

# 有限码长域下针对多用户大规模MIMO系统速率优化的高效功率分配算法

胡钰林\* 肖志成 徐浩  
(武汉大学电子信息学院 武汉 430000)

**摘要:** 第六代(6G)移动通信网络需要为大规模节点提供高可靠低时延通信(URLLC)服务。为此, 该文针对多用户大规模多输入多输出(MIMO)技术辅助的URLLC下行通信场景, 基于有限码长(FBL)域理论表征系统性能, 以用户速率公平性为目标, 提出一种高效的功率分配算法。具体而言, 该文首先针对传统MIMO中基于全局奇异值分解(SVD)的线性预编码方案复杂度高、不能兼顾用户公平性等问题, 设计基于局部SVD的预编码方案, 以相对较低的复杂度实现对MIMO用户间干扰和用户内干扰的有效抑制。其次, 该文以功率分配因子为优化变量、以最大化最小用户速率(MMR)为目标构建优化问题。为解决所构建的高维变量耦合非凸问题, 该文通过引入辅助变量、分段McCormick包络将目标函数中香农容量相关项凸松弛处理, 实现MMR问题重构。进而该文提出基于连续凸近似(SCA)的优化算法有效求解MMR问题。仿真结果验证了所提优化算法的收敛性与准确性, 同时也表明所提优化方案相比于现有方案在系统MMR性能和鲁棒性上均具有优势。

**关键词:** 超可靠低时延通信(URLLC); 多用户MIMO; 有限码长(FBL); 连续凸近似(SCA)

**中图分类号:** TN929.5

**文献标识码:** A

**文章编号:** 1009-5896(2025)01-0035-13

**DOI:** 10.11999/JEIT240241

## 1 引言

随着工业4.0的发展, 工业物联网、车联网等应用亟需新型通信网络以满足大规模节点的超可靠低时延通信(Ultra-Reliable and Low-Latency Communications, URLLC)需求<sup>[1,2]</sup>。预期在第六代(The 6th Generation, 6G)移动通信系统中, 将提供超大规模的无线连接与高速可靠的无线服务, 集成传感、通信、计算等功能, 支持各类垂直应用<sup>[3,4]</sup>。

面对上述大规模接入能力以及URLLC需求, 大规模MIMO(Multi-Input Multiple-Output)技术能够通过空分复用, 空间分集和波束赋形等优势实现系统性能增强, 因而成为极具前景的关键技术之一<sup>[3,5]</sup>。目前, 已有一些学者针对MIMO增强6G系统性能展开了研究。例如, Ma等人<sup>[6]</sup>通过联合优化收发天线的空间部署和发射信号的协方差, 基于点对点MIMO提升了通信系统的容量。Chi等人<sup>[7]</sup>通过联合设计编码、调制、检测和解码方案, 有效提升了通信系统在应对海量用户接入时的总容量。相关研究也证实大规模MIMO在提升通信系统的超大规模连接能力<sup>[8]</sup>、可靠性<sup>[9]</sup>等方面的显著优势, 进一步证实在面向6G通信系统需求时研究大规模MIMO辅助的URLLC系统的必要性。

然而, 现有关于大规模MIMO辅助通信系统的

研究<sup>[3,6-9]</sup>大都聚焦于在传统香农域下提升系统容量和可靠性。为保障超低时延传输, URLLC系统通常基于有限长的编码进行通信。在有限码长(Finite BlockLength, FBL)域下, 信息传输不再任意可靠, 同时信道容量也无法准确刻画系统的可达通信速率。为表征有限长编码通信性能, Polyanskiy等人<sup>[10]</sup>研究了FBL域下单用户单天线的通信性能, 推导出可达速率和误码率的非渐近表达。基于Polyanskiy的研究, Yang等人<sup>[11]</sup>进一步将该模型扩展至多天线的通信场景, 并推导出准静态衰落信道下MIMO辅助FBL通信的最大可达速率。该表达式揭示了可达速率、码长、信噪比(Signal-to-Noise Ratio, SNR)与译码错误概率之间的相互作用关系。基于上述FBL域下的通信系统性能刻画, Zhu等人<sup>[12]</sup>对译码错误概率和信道色散进行精准近似, 证明在具有实际价值的区域内译码错误概率是关于码长和发送功率的联合凸函数, 并进一步基于所刻画的问题凸性进行系统资源分配优化, 有效降低了用户误码率。Zhao等人<sup>[13]</sup>对MIMO辅助上行URLLC系统进行研究, 分析了此场景下时延和可靠性之间的耦合关系, 并精准刻画了系统能效。然而, 上述工作<sup>[12,13]</sup>尚未涉及设计功率分配方案, 对多用户干扰系统中用户速率和以及用户公平性速率进行优化。针对速率和性能提升与系统功率资源分配, Peng等人<sup>[14]</sup>研究了多物联网设备URLLC场景下的速率和问题, 在FBL域下推导出了上行通信链路速率和的下界, 并证明此下界为关于功率的凸函数, 进而联合优化导频功率和负载功率以最大化系统通信速率和。尽管

收稿日期: 2024-04-08; 改回日期: 2024-12-09; 网络出版: 2024-12-12

\*通信作者: 胡钰林 yulin.hu@whu.edu.cn

基金项目: 国家重点研发计划(2023YFE0206600)

Foundation Item: The National Key R&D Plan (2023YFE0206600)

如此, Peng等人<sup>[14]</sup>的方案难以保障用户间的速率公平性。面向公平性, He等人<sup>[15]</sup>研究多入单出(Multiple Input Single Output, MISO)辅助的URLLC系统, 分析了该模型下可达速率关于信号与干扰加噪声比(Signal to Interference plus Noise Ratio, SINR)的单调性和凹凸性, 进而优化用户功率分配以最大化所有用户间的最小通信速率(Maximum Minimum Rate, MMR)。Fang等人<sup>[16]</sup>研究无蜂窝大规模MIMO系统, 提出一种路径跟踪方法, 利用对数函数近似MMR问题的目标函数, 并将多项式约束转换为单项式约束, 从而将原始问题重构为一系列几何规划问题, 通过迭代的方式进行求解, 有效提升了系统的MMR。He等人<sup>[15]</sup>和Fang等人<sup>[16]</sup>的模型中每个用户配备单根天线, 在实际的大规模MIMO场景下, 增加用户端配备的天线数可以提高天线的分集增益, 进而提高传输速率。然而, 多天线场景下显著的用户间干扰对系统性能造成损失, 同时多天线场景下优化问题具有高维变量耦合、目标函数和约束非凸等特点, 增加了问题的求解难度。Sheng等人<sup>[17]</sup>考虑具有窃听者的多用户MIMO干扰网络, 在香农域下兼顾用户公平实现了传输速率的提升, 然而该工作并未考虑到URLLC应用的高可靠性和超低时延传输需求。

综上, 针对MIMO辅助的URLLC系统, 兼顾用户速率公平性的功率分配设计仍然是一个开放性问题。为了解决此问题, 该文研究大规模MIMO辅助的URLLC下行通信, 设计高效的功率分配策略来最大化所有用户间最小通信速率。具体贡献如下:

