物理网络 $\mathcal{G}$ 要求 (i) 将每个功能单元 f_s^k 映射到物理网络中特定的功能节点 $v_s^k \in \mathcal{V}$ ($s \in \mathcal{F}(k) := \{1, 2, \dots, \ell_k\}$); (ii) 将虚拟边 (f_s^k, f_{s+1}^k) 映射到物理网络中从节点 v_s^k 到节点 v_{s+1}^k 的网络流, 记为 (k, s) . 注意, 这里为了符号的方便, 假设 f_0^k 和 $f_{\ell_k}^k$ 是虚拟的功能单元且 $v_0^k = S(k)$ 和 $v_{\ell_k}^k = D(k)$. 如何通过合理地分配网络资源 (功能节点的计算能力、边的带宽等) 将所有的功能服务链同时映射到共享的物理网络 $\mathcal{G}$ 是一个十分重要的研究问题. 下面介绍该问题的数学规划模型.

引入 0-1 变量 $x_{v,s}(k)$ 表示功能单元 f_s^k 是否由功能节点 v 处理, 要求每一个功能单元只能由网络中的一个功能节点处理:

$$\sum_{v \in \mathcal{V}} x_{v,s}(k) = 1, \quad \forall k \in \mathcal{K}, \quad \forall s \in \mathcal{F}(k). \quad (3.30)$$

特别地, 如果功能节点 v 不能处理功能单元 f_s^k , 则可简单设置 $x_{v,s}(k) = 0$. 令 $\lambda(k)$ 表示数据流 $k \in \mathcal{K}$ 的流量. 设功能节点每处理一单位的数据流需要消耗 (标准化的) 一单位的计算能力. 我们要求在节点 v 处理的数据流消耗的总计算能力不能超过 μ_v :

$$\sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{F}(k)} \lambda(k) x_{v,s}(k) \leq \mu_v, \quad \forall v \in \mathcal{V}. \quad (3.31)$$

令变量 $r_{ij}(k, s)$ 表示网络流 (k, s) 在边 (i, j) 上的流量, 则流守恒约束可表示为

$$\sum_{j: (j,i) \in \mathcal{L}} r_{ji}(k, s) - \sum_{j: (i,j) \in \mathcal{L}} r_{ij}(k, s) = \lambda(k) * b_{i,s}(k), \quad \forall i \in \mathcal{I}, \quad \forall k \in \mathcal{K}, \quad \forall s \in \mathcal{F}(k) \cup \{0\}, \quad (3.32)$$

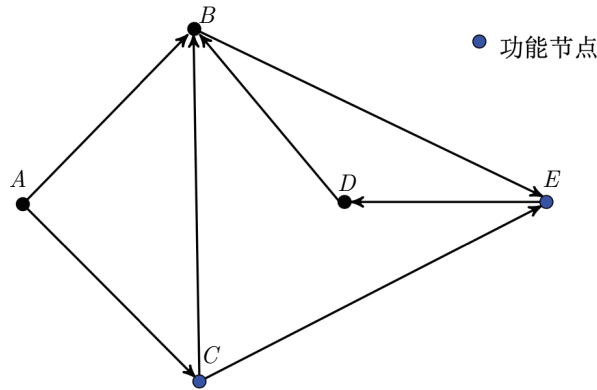


图 3 网络切片问题物理网络示意图

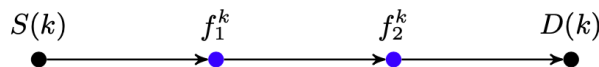


图 4 网络切片问题业务功能链示意图

其中

$$b_{i,s}(k) = \begin{cases} -1, & \text{如果 } s = 0 \text{ 且 } i = S(k), \\ x_{i,s+1}(k), & \text{如果 } s = 0 \text{ 且 } i \in \mathcal{V}, \\ x_{i,s+1}(k) - x_{i,s}(k), & \text{如果 } 1 \leq s < \ell_k \text{ 且 } i \in \mathcal{V}, \\ -x_{i,s}(k), & \text{如果 } s = \ell_k \text{ 且 } i \in \mathcal{V}, \\ 1, & \text{如果 } s = \ell_k \text{ 且 } i = D(k), \\ 0, & \text{其他.} \end{cases}$$

我们要求在边 (i, j) 上的总流量不能超过其容量 C_{ij} :

$$\sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{F}(k)} r_{ij}(k, s) \leq C_{ij}, \quad \forall (i, j) \in \mathcal{L}. \quad (3.33)$$

基于上述变量和约束条件, 文献 [190, 191] 将网络切片问题建模为

$$\begin{aligned} \min_{\{\mathbf{x}, \mathbf{r}\}} \quad & \sum_{(i,j) \in \mathcal{L}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{F}(k)} r_{ij}(k, s) \\ \text{s.t.} \quad & (3.30) - (3.33), \\ & x_{v,s}(k) \in \{0, 1\}, \quad \forall v \in \mathcal{V}, \quad \forall k \in \mathcal{K}, \quad \forall s \in \mathcal{F}(k), \\ & r_{ij}(k, s) \in \{0, 1\}, \quad \forall (i, j) \in \mathcal{L}, \quad \forall k \in \mathcal{K}, \quad \forall s \in \mathcal{F}(k), \end{aligned} \quad (3.34)$$

其中目标函数为最小化整个网络的流量. 特别地, 它可避免网络流 (k, s) 有圈. 关于网络切片问题的其他目标函数, 参见文献 [24, 166]. 问题 (3.34) 是一个线性 (混合) 整数规划问题, 借助目前成熟的线性整数规划求解器 (如 CPLEX^[39] 和 GUROBI^[60]), 可得到其全局最优解. 然而, 问题 (3.34) 是强 NP-困难的^[190], 因而目前的算法大部分都是启发式算法, 如基于线性规划松弛的圆整启发式算法、模拟退火算法和禁忌算法等.

