

无线通信系统设计中的两个优化问题和相关优化方法^{*}

刘亚峰^{1,†}

摘要 无线通信系统设计中的许多问题可建模为优化问题。一方面，这些优化问题常常具有高度的非线性性，一般情况下难于求解；另一方面，它们又有自身的特殊结构，例如隐含的凸性、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。本文重点讨论无线通信系统设计中的两个优化问题和相关优化方法，包括多用户干扰信道最大最小准则下的联合传输/接收波束成形设计和多输入多输出（Multi-Input Multi-Output, MIMO）检测问题，主要介绍现代优化技术结合问题的特殊结构在求解和处理上述两个问题的最新进展。

关键词 半正定松弛，变量交替最优化，多用户干扰信道，联合传输/接收波束成形设计，计算复杂性，紧松弛，MIMO 检测，无线通信系统

中图分类号 O221, TN92

2010 数学分类号 65K05, 90B18, 90C22, 90C30, 90C90

Two optimization problems in wireless communication system design and related optimization methods^{*}

LIU Yafeng^{1,†}

Abstract Many problems arising from wireless communication system design can be formulated as optimization problems. On the one hand, these optimization problems are often non-convex and highly nonlinear and thus are difficult to solve; on the other hand, these problems have their own special structures such as (hidden) convexity and separability. Recently applying mathematical optimization methods to solve/deal with these problems while judiciously taking care of their special structures is a hot research topic. This (survey) paper aims to introduce two optimization problems in wireless communication system design, max-min fairness linear transceiver design problem and MIMO detection problem, and related optimization methods. This paper will focus on the above two problems and overview recent advances of applying mathematical optimization techniques to solve/deal with them by exploiting their special structures.

Keywords semi-definite relaxation, alternating optimization, multiuser interference channel, transceiver beamforming design, computational complexity, tight relaxation, MIMO detection, wireless communication system

收稿日期: 2019-03-01

*基金项目: 国家自然科学基金 (Nos. 11688101, 11671419, 11631013, 11571221), 北京市自然基金重点项目专项 (No. L172020), 国家发改委科学大数据公共服务平台与创新应用示范项目 (No. 2016-999999-65-01-000696-01)

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, 科学与工程计算国家重点实验室, 北京 100190; State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

† 通信作者 E-mail: yafliu@lsec.cc.ac.cn

Chinese Library Classification O221, TN92

2010 Mathematics Subject Classification 65K05, 90B18, 90C22, 90C30, 90C90

0 引言

最优化^[1-3]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、人工智能等许多领域有着广泛的应用. 无线通信技术^[4-6]主要利用无线电波来传输信息. 现代意义上的无线通信始于 19 世纪 90 年代意大利人马可尼 (Guglielmo Marconi) 的一系列电磁波实验, 他成功地利用电磁波实现了无线数据传输. 1948 年, 香农 (Claude Elwood Shannon) 发表了题为“*A Mathematical Theory of Communication*”的论文, 建立了通信的数学理论^[7]. 几十年来, 人们不断经历着新的无线通信方法的产生和变革, 并享受着多种多样的无线通信服务. 无线通信技术的发展极大地改善了人们的出行、通信和文化传播等方式, 为人们提供了即时便捷的服务. 无线通信技术的发展也是当前国家战略发展计划的重要组成部分. 本文重点讨论无线通信系统设计中的优化问题和优化方法, 主要介绍现代优化技术结合问题的特殊结构在无线通信系统性能优化的两个案例及其最新进展.

无线通信系统设计中的许多问题归根结底都是优化问题, 可运用优化方法加以求解, 比如通信网络的拓扑结构设计、信号的编解码过程、最优资源配置等. 目前, 相对于通信技术在现实生活中的蓬勃发展, 通信系统优化的数学理论与方法显得相对滞后, 在某些方面已经成为影响其发展和应用的关键因素. 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非凸非线性约束优化问题. 一方面, 这些优化问题常常具有高度的非线性性, 一般情况下难于求解; 另一方面, 它们又有自身的特殊结构, 例如隐含的凸性、可分性、稀疏性等. 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, 近二十年平均每年都有这方面的工作获 IEEE 信号处理学会或者 IEEE 通信学会最佳论文奖. 世界著名优化专家、加拿大皇家科学院院士、信号处理和无线通信顶级期刊 IEEE Transactions on Signal Processing 前任主编罗智泉教授的主要研究方向是优化算法的设计、分析及其在无线通信和信号处理中的应用. 罗智泉教授在上述方向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, 2015 年 IEEE Signal Processing Magazine 最佳论文奖, 2011 年欧洲信号处理学会最佳论文奖, 和 2011 年国际通信大会最佳论文奖. 优化方法在无线通信系统设计中的重要性由此可见一斑.

本文着重介绍现代优化技术结合问题的特殊结构在无线通信系统性能优化的两个案例及其最新进展. 我们将在第 1 节介绍多输入多输出 (Multi-Input Multi-Output, MIMO) 干扰信道中最大最小准则下的联合传输/接收波束成形设计问题及其最新研究进展; 在第 2 节介绍 MIMO 检测 (MIMO Detection) 问题及其最新研究进展. 本文侧重于从优化的角度审视无线通信系统设计, 有如下两个侧重点. 本文的第一个侧重点是关于无线通信系统设计中优化问题计算复杂性的刻画^[8]. 计算复杂性可以刻画所考虑问题的本质难易程度, 是解决和分析实际问题的关键步骤. 本文的第二个侧重点是根据问题的特殊结构设计有效的方法处理和求解相应的无线通信系统优化设计问题, 并对相关方法进行理论分析等. 关于无线通信系统优化设计的综述文章, 可参见文献 [9-16].

符号约定: 本文用小写黑体字母表示列向量, 大写黑体字母表示矩阵. 给定矩阵 \mathbf{A} ,

\mathbf{A}^\top 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, $A_{i,j}$ 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 $\mathbf{x} = [x_1, x_2, \dots, x_K]^\top$, $\|\mathbf{x}\|_p \triangleq \sum_{k=1}^K |x_k|^p$ 表示它的 p 范数, 其中 $p \in \{2, \infty\}$; $\text{Diag}(\mathbf{x})$ 表示由向量 \mathbf{x} 生成的对角矩阵. 给定(适当维数的) Hermitian 矩阵 \mathbf{A} 和 \mathbf{B} , $\mathbf{A} \succeq \mathbf{0}$ 表示 \mathbf{A} 是一个半正定矩阵; $\mathbf{A} \bullet \mathbf{B}$ 表示 \mathbf{AB} 的迹, 即 $\sum_i \sum_j A_{i,j} B_{j,i}$; $\mathbf{A} \otimes \mathbf{B}$ 表示它们的 Kronecker 乘积. 最后, 本文用 \mathbf{e} 表示适当维数的全 1 向量, 用 \mathbf{I} 表示适当维数的单位矩阵, 用 $\mathbf{0}$ 表示适当维数的全 0 矩阵或向量.

1 最大最小准则下联合传输/接收波束成形设计

1.1 优化模型

多用户干扰信道如图 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.