(1)该文通过局部奇异值分解(Singular Value Decomposition, SVD)设计低复杂度预编码方案, 有效消除用户内各天线间干扰及减少各用户间干扰, 弥补传统的SVD线性预编码方案时间复杂度高, 不能兼顾用户公平性的缺陷。基于此, 在FBL域下以最大化最小用户速率(Maximum Minimum Rate, MMR)为目标, 构建了发射总功率受限时最优的功率分配问题。

(2)针对MMR问题的非凸性及高维变量耦合难点, 该文引入一种基于连续凸近似(Successive Convex Approximation, SCA)的优化算法。首先, 引入辅助变量对香农容量项凸松弛处理, 实现MMR问题重构。其次, 针对重构问题中信道色散相关项复杂非凸的挑战, 构造对优化变量联合凸性的近似函数, 进而将重构问题转换为可直接求解的凸子问题, 最终基于SCA算法迭代求解凸子问题以获得原始问题的有效次优解。

(3)该文通过仿真验证了所设计预编码方案和优化算法的低复杂度特性与高效性, 同时证实了所设计优化方案相较于现有对比方案的性能优势。最后, 该文分析了码长、天线数、用户数等参量对系统性能的影响。

符号定义:  $\mathbb{E}[\cdot]$ 表示随机变量的期望,  $(\cdot)^T$ 表示矩阵或者向量的转置操作,  $(\cdot)^H$ 表示矩阵或者向量的共轭转置操作,  $\mathbb{C}^{N \times M}$ 表示 $N \times M$ 维复数矩阵,  $\text{diag}\{\cdot\}$ 表示对角矩阵。

## 2 系统模型与问题描述

### 2.1 系统模型

如图1所示, 考虑一个大规模MIMO辅助的URLLC下行通信场景, 其中一个基站通过下行链路对 $M$ 个用户进行服务。基站与所有用户均配备多天线, 其中基站的发射天线数量记为 $N$ , 用户 $g(1 \leq g \leq M)$ 的接收天线数量记为 $K_g$ 。为了支持URLLC, 基站以有限长编码的方式向用户发送数据。在一个服务周期内, 基站将编码后的数据下行传输至用户端。数据经过编码后的符号长度, 即码长, 记为 $n$ 。一次传输中, 基站传输给用户 $g$ 的原始符号向量定义为 $\mathbf{d}_g = (d_1, \dots, d_i, \dots, d_{K_g})^T$ , 其中 $d_i$ 为传输给用户 $g$ 第 $i$ 根天线的符号, 且满足单位功率条件, 即 $\mathbb{E}(d_i^2) = 1$ 。

用户 $g$ 的功率矩阵定义为 $\mathbf{W}_g = \text{diag}\{\sqrt{P_t \beta_1}, \dots, \sqrt{P_t \beta_i}, \dots, \sqrt{P_t \beta_{K_g}}\}$ , 其中 $P_t$ 为总功率,  $\beta_i$ 为用户 $i$ 的功率分配因子。经功率分配后, 用户 $g$ 的发射符号向量可写为 $\mathbf{x}_g = (x_1, x_2, \dots, x_{K_g})^T = \mathbf{W}_g \mathbf{d}_g$ 。进一步, 可将所有用户的发射符号向量写为 $\mathbf{x} = (\mathbf{x}_1^T, \dots, \mathbf{x}_M^T)^T$ 。定义基站对所有用户进行处理的预编码矩阵 $\mathbf{S}_T = \text{diag}\{\mathbf{S}_{T1}, \dots, \mathbf{S}_{TM}\}$ , 其中用户 $g$ 的预编码矩阵 $\mathbf{S}_{Tg} = \text{diag}\{s_{Tg}^{(1)}, s_{Tg}^{(2)}, \dots, s_{Tg}^{(K_g)}\} \in \mathbb{C}^{N \times K_g}$ 。经功率分配和预编码处理后信号进入信道传输, 用户 $g$ 的信道系数矩阵定义为 $\mathbf{H}_g = \{\mathbf{h}_{g1}, \mathbf{h}_{g2}, \dots, \mathbf{h}_{gK_g}\}$ ,  $\mathbf{H}_g$ 的最大秩 $\min\{N, \sum_g K_g\}$ 表示MIMO系统数据传输所能支持的最大流数。每一个流对应着发射端多个信号的线性组合, 在数值上流的信道增益对应

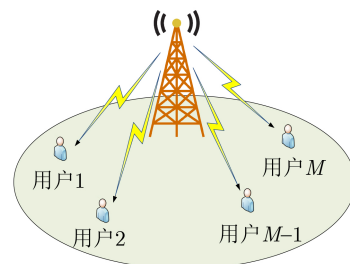


图1 一种多用户大规模MIMO辅助的URLLC下行链路模型

着信道矩阵经奇异值分解后的奇异值大小，如果MIMO系统的信道矩阵具有 $r$ 个非零奇异值，那么系统最多可以支持 $r$ 个独立的数据流，即奇异值决定了系统支持的数据流数量和每个数据流的信道增益。通过设计合适的预编码和功率分配方案，可以充分利用MIMO中空间自由度增加可利用的流数，从而提升传输速率。假设基站与用户间信道为准静态衰落信道，且服从莱斯因子为 $L$ 的莱斯衰落模型，则用户 $g$ 的信道系数矩阵可表示为<sup>[17-19]</sup>

$$\mathbf{H}_g = \sqrt{\frac{L}{L+1}} \mathbf{H}_{Dg} + \sqrt{\frac{1}{L+1}} \mathbf{H}_{Sg} \quad (1)$$

其中， $\mathbf{H}_{Dg}$ 和 $\mathbf{H}_{Sg}$ 分别表示视线传输(Line Of Sight, LOS)和非视线传输(Non Line Of Sight, NLOS)分量， $\mathbf{H}_{Dg}$ 的元素均为1， $\mathbf{H}_{Sg}$ 由Kronecker模型<sup>[20]</sup>给出，即

$$\mathbf{H}_{Sg} = \mathbf{I}_{R_g}^{\frac{1}{2}} \mathbf{B}_g \mathbf{I}_{T_g}^{\frac{1}{2}} \quad (2)$$