在 5G 网络中, 功能服务链的端对端延迟控制非常重要 (参见文献 [2]). 特别地, 功能服务链的端对端延迟主要包括两部分: (i) 功能单元在功能节点的计算延迟; (ii) 处理相邻两个功能单元的两个物理节点之间的网络流的传播延迟. 实际应用中往往要求功能服务链的端对端延迟不超过给定的阈值 (参见文献 [153]). 文献 [24, 25, 124] 建立了考虑功能服务链的端对端延迟约束的网络切片问题的线性 (混合) 整数规划模型. 文献 [22] 设计了基于线性规划松弛的圆整和改进算法求解该问题. 关于网络切片问题的更多工作, 可参见文献 [23, 26, 43, 115, 153].

4 未来发展方向和展望

随着 5G 时代的到来和后 5G 预研的开始, 无线通信技术进入了新的快速发展阶段. 新技术、新设备的引入为通信的资源配置问题带来了新的挑战. 本节仅列举几个无线通信发展的重要方向以供读者参考.

通信感知一体化: 通信与雷达感知被视为下一代无线网络的两种基本功能. 现有雷达与通信系统的硬件架构、信道特性和信号处理方法已经十分接近. 从技术角度看, 两者的一体化已成必然趋势.

中国工程院在 2020 年年底发布的《全球工程前沿 2020》报告中将“无线通信与感知一体化技术”列为信息电子领域工程开发前沿 Top 2。华为在 2020 全球移动宽带论坛上提出了其对 5.5G 的三大愿景,其中之一即为“融合感知通信”,要求蜂窝网络既能提供通信能力,也能提供感知能力。通信感知一体化是一个新的研究方向,即在同一系统上实现通信与感知功能。通信与感知任务通常具有不同的性能指标,例如,通信的性能指标主要是误码率和传输速率,而感知的性能指标主要是最小均方差和检测概率等。因此,通信感知一体化中一个重要的问题就是两者的物理层性能优化。一个典型例子是通信感知一体化波束成形,需要综合两者的性能指标建模并求解优化问题。通信感知一体化中的另一个重要问题是从信息论角度刻画两者之间的基本理论性能边界。IEEE 通信学会于 2021 年新成立了一个新兴技术倡议 (Emerging Technology Initiative) 委员会,其主题即为通信感知一体化,详见 <https://isac.committees.comsoc.org/>。

数据驱动的未来通信网络: 在未来通信网络中,在数据驱动下,融合机器学习和人工智能技术是一个热门研究方向。与以往基于模型的通信网络的研究相比,该技术最大的特点在于去模型化 (model free)、数据驱动 (data driven) 和自适应性 (adaptivity)。无线系统设计所面临的一个基本挑战就是在困难的条件下管理并且分配资源,以满足流量需求。传统的解决办法是基于领域知识和经验对系统进行建模、分析和模拟,从而制定出具有针对性的规则。在当今的网络环境下,流量呈现出多样化和动态化的发展趋势,并且网络和资源的架构都变得更加复杂,因此该方法面临着越来越大的挑战。未来的网络必须具备自学习和自适应能力,从而能满足用户设备、无线条件和各种应用生成的流量特征所提出的需求。机器学习特别是深度学习和强化学习技术可以通过对过去的行为模式、输出成果以及统一网络或者其他网络上的类似实体的行为进行学习,使得网络的决策质量持续提高。当前和未来网络行为和性能因素比过去更加动态和不可预测。随着网络运营与优化的焦点从网络性能转变为用户体验,传统的方法和工具已经无法跟上时代的发展脚步,必须采用新的基于大数据分析的技术来支持不同的网络使用模型。如何从海量无线网络数据中分析网络行为以提高用户的体验以及提取有价值的信息是一个非常值得研究的课题。

无线通信应用驱动的优化方法: 随着新技术 (如超大规模天线阵列技术等)、新设备 (如智能反射面等) 的不断涌现,无线通信系统优化设计中的问题也呈现出新的趋势。第一,问题的规模越来越大。例如,现有的基站只有几根天线,而未来的超大规模天线阵列场景下的基站可能装有成百上千根天线;智能反射面中反射单元的个数也是成百上千。又由于实时性的要求,这给系统信号处理和优化带来了不小的挑战。第二,问题呈现新的结构,非线性程度越来越高。例如,带有智能反射面的波束成形设计问题中具有最优化变量相乘的双线性项,这一特殊结构与分式结构和对数变换进行复合使得问题求解非常困难。最后,最优化问题的自变量呈现组合趋势,即问题中离散变量越来越多,很多问题最后建模为带有整数和连续变量的混合整数非线性规划问题。与文中介绍的大多数连续优化问题相比,混合整数非线性规划问题求解起来更加困难且更加具有挑战性,混合整数非线性规划问题也是近年来在国际上非常前沿和热门的优化问题。总之,一方面,这些新的问题和挑战为优化的研究提供了新的素材,也必将驱动新的优化理论和高效优化算法的诞生;另一方面,新的高效优化算法也将大大提升相关问题求解的效率,推动相关通信技术的发展。

致谢 作者感谢北京邮电大学孙聪博士在本文前期撰写过程中提供的宝贵意见和建议。

参考文献

- 1 Aein J M. Power balancing in systems employing frequency reuse. COMSAT Tech Rev, 1973, 3: 277-299