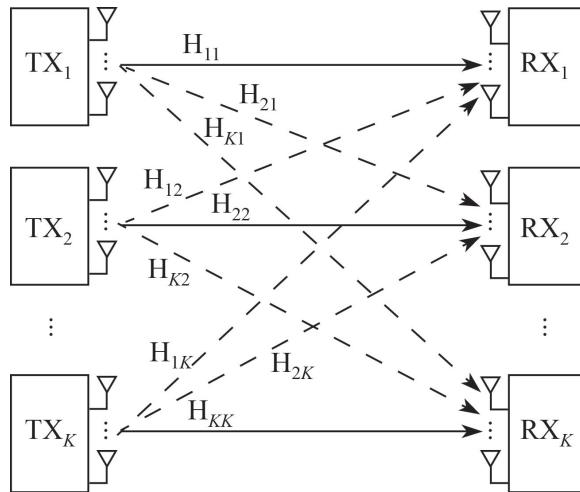


图1 多用户干扰信道示意图

假设系统中有 K 个用户, 我们用 $\mathcal{K} = \{1, 2, \dots, K\}$ 表示系统中所有用户的集合. 考虑单载波通信系统, 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 . 我们用 $\mathbf{H}_{kj} \in \mathbb{C}^{M_k \times N_j}$ 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_k 分别表示第 k 个接收端和第 k 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, $\mathbf{z}_k \in \mathbb{C}^{M_k \times 1}$ 是加性高斯白噪声, 服从正态分布 $\mathcal{CN}(\mathbf{0}, \sigma_k^2 \mathbf{I})$. 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 (Signal-to-Interference-plus-Noise-Ratio, SINR) 可表示为

$$\text{SINR}_k = \frac{|\mathbf{u}_k^\dagger \mathbf{H}_{kk} \mathbf{v}_k|^2}{\sigma_k^2 \|\mathbf{u}_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^\dagger \mathbf{H}_{kj} \mathbf{v}_j|^2}, \quad k \in \mathcal{K}. \quad (1.1)$$

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: 当系统中每个传输端和接收端都装有多根天线时, 即 $M_k \geq 2, N_k \geq 2 (k \in \mathcal{K})$ 时, 此信道称为 MIMO 干扰信道, 其 SINR 表达式如式 (1.1) 所示; 当系统中每个传输端装有多根天线, 每个接收端装有 1 根天线, 即 $M_k = 1, N_k \geq 2 (k \in \mathcal{K})$ 时, 此信道称为多输入单输出 (Multi-Input Single-Output, MISO) 干扰信道.

在 MIMO 和 MISO 干扰信道中, 每个传输端的功率约束可表示为 $\|\mathbf{v}_k\|^2 \leq \bar{p}_k, k \in \mathcal{K}$, 其中 \bar{p}_k 表示第 k 个传输端的功率上限. 当系统中每个接收端装有多根天线, 每个传输端装有 1 根天线时, 即 $M_k \geq 2, N_k = 1 (k \in \mathcal{K})$ 时, 此信道为单输入多输出 (Single-Input Multi-Output, SIMO) 干扰信道.

为了保证用户之间的公平性, 我们考虑 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:

$$\begin{aligned} \max_{\{\mathbf{u}, \mathbf{v}\}} \quad & G(\mathbf{u}, \mathbf{v}) \triangleq \min_{k \in \mathcal{K}} \{\text{SINR}_k(\mathbf{u}_k, \mathbf{v})\} \\ \text{s.t.} \quad & \|\mathbf{u}_k\|^2 = 1, \|\mathbf{v}_k\|^2 \leq \bar{p}_k, k \in \mathcal{K}, \end{aligned} \quad (1.2)$$

其中 $\text{SINR}_k(\mathbf{u}_k, \mathbf{v})$ 由式 (1.1) 给出. 注意到第 k 个用户的 SINR 表达式 (1.1) 不依赖于 $\{\mathbf{u}_j\}_{j \neq k}$, 所以我们用 $\text{SINR}_k(\mathbf{u}_k, \mathbf{v})$ 表示第 k 个用户的 SINR. 引入辅助变量 $\text{SINR} = \min_{k \in \mathcal{K}} \{\text{SINR}_k\}$, 上述最小 SINR 最大化问题 (1.2) 可以等价地写为

$$\begin{aligned} \max_{\{\mathbf{u}, \mathbf{v}\}} \quad & \text{SINR} \\ \text{s.t.} \quad & \text{SINR} \leq \text{SINR}_k, \|\mathbf{u}_k\|^2 = 1, \|\mathbf{v}_k\|^2 \leq \bar{p}_k, k \in \mathcal{K}. \end{aligned} \quad (1.3)$$

1.2 最新进展

当每个接收端装有 1 根天线 ($M_k = 1, k \in \mathcal{K}$) 时, 即 MISO 干扰信道: 文献 [17] 提出通过最小化逆信干比的加权和来逼近问题 (1.2), 并证明可以通过选取适当的加权系数获得问题 (1.2) 的最优解, 此工作获得了 2007 年 IEEE 信号处理学会最佳论文奖; 文献 [18,19] 提出多项式时间算法可求得问题 (1.2) 的全局最优解.

当每个传输端装有 1 根天线 ($N_k = 1, k \in \mathcal{K}$) 时, 即 SIMO 干扰信道, 我们首次得到如下多项式时间可解的结果, 这一结果主要是由于问题中隐含的凸性^[20]. 由于其证明是构造性的, 所以同时也给出了问题 (1.2) 当传输端有 1 根天线时的多项式时间算法; 详见文献 [21].

定理 1.1 当每个传输端装有 1 根天线 ($N_k = 1, k \in \mathcal{K}$) 时, 最小 SINR 最大化问题 (1.2) 可在多项式时间内求解.

同时, 我们也首次得到了传输端和接收端都装有多根天线时问题 (1.2) 的复杂性结果^[23]: 即当每个传输端装有的天线数大于或等于 2(3), 每个接收端装有的天线数大于或等于 3(2) 时, 判断给定 SINR 目标的可行性问题是强 NP-难的. 关于传输端和接收端装有的天线数都等于 2 时问题 (1.2) 的复杂性, 这一问题一直悬而未决. 最近, 我们回答了这一问题; 详见文献^[22].

定理 1.2 给定问题 (1.3) 中 SINR 目标 ζ . 当每个传输端/接收端装有的天线数大于或等于 2 时, 判断给定 ζ 的可行性问题是强 NP-难的. 因此, 最优化问题 (1.2) 也是强 NP-难的.

表 1 中总结了 MIMO 干扰信道最小 SINR 最大化问题 (1.2) 在传输端和接收端装有不同天线数时的计算复杂性. 从表 1 中我们可知, 问题 (1.2) 的复杂性随着传输端和接收端装有的天线数目的变化而变化. 一般来讲, 增加天线数目可以提高通信系统的性能, 但也给相应优化问题的求解带来了困难.

表 1 最大最小准则下联合传输/接收波束成形设计问题 (1.2) 的计算复杂性

| 接收端 \ 传输端 | | |
|-----------|-----------------------|--------------------------|
| | $N_k = 1$ | $N_k \geq 2$ |
| $M_k = 1$ | 多项式可解 | 多项式可解 ^[18,19] |
| | 多项式可解 ^[21] | 强 NP-难 ^[22] |

我们也提出了两个变量交替最优化方法求解最小 SINR 最大化问题 (1.2): 第一个算法是精确循环坐标上升算法 (Exact Cyclic Coordinate Ascent Algorithm, ECCAA), 每一步迭代精确求解子问题; 第二个算法是非精确循环坐标上升算法 (Inexact Cyclic Coordinate Ascent Algorithm, ICCAA), 其中有些子问题非精确求解. 变量交替最优化算法的基本思想是首先适当地将问题的变量分块, 然后 (精确或者非精确) 交替地在各个块之间进行优化/迭代, 每次优化/迭代只更新其中一块变量, 其他块变量固定不变.