其中， $\mathbf{B}_g$ 是 $N \times K_g$ 维矩阵，其元素 $b_{(m,n,g)}$  ( $1 \leq m \leq N, 1 \leq n \leq K_g$ )均为独立同分布的复高斯随机变量，满足 $b_{(m,n,g)} \sim \mathcal{CN}(0, 1)$ 。 $\mathbf{I}_{T_g}$ 和 $\mathbf{I}_{R_g}$ 分别表示发送端天线之间的空间相关性和接收端天线之间的空间相关性。假设基站和用户端的天线为均匀线性阵列，记相邻天线的空间相关系数为 $\rho_0$ ，则根据指数相关性模型<sup>[21,22]</sup>，基站第 $m$ 根天线和第 $n$ 根天线 ( $1 \leq m \leq N, 1 \leq n \leq N_g$ )之间的空间相关性可表示为 $r_{m,n} = \rho_0^{|m-n|}$ 。同理，用户 $g$ 第 $i$ 根天线和第 $j$ 根天线之间的空间相关性为 $r_{i,j} = \rho_0^{|i-j|}$  ( $1 \leq i \leq K_g, 1 \leq j \leq K_g$ )。基站与所有用户间的信道系数矩阵记为 $\mathbf{H} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_M\}$ 。根据上述建模，用户端所有接收信号向量表征为

$$\mathbf{y} = \mathbf{H}^H \mathbf{S}_T \mathbf{x} + \mathbf{n} \quad (3)$$

其中 $\mathbf{n}$ 为噪声向量，其第 $g$ 个元素 $\mathbf{n}_g = (n_{g1}, n_{g2}, \dots, n_{gK_g})^T$ 表示用户 $g$ 的信号信道中受到的噪声干扰，满足 $\mathbb{E}(n_{gi}^2) = 1$ 。特别地，用户 $g$ 端的接收信号向量为

$$\mathbf{y}_g = \mathbf{H}_g^H \mathbf{S}_{Tg} \mathbf{x}_g + \mathbf{n}_g \quad (4)$$

用户 $g$ 对信号进行解码处理后，接收信号可表示为

$$\mathbf{y}_g = \mathbf{S}_{Rg}^H (\mathbf{H}_g^H \mathbf{S}_{Tg} \mathbf{W}_g \mathbf{d}_g + \mathbf{H}_g^H \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c + \mathbf{n}_g), \quad (5)$$

其中， $\mathbf{S}_{Rg}^H = \{\mathbf{s}_{Rg}^{(1)}, \mathbf{s}_{Rg}^{(2)}, \dots, \mathbf{s}_{Rg}^{(K_g)}\}$ 为解码矩阵。进一步，对于用户 $g$ 的第 $k$ 根天线，其接收信号表示为

$$\begin{aligned} \mathbf{y}_{gk} = & \mathbf{S}_{Rg}^H \left( \mathbf{h}_{gk}^H \sum_{i=1}^{K_g} s_{Tg}^{(i)} \sqrt{P_t \beta_{gi}} d_{gi} \right. \\ & + \sum_{j \neq k} \mathbf{h}_{gj}^H \sum_{i=1}^{K_g} s_{Tg}^{(i)} \sqrt{P_t \beta_{gi}} d_{gi} \\ & \left. + \mathbf{H}_g^H \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c + \mathbf{n}_g \right), \quad (6) \end{aligned}$$

其中，式(6)右边(Right Hand Side, RHS)第2项、第3项分别表示用户内干扰和用户间干扰。若不加处理，此类干扰会显著降低接收端的SINR，限制系统容量<sup>[23,24]</sup>。后续优化设计中，本文将针对此类干扰项设计预编码方案进行有效处理。

## 2.2 问题描述

根据式(6)，用户 $g$ 第 $k$ 根天线处的接收信号SINR可表示为

$$\gamma_{gk} = \frac{\lambda_{gk} \beta_{gk}}{\varphi_{gk}(\boldsymbol{\beta}) + c_{gk}} \quad (7)$$

其中功率分配因子向量 $\boldsymbol{\beta} = \{\beta_{11}, \beta_{12}, \dots, \beta_{MK_M}\}$ ， $\varphi_{gk}(\boldsymbol{\beta}) = \sum_{c=1}^M \sum_{j=1}^{K_c} \lambda_{gk} \beta_{cj} ((\mathbf{s}_{Tg}^{(k)})^H \mathbf{s}_{Tc}^{(j)})^2$ ， $c_{gk} = (\mathbf{s}_{Rg}^{(k)})^H \mathbf{s}_{Rg}^{(k)} / P_t$ ，且 $\gamma_{gk}$ 满足

$$\gamma_{gk} \leq \hat{\gamma}_{gk} = \frac{\lambda_{gk} P_t}{(\mathbf{s}_{Rg}^{(k)})^H \mathbf{s}_{Rg}^{(k)}} \quad (8)$$

在得到SINR的条件下，本文采用文献<sup>[10,11]</sup>的模型刻画系统性能。根据上述模型，用户 $g$ 可达通信速率表示为

$$R_g(\boldsymbol{\beta}) \approx \sum_{k=1}^{K_g} \log_2(1 + \gamma_{gk}) - Q^{-1}(\varepsilon_g) \sqrt{\frac{V_g(\boldsymbol{\beta})}{nK_g}} \log_2 e \quad (9)$$

其中， $\varepsilon_g$ 是用户 $g$ 接收信号的译码错误概率， $Q^{-1}(\cdot)$ 是高斯 $Q$ 函数的反函数， $V_g(\boldsymbol{\beta})$ 是信道色散，在典型的加性高斯白噪声(Additive White Gaussian Noise, AWGN)信道中，其定义为<sup>[25]</sup>

$$V_g(\boldsymbol{\beta}) = K_g - \sum_{k=1}^{K_g} \frac{1}{(1 + \gamma_{gk})^2}. \quad (10)$$

根据式(8)-式(10)可知，用户端的可达速率得到了有效表征。考虑到系统总功率受限，若将总吞吐量最大化作为优化目标，部分用户会因为其较好的信道质量而过多占用功率资源，进而难以满足系统中剩余信道较差用户的通信速率需求。此外，由式(7)可知，不恰当的功率资源分配也会增加用户端受到的干扰，即 $\varphi_{gk}(\boldsymbol{\beta})$ 项，进而显著降低用户SINR。因此，以用户公平性为目标进行资源分配方案设计对保障系统公平性和提升资源利用效率具

有重要意义。综合式(7)–式(10)，以功率分配因子为优化变量，在总发射功率受限的条件下，MMR优化问题(OP1)可建模为

$$\begin{aligned} \text{(OP1)} : \max_{\boldsymbol{\beta}} R_{\min}(\boldsymbol{\beta}) &= \min_{1 \leq g \leq M} R_g(\boldsymbol{\beta}) \\ \text{s.t. } 0 \leq \beta_{g,k} \leq 1 & (1 \leq g \leq M, 1 \leq k \leq K_g), \\ \sum_{g=1}^M \sum_{k=1}^{K_g} \beta_{gk} &= 1 \end{aligned} \quad (11)$$

其中，第1项约束给出功率分配因子的取值范围为 $[0, 1]$ ，第2项约束规定所有用户的功率分配因子之和为1。观察目标函数中用户可达通信速率 $R_g(\boldsymbol{\beta})$ 的表达式，可知该问题为非凸问题，难以直接求解。因而，后续将设计一种低复杂度预编码方案消除干扰，并通过高效的功率分配方案求解MMR问题。