- 2 Agyapong P K, Iwamura M, Staehle D, et al. Design considerations for a 5G network architecture. *IEEE Commun Mag*, 2014, 52: 65–75
- 3 Ai W, Huang Y W, Zhang S. New results on Hermitian matrix rank-one decomposition. *Math Program*, 2011, 128: 253–283
- 4 Andersin M, Rosberg Z, Zander J. Gradual removals in cellular PCS with constrained power control and noise. *Wireless Netw*, 1996, 2: 27–43
- 5 Andrews J G, Buzzi S, Choi W, et al. What will 5G be? *IEEE J Sel Areas Commun*, 2014, 32: 1065–1082
- 6 Ayach O E, Rajagopal S, Abu-Surra S, et al. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans Wireless Commun*, 2014, 13: 1499–1513
- 7 Bengtsson M, Ottersten B. Optimal and suboptimum transmit beamforming. In: *Handbook of Antennas in Wireless Communications*. Boca Raton: CRC Press, 2002, 18:1–18:33
- 8 Bertsekas D P. *Nonlinear Programming*, 2nd ed. Belmont: Athena Scientific, 1999
- 9 Björnson E, Hoydis J, Sanguinetti L. Massive MIMO networks: Spectral, energy, and hardware efficiency. *FNT Signal Processing*, 2017, 11: 154–655
- 10 Björnson E, Jorswieck E. Optimal resource allocation in coordinated multi-cell systems. *FNT Commun Inf Theor*, 2013, 9: 113–381
- 11 Björnson E, Sanguinetti L, Kountouris M. Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO. *IEEE J Sel Areas Commun*, 2016, 34: 832–847
- 12 Bornhorst N, Pesavento M, Gershman A B. Distributed beamforming for multi-group multicasting relay networks. *IEEE Trans Signal Process*, 2012, 60: 221–232
- 13 Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge: Cambridge University Press, 2004
- 14 Burer S, Monteiro R D C, Zhang Y. Rank-two relaxation heuristics for MAX-CUT and other binary quadratic programs. *SIAM J Optim*, 2002, 12: 503–521
- 15 Burke J V, Curtis F E, Wang H. A sequential quadratic optimization algorithm with rapid infeasibility detection. *SIAM J Optim*, 2014, 24: 839–872
- 16 Cao P, Liu W J, Thompson J S, et al. Semidynamic green resource management in downlink heterogeneous networks by group sparse power control. *IEEE J Sel Areas Commun*, 2016, 34: 1250–1266
- 17 Cendrillon R, Yu W, Moonen M, et al. Optimal multiuser spectrum balancing for digital subscriber lines. *IEEE Trans Commun*, 2006, 54: 922–933
- 18 Chang J H, Tassiulas L, Rashid-Farrokhi F. Joint transmitter receiver diversity for efficient space division multiaccess. *IEEE Trans Wireless Commun*, 2002, 1: 16–27
- 19 Chang T H, Liu Y F, Lin S C. QoS-based linear transceiver optimization for full-duplex multiuser communications. *IEEE Trans Signal Process*, 2018, 66: 2300–2313
- 20 Charnes A, Cooper W W. Programming with linear fractional functionals. *Naval Res Logist*, 1962, 9: 181–186
- 21 Chen E, Tao M, Liu Y F. Joint base station clustering and beamforming for non-orthogonal multicast and unicast transmission with backhaul constraints. *IEEE Trans Wireless Commun*, 2018, 17: 6265–6279
- 22 Chen W K, Liu Y F, Dai Y H, et al. An efficient linear programming rounding-and-refinement algorithm for large-scale network slicing problem. In: *Proceedings of IEEE ICASSP'21*. Toronto: IEEE, 2021, 4735–4739
- 23 Chen W K, Liu Y F, Dai Y H, et al. Optimal QoS-aware network slicing for service-oriented networks with flexible routing. In: *Proceedings of IEEE ICASSP'22*. Singapore: IEEE, 2022, 5288–5292
- 24 Chen W K, Liu Y F, De Domenico A, et al. Network slicing for service-oriented networks with flexible routing and guaranteed E2E latency. In: *Proceedings of IEEE SPAWC'20*. Atlanta: IEEE, 2020, 1–5
- 25 Chen W K, Liu Y F, De Domenico A, et al. Optimal network slicing for service-oriented networks with flexible routing and guaranteed E2E latency. *IEEE Trans Netw Serv Manage*, 2021, 18: 4337–4352
- 26 Chen W K, Liu Y F, Liu F, et al. Towards efficient large-scale network slicing: An LP rounding-and-refinement approach. *arXiv:2107.14404*, 2022
- 27 Chen X, Ng D W K, Yu W, et al. Massive access for 5G and beyond. *IEEE J Sel Areas Commun*, 2021, 39: 615–637
- 28 Chen Z, Sohrabi F, Liu Y F, et al. Covariance based joint activity and data detection for massive random access with massive MIMO. In: *Proceedings of IEEE ICC'19*. Shanghai: IEEE, 2019, 1–6
- 29 Chen Z, Sohrabi F, Liu Y F, et al. Phase transition analysis for covariance-based massive random access with massive MIMO. *IEEE Trans Inform Theory*, 2022, 68: 1696–1715
- 30 Chen Z, Sohrabi F, Yu W. Sparse activity detection for massive connectivity. *IEEE Trans Signal Process*, 2018, 66: 1890–1904
- 31 Chen Z, Sohrabi F, Yu W. Sparse activity detection in multi-cell massive MIMO exploiting channel large-scale fading.