特别地, 对于问题(1.2), 将其变量分为两块 $\mathbf{u}^n = (\mathbf{u}_1^n, \dots, \mathbf{u}_K^n)$ 和 $\mathbf{v}^n = (\mathbf{v}_1^n, \dots, \mathbf{v}_K^n)$, 其中 $n \geq 0$ 表示迭代指标. 固定 \mathbf{v}^n , 因为 SINR_k 仅与 \mathbf{u}_k 相关, 所以可以通过求解 K 个小问题

$$\begin{aligned} \max_{\{\mathbf{u}_k\}} \quad & \frac{|\mathbf{u}_k^\dagger \mathbf{H}_{kk} \mathbf{v}_k^n|^2}{\sigma_k^2 \|\mathbf{u}_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^\dagger \mathbf{H}_{kj} \mathbf{v}_j^n|^2} \\ \text{s.t.} \quad & \|\mathbf{u}_k\|^2 = 1 \end{aligned} \quad (1.4)$$

得到问题 (1.2) 当 \mathbf{v}^n 固定时的解. 定义

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

则问题 (1.4) 的最优解 \mathbf{u}_k^n 为线性最小均方误差 (Linear Minimum Mean Square Error, LMMSE) 接收波束成形向量

$$\mathbf{u}_k^n = \tilde{\mathbf{u}}_k^n / \|\tilde{\mathbf{u}}_k^n\|, \quad \tilde{\mathbf{u}}_k = \mathbf{M}_k(\mathbf{v}^n) \mathbf{H}_{kk} \mathbf{v}_k^n. \quad (1.5)$$

固定 \mathbf{u}^n , 问题 (1.2) 关于变量 \mathbf{v} 变为如下 MISO 干扰信道中的优化问题:

$$\begin{aligned} \max_{\{\mathbf{v}\}} \quad & \min_{k \in \mathcal{K}} \left\{ \frac{|(\mathbf{u}_k^n)^\dagger \mathbf{H}_{kk} \mathbf{v}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |(\mathbf{u}_k^n)^\dagger \mathbf{H}_{kj} \mathbf{v}_j|^2} \right\} \\ \text{s.t.} \quad & \|\mathbf{v}_k\|^2 \leq \bar{p}_k, \quad k \in \mathcal{K}. \end{aligned} \quad (1.6)$$

由表 1 可知, 问题 (1.6) 可以在多项式时间内求得全局最优解.

算法 1.1 ECCAA

步 1 初始化: 给定初始点 \mathbf{v}^0 和停止准则 ε . 令 $n = 0$.

步 2 计算 \mathbf{u}^n : 固定 \mathbf{v}^n , 计算最优的 LMMSE 接收波束成形向量 \mathbf{u}^n (见式 (1.5)).

步 3 计算 \mathbf{v}^{n+1} : 固定 \mathbf{u}^n , 求解问题 (1.6) 得到最优的传输波束成形 \mathbf{v}^{n+1} .

步 4 终止检验: 当

$$\frac{G(\mathbf{u}^n, \mathbf{v}^{n+1}) - G(\mathbf{u}^n, \mathbf{v}^{n-1})}{\max \{G(\mathbf{u}^n, \mathbf{v}^{n-1}), 1\}} \leq \varepsilon \quad (1.7)$$

满足时, 终止算法; 否则, 置 $n := n + 1$, 转步 2.

一般情况下, 即使每个子问题都精确求解, 仍然不能保证变量交替算法的收敛性^[23]. 目标函数非光滑部分以及约束中变量的可分性对于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, 变量交替算法产生的点列可能收敛到一个无意义的点. 尽管问题 (1.2) 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, 我们利用变量 \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 ECCAA 产生的点列收敛到问题 (1.2) 的一个 KKT (Karush-Kuhn-Tucker) 点^[24].

定理 1.3 令 ECCAA 算法中 $\varepsilon = 0$. ECCAA 产生的点列 $\{(\mathbf{u}^n, \mathbf{v}^n)\}$ 或终止于问题 (1.2) 的一个 KKT 点, 或它的任意聚点 $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ 是问题 (1.2) 的 KKT 点.

为了进一步提高 ECCAA 算法的计算有效性, 我们在文 [25] 中提出了一个非精确的变量交替最优化方法求解问题 (1.2), 其每次迭代非精确地求解关于变量 \mathbf{v} 的优化问题. 具体地, 文 [25] 提出通过求解子问题

$$\begin{aligned} \max_{\{\mathbf{v}, \theta\}} \quad & \theta \\ \text{s.t.} \quad & \frac{(\mathbf{u}_k^n)^\dagger \mathbf{H}_{kk} \mathbf{v}_k - \theta}{\sqrt{\sigma_k^2 + \sum_{j \neq k} |(\mathbf{u}_k^n)^\dagger \mathbf{H}_{kj} \mathbf{v}_j|^2}} \geq \sqrt{G_{2n}}, \quad k \in \mathcal{K}, \\ & \|\mathbf{v}_k\|^2 \leq \bar{p}_k, \quad k \in \mathcal{K}, \end{aligned} \quad (1.8)$$

更新波束成形向量 \mathbf{v}^{n+1} , 其中 G_{2n} 是问题 (1.2) 在点 $(\mathbf{u}^n, \mathbf{v}^n)$ 处的函数值. 注意到问题 (1.8) 是一个二阶锥规划 (Second Order Cone Program, SOCP) 问题, 所以可在多项式时间内求得其全局最优解. 为了更新传输波束成形向量 \mathbf{v}^{n+1} , ECCAA 需要求解一系列 SOCP 可行性问题, 而 ICCAA 只需要求解一个 SOCP 问题, 所以极大地降低了每次更新变量 \mathbf{v}^{n+1} 的复杂性. 与 ECCAA 算法类似, 我们可证明 ICCAA 的全局收敛性^[25].

关于 MIMO 干扰信道联合传输/接收波束成形设计和功率控制进一步的工作, 包括更加复杂的场景 (例如 MIMO 广播信道、异构网络、全双工系统) 和设计 (例如基站自适应合作、基站激活、基站-用户关联) 等, 请参见 [26-39]; 关于变量交替最优化方法求解无线通信系统设计问题的更多工作, 请参见 [34,37,38,40-43]; 关于交替最优化方法的综述文章及其在各个领域中的应用, 请参见 [3,44,45].

2 MIMO 检测

2.1 优化模型

MIMO 检测问题是现代数字通信中的基本问题之一^[46,47]. 数学上, MIMO 信道的输入输出关系可表示为

$$\mathbf{r} = \mathbf{H}\mathbf{x}^* + \mathbf{v}, \quad (2.1)$$

其中 $\mathbf{r} \in \mathbb{C}^m$ 为接收端接收到的向量; $\mathbf{H} \in \mathbb{C}^{m \times n}$ 为复的信道矩阵; $\mathbf{x}^* \in \mathbb{C}^n$ 表示传输的真实信号向量; $\mathbf{v} \in \mathbb{C}^m$ 为加性噪声. 如果使用 M -Phase-Shift Keying (M -PSK) 的调制方式, 那么 \mathbf{x}^* 的每一个分量 x_i^* 应属于一个有限集合, 即

$$\mathbf{x}_i^* \in \left\{ \exp(i\theta) \mid \theta = \frac{2j\pi}{M}, j = 0, 1, \dots, M-1 \right\}, \quad i = 1, 2, \dots, n,$$

其中 i 为虚部单位满足 $i^2 = -1$. 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

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{C}^n} \quad & \|\mathbf{H}\mathbf{x} - \mathbf{r}\|_2^2 \\ \text{s.t.} \quad & |\mathbf{x}_i|^2 = 1, \arg(\mathbf{x}_i) \in \mathcal{A}, \quad i = 1, 2, \dots, n, \end{aligned} \quad (2.2)$$

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

当式 (2.1) 中的噪声 \mathbf{v} 服从高斯分布时, 上述极小化模型 (2.2) 等价于最大似然估计模型. 容易验证, MIMO 检测问题 (2.2) 的可行点个数为 M^n . 实际上, MIMO 检测问题 (2.2) 是强 NP-难的^[48], 所以不存在多项式时间的算法能求解到其全局最优解 (除非 P=NP).