### 3 预编码方案与面向用户公平性的高效功率分配算法

为了有效解决第2节构建的MMR问题(OP1)，本节将设计一种有效的系统资源分配方法。首先，本节引入一种基于局部SVD的低复杂度预编码方案，抑制用户内干扰和用户间天线间干扰。其次，考虑到MMR问题非凸且具有多维优化变量的挑战，本节设计一种基于SCA的功率分配算法。其具体思路为：先松弛处理原始问题中香农容量项以实现MMR问题重构，随后利用凸函数精准近似信道色散项，进而在局部点将优化问题转换为关于功率分配因子的凸子问题，通过算法迭代构建并求解凸子问题，最终收敛至原始问题的次优解。

#### 3.1 低复杂度预编码方案

为实现低复杂度的同时抑制用户内干扰和用户间干扰，本节首先引入一种基于局部SVD的预编码策略。该文考虑到发送端天线数一般大于单个用户的天线数，即假设 $N \geq K_g (\forall g \in [0, M])$ 。该方案首先对每个用户的信道矩阵 $\mathbf{H}_g^H$ 进行SVD，即将 $\mathbf{H}_g^H$ 分解为

$$\mathbf{H}_g^H = \mathbf{D}_{Rg} \mathbf{A}'_g \mathbf{D}_{Tg}^H \quad (12)$$

$$\mathbf{D}_{Rg} = \left\{ \mathbf{d}_{Rg}^{(1)}, \dots, \mathbf{d}_{Rg}^{(K_g)} \right\} \in \mathbb{C}^{K_g \times K_g}$$

$$\mathbf{D}_{Tg} = \left\{ \mathbf{d}_{Tg}^{(1)}, \mathbf{d}_{Tg}^{(2)}, \dots, \mathbf{d}_{Tg}^{(N)} \right\} \in \mathbb{C}^{N \times N} \quad (13)$$

其中， $\mathbf{D}_{Rg}$ 、 $\mathbf{D}_{Tg}$ 分别为左奇异值矩阵和右奇异值矩阵， $\mathbf{A}'_g = (\mathbf{A}_g, \mathbf{A}_g^{\text{zero}})$ ， $\mathbf{A}_g^{\text{zero}}$ 为全零矩阵， $\mathbf{A}_g = \text{diag}\{\sqrt{\lambda_{g1}}, \sqrt{\lambda_{g2}}, \dots, \sqrt{\lambda_{gK_g}}\}$ 。令预编码矩阵和其对应的解码矩阵为

$$\mathbf{S}_{Tg} = \left\{ \mathbf{d}_{Tg}^{(1)}, \dots, \mathbf{d}_{Tg}^{(K_g)} \right\} \quad (14)$$

$$\mathbf{S}_{Rg} = \mathbf{D}_{Rg} \quad (15)$$

则可基于该预编码矩阵和解码矩阵对信道矩阵 $\mathbf{H}_g$ 进行处理，进而得到

$$\mathbf{S}_{Rg}^H \mathbf{H}_g^H \mathbf{S}_{Tg} = \begin{bmatrix} \sqrt{\lambda_{g1}} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_{g2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_{gK_g}} \end{bmatrix} = \mathbf{A}_g \quad (16)$$

$$\mathbf{S}_{Rg}^H \mathbf{H}_g^H \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c = \mathbf{A}'_g \mathbf{D}_{Tg}^H \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c \quad (17)$$

根据式(16)和式(17)，将式(5)中的接收信号简化为

$$\mathbf{y}_g = \mathbf{A}_g \mathbf{W}_g \mathbf{d}_g + \mathbf{A}'_g \mathbf{D}_{Tg}^H \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c + \mathbf{S}_{Rg}^H \mathbf{n}_g \quad (18)$$

其中，式(18) RHS第2项可以进一步展开并简化，即

$$\mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c = \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj} \quad (19)$$

进一步，将式(18)中矩阵 $\mathbf{D}_{Tg}$ 按列拆分为 $(\mathbf{d}_{Tg}^{(1)}, \mathbf{d}_{Tg}^{(2)}, \dots, \mathbf{d}_{Tg}^{(N)})$ ，其中 $\mathbf{d}_{Tg}^{(i)} (1 \leq i \leq N)$ 是矩阵 $\mathbf{D}_{Tg}$ 的第 $i$ 列向量。由于 $\mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c = \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj}$ 是列向量， $(\mathbf{d}_{Tg}^{(i)})^H$ 是行向量，因此二者相乘得到 $(\mathbf{d}_{Tg}^{(i)})^H \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj}$ ，即 $\mathbf{D}_{Tg}^H \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c$ 的第 $i$ 个元素， $\mathbf{D}_{Tg}^H \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c$ 项进而可展开为

$$\begin{aligned} \mathbf{D}_{Tg}^\dagger \sum_{c \neq g} \mathbf{S}_{Tc} \mathbf{W}_c \mathbf{d}_c &= \begin{pmatrix} (\mathbf{d}_{Tg}^{(1)})^H \\ (\mathbf{d}_{Tg}^{(2)})^H \\ \vdots \\ (\mathbf{d}_{Tg}^{(N)})^H \end{pmatrix} \\ &\cdot \sum_{c \neq g} \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj} \\ &= \begin{pmatrix} (\mathbf{d}_{Tg}^{(1)})^H \sum_{c \neq g} \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj} \\ (\mathbf{d}_{Tg}^{(2)})^H \sum_{c \neq g} \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj} \\ \vdots \\ (\mathbf{d}_{Tg}^{(N)})^\dagger \sum_{c \neq g} \sum_{j=1}^{K_c} \sqrt{P_t \beta_{cj}} \mathbf{s}_{Tc}^{(j)} d_{cj} \end{pmatrix} \quad (20) \end{aligned}$$

结合式(19)和式(20)，用户 $g$ 第 $k$ 根天线处的接收信号可表示为



$$\begin{aligned} \mathbf{y}_{gk} &= \sqrt{\lambda_{gk} P_t \beta_{gk}} \mathbf{d}_{gk} \\ &+ \sum_{c \neq g} \sum_{j=1}^{K_c} \sqrt{\lambda_{gk} P_t \beta_{cj}} \left( \mathbf{d}_{Tg}^{(k)} \right)^H \mathbf{d}_{Tc}^{(j)} \mathbf{d}_{cj} \\ &+ \left( \mathbf{s}_{Rg}^{(k)} \right)^H \mathbf{n}_g \end{aligned} \quad (21)$$

对比式(21)和式(6)中接收信号的表达式，可以观察到式(6)中RHS第2项，即用户内干扰已经得到消除。另外，注意到式(21) RHS第2项满足

$$\left( \mathbf{d}_{Tg}^{(k)} \right)^H \mathbf{d}_{Tc}^{(j)} \leq \left( \mathbf{d}_{Tg}^{(k)} \right)^H \mathbf{d}_{Tg}^{(k)} \left( \mathbf{d}_{Tc}^{(j)} \right)^H \mathbf{d}_{Tc}^{(j)} = 1 \quad (22)$$