- IEEE Trans Signal Process, 2021, 69: 3768–3781
- 32 Cheng Y, Pesavento M, Philipp A. Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming. IEEE Trans Signal Process, 2013, 61: 3972–3987
- 33 Chiang M, Hande P, Lan T, et al. Power control in wireless cellular networks. FNT Networking, 2008, 2: 381–533
- 34 Christensen S S, Agarwal R, De Carvalho E, et al. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. IEEE Trans Wireless Commun, 2008, 7: 4792–4799
- 35 Christopoulos D, Chatzinotas S, Ottersten B. Weighted fair multicast multigroup beamforming under per-antenna power constraints. IEEE Trans Signal Process, 2014, 62: 5132–5142
- 36 Codreanu M, Tölli A, Juntti M, et al. Joint design of Tx-Rx beamformers in MIMO downlink channel. IEEE Trans Signal Process, 2007, 55: 4639–4655
- 37 Conforti M, Cornuéjols G, Zambelli G. Integer Programming. New York: Springer, 2014
- 38 Cover T M, Thomas J A. Elements of Information Theory. New York: John Wiley & Sons, 2006
- 39 CPLEX. CPLEX User's Manual. 2022
- 40 Dai B B, Yu W. Sparse beamforming and user-centric clustering for downlink cloud radio access network. IEEE Access, 2014, 2: 1326–1339
- 41 Dai Y H, Liu X W, Sun J. A primal-dual interior-point method capable of rapidly detecting infeasibility for nonlinear programs. J Ind Manag Optim, 2020, 16: 1009–1035
- 42 Dai Y H, Zhang L W. Optimization with least constraint violation. CSIAM Trans Appl Math, 2021, 2: 551–584
- 43 De Domenico A, Liu Y F, Yu W. Optimal virtual network function deployment for 5G network slicing in a hybrid cloud infrastructure. IEEE Trans Wireless Commun, 2020, 19: 7942–7956
- 44 Dinkelbach W. On nonlinear fractional programming. Manage Sci, 1967, 13: 492–498
- 45 Evangelinakis D I, Sidiropoulos N D, Swami A. Joint admission and power control using branch & bound and gradual admissions. In: Proceedings of IEEE SPAWC'10. Marrakech: IEEE, 2010, 1–5
- 46 Fan C M, Zhang Y J, Yuan X. Dynamic nested clustering for parallel PHY-layer processing in cloud-RANs. IEEE Trans Wireless Commun, 2016, 15: 1881–1894
- 47 Fan X, Liu Y F, Liu L. Efficiently and globally solving joint beamforming and compression problem in the cooperative cellular network via Lagrangian duality. In: Proceedings of IEEE ICASSP'22. Singapore: IEEE, 2022, 5388–5392
- 48 Fengler A, Haghighatshoar S, Jung P, et al. Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver. IEEE Trans Inform Theory, 2021, 67: 2925–2951
- 49 Fincke U, Pohst M. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. Math Comp, 1985, 44: 463–471
- 50 Foschini G J, Miljanic Z. A simple distributed autonomous power control algorithm and its convergence. IEEE Trans Veh Technol, 1993, 42: 641–646
- 51 Ganesan U K, Björnson E, Larsson E G. An algorithm for grant-free random access in cell-free massive MIMO. In: Proceedings of IEEE SPAWC'20. Atlanta: IEEE, 2020, 1–5
- 52 Ge D, Jiang X Y, Ye Y. A note on the complexity of L_p minimization. Math Program, 2011, 129: 285–299
- 53 Ge X, Yang B, Ye J, et al. Spatial spectrum and energy efficiency of random cellular networks. IEEE Trans Commun, 2015, 63: 1019–1030
- 54 Gershman A B, Sidiropoulos N D, Shahbazpanahi S, et al. Convex optimization-based beamforming. IEEE Signal Process Mag, 2010, 27: 62–75
- 55 Gesbert D, Hanly S, Huang H, et al. Multi-cell MIMO cooperative networks: A new look at interference. IEEE J Sel Areas Commun, 2010, 28: 1380–1408
- 56 Goldsmith A. Wireless Communications. Singapore: World Scientific, 2006
- 57 Golub G H, Van Loan C F. Matrix Computations, 3rd ed. Baltimore: The Johns Hopkins Press, 1996
- 58 Gopalakrishnan B, Sidiropoulos N D. High performance adaptive algorithms for single-group multicast beamforming. IEEE Trans Signal Process, 2015, 63: 4373–4384
- 59 Guo H Y, Liang Y C, Chen J, et al. Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks. IEEE Trans Wireless Commun, 2020, 19: 3064–3076
- 60 Gurobi Optimization. Gurobi Optimizer Reference Manual. https://www.gurobi.com/wp-content/plugins/hd_documentations/documentation/9.0/refman.pdf, 2022
- 61 Haghighatshoar S, Jung P, Caire G. Improved scaling law for activity detection in massive MIMO systems. In: Proceedings of IEEE ISIT'18. Vail: IEEE, 2018, 381–385
- 62 Hong M, Li Q, Liu Y F. Decomposition by successive convex approximation: A unifying approach for linear transceiver design in heterogeneous networks. IEEE Trans Wireless Commun, 2016, 15: 1377–1392