2.2 最新进展

基于半正定松弛 (SemiDefinite Relaxation, SDR) 的方法是求解问题 (2.2) 非常有效的方法之一. 基于 SDR 的方法不但具有多项式时间的最坏复杂度, 而且在实际中也展现出很好的检测误码率. 基于 SDR 的方法的基本思想是将问题 (2.2) 先松弛为一个 (复或者实的) SDR, 然后通过一些方法 (随机化的方法或者启发式的方法) 得到原问题的一个可行解.

本章中我们着重介绍关于 MIMO 检测问题 (2.2) SDR 的最新进展. 为了叙述方便, 我们引入如下的记号: 令

$$\mathbf{Q} = \mathbf{H}^\dagger \mathbf{H}, \mathbf{c} = -\mathbf{H}^\dagger \mathbf{r};$$

令 $\mathbf{s} = [s_1, s_2, \dots, s_M]^\top \in \mathbb{C}^M$, 其中

$$s_j = \cos\left(\frac{2(j-1)\pi}{M}\right) + i \sin\left(\frac{2(j-1)\pi}{M}\right), \quad j = 1, 2, \dots, M;$$

令 $\mathbf{s}_R = \text{Re}(\mathbf{s})$, $\mathbf{s}_I = \text{Im}(\mathbf{s})$, 其中符号 $\text{Re}(\cdot)$ 和 $\text{Im}(\cdot)$ 分别表示相应复矩阵/复向量/复数的实部和虚部.

引入 $n \times n$ 复矩阵 $\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger$, 问题 (2.2) 可等价变形为

$$\min_{\mathbf{x}, \mathbf{X}} \mathbf{Q} \bullet \mathbf{X} + 2\text{Re}(\mathbf{c}^\dagger \mathbf{x})$$

$$\begin{aligned} \text{s.t. } & X_{i,i} = 1, \quad i = 1, \dots, n, \\ & \arg(x_i) \in \mathcal{A}, \quad i = 1, \dots, n, \\ & \mathbf{X} = \mathbf{x}\mathbf{x}^\dagger, \end{aligned}$$

其中变量 $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{X} \in \mathbb{C}^{n \times n}$. 问题 (2.2) 的传统 (但质量较差) 复 SDR 是

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{X}} & \quad \mathbf{Q} \bullet \mathbf{X} + 2\operatorname{Re}(\mathbf{c}^\dagger \mathbf{x}) \\ \text{s.t. } & X_{i,i} = 1, \quad i = 1, \dots, n, \\ & \mathbf{X} \succeq \mathbf{x}\mathbf{x}^\dagger, \end{aligned} \tag{2.3}$$

其中变量 $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{X} \in \mathbb{C}^{n \times n}$. 由于其对角特殊结构, 内点算法可在 $\mathcal{O}(n^{3.5})$ 时间内求得 SDR (2.3) 的最优解^[49]. 数值实验表明, 与球解码 (Sphere Decoding) 算法^[50-52]相比, 基于 SDR 的方法在误码率性能和时间复杂性上有非常好的平衡^[49]. 球解码算法是求解 MIMO 检测问题 (2.2) 最著名的算法之一. 从数学角度, 球解码算法是一种充分利用问题 (2.2) 特殊结构的分支定界算法, 所以其可以保证求得问题 (2.2) 的全局最优解, 但是球解码算法的最坏和平均复杂度都是指数的^[48,53].

比较式 (2.2) 和 (2.3) 可知, 式(2.3) 将约束 $\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger$ 松弛为约束 $\mathbf{X} \succeq \mathbf{x}\mathbf{x}^\dagger$ 并丢掉了所有的辐角约束 $\arg(x_i) \in \mathcal{A}, \quad i = 1, 2, \dots, n$. 容易验证

$$\mathbf{X} \succeq \mathbf{x}\mathbf{x}^\dagger \iff \begin{bmatrix} 1 & \mathbf{x}^\dagger \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}.$$

为了提高上述复 SDR 的质量, 我们在文 [54] 中提出在其基础上添加有效辐角割^[55,56]. 具体地, 根据 \mathbf{x} 的定义, 理想中 x_i 应当等于向量 \mathbf{s} 中的某一个分量. 通过松弛上述的组合约束, 我们得到如下加强的复 SDR:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{X}, \mathbf{t}} & \quad \mathbf{Q} \bullet \mathbf{X} + 2\operatorname{Re}(\mathbf{c}^\dagger \mathbf{x}) \\ \text{s.t. } & X_{i,i} = 1, \quad i = 1, \dots, n, \\ & \mathbf{X} \succeq \mathbf{x}\mathbf{x}^\dagger, \\ & \mathbf{x} = \mathbf{S}\mathbf{t}, \\ & \mathbf{A}\mathbf{t} = \mathbf{e}_n, \quad \mathbf{t} \geq \mathbf{0}, \end{aligned} \tag{2.4}$$

其中变量 $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{X} \in \mathbb{C}^{n \times n}$, $\mathbf{t} \in \mathbb{R}^{nM}$,

$$\mathbf{S} = \mathbf{I}_n \otimes \mathbf{s}^\top, \quad \mathbf{A} = \mathbf{I}_n \otimes \mathbf{e}_M^\top. \tag{2.5}$$

在式 (2.4) 中, $\mathbf{t} = [\mathbf{t}_1^\top, \mathbf{t}_2^\top, \dots, \mathbf{t}_n^\top]^\top$ 和 $\mathbf{t}_i = [t_{i,1}, t_{i,2}, \dots, t_{i,M}]^\top \in \mathbb{R}^M$.

文献 [54] 证明上述复 SDR 和如下的实 SDR 等价

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{Y}} & \quad \hat{\mathbf{Q}} \bullet \mathbf{Y} + 2\hat{\mathbf{c}}^\top \mathbf{y} \\ \text{s.t. } & Y_{i,i} + Y_{n+i,n+i} = 1, \quad i = 1, 2, \dots, n, \\ & \mathbf{Y} \succeq \mathbf{y}\mathbf{y}^\top, \end{aligned} \tag{2.6}$$

其中变量 $\mathbf{y} \in \mathbb{R}^{2n}$, $\mathbf{Y} \in \mathbb{R}^{2n \times 2n}$,

$$\hat{\mathbf{Q}} = \begin{bmatrix} \operatorname{Re}(\mathbf{Q}) & -\operatorname{Im}(\mathbf{Q}) \\ \operatorname{Im}(\mathbf{Q}) & \operatorname{Re}(\mathbf{Q}) \end{bmatrix}, \quad \hat{\mathbf{c}} = \begin{bmatrix} \operatorname{Re}(\mathbf{c}) \\ \operatorname{Im}(\mathbf{c}) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \operatorname{Re}(\mathbf{x}) \\ \operatorname{Im}(\mathbf{x}) \end{bmatrix}. \quad (2.7)$$

基于式 (2.6), 文献 [54] 进一步提出了一个加强的实 SDR. 定义 3×3 的矩阵

$$\mathbf{Y}_i = \begin{bmatrix} 1 & y_i & y_{n+i} \\ y_i & Y_{i,i} & Y_{i,n+i} \\ y_{n+i} & Y_{n+i,i} & Y_{n+i,n+i} \end{bmatrix}, \quad i = 1, 2, \dots, n \quad (2.8)$$