这表明该预编码方案也有效降低了用户间干扰。特别地，当各用户间的信道差异显著，即 $\mathbf{d}_{Tg}^{(k)}$ 和 $\mathbf{d}_{Tc}^{(j)}$ 趋于线性无关时，由柯西-施瓦茨不等式可得 $\left( \mathbf{d}_{Tg}^{(k)} \right)^H \mathbf{d}_{Tc}^{(j)}$ 项趋于0，表明该情形下所设计的预编码方案能够最大限度降低用户间干扰。此外，分析复杂度可知，所设计的预编码方案时间复杂度为 $O\left(N^2 \sum_{c=1}^M K_c + \sum_{c=1}^M (K_c)^3\right)$ ，相比于全局SVD预编码方案的 $O\left(N^2 \sum_{c=1}^M K_c + \left(\sum_{c=1}^M K_c\right)^3\right)$ 时间复杂度<sup>[26]</sup>降低了 $O\left(\sum_{i=1}^M \sum_{j=1}^M \sum_{l=1}^M K_i K_j K_l - \sum_{c=1}^M K_c^3\right)$ ，具有显著的低复杂度优势。在基站端通过预编码设计有效抑制干扰之后，该文将进一步针对MMR问题进行求解，在3.2节中对香农容量项进行凸松弛

$$\begin{aligned} \text{(SP1.1): } \max_{\beta} R_{\min}(\beta) &= \min_{1 \leq g \leq M} \left( \sum_{k=1}^{K_g} \log_2(1 + t_{gk}) - Q^{-1}(\varepsilon_g) \sqrt{\frac{V_g(\beta)}{nK_g}} \log_2 e \right) \\ \text{s.t. } &0 \leq \beta_{g,k} \leq 1 (1 \leq g \leq M, 1 \leq k \leq K_g), \\ &\sum_{g=1}^M \sum_{k=1}^{K_g} \beta_{gk} = 1, \\ &t_{gk} \lambda_{gk} \sum_{c \neq g} \sum_{j=1}^{K_c} \beta_{cj} \left( \left( \mathbf{s}_{Tg}^{(k)} \right)^H \mathbf{s}_{Tc}^{(j)} \right)^2 + c_{gk} t_{gk} - \lambda_{gk} \beta_{gk} \leq 0 \end{aligned} \quad (24)$$

由式(24)可知所述凸松弛的引入解决了目标函数中香农容量项的非凸性。然而，由于 $t_{gk} \beta_{cj}$ 是二元变量的乘积，式(23)仍为非凸约束。为此，本节利用分段McCormick包络(Piecewise McCormick Envelope, PME)<sup>[26,27]</sup>进一步将其转化为凸约束。引入辅助变量 $\mu_{(g,k,c,j)} = t_{gk} \beta_{cj}$ ，将问题(SP1.1)重构为

$$\begin{aligned} \text{(SP1.2): } \max_{\beta} R_{\min}(\beta) \\ &= \min_{1 \leq g \leq M} \left( \sum_{k=1}^{K_g} \log_2(1 + t_{gk}) \right. \\ &\quad \left. - Q^{-1}(\varepsilon_g) \sqrt{\frac{V_g(\beta)}{nK_g}} \log_2 e \right) \end{aligned} \quad (25a)$$

处理并重构原问题(OP1)。本节设计的预编码方案具有一般性。根据Jiang等人<sup>[19]</sup>的结论，当发射天线数小于等于用户 $g$ 的接收天线数时，式(12)–式(22)同样成立。两者的区别仅在于发送端天线数大于等于接收端天线数时，最大非零奇异值数目为 $K_g$ ；而发送端天线数小于接收端天线数时，最大非零奇异值数目为 $N$ 。

### 3.2 香农容量凸松弛与原问题等效

优化问题(OP1)中的约束均为凸约束，而非凸性由目标函数中香农容量和信道色散相关项引入，这使得该问题难以直接求解。因此，本节引入一种针对香农容量项的凸松弛方法，重构问题(OP1)。首先考虑问题(OP1)中目标函数的香农容量项，即 $\log_2(1 + \gamma_{gk})$ ，由于此项为关于功率分配因子向量 $\beta$ 的非凸函数。为了解决此非凸问题，引入辅助变量 $t_{gk}$  ( $1 \leq g \leq M, 1 \leq k \leq K_g$ ) 对问题(OP1)进行变换。为保证问题结构的等价性， $t_{gk}$ 的值应满足 $t_{gk} \leq \beta_{gk}$ ，利用此条件得到

$$t_{gk} \lambda_{gk} \sum_{c \neq g} \sum_{j=1}^{K_c} \beta_{cj} \left( \left( \mathbf{s}_{Tg}^{(k)} \right)^H \mathbf{s}_{Tc}^{(j)} \right)^2 + c_{gk} t_{gk} - \lambda_{gk} \beta_{gk} \leq 0 \quad (23)$$

结合辅助变量 $t_{gk}$  ( $1 \leq g \leq M, 1 \leq k \leq K_g$ ) 和式(23)，优化问题(OP1)可等价表示为

$$\text{s.t. } 0 \leq \beta_{g,k} \leq 1 (1 \leq g \leq M, 1 \leq k \leq K_g) \quad (25b)$$

$$\sum_{g=1}^M \sum_{k=1}^{K_g} \beta_{gk} = 1 \quad (25c)$$

$$\begin{aligned} \lambda_{gk} \sum_{c \neq g} \sum_{j=1}^{K_c} \mu_{(g,k,c,j)} \left( \left( \mathbf{s}_{Tg}^{(k)} \right)^H \mathbf{s}_{Tc}^{(j)} \right)^2 + c_{gk} t_{gk} \\ - \lambda_{gk} \beta_{gk} \leq 0 \end{aligned} \quad (25d)$$

$$\mu_{(g,k,c,j)} \geq \beta_{cj} t_{gk,n}^L + \beta_{cj}^L t_{gk} - \beta_{cj}^L t_{gk,n}^L \quad (25e)$$

$$\mu_{(g,k,c,j)} \geq \beta_{cj} t_{gk,n}^U + \beta_{cj}^U t_{gk} - \beta_{cj}^U t_{gk,n}^U \quad (25f)$$

$$\mu_{(g,k,c,j)} \leq \beta_{cj} t_{gk,n}^L + \beta_{cj}^U t_{gk} - \beta_{cj}^U t_{gk,n}^L \quad (25g)$$

$$\mu_{(g,k,c,j)} \leq \beta_{cj} t_{gk,n}^U + \beta_{cj}^L t_{gk} - \beta_{cj}^L t_{gk,n}^U \quad (25h)$$

$$t_{gk,n}^L \leq t_{gk} \leq t_{gk,n}^U, \beta_{cj,n}^L \leq \beta_{cj} \leq \beta_{cj,n}^U \quad (25i)$$

$$t_{gk,n}^L = t_{gk}^L + (t_{gk}^U - t_{gk}^L) \cdot (n-1)/N \quad (25j)$$

$$t_{gk,n}^U = t_{gk}^L + (t_{gk}^U - t_{gk}^L) \cdot (n)/N \quad (25k)$$

其中  $t_{gk}^L = 0$ ,  $\beta_{cj}^L = 0$  为变量  $t, \beta_{c,j}$  取值下限,  $t_{gk}^U = \hat{\gamma}_{gk}$ ,  $\beta_{cj}^U = 1$  为变量  $t, \beta_{c,j}$  取值上限, 将变量的定义域均等划分为  $N$  个子区间后,  $t_{gk,n}^L$  表示第  $n$  个子区间取值的下限,  $t_{gk,n}^U$  表示第  $n$  个子区间取值上限, 且  $\exists n \in (1, N)$  使得式(25e)–式(25k)同时成立。注意到由于PME的引入, 非凸约束项式(23)等价转化为凸约束项式(25d)–式(25k)。因此, 优化问题(SP1.2)中仅剩余目标函数的信道色散相关项非凸。针对此复合非凸函数, 下一节将进一步引入一种凸近似方法实现信道色散项的精确近似。