- 63 Hong M, Luo Z Q. Signal processing and optimal resource allocation for the interference channel. In: Academic Press Library in Signal Processing. Amsterdam: Elsevier, 2014, 2: 409–469
- 64 Hong M, Razaviyayn M, Luo Z Q, et al. A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing. *IEEE Signal Process Mag*, 2016, 33: 57–77
- 65 Hong M, Sun R, Baligh H, et al. Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks. *IEEE J Sel Areas Commun*, 2013, 31: 226–240
- 66 Jaldén J. Detection for multiple input multiple output channels: Analysis of sphere decoding and semidefinite relaxation. PhD Thesis. Stockholm: KTH, 2006
- 67 Jaldén J, Martin C, Ottersten B. Semidefinite programming for detection in linear systems-optimality conditions and space-time decoding. In: *Proceedings of IEEE ICASSP'03*. Hong Kong: IEEE, 2003, 9–12
- 68 Jaldén J, Ottersten B. On the complexity of sphere decoding in digital communications. *IEEE Trans Signal Process*, 2005, 53: 1474–1484
- 69 Ji S, Sze K F, Zhou Z, et al. Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In: *Proceedings of INFOCOM'13*. Turin: IEEE, 2013, 2499–2507
- 70 Jiang D, Cui Y. ML estimation and MAP estimation for device activities in grant-free random access with interference. In: *Proceedings of IEEE WCNC'20*. Seoul: IEEE, 2020, 1–6
- 71 Jiang R, Liu Y F, Bao C, et al. Tightness and equivalence of semidefinite relaxations for MIMO detection. *arXiv:2102.04586*, 2020
- 72 Karipidis E, Sidiropoulos N D, Luo Z Q. Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups. *IEEE Trans Signal Process*, 2008, 56: 1268–1279
- 73 Kim J W, Lee H W, Chong S. Virtual cell beamforming in cooperative networks. *IEEE J Sel Areas Commun*, 2014, 32: 1126–1138
- 74 Kisialiou M, Luo X, Luo Z Q. Efficient implementation of quasi- maximum-likelihood detection based on semidefinite relaxation. *IEEE Trans Signal Process*, 2009, 57: 4811–4822
- 75 Kisialiou M, Luo Z Q. Probabilistic analysis of semidefinite relaxation for binary quadratic minimization. *SIAM J Optim*, 2010, 20: 1906–1922
- 76 Li H, Li M, Liu Q. Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmwave MU-MISO systems. *IEEE Trans Commun*, 2020, 68: 602–614
- 77 Li Q, Hong M, Wai H T, et al. Transmit solutions for MIMO wiretap channels using alternating optimization. *IEEE J Sel Areas Commun*, 2013, 31: 1714–1727
- 78 Li R, Guo B, Tao M, et al. Joint design of hybrid beamforming and reflection coefficients in RIS-aided mmwave MIMO systems. *IEEE Trans Commun*, 2022, 70: 2404–2416
- 79 Li Y, Lin Q, Liu Y F, et al. Asynchronous activity detection for cell-free massive MIMO: From centralized to distributed algorithm. *IEEE Trans Wireless Commun*, 2022, doi: 10.1109/TWC.2022.3211967
- 80 Liao W C, Hong M, Liu Y F, et al. Base station activation and linear transceiver design for optimal resource management in heterogeneous networks. *IEEE Trans Signal Process*, 2014, 62: 3939–3952
- 81 Liu F, Liu Y F, Li A, et al. Cramér-Rao bound optimization for joint radar-communication beamforming. *IEEE Trans Signal Process*, 2022, 70: 240–253
- 82 Liu H, Yue M C, So A M C, et al. A discrete first-order method for large-scale MIMO detection with provable guarantees. In: *Proceedings of IEEE SPAWC'17*. Sapporo: IEEE, 2017, 669–673
- 83 Liu J, Shroff N B, Sherali H D. Optimal power allocation in multi-relay MIMO cooperative networks: Theory and algorithms. *IEEE J Sel Areas Commun*, 2012, 30: 331–340
- 84 Liu L, Larsson E G, Yu W, et al. Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things. *IEEE Signal Process Mag*, 2018, 35: 88–99
- 85 Liu L, Liu Y F. An efficient algorithm for device detection and channel estimation in asynchronous IoT systems. In: *Proceedings of IEEE ICASSP'21*. Toronto: IEEE, 2021, 4815–4819
- 86 Liu L, Liu Y F, Patil P, et al. Uplink-downlink duality between multiple-access and broadcast channels with compressing relays. *IEEE Trans Inform Theory*, 2021, 67: 7304–7337
- 87 Liu L, Yu W. Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation. *IEEE Trans Signal Process*, 2018, 66: 2933–2946
- 88 Liu Y F. Optimal resource allocation for wireless communications: Complexity analysis and algorithm design (in Chinese). *Sci Sin Math*, 2013, 43: 953–964 [刘亚锋. 无线通信中的最优资源分配—复杂性分析与算法设计. *中国科学: 数学*, 2013, 43: 953–964]
- 89 Liu Y F. Dynamic spectrum management: A complete complexity characterization. *IEEE Trans Inform Theory*,

- 2017, 63: 392–403
- 90 Liu Y F, Dai Y H, Luo Z Q. Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms. *IEEE Trans Signal Process*, 2011, 59: 1142–1157
- 91 Liu Y F, Dai Y H, Luo Z Q. Max-min fairness linear transceiver design for a multi-user MIMO interference channel. In: *Proceedings of IEEE ICC'11*. Kyoto: IEEE, 2011, 1–5
- 92 Liu Y F, Dai Y H, Luo Z Q. Joint power and admission control via linear programming deflation. *IEEE Trans Signal Process*, 2013, 61: 1327–1338
- 93 Liu Y F, Dai Y H, Luo Z Q. Max-min fairness linear transceiver design for a multi-user MIMO interference channel. *IEEE Trans Signal Process*, 2013, 61: 2413–2423
- 94 Liu Y F, Dai Y H, Ma S. Joint power and admission control: Non-convex L_q approximation and an effective polynomial time deflation approach. *IEEE Trans Signal Process*, 2015, 63: 3641–3656
- 95 Liu Y F, Hong M, Dai Y H. Max-min fairness linear transceiver design problem for a multi-user SIMO interference channel is polynomial time solvable. *IEEE Signal Process Lett*, 2013, 20: 27–30
- 96 Liu Y F, Ma S, Dai Y H, et al. A smoothing SQP framework for a class of composite L_q minimization over polyhedron. *Math Program*, 2016, 158: 467–500
- 97 Liu Y F, Yu W. Wireless multicast for cloud radio access network with heterogeneous backhaul. In: *Proceedings of ASILOMAR'17*. Pacific Grove: IEEE, 2018, 2576–2303
- 98 Lu C, Deng Z, Zhang W Q, et al. Argument division based branch-and-bound algorithm for unit-modulus constrained complex quadratic programming. *J Global Optim*, 2018, 70: 171–187
- 99 Lu C, Liu Y F. An efficient global algorithm for single-group multicast beamforming. *IEEE Trans Signal Process*, 2017, 65: 3761–3774
- 100 Lu C, Liu Y F, Zhang W Q, et al. Tightness of a new and enhanced semidefinite relaxation for MIMO detection. *SIAM J Optim*, 2019, 29: 719–742
- 101 Lu C, Liu Y F, Zhou J. An enhanced SDR based global algorithm for nonconvex complex quadratic programs with signal processing applications. *IEEE Open J Signal Process*, 2020, 1: 120–134
- 102 Luo Z Q. Applications of convex optimization in signal processing and digital communication. *Math Program Ser B*, 2003, 97: 177–207
- 103 Luo Z Q, Luo X, Kisiailiou M. An efficient quasi-maximum likelihood decoder for PSK signals. In: *Proceedings of IEEE ICASSP'03*. Hong Kong: IEEE, 2003, 561–564
- 104 Luo Z Q, Ma W K, So A M C, et al. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Process Mag*, 2010, 27: 20–34
- 105 Luo Z Q, Pang J S. Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines. *EURASIP J Adv Signal Process*, 2006, Article ID 24012
- 106 Luo Z Q, Sidiropoulos N D, Tseng P, et al. Approximation bounds for quadratic optimization with homogeneous quadratic constraints. *SIAM J Optim*, 2007, 18: 1–28
- 107 Luo Z Q, Zhang S. Dynamic spectrum management: Complexity and duality. *IEEE J Sel Top Signal Process*, 2008, 2: 57–73
- 108 Ma W K, Pan J, So A M C, et al. Unraveling the rank-one solution mystery of robust MISO downlink transmit optimization: A verifiable sufficient condition via a new duality result. *IEEE Trans Signal Process*, 2017, 65: 1909–1924
- 109 Ma X, Guo S, Zhang H, et al. Joint beamforming and reflecting design in reconfigurable intelligent surface-aided multi-user communication systems. *IEEE Trans Wireless Commun*, 2021, 20: 3269–3283
- 110 Mahdavi-Doost H, Ebrahimi M, Khandani A K. Characterization of SINR region for interfering links with constrained power. *IEEE Trans Inform Theory*, 2010, 56: 2816–2828
- 111 Marks B R, Wright G P. Technical note—A general inner approximation algorithm for nonconvex mathematical programs. *Oper Res*, 1978, 26: 681–683
- 112 Matskani E, Sidiropoulos N D, Luo Z Q, et al. Convex approximation techniques for joint multiuser downlink beamforming and admission control. *IEEE Trans Wireless Commun*, 2008, 7: 2682–2693
- 113 Matskani E, Sidiropoulos N D, Luo Z Q, et al. Efficient batch and adaptive approximation algorithms for joint multicast beamforming and admission control. *IEEE Trans Signal Process*, 2009, 57: 4882–4894
- 114 Mijumbi R, Serrat J, Gorricho J L, et al. Network function virtualization: State-of-the-art and research challenges. *IEEE Commun Surv Tutor*, 2016, 18: 236–262
- 115 Mirjalily G, Luo Z Q. Optimal network function virtualization and service function chaining: A survey. *Chin J Electron*, 2018, 27: 704–717