和

$$\mathbf{P}_j = \begin{bmatrix} 1 \\ \operatorname{Re}(s_j) \\ \operatorname{Im}(s_j) \end{bmatrix} [1 \quad \operatorname{Re}(s_j) \quad \operatorname{Im}(s_j)], \quad j = 1, 2, \dots, M. \quad (2.9)$$

根据式 (2.7) 中 \mathbf{y} 的定义, 理想中式 (2.8) 中定义的 \mathbf{Y}_i 应当等于 \mathbf{P}_j ($j = 1, 2, \dots, M$) 中的某一个, 即

$$\mathbf{Y}_i \in \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_M\}, \quad i = 1, 2, \dots, n.$$

通过松弛上述的组合约束以及丢掉一些冗余的约束, 文献 [54] 提出如下的关于 MIMO 检测问题 (2.2) 的 SDR:

$$\begin{aligned} & \min_{\mathbf{y}, \mathbf{Y}, \mathbf{t}} \quad \hat{\mathbf{Q}} \bullet \mathbf{Y} + 2\hat{\mathbf{c}}^T \mathbf{y} \\ \text{s.t.} \quad & \mathbf{Y}_i = \sum_{j=1}^M t_{i,j} \mathbf{P}_j, \quad i = 1, 2, \dots, n, \\ & \mathbf{A}\mathbf{t} = \mathbf{e}_n, \quad \mathbf{t} \geq \mathbf{0}, \\ & \mathbf{Y} \succeq \mathbf{y}\mathbf{y}^T, \end{aligned} \quad (2.10)$$

其中变量 $\mathbf{y} \in \mathbb{R}^{2n}$, $\mathbf{Y} \in \mathbb{R}^{2n \times 2n}$, $\mathbf{t} \in \mathbb{R}^{Mn}$, $\hat{\mathbf{Q}}$ 和 $\hat{\mathbf{c}}$ 在式 (2.7) 中给出, \mathbf{Y}_i 在式 (2.8) 中给出, \mathbf{P}_j 在式 (2.9) 中给出, \mathbf{A} 在式 (2.5) 中给出. 由于 \mathbf{Y}_i 的对称性, 约束 $\mathbf{Y}_i = \sum_{j=1}^M t_{i,j} \mathbf{P}_j$ 可等价地表示为如下 5 个线性约束:

$$\begin{aligned} y_i &= \mathbf{t}_i^T \mathbf{s}_R, \quad y_{n+i} = \mathbf{t}_i^T \mathbf{s}_I, \quad Y_{i,i} = \mathbf{s}_R^T \operatorname{Diag}(\mathbf{t}_i) \mathbf{s}_R, \\ & Y_{n+i,n+i} = \mathbf{s}_I^T \operatorname{Diag}(\mathbf{t}_i) \mathbf{s}_I, \quad Y_{i,n+i} = \mathbf{s}_R^T \operatorname{Diag}(\mathbf{t}_i) \mathbf{s}_I. \end{aligned} \quad (2.11)$$

关于 MIMO 检测问题 (2.2) 的 SDR, 一个重要的理论问题是其在什么条件下是紧的, 即在什么条件下求解相应的 SDR 可以得到原问题的最优解. 关于 $M = 2$ 时 SDR 紧性的结果很多, 相关结果请参见文献 [49,57-60]. 特别地, 文献 [60] 证明了如下结论.

定理 2.1 对于 $M = 2$, 如果式 (2.1) 中的参数 \mathbf{H} 和 \mathbf{v} 满足

$$\lambda_{\min}(\operatorname{Re}(\mathbf{H}^\dagger \mathbf{H})) > \|\operatorname{Re}(\mathbf{H}^\dagger \mathbf{v})\|_\infty, \quad (2.12)$$

其中 $\lambda_{\min}(\cdot)$ 表示相关矩阵的最小特征值, 则如下实 SDR 关于问题 (2.2) 是紧的:

$$\begin{aligned} & \min_{\mathbf{x}, \mathbf{X}} \quad \operatorname{Re}(\mathbf{Q}) \bullet \mathbf{X} + 2\operatorname{Re}(\mathbf{c})^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{X}_{i,i} = 1, \quad i = 1, 2, \dots, n, \\ & \mathbf{X} \succeq \mathbf{x}\mathbf{x}^\dagger. \end{aligned} \quad (2.13)$$

关于 $M > 2$ 时 SDR 紧性分析的结果在文献中非常之少. 特别地, 文献 [60] 中提出了一个公开问题, 即对于 $M > 2$ 的情形, 是否存在关于问题 (2.2) 紧的 SDR. 我们最近在文献 [54] 中回答了这一公开问题, 证明了式 (2.4) 和 (2.10) 中的 SDR 都是紧的. 这两个 SDR 也是关于一般 MIMO 检测问题 (2.2) 已知最早的在理论上有紧性保证的 SDR. 具体结果如下.

定理 2.2 对于任意的 $M \geq 3$, 如果式 (2.1) 中的参数 \mathbf{H} 和 \mathbf{v} 满足

$$\lambda_{\min}(\mathbf{H}^\dagger \mathbf{H}) \sin\left(\frac{\pi}{M}\right) > \|\mathbf{H}^\dagger \mathbf{v}\|_\infty, \quad (2.14)$$

则式 (2.4) 和 (2.10) 中的 SDR 关于问题 (2.2) 都是紧的.

关于式 (2.4) 和 (2.10) 中的 SDR 以及定理 2.2 的一些讨论如下. 首先, 当 $M = 2$ 时, (2.9) 中定义的矩阵退化为

$$\mathbf{P}_0 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ 和 } \mathbf{P}_1 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

所以 (2.11) 中的线性约束退化为

$$Y_{i,i} = 1, \quad y_i = t_{i,1} - t_{i,2}, \quad \text{和} \quad y_{n+i} = Y_{n+i,i} = Y_{n+i,n+i} = 0.$$

忽略矩阵 \mathbf{Y} 中的零块, 容易验证式 (2.10) 中的 SDR 退化为式 (2.13) 中的 SDR. 所以当 $M = 2$ 时, SDR (2.10) 在条件式 (2.12) 成立时关于问题 (2.2) 也是紧的.

其次, 式(2.12) 和 (2.14) 中的 (充分) 条件都有很好的工程和物理解释. 以式 (2.14) 为例: (2.14) 中的项 $\lambda_{\min}(\mathbf{H}^\dagger \mathbf{H})$ 表征了信道的好坏, 当信道矩阵 \mathbf{H} 中各列差别比较小时, $\lambda_{\min}(\mathbf{H}^\dagger \mathbf{H})$ 越小, 条件 (2.14) 越难满足; (2.14) 中的项 $\sin(\pi/M)$ 刻画了 M 如何影响问题的难度, M 越大, 相邻两个符号的距离 $2\sin(\pi/M)$ 越小, 条件 (2.14) 越难满足; (2.14) 中的项 $\|\mathbf{H}^\dagger \mathbf{v}\|_\infty$ 刻画了噪声的影响, 噪声 \mathbf{v} 对信道矩阵 \mathbf{H} 的列空间影响越大 (即 $\|\mathbf{H}^\dagger \mathbf{v}\|_\infty$ 越大), 条件 (2.14) 越难满足. 简而言之, 当调制的阶数越大、信道越差、噪声 (对信道矩阵列空间影响) 越大, (2.14) 中的条件越难满足; 反之 (2.14) 中的条件越容易满足.