### 3.3 基于连续凸近似的求解算法

本节首先对复杂非凸的信道色散项凸近似处理, 将原问题重构为子凸问题, 进而提出一种基于连续凸近似的功率分配算法进行求解。

结合式(8)和式(10), 注意到信道色散相关项  $\sqrt{V_g(\beta)}/K_g$  涉及多变量耦合, 难以直接处理。为此, 给出命题:  $e_g(\gamma) = \sqrt{V_g(\beta)}/K_g$  是关于  $\gamma_{gk}$ ,

$$\gamma_{gk} = \frac{\lambda_{gk} \beta_{gk}}{\varphi_{gk}(\beta) + c_{gk}} \leq \frac{\lambda_{gk} \left( \beta_{gk}^{(m)} + \frac{\beta_{gk}^2}{\beta_{gk}^{(m)}} \right)}{\varphi_{gk}(\beta) + c_{gk}} = \frac{\lambda_{gk} \left( \beta_{gk}^{(m)} \right)}{\varphi_{gk}(\beta) + c_{gk}} + \frac{\lambda_{gk} \beta_{gk}^2}{(\varphi_{gk}(\beta) + c_{gk}) \beta_{gk}^{(m)}} = \hat{\gamma}_{gk}^{(m)}(\beta) \quad (28)$$

其中  $\hat{\gamma}_{gk}^{(m)}(\beta)$  为关于  $\beta$  的联合凸函数。

该命题证明如下: 要证该命题成立, 只需证明  $f(\beta) = \beta_j^2 / (\sum_{i=1}^n \beta_i a_i + c)$  是关于  $\beta$  的联合凸函数。为此, 计算  $f(\beta)$  的2阶偏导数

$$\frac{\partial^2 f(\beta)}{\partial \beta_j^2} = \frac{2 \left( \sum_{i=1}^n a_i \beta_i + c \right)^2 - 2a_j \left( 2\beta_j \sum_{\substack{i=1 \\ i \neq j}}^n a_i \beta_i + a_j \beta_j^2 + 2c\beta_j \right)}{\left( \sum_{i=1}^n a_i \beta_i + c \right)^3} = \frac{2 \left( \sum_{\substack{i=1 \\ i \neq j}}^n a_i \beta_i + c \right)^2}{\left( \sum_{i=1}^n a_i \beta_i + c \right)^3} \quad (29)$$

$$\frac{\partial^2 f(\beta)}{\partial \beta_k \partial \beta_j} = \frac{2\beta_k a_k \left( \sum_{i=1}^n a_i \beta_i + c \right) - 2 \left( 2\beta_j \sum_{\substack{i=1 \\ i \neq j}}^n a_i \beta_i + a_j \beta_j^2 + 2c\beta_j \right)}{\left( \sum_{i=1}^n a_i \beta_i + c \right)^3} = \frac{-2\beta_j a_k \left( \sum_{\substack{i=1 \\ i \neq j}}^n a_i \beta_i + c \right)}{\left( \sum_{i=1}^n a_i \beta_i + c \right)^3} (k \neq j) \quad (30)$$

$\forall 1 \leq k \leq K_g, 1 \leq g \leq M$  的联合凸函数。

该命题证明如下: 由  $\partial^2(1/(1+\gamma_{gk})^2)/\partial x^2 = 6/(1+\gamma_{gk})^6 \geq 0$  可得  $1/(1+\gamma_{gk})^2$  是关于  $\gamma_{gk}$  的凸函数。进一步得到  $f_{gk}(\gamma) = \sum_{k=1}^{K_g} 1/K_g(1+\gamma_{gk})^2$  是关于  $\gamma = (\gamma_{1,1}, \gamma_{1,2}, \dots, \gamma_{1,K_1}, \dots, \gamma_{M,K_M})$  的联合凸函数。由于  $-\sqrt{1-x}$  是关于  $x$  的凸函数, 且单调递增, 根据保凸运算定理<sup>[28]</sup>,  $e_g(\gamma)$  是关于  $\gamma$  的联合凸函数, 上述命题得证。

基于此命题, 对凸函数  $e_g(\gamma)$  在  $\gamma^{(m)}$  处1阶泰勒展开, 得到

$$e_g(\gamma) \geq e_g(\gamma^{(m)}) + \sum_{k=1}^{K_g} \frac{\gamma_{gk} - \gamma_{gk}^{(m)}}{K_g(1+\gamma_{gk}^{(m)})^3} e_g(\gamma^{(m)}) = e_g^{(m)}(\gamma^{(m)}) \quad (26)$$

信道色散项  $e_g(\gamma)$  可由式(26)给出的函数  $e_g^{(m)}(\gamma^{(m)})$  有效近似。进一步利用函数  $e_g^{(m)}(\gamma^{(m)})$ , 问题(SP1.2)中的可达速率项可近似为

$$\hat{R}_g^{(m)}(\beta) = \log_2(1+t_{gk}) + \frac{Q^{-1}(\varepsilon_g) e_g^{(m)}(\beta^{(m)})}{\sqrt{n}} \log_2 e \quad (27)$$

考虑到  $\gamma_{gk}$  为非线性分式, 需要进一步凸近似处理。为此, 给出如下命题:  $\gamma_{gk}$  在  $\beta_{gk}^{(m)}$  处的上界可由  $\hat{\gamma}_{gk}^{(m)}(\beta)$  给出, 具体定义为

$$\frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_l} = \frac{2a_l a_k \beta_j^2}{\left(\sum_{i=1}^n a_i \beta_i + c\right)^3} (k \neq j, l \neq j). \quad (31)$$

则  $f(\boldsymbol{\beta})$  的黑塞矩阵可表示为  $\mathbf{A} = \mathbf{A}_0 / \left(\sum_{i=1}^n a_i \beta_i + c\right)^3$ ，其中  $\mathbf{A}_0$  定义为

$$\mathbf{A}_0 = \begin{bmatrix} a_1^2 \beta_j^2 & \cdots & -\beta_j a_1 \left(\sum_{i \neq j}^n a_i \beta_i + c\right) & \cdots & a_1 a_n \beta_j^2 \\ a_1 a_2 \beta_j^2 & \cdots & -\beta_j a_2 \left(\sum_{i \neq j}^n a_i \beta_i + c\right) & \cdots & a_2 a_n \beta_j^2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_1 a_n \beta_j^2 & \cdots & -\beta_j a_n \left(\sum_{i \neq j}^n a_i \beta_i + c\right) & \cdots & a_n^2 \beta_j^2 \end{bmatrix}. \quad (32)$$