- 116 Mitliagkas I, Sidiropoulos N D, Swami A. Joint power and admission control for ad-hoc and cognitive underlay networks: Convex approximation and distributed implementation. *IEEE Trans Wireless Commun*, 2011, 10: 4110–4121
- 117 Ng C T K, Huang H. Linear precoding in cooperative MIMO cellular networks with limited coordination clusters. *IEEE J Sel Areas Commun*, 2010, 28: 1446–1454
- 118 Nguyen D, Tran L N, Pirinen P, et al. Precoding for full duplex multiuser MIMO systems: Spectral and energy efficiency maximization. *IEEE Trans Signal Process*, 2013, 61: 4038–4050
- 119 Nocedal J, Wright S J. *Numerical Optimization*, 2nd ed. New York: Springer, 2006
- 120 Pan J, Ma W K, Jaldén J. MIMO detection by Lagrangian dual maximum-likelihood relaxation: Reinterpreting regularized lattice decoding. *IEEE Trans Signal Process*, 2014, 62: 511–524
- 121 Papandriopoulos J, Evans J S. SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks. *IEEE Trans Inform Theory*, 2009, 55: 3711–3724
- 122 Peters S W, Heath R W. Interference alignment via alternating minimization. In: *Proceedings of IEEE ICASSP'09*. Taipei: IEEE, 2009, 2445–2448
- 123 Powell M J D. On search directions for minimization algorithms. *Math Program*, 1973, 4: 193–201
- 124 Promwongsa N, Abu-Lebdeh M, Kianpisheh S, et al. Ensuring reliability and low cost when using a parallel VNF processing approach to embed delay-constrained slices. *IEEE Trans Netw Serv Manage*, 2020, 17: 2226–2241
- 125 Rashid-Farrokhi F, Liu K J R, Tassiulas L. Transmit beamforming and power control for cellular wireless systems. *IEEE J Sel Areas Commun*, 1998, 16: 1437–1450
- 126 Saad Y. *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia: SIAM, 2003
- 127 Sanjabi M, Razaviyayn M, Luo Z Q. Optimal joint base station assignment and beamforming for heterogeneous networks. *IEEE Trans Signal Process*, 2014, 62: 1950–1961
- 128 Schad A, Law K L, Pesavento M. A convex inner approximation technique for rank-two beamforming in multicasting relay networks. In: *Proceedings of EUSIPO'12*. Bucharest: IEEE, 2012, 1369–1373
- 129 Schaible S. Parameter-free convex equivalent and dual programs of fractional programming problems. *Z Oper Res*, 1974, 18: 187–196
- 130 Schmidt D A, Shi C X, Berry R A, et al. Minimum mean squared error interference alignment. In: *Proceedings of ACSSC'09*. Pacific Grove: IEEE, 2009, 1106–1110
- 131 Schubert M, Boche H. Solution of the multiuser downlink beamforming problem with individual SINR constraints. *IEEE Trans Veh Technol*, 2004, 53: 18–28
- 132 Senel K, Larsson E G. Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach. *IEEE Trans Commun*, 2018, 66: 6164–6175
- 133 Shannon C E. A mathematical theory of communication. *Bell Syst Tech J*, 1948, 27: 379–423
- 134 Shao X, Chen X, Ng D W K, et al. Cooperative activity detection: Sourced and unsourced massive random access paradigms. *IEEE Trans Signal Process*, 2020, 68: 6578–6593
- 135 Shen C, Chang T H, Wang K Y, et al. Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach. *IEEE Trans Signal Process*, 2012, 60: 2988–3003
- 136 Shen K, Liu Y F, Ding D Y, et al. Flexible multiple base station association and activation for downlink heterogeneous networks. *IEEE Signal Process Lett*, 2017, 24: 1498–1502
- 137 Shen K, Yu W. Fractional programming for communication systems—Part I: Power control and beamforming. *IEEE Trans Signal Process*, 2018, 66: 2616–2630
- 138 Shen K, Yu W. Fractional programming for communication systems—Part II: Uplink scheduling via matching. *IEEE Trans Signal Process*, 2018, 66: 2631–2644
- 139 Shen K, Yu W, Zhao L C, et al. Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization-maximization approach. *IEEE ACM Trans Netw*, 2019, 27: 2164–2177
- 140 Shi Q, Razaviyayn M, Hong M, et al. SINR constrained beamforming for a MIMO multi-user downlink system: Algorithms and convergence analysis. *IEEE Trans Signal Process*, 2016, 64: 2920–2933
- 141 Shi Q, Razaviyayn M, Luo Z Q, et al. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Trans Signal Process*, 2011, 59: 4331–4340
- 142 Shi Y, Cheng J, Zhang J, et al. Smoothed L_p -minimization for green cloud-RAN with user admission control. *IEEE J Sel Areas Commun*, 2016, 34: 1022–1036
- 143 Shi Y, Zhang J, Letaief K B. Group sparse beamforming for green cloud-RAN. *IEEE Trans Wireless Commun*, 2014, 13: 2809–2823
- 144 Sidiropoulos N D, Davidson T N, Luo Z Q. Transmit beamforming for physical-layer multicasting. *IEEE Trans Signal*