再次, 我们讨论目前为止本文中列出的 SDR 之间以及与文献 [61] 中 SDR 的关系. 容易验证, SDR (2.10) 比 SDR (2.4) 强, SDR (2.4) 比 SDR (2.3) 和 (2.6) 强, 而 SDR (2.3) 和 (2.6) 等价. 文献 [61] 中提出的 SDR 是基于如下的观察: 对于 \mathbf{x}^* 中的任何 x_i^* 都存在 $\mathbf{t}_i \in \mathbb{R}^M$ 使得其中只有一个分量为 1 其余分量都为 0 且满足 $x_i^* = \mathbf{t}_i^\top \mathbf{s}$. 基于此, 文献 [61] 将问题 (2.2) 等价写为

$$\begin{aligned} \min_{\mathbf{t}} \quad & \mathbf{t}^\top \bar{\mathbf{Q}} \mathbf{t} + 2\bar{\mathbf{c}}^\top \mathbf{t} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{t} = \mathbf{e}_n, \quad \mathbf{t} \geq \mathbf{0}, \\ & \mathbf{t} \in \{0, 1\}^{Mn}, \end{aligned}$$

其中

$$\bar{\mathbf{Q}} = \hat{\mathbf{S}}^\top \hat{\mathbf{Q}} \hat{\mathbf{S}}, \quad \bar{\mathbf{c}} = \hat{\mathbf{S}}^\top \hat{\mathbf{c}}, \quad \hat{\mathbf{S}} = \begin{pmatrix} \operatorname{Re}(\mathbf{S}) \\ \operatorname{Im}(\mathbf{S}) \end{pmatrix}. \quad (2.15)$$

基于上述等价变形并利用向量 \mathbf{t} 的特殊结构, 文献 [61] 提出了如下的关于问题 (2.2) 的 SDR*:

$$\begin{aligned} \min_{\mathbf{t}, \mathbf{T}} \quad & \bar{\mathbf{Q}} \bullet \mathbf{T} + 2\bar{\mathbf{c}}^T \mathbf{t} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{t} = \mathbf{e}_n, \quad \mathbf{t} \geq \mathbf{0}, \\ & \mathbf{T} \succeq \mathbf{t}\mathbf{t}^T, \\ & \mathbf{T}_{i,i} = \text{Diag}(\mathbf{t}_i), \quad i = 1, 2, \dots, n, \end{aligned} \quad (2.16)$$

其中变量 $\mathbf{t} \in \mathbb{R}^{Mn}$, $\mathbf{T} \in \mathbb{R}^{Mn \times Mn}$, $\mathbf{T}_{i,i} \in \mathbb{R}^{M \times M}$ 是矩阵 \mathbf{T} 的第 i 个对角块. 最后一个约束 $\mathbf{T}_{i,i} = \text{Diag}(\mathbf{t}_i)$ 要求 $\mathbf{T}_{i,i}$ 是一个对角矩阵且其对角元素为 \mathbf{t}_i . 无论是在形式还是推导方式上, SDR (2.10) 和 SDR (2.16) 都不一样. 有意思的是, 我们在文献 [62] 中证明了这两个 SDR 之间的等价关系.

定理 2.3 SDR (2.10) 和 SDR (2.16) 的可行集之间满足如下的关系:

$$\hat{\mathbf{S}}\mathbf{T}\hat{\mathbf{S}}^T = \mathbf{Y} \text{ 和 } \hat{\mathbf{S}}\mathbf{t} = \mathbf{y}, \quad (2.17)$$

其中 $\hat{\mathbf{S}}$ 在 (2.15) 中给出. 所以 SDR (2.10) 和 SDR (2.16) 是等价的.

上述条件 (2.17) 保证相应的两个 SDR 的目标函数值相等. 结合定理 2.2 和 2.3, 我们可得: 当 (2.1) 中的参数 \mathbf{H} 和 \mathbf{v} 满足 (2.14) 时, SDR (2.16) 关于问题 (2.2) 是紧的. 定理 2.3 同时揭示 SDR (2.16) 中具有一定冗余性. 从 (2.17) 中可以得知, 高维空间 (\mathbf{T}, \mathbf{t}) 中的有用信息在低维空间 $(\mathbf{y}, \mathbf{Y}, \mathbf{t})$ 中都得到了保持. 特别地, SDR (2.16) 中矩阵变量的维数是 $Mn \times Mn$ 而 SDR (2.10) 中矩阵变量的维数是 $2n \times 2n$. 所以求解 SDR (2.10) 将会比求解 SDR (2.16) 在数值上有效的多 (特别是当 M 远大于 2 时).

基于上述加强的 SDR, 我们也发展了有效的分支定界算法求解 MIMO 检测问题 (2.2). 关于这方面的进展, 请参见文献 [63]. 值得一提的是, 在坏条件下, 例如输入和输出的数目相等或系统的信噪比较低时, 我们发展的基于加强 SDR 的分支定界算法比球解码算法在求解速度上有数量级的提高. 我们发展的分支定界算法可以推广用来求解如下更一般的非凸复二次规划问题:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{C}^n} \quad & \frac{1}{2} \mathbf{x}^\dagger \mathbf{Q} \mathbf{x} + \text{Re}(\mathbf{c}^\dagger \mathbf{x}) \\ \text{s.t.} \quad & \ell_i \leq |x_i| \leq u_i, \quad i = 1, 2, \dots, n, \\ & \arg(x_i) \in \mathcal{A}_i, \quad i = 1, 2, \dots, n, \end{aligned}$$

其中 $|x_i|$ 表示复数 x_i 的幅度, u_i 和 ℓ_i ($i = 1, 2, \dots, n$) 是 $2n$ 个实数满足 $u_i \geq \ell_i \geq 0$, \mathcal{A}_i ($i = 1, 2, \dots, n$) 是 n 个离散/连续的集合.

关于求解 MIMO 检测问题 (2.2) 的其他有效算法、最新进展以及综述, 请参见文献 [47,48,50-53,63-68]; 关于半正定规划/半正定松弛及其在信号处理和无线通信中应用的综述和最新进展, 请参见文献 [11,69,70],

*SDR (2.16) 和 [61] 中的 SDR (Model III) 的微小区别在于 SDR (Model III) 中利用每个向量 \mathbf{t}_i 求和等于 1 的性质消去了一个变量.

3 结束语

无线通信系统设计中的许多问题归根结底都是优化问题。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。本文着重介绍了两类无线通信系统设计中的问题，即多用户干扰信道最大最小准则下的联合传输/接收波束成形设计问题和 MIMO 检测问题，主要介绍了这两类问题的优化模型以及利用优化技术结合问题的特殊结构在求解和处理上述两类问题的最新进展。

近年来，第5代(5G)移动通信系统已成为世界范围内学术界和工业界的研究热点^[71,72]。相比于当前的移动通信系统，5G 除了需要进一步提高用户体验速率外，还需要支持高连接密度以实现移动数据业务量千倍增长、传输端与接收端的通信时延降低 90%、且实现频谱效率和功率效率提升十倍的目标。为实现此目标，5G 移动通信在无线传输技术和无线网络架构方面进行了革命性的技术创新。这些变革一方面能够实现上述目标，另一方面也给研究者们提出了新的更具挑战性的基础理论和关键科学问题。这些问题中很多都是优化问题，例如云无线接入网 (Cloud Radio Access Network) 中的缓存优化^[73,74]、大规模天线阵列 (Massive MIMO) 中的 1 比特预编码 (One-Bit Precoding) 设计^[75,76]、服务导向网络的网络切片 (Network Slicing)^[77,78]、物联网中的随机接入检测 (Random Access Detection)^[79,80] 等，解决这些问题的主要困难和挑战在于设计高效核心决策优化算法对通信资源进行动态分配以适应快速变化的无线环境、用户以及流量状态。总而言之，优化方法将继续在 5G 移动通信系统设计中扮演不可替代的重要角色。