为证明矩阵  $\mathbf{A}$  半正定，引入变量  $b_k = a_k \beta_j$  ( $k \in (1, n), k \neq j$ )， $b_j = -\left(\sum_{i=1}^n a_i \beta_i + c\right)$ ，将矩阵  $\mathbf{A}$  等价表示为

$$\mathbf{A}_0 = \mathbf{b} \mathbf{b}^T \geq 0, \quad \mathbf{b} = (b_1, b_2, \dots, b_n)^T \quad (33)$$

$$\mathbf{A} = \frac{\mathbf{A}_0}{\left(\sum_{i=1}^n a_i \beta_i + c\right)^3} \geq 0 \quad (34)$$

结合式(33)和式(34)可知， $\mathbf{A}$  为半正定阵，则  $f(\boldsymbol{\beta})$  是关于  $\boldsymbol{\beta}$  的联合凸函数。进而可证明  $\hat{\gamma}_{gk}^{(m)}(\boldsymbol{\beta})$  是关于  $\boldsymbol{\beta}$  的联合凸函数。命题证毕。

根据式(26)和式(28)，在  $\boldsymbol{\beta}^{(m)}$  点信道色散项的凸近似可由  $\hat{e}_g^{(m+1)}(\boldsymbol{\beta}^{(m)})$  给出，即

$$\begin{aligned} e_g(\boldsymbol{\beta}) &\geq e_g(\boldsymbol{\beta}^{(m)}) + \sum_{k=1}^{K_g} \frac{\hat{\gamma}_{gk}^{(m)} - \gamma_{gk}}{K_g (1 + \gamma_{gk}^{(m)})^3} e_g(\boldsymbol{\beta}^{(m)}) \\ &= \hat{e}_g^{(m+1)}(\boldsymbol{\beta}^{(m)}). \end{aligned} \quad (35)$$

利用  $\hat{e}_g^{(m+1)}(\boldsymbol{\beta}^{(m)})$  近似问题(SP1.2)中的  $e_g(\boldsymbol{\beta})$  项，将问题(SP1.2)的目标函数近似为

$$\hat{R}_g^{(m+1)}(\boldsymbol{\beta}) = \log_2(1 + t_{gk}) + \frac{Q^{-1}(\varepsilon_g) \hat{e}_g^{(m+1)}(\boldsymbol{\beta}^{(m)})}{\sqrt{n}} \log_2 e \quad (36)$$

基于上式，在任一可行局部点  $\boldsymbol{\beta}^{(m)}$  处，可构建问题(OP1)的凸子问题为

$$(\text{SP1.3}) : \max_{\boldsymbol{\beta}} R_{\min}^{(m+1)}(\boldsymbol{\beta}) = \min_{1 \leq g \leq M} \hat{R}_g^{(m+1)}(\boldsymbol{\beta}) \quad (37a)$$

$$\text{s.t. (20b)} \quad (37b)$$

根据SCA的基本原理<sup>[29]</sup>，可基于该凸子问题(SP1.3)对原问题(OP1)进行迭代求解。在第  $m+1$  次迭代中，所设计的功率分配算法基于式(37)构建凸子问题并进行求解，并得到此次迭代中的功率系数分配最优解  $\boldsymbol{\beta}^{(m+1)}$ ，对应的SINR为  $\boldsymbol{\gamma}^{(m+1)}$ 。进而，在下次(即第  $m+2$  次)迭代中，算法以  $\boldsymbol{\beta}^{(m+1)}$

作为凸近似的局部点，再次基于式(37)构建凸子问题进行求解，如此往复。依此，算法将原始非凸优化问题转化为包含若干个凸子问题的序列，不断地求解序列中的凸子问题并基于当次迭代的解构建下次迭代中的子问题，直至算法前后两次迭代所得解的差小于给定阈值。

注意到在第  $m+1$  次迭代过程中， $\boldsymbol{\beta}^{(m+1)}$  是问题(SP1.2)的最优解，而  $\boldsymbol{\beta}^{(m)}$  问题(SP1.2)的一个可行解，则根据最优性定义可得

$$\begin{aligned} R_{\min}(\boldsymbol{\beta}^{(m+1)}) &\geq R_{\min}^{(m+1)}(\boldsymbol{\beta}^{(m+1)}) \geq R_{\min}^{(m+1)}(\boldsymbol{\beta}^{(m)}) \\ &= R_{\min}(\boldsymbol{\beta}^{(m)}) \end{aligned} \quad (38)$$

式(38)保证了所设计的MMR功率分配算法通过迭代至少能收敛到原始问题的局部最优解。算法1给出了该功率分配算法的伪代码。

## 4 仿真结果与分析

本章通过仿真结果评估所提优化算法的性能。仿真参数设置如下：基站端天线空间相关性和用户端天线空间相关性分别设置为  $\rho_0 = 0.4$ ， $\rho_1 = 0.2$ 。信道噪声满足  $\mathbb{E}(n_{gi}^2) = 1$  ( $1 \leq g \leq M, 1 \leq i \leq K_g$ )，信道莱斯因子  $L = 1$ ，码长  $n = 200$ ，发射平均功率设置为  $P_t/N = 1$ 。对于所提MMR功率分配算法，将其收敛阈值设置为  $\varepsilon = 0.05$ 。

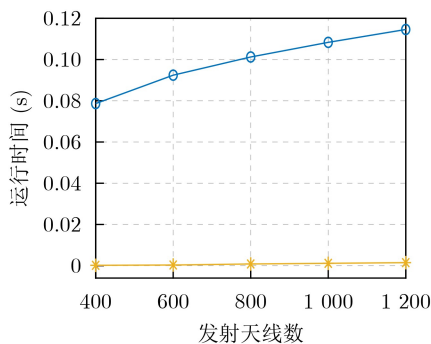
为验证所提局部SVD预编码算法的低复杂度优势，除了3.1节对两种预编码方案进行复杂度分析外，本节将局部SVD预编码与全局SVD预编码的运行时间进行对比。本章进一步引入香农速率最大化<sup>[17]</sup>、大规模MIMO-URLLC<sup>[30]</sup>两种对比方案评估香农容量凸松弛的精度。此外，为验证所提MMR功率分配算法得到的次优解性能，本节将其与遍历搜索以及多用户单天线速率最大化算法<sup>[15]</sup>进行对比分析。上述对比方案简要介绍如下：(1)香农速率最大化算法通过引入凸函数，并将其进行一阶泰勒展开，得到香农容量项的近似函数，基于此近似函数替换原问题中香农容量项，将原问题转换为可直