- Process, 2006, 54: 2239–2251
- 145 So A M C. Probabilistic analysis of the semidefinite relaxation detector in digital communications. In: Proceedings of ACM-SIAM SODA'10. Austin: SIAM, 2010, 698–711
- 146 Song E, Shi Q, Sanjabi M, et al. Robust SINR-constrained MISO downlink beamforming: When is semidefinite programming relaxation tight? EURASIP J Wireless Commun Netw, 2012, 243: 1–11
- 147 Stancu-Minasian I M. Fractional Programming: Theory, Methods and Applications. Mathematics and Its Applications, vol. 409. Dordrecht: Springer, 2012
- 148 Sun C, Jorswieck E A, Yuan Y X. Sum rate maximization for non-regenerative MIMO relay networks. IEEE Trans Signal Process, 2016, 64: 6392–6405
- 149 Tao M, Chen E, Zhou H, et al. Content-centric sparse multicast beamforming for cache-enabled cloud RAN. IEEE Trans Wireless Commun, 2016, 15: 6118–6131
- 150 Tran L N, Hanif M F, Juntti M. A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays. IEEE Signal Process Lett, 2014, 21: 114–117
- 151 Tse D, Viswanath P. Fundamental of Wireless Communications. Cambridge: Cambridge University Press, 2005
- 152 Vandenberghe L, Boyd S. Semidefinite programming. SIAM Rev, 1995, 38: 49–95
- 153 Vassilaras S, Gkatzikis L, Liakopoulos N, et al. The algorithmic aspects of network slicing. IEEE Commun Mag, 2017, 55: 112–119
- 154 Venturino L, Zappone A, Risi C, et al. Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination. IEEE Trans Wireless Commun, 2015, 14: 1–14
- 155 Visotsky E, Madhow U. Optimum beamforming using transmit antenna arrays. In: Proceedings of IEEE VTC'99. Houston: IEEE, 2002, 851–856
- 156 Viterbo E, Bours J. A universal lattice code decoder for fading channels. IEEE Trans Inform Theory, 1999, 45: 1639–1642
- 157 Wai H T, Ma W K. A decentralized method for joint admission control and beamforming in coordinated multicell downlink. In: Proceedings of ASIOMAR'12. Pacific Grove: IEEE, 2012, 559–563
- 158 Wang Z, Chen Z, Liu Y F, et al. An efficient active set algorithm for covariance based joint data and activity detection for massive random access with massive MIMO. In: Proceedings of IEEE ICASSP'21. Toronto: IEEE, 2021, 4840–4844
- 159 Wang Z, Liu Y F, Chen Z, et al. Accelerating coordinate descent via active set selection for device activity detection for multi-cell massive random access. In: Proceedings of IEEE SPAWC'21. Lucca: IEEE, 2021, 366–370
- 160 Wang Z, Liu Y F, Liu L. Covariance-based joint device activity and delay detection in asynchronous mMTC. IEEE Signal Process Lett, 2022, 29: 538–542
- 161 Wang Z, Liu Y F, Wang Z, et al. Scaling law analysis for covariance based activity detection in cooperative multi-cell massive MIMO. arXiv:2210.14688, 2022
- 162 Weeraddana P C, Codreanu M, Latva-aho M, et al. Weighted sum-rate maximization in wireless networks: A review. Found Trends Netw, 2012, 6: 1–163
- 163 Wen X, Law K L, Alabed S J, et al. Rank-two beamforming for single-group multicasting networks using OSTBC. In: Proceedings of IEEE SAM'12. Hoboken: IEEE, 2012, 69–72
- 164 Wiesel A, Eldar Y C, Shamai S. Linear precoding via conic optimization for fixed MIMO receivers. IEEE Trans Signal Process, 2006, 54: 161–176
- 165 Wipf D P, Rao B D. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Trans Signal Process, 2007, 55: 3704–3716
- 166 Woldeyohannes Y T, Mohammadkhan A, Ramakrishnan K K, et al. ClusPR: Balancing multiple objectives at scale for NFV resource allocation. IEEE Trans Netw Serv Manage, 2018, 15: 1307–1321
- 167 Wright S J. Coordinate descent algorithms. Math Program, 2015, 151: 3–34
- 168 Wu Q, Zhang R. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming. IEEE Trans Wireless Commun, 2019, 18: 5394–5409
- 169 Wu S X, Ma W K, So A M C. Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding. IEEE Trans Signal Process, 2013, 61: 4230–4245
- 170 Wu S X, So A M C, Ma W K. Rank-two transmit beamformed Alamouti space-time coding for physical-layer multicasting. In: Proceedings of IEEE ICASSP'12. Kyoto: IEEE, 2012, 2793–2796
- 171 Wu Z, Jiang B, Liu Y F, et al. A novel negative ℓ_1 penalty approach for multiuser one-bit massive MIMO downlink with PSK signaling. In: Proceedings of IEEE ICASSP'22. Singapore: IEEE, 2022, 5323–5327
- 172 Wu Z, Liu Y F, Jiang B, et al. Efficient quantized constant envelope precoding for multiuser downlink massive MIMO