限于作者水平有限，本文难免有遗漏和偏颇之处，请读者批评指正。

致谢

本文所涉及作者的研究工作都是与合作者一起完成的。特别地，本文关于 MIMO 干扰信道最大最小准则下的联合传输/接收波束成形设计的工作主要是和作者的导师戴彧虹研究员和香港中文大学 (深圳) 罗智泉教授合作完成的；本文关于 MIMO 检测问题半正定松弛的工作主要是和华北电力大学路程博士合作完成的。在此向所有合作者一并表示感谢。作者也借此机会特别感谢中国科学院数学与系统科学研究院袁亚湘院士和优化课题组的老师同学们。

参考文献

- [1] 袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京：科学出版社, 1997.
- [2] Nocedal J, Wright S J. *Numerical Optimization* [M]. New York: Springer Science+Business Media, 2006.
- [3] Bertsekas D. *Nonlinear Programming* [M]. Belmont: Athena Scientific, 1999.
- [4] Tse D, Viswanath P. *Fundamental of Wireless Communications* [M]. New York: Cambridge University Press, 2005.
- [5] Cover T M, Thomas J A. *Elements of Information Theory* [M]. New York: John Wiley & Sons, Inc., 2006.
- [6] Goldsmith A. *Wireless Communications* [M]. New York: Cambridge University Press, 2006.
- [7] Shannon C E. A mathematical theory of communication [J]. *The Bell System Technical Journal*, 1948, 5: 379–423, 623–656.

- [8] Garey M R, Johnson D S. *Computers and Intractability: A Guide to the Theory of NP-Completeness* [M]. San Francisco: W. H. Freeman and Company, 1979.
- [9] Luo Z Q, Yu W. An introduction to convex optimization for communications and signal processing [J]. *IEEE Journal on Selected Areas in Communications*, 2006, **24**(8): 1426-1438.
- [10] Luo Z Q. Applications of convex optimization in signal processing and digital communication [J]. *Mathematical Programming*, 2003, **97**(1-2): 177-207.
- [11] Luo Z Q, Ma W K, So A M C, et al. Semidefinite relaxation of quadratic optimization problems [J]. *IEEE Signal Processing Magazine*, 2010, **27**(3): 20-34.
- [12] Gershman A B, Sidiropoulos N D, Shahbazpanahi S, et al. Convex optimization-based beamforming [J]. *IEEE Signal Processing Magazine*, 2010, **27**(3): 62-75.
- [13] Björnson E, Jorswieck E. Optimal resource allocation in coordinated multi-cell systems [J]. *Foundations and Trends® in Communications and Information Theory*, 2013, **9**(2-3): 113-381.
- [14] Hong M, Luo Z Q. *Signal processing and optimal resource allocation for the interference channel* [M]//Academic Press Library in Signal Processing. Elsevier, 2014, 409-469.
- [15] 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.
- [16] 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 [J]. 中国科学: 数学, 2013, **43**(10): 953-964.
- [17] Schubert M, Boche H. Solution of the multiuser downlink beamforming problem with individual SINR constraints [J]. *IEEE Transactions on Vehicular Technology*, 2004, **53**(1): 18-28.
- [18] Wiesel A, Eldar Y C, Shamai S. Linear precoding via conic optimization for fixed MIMO receivers [J]. *IEEE Transactions on Signal Processing*, 2006, **54**(1): 161-176.
- [19] Liu Y F, Dai Y H, Luo Z Q. Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms [J]. *IEEE Transactions on Signal Processing*, 2011, **59**(3): 1142-1157.
- [20] Xia Y. A survey of hidden convex optimization [J]. 2019. Available: <https://arxiv.org/abs/1902.10921>
- [21] 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 [J]. *IEEE Signal Processing Letters*, 2013, **20**(1): 27-30.
- [22] Liu Y F. Dynamic spectrum management: A complete complexity characterization [J]. *IEEE Transactions on Information Theory*, 2017, **63**(1): 392-403.
- [23] Powell M J. On search directions for minimization algorithms [J]. *Mathematical Programming*, 1973, **4**(1): 193-201.
- [24] Liu Y F, Dai Y H, Luo Z Q. Max-min fairness linear transceiver design for a multi-user MIMO interference channel [C]//*Proceedings of IEEE International Conference on Communications (ICC)*, 2011, 1-5.
- [25] Liu Y F, Dai Y H, Luo Z Q. Max-min fairness linear transceiver design for a multi-user MIMO interference channel [J]. *IEEE Transactions on Signal Processing*, 2013, **61**(9): 2413-2423.
- [26] Yu W, Kwon T, Shin C. *Adaptive resource allocation in cooperative cellular networks* [M]//*Cooperative Cellular Wireless Networks*. New York: Cambridge University Press, 2010, 233-256.
- [27] Zhang J, Chen R, Andrews J G, et al. Networked MIMO with clustered linear precoding [J]. *IEEE Transactions on Wireless Communications*, 2009, **8**(4): 1910-1921.
- [28] Chen E, Tao M, Liu Y F. Joint base station clustering and beamforming for non-orthogonal multicast and unicast transmission with backhaul constraints [J]. *IEEE Transactions on Wireless Communications*, 2018, **17**(9): 6265-6279.

- [29] Ng C T, Huang H. Linear precoding in cooperative MIMO cellular networks with limited coordination clusters [J]. *IEEE Journal on Selected Areas in Communications*, 2010, **28**(9): 1446-1454.
- [30] Hong M, Sun R, Baligh H, et al. Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks [J]. *IEEE Journal on Selected Areas in Communications*, 2013, **31**(2): 226-240.
- [31] Dai B, Yu W. Sparse beamforming and user-centric clustering for downlink cloud radio access network [J]. *IEEE Access*, 2014, **2**: 1326-1339.
- [32] Shi Y, Zhang J, Letaief K. Group sparse beamforming for green cloud-RAN [J]. *IEEE Transactions on Wireless Communications*, 2014, **13**(5): 2809-2823.
- [33] 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.
- [34] Liao W C, Hong M, Liu Y F, et al. Base station activation and linear transceiver design for optimal resource management in heterogeneous networks [J]. *IEEE Transactions on Signal Processing*, 2014, **62**(15): 3939-3952.
- [35] Shen K, Yu W. Distributed pricing-based user association for downlink heterogeneous cellular networks [J]. *IEEE Journal on Selected Areas in Communications*, 2014, **32**(6): 1100-1113.
- [36] Shen K, Liu Y F, Ding D Y, et al. Flexible multiple base station association and activation for downlink heterogeneous networks [J]. *IEEE Signal Processing Letters*, 2017, **24**(10): 1498-1502.
- [37] Sanjabi M, Razaviyayn M, Luo Z Q. Optimal joint base station assignment and beamforming for heterogeneous networks [J]. *IEEE Transactions on Signal Processing*, 2014, **62**(8): 1950-1961.
- [38] 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 [J]. *IEEE Transactions on Signal Processing*, 2011, **59**(9): 4331-4340.
- [39] Zhang X, Chang T H, Liu Y F, et al. Max-min fairness user scheduling and power allocation in full-duplex OFDMA systems [J]. *IEEE Transactions on Wireless Communications*, 2019, **18**(6): 3078 - 3092.
- [40] Codreanu M, Tölli A, Juntti M, et al. Joint design of Tx-Rx beamformers in MIMO downlink channel [J]. *IEEE Transactions on Signal Processing*, 2007, **55**(9): 4639-4655.
- [41] Chang T H, Liu Y F, Lin S C. QoS-based linear transceiver optimization for full-duplex multi-user communications [J]. *IEEE Transactions on Signal Processing*, 2018, **66**(9): 2300-2313.
- [42] Li Q, Hong M, Wai H T, et al. Transmit solutions for MIMO wiretap channels using alternating optimization [J]. *IEEE Journal on Selected Areas in Communications*, 2013, **31**(9): 1714-1727.
- [43] Hong M, Li Q, Liu Y F. Decomposition by successive convex approximation: A unifying approach for linear transceiver design in heterogeneous networks [J]. *IEEE Transactions on Wireless Communications*, 2016, **15**(2): 1377-1392.
- [44] Wright S J. Coordinate descent algorithms [J]. *Mathematical Programming*, 2015, **151**(1): 3-34.
- [45] 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 [J]. *IEEE Signal Processing Magazine*, 2016, **33**(1): 57-77.
- [46] Verdú S. *Multiuser Detection* [M]. New York: Cambridge University Press, 1998.
- [47] Yang S, Hanzo L. Fifty years of MIMO detection: The road to large-scale MIMOs [J]. *IEEE Communication Surveys & Tutorials*, 2015, **17**(4): 1941-1988.
- [48] Verdú S. Computational complexity of optimum multiuser detection [J]. *Algorithmica*, 1989, **4**(1-4): 303-312.
- [49] Kisialiou M, Luo X, Luo Z Q. Efficient implementation of quasi-maximum-likelihood detection based on semidefinite relaxation [J]. *IEEE Transactions on Signal Processing*, 2009, **57**(12): 4811-4822.