- systems. arXiv:2210.14534, 2022
- 173 Xie H, Xu J, Liu Y F, et al. User grouping and reflective beamforming for IRS-aided URLLC. *IEEE Wireless Commun Lett*, 2021, 10: 2533–2537
 - 174 Xu W W, Zhang G, Bai B, et al. Ten key ICT challenges in the post-Shannon era (in Chinese). *Sci Sin Math*, 2021, 51: 1095–1138 [徐文伟, 张弓, 白铂, 等. 后香农时代 ICT 领域的十大挑战问题. *中国科学: 数学*, 2021, 51: 1095–1138]
 - 175 Yang S, Hanzo L. Fifty years of MIMO detection: The road to large-scale MIMOs. *IEEE Commun Surv Tutor*, 2015, 17: 1941–1988
 - 176 Yang Z, Li J, Stoica P, et al. Sparse methods for direction-of-arrival estimation. In: *Academic Press Library in Signal Processing*. Amsterdam: Elsevier, 2018, 7: 509–581
 - 177 Ye Y. On the complexity of approximating a KKT point of quadratic programming. *Math Program*, 1998, 80: 195–211
 - 178 You X H, Pan Z W, Cao S M, et al. The 5G mobile communication: The development trends and its emerging key techniques (in Chinese). *Sci Sin Inform*, 2014, 44: 551–563 [尤肖虎, 潘志文, 高西奇, 等. 5G 移动通信发展趋势与若干关键技术. *中国科学: 信息科学*, 2014, 44: 551–563]
 - 179 Yu W, Ginis G, Cioffi J M. Distributed multiuser power control for digital subscriber lines. *IEEE J Sel Areas Commun*, 2002, 20: 1105–1115
 - 180 Yu W, Kwon T, Shin C. Adaptive resource allocation in cooperative cellular networks. In: *Cooperative Cellular Wireless Networks*, 2010, 233–256
 - 181 Yu W, Lan T. Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Trans Signal Process*, 2007, 55: 2646–2660
 - 182 Yu W, Lui R. Dual methods for nonconvex spectrum optimization of multicarrier systems. *IEEE Trans Commun*, 2006, 54: 1310–1322
 - 183 Yu W, Patil P, Dai B, et al. Cooperative beamforming and resource optimization in C-RAN. In: *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge: Cambridge University Press, 2017, 1–30
 - 184 Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc Ser B Stat Methodol*, 2005, 68: 49–67
 - 185 Yuan Y X, Sun W Y. *Optimization Theory and Methods: Nonlinear Programming* (in Chinese). Beijing: Science Press, 1997 [袁亚湘, 孙文瑜. 最优化理论与方法. 北京: 科学出版社, 1997]
 - 186 Zander J. Performance of optimum transmitter power control in cellular radio systems. *IEEE Trans Veh Technol*, 1992, 41: 57–62
 - 187 Zappone A, Jorswieck E A. Energy efficiency in wireless networks via fractional programming theory. *FNT Commun Inf Theor*, 2015, 11: 185–396
 - 188 Zappone A, Jorswieck E A, Buzzi S. Energy efficiency and interference neutralization in two-hop MIMO interference channels. *IEEE Trans Signal Process*, 2014, 62: 6481–6495
 - 189 Zhang J, Chen R H, Andrews J G, et al. Networked MIMO with clustered linear precoding. *IEEE Trans Wireless Commun*, 2009, 8: 1910–1921
 - 190 Zhang N, Liu Y F, Farmanbar H, et al. Network slicing for service-oriented networks under resource constraints. *IEEE J Sel Areas Commun*, 2017, 35: 2512–2521
 - 191 Zhang N, Liu Y F, Farmanbar H, et al. System and method for network slicing for service-oriented networks. <https://patents.glgoo.top/patent/US20190394132A1/en>, 2021
 - 192 Zhang R, Liang Y C, Chai C C, et al. Optimal beamforming for two-way multi-antenna relay channel with analogue network coding. *IEEE J Sel Areas Commun*, 2009, 27: 699–712
 - 193 Zhang Y, Beheshti N, Beliveau L, et al. StEERING: A software-defined networking for inline service chaining. In: *Proceedings of IEEE ICNP'13*. Goettingen: IEEE, 2014, 1–10
 - 194 Zhao P F, Li Q N, Chen W K, et al. An efficient quadratic programming relaxation based algorithm for large-scale MIMO detection. *SIAM J Optim*, 2021, 31: 1519–1545
 - 195 Zhao Y B. *Sparse Optimization Theory and Methods*. Boca Raton: CRC Press, 2018
 - 196 Zou Y, Zhu J, Zhang R. Exploiting network cooperation in green wireless communication. *IEEE Trans Commun*, 2013, 61: 999–1010

Optimization problems and methods for resource allocation in wireless communications

Ya-Feng Liu, Zheyu Wu, Wei-Kun Chen & Yu-Hong Dai

Abstract Many problems arising from wireless communication system design can be formulated into optimization problems. On the one hand, these optimization problems are often highly nonlinear, and thus generally difficult to solve. On the other hand, they have their own special structures, such as hidden convexity and separability. Designing efficient optimization algorithms to solve these problems based on their special structure has been a hot research topic in recent years. In this paper, we focus on optimization methods for resource allocation problems in wireless communication system design. Taking optimization theories and algorithms as the main line, we introduce their applications in solving resource allocation problems, including how conic programming reveals the hidden convexity in specific non-convex problems, how Lagrangian duality helps to characterize the structure of their optimal solutions, how sparse optimization and integer programming techniques help to formulate the related problems, and how semidefinite relaxation, alternating optimization, and fractional programming help to design efficient algorithms. Finally, we give a prospect of some future research directions and the key problems in wireless communication system design.

Keywords semidefinite relaxation, alternating optimization, duality theory, multiuser interference channel, fractional programming, wireless communication system design, sparse optimization, hidden convexity, integer programming

MSC(2020) 90B18, 90C90, 90C25, 90C26

doi: 10.1360/SSM-2021-0188