- [50] Fincke U, Pohst M. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis [J]. *Mathematics of Computation*, 1985, **44**(170): 463-471.
- [51] Viterbo E, Boutros J. A universal lattice code decoder for fading channels [J]. *IEEE Transactions on Information Theory*, 1999, **45**(5): 1639-1642.
- [52] Damen M O, Gamal H E, Caire G. On maximum-likelihood detection and the search for the closest lattice point [J]. *IEEE Transactions on Information Theory*, 2003, **49**(10): 2389-2402.
- [53] Jaldén J, Ottersten B. On the complexity of sphere decoding in digital communications [J]. *IEEE Transactions on Signal Processing*, 2005, **53**(4): 1474-1484.
- [54] Lu C, Liu Y F, Zhang W Q, et al. Tightness of a new and enhanced semidefinite relaxation for MIMO detection [J]. *SIAM Journal on Optimization*, 2019, **29**(1): 719-742.
- [55] 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.
- [56] Lu C, Deng Z, Zhang W Q, et al. Argument division based branch-and-bound algorithm for unit-modulus constrained complex quadratic programming [J]. *Journal of Global Optimization*, 2018, **70**(1): 171-187.
- [57] Jaldén J. Detection for multiple input multiple output channels: Analysis of sphere decoding and semidefinite relaxation. Ph.D. thesis. Stockholm: KTH, 2006.
- [58] Jaldén J, Martin C, Ottersten B. Semidefinite programming for detection in linear systems—Optimality conditions and space-time decoding [C]//*Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011, 9-12.
- [59] Kisialiou M, Luo Z Q. Probabilistic analysis of semidefinite relaxation for binary quadratic minimization [J]. *SIAM Journal on Optimization*, 2010, **20**(4): 1906-1922.
- [60] So A M C. Probabilistic analysis of the semidefinite relaxation detector in digital communications [C]//*Proceedings of twenty-first annual ACM-SIAM symposium on Discrete Algorithms (SODA)*, 2010, 698-711.
- [61] Mobasher A, Taherzadeh M, Sotirov R, et al. A near-maximum-likelihood decoding algorithm for MIMO systems based on semi-definite programming [J]. *IEEE Transactions on Information Theory*, 2007, **53**(11): 3869-3886.
- [62] Liu Y F, Xu Z, Lu C. On the equivalence of semidefinite relaxations for MIMO detection with general constellations [C]//*Proceedings of ICASSP*, 2019, 4549-4553.
- [63] Lu C, Liu Y F, Zhou J. An enhanced SDR based global algorithm for nonconvex complex quadratic programs with signal processing applications [J]. arXiv:1902.04287.
- [64] Liu H, Yue M C, So A M C, et al. A discrete first-order method for large-scale MIMO detection with provable guarantees [C]//*Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017, 669-673.
- [65] Burer S, Monteiro R D, Zhang Y. Rank-two relaxation heuristics for max-cut and other binary quadratic programs [J]. *SIAM Journal on Optimization*, 2002, **12**(2): 503-521.
- [66] Luo Z Q, Luo X, Kisialiou M. An efficient quasi-maximum likelihood decoder for PSK signals [C]//*Proceedings of ICASSP*, 2003, 561-564.
- [67] Pan J, Ma W K, Jaldén J. MIMO detection by Lagrangian dual maximum-likelihood relaxation: Reinterpreting regularized lattice decoding [J]. *IEEE Transactions on Signal Processing*, 2014, **62**(2): 511-524.
- [68] Lu C, Liu Y F, Zhou J. An efficient global algorithm for nonconvex complex quadratic problems with applications in wireless communications [C]//*Proceedings of IEEE/CIC International Conference on Communications in China (ICCC)*, 2017, 1-5.
- [69] Vandenberghe L, Boyd S. Semidefinite programming [J]. *SIAM Review*, 1996, **38**(1): 49-95.
- [70] Pu W, Liu Y F, Yan J, et al. Optimal estimation of sensor biases for asynchronous multi-sensor data fusion [J]. *Mathematical Programming*, 2018, **170**(1): 357-386.

-
- [71] Andrews J G, Buzzi S, Choi W, et al. What will 5G be? [J]. *IEEE Journal on Selected Areas in Communications*, 2014, **32**(6): 1065-1082.
 - [72] 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 [J]. 中国科学: 信息科学, 2014, **44**: 551-563.
 - [73] Tao M, Chen E, Yu W, et al. Cache-enabled cloud radio access networks [M]// Wireless Edge Caching: Modeling, Analysis and Optimization, Cambridge University Press, 2019.
 - [74] Dai B, Liu Y F, Yu W. Optimized base-station cache allocation for cloud radio access network with multicast backhaul [J]. *IEEE Journal on Selected Areas in Communications*, 2018, **36**(8): 1737-1750.
 - [75] Jacobsson S, Durisi G, Coldrey M, et al. Quantized precoding for massive MU-MIMO [J]. *IEEE Transactions on Communications*, 2017, **65**(11): 4670-4684.
 - [76] Sohrabi F, Liu Y F, Yu W. One-bit precoding and constellation range design for massive MIMO with QAM signaling [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2018, **12**(3): 557-570.
 - [77] Zhang N, Liu Y F, Farmanbar H, et al. Network slicing for service-oriented networks under resource constraints [J]. *IEEE Journal on Selected Areas in Communications*, 2017, **35**(11): 2512-2521.
 - [78] Domenico A D, Liu Y F, Yu W. Optimal computational resource allocation and network slicing deployment in 5G hybrid C-RAN [C]// *Proceedings of IEEE International Conference on Communications (ICC)*, 2019.
 - [79] 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 [J]. *IEEE Signal Processing Magazine*, 2018, **35**(5): 88-99.
 - [80] Chen Z, Sohrabi F, Liu Y F, et al. Covariance based joint activity and data detection for massive random access with massive MIMO [C]// *Proceedings of IEEE International Conference on Communications (ICC)*, 2019.