接求解的凸问题序列, 通过迭代求解凸问题序列, 得到MMR问题的解; (2)大规模MIMO-URLLC算法分别对原问题优化目标中香农容量项和信道色散相关项进行不同的凸近似处理, 一方面引入凸近似函数替换香农容量项, 另一方面针对信道色散相关项复杂非凸的问题, 引入辅助变量并对其进行两次近似处理, 从而将原始非凸问题转换为凸问题, 提出一种路径跟踪算法迭代求解; (3)多用户单天线速率最大化算法以预编码向量和功率分配因子向量为优化变量, 构建MMR问题。该算法以迭代的方式交替优化预编码向量和功率分配因子向量, 获得MMR问题的有效解。具体而言, 该算法利用最小均方误差准则推导出功率分配因子为定值时预编码向量的最优解。基于此最优解, 进一步引入辅助变量并将MMR问题重构为凸问题, 求出功率分配因子的有效解, 并通过迭代的方式交替优化预编码向量和功率分配因子向量, 直至算法收敛到指定阈值。

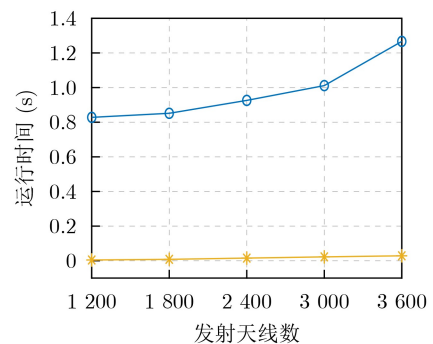
首先, 为了显示所设计局部预编码方案的低复杂度优势, 给出该局部SVD方案与全局预编码方案的运行时间对比图, 如图2(a)和图2(b)所示, 其中用户数 $M = 6$ ,  $K_i$ 表示第 $i$ 个用户的天线数。从图2(a)

算法 1 基于连续凸近似的MMR功率分配算法

- (1) 初始化功率分配因子向量 $\beta^{(0)}$ , 设置阈值 $\varepsilon$ ;
- (2) 初始化迭代次数 $m = 0$ , 根据 $\beta^{(0)}$ 计算 $\gamma_{gk}$ , 得到 $\gamma_{gk}^{(0)}$ ;
- (3) 根据 $\beta_{gk}^{(0)}$ ,  $\gamma_{gk}^{(0)}$ 求解凸子优化问题(SP1.3), 得到 $\gamma_{gk}^{(1)}$ ,  $\beta_{gk}^{(1)}$ ,  $R_{\min}^{(1)}$ ;
- (4) for  $m = 2, 3, \dots$
- (5) 根据 $\beta_{gk}^{(m-1)}$ ,  $\gamma_{gk}^{(m-1)}$ 求解优化问题(SP1.3), 得到 $\gamma_{gk}^{(m)}$ ,  $\beta_{gk}^{(m)}$ ,  $R_{\min}^{(m)}$ ;
- (6) if  $\|R_{\min}^{(m)} - R_{\min}^{(m-1)}\| \geq \varepsilon$ , then 更新 $m \leftarrow m + 1$ 再次进入步骤5并继续计算;
- (7) else 根据 $\beta^{(m)}$ 计算目标函数 $R_{\min}(\beta)$ , 得到问题(OP1)的有效解, 终止迭代并返回结果;
- (8) end



(a) 低数量级天线运行时间,  $K_i=60$



(b) 高数量级运行天线时间,  $K_i=60$

图2 预编码运行时间

观察到, 该文设计的局部SVD预编码运行时间远小于全局SVD预编码运行时间, 且随着发射天线数的增加, 全局SVD预编码运行时间显著增加。而在用户数以及用户天线数较多的场景中, 如图2(b)所示, 局部SVD预编码方案运行时间始终较短, 而全局SVD预编码方案运行时间快速增加, 难以适用于大规模MIMO辅助的超低时延通信场景, 并验证了局部SVD预编码方案的低复杂度特性。

为了评估功率分配算法所得次优解的性能, 图3分析了MMR功率分配算法相比于遍历搜索的性能差距, 其中 $M$ 表示用户数,  $K_g$ 表示第 $g$ 个用户的天线数。由图3观察到, 在不同的发射天线数下, 该文提出的面向用户公平性的MMR功率分配算法所得次优解均具有和遍历搜索得到的全局最优解接近的性能, 仅有1.2%~2.1%的性能损失, 验证所提MMR功率分配算法的高效性。

图4展示了基站天线数对MMR以及系统总吞吐量的影响, 并对比了所设计局部SVD预编码方案与传统SVD线性预编码方案的性能表现。结果显示, 通过局部SVD预编码方案得到的MMR随发射端天线数呈现线性增加的趋势, 而SVD线性预编码方案则始终接近于0。这是因为SVD线性预编码方案需要对基站与所有用户之间的信道矩阵进行SVD分解, 而大规模MIMO系统中天线的差异性、空间相关性导致各个流的信道增益差异化严重。在数值上, 这表明信道矩阵 $H$ 经过SVD后, 有一部分数据流对应的奇异值较小, 甚至趋近于0。此情况下为保障用户间公平性, SVD线性预编码方案在处理MMR问题时, 需要将大部分功率分配给信道增益较差的流。由于该部分流对应的奇异值(即式(7)中的 $\sqrt{\lambda_{gk}}$ )趋近于0, 因此分配功率后其SINR接近于0, 最终导致其MMR性能接近于0。这也反映出SVD线性预编码方案在处理MMR问题时, 系统资源未得到充分利用。另外, 在以总吞吐量为优化目标时, 线性预编码则具有更好的表现。由于SVD线



性预编码方案可完全消除用户干扰，当部分用户信道质量较好时，其SINR可以近似为SNR。此时，功率资源被完全分配至该部分流上，以尽可能提高总吞吐量。与SVD线性预编码方案相比，该文的预编码方案在处理总吞吐量最大化问题时仅损失11%的性能，而在MMR指标上，所提局部SVD预编码方案展现了显著的性能优势，进一步表明所提预编码方案适用性强的优点。

图5(a)和图5(b)评估了香农容量凸松弛的性能，其中用户数 $M$ 均为5，图5(a)中基站天线数 $N = 256$ ，图5(b)中 $K_g = 32(1 \leq g \leq M)$ 。观察图5(a)和图5(b)可知信道最差用户香农容量随着发射端与接收端天线的增加，均呈现近似线性增长的趋势，且接收端天线数增加带来更大的系统性能增益，表明该场景下优先提高用户端天线数对系统而言具有更高收益。此外，图5(a)和图5(b)中，香农速率最大化算法相比于大规模MIMO-URLLC算法具有性能优势，而所提优化算法在相同指标上进一步超越了香农速率最大化算法，具有三者中最佳的性能。特别地，由图中的曲线趋势可知，随着发射天线数和用户天线数的增加，本文提出的算法性能优势更为显著，表明本文所提算法能充分提升分集增益。进一步验证了所提优化算法的优越性。

图6给出了几种优化方案下MMR与码长的关系，其中用户数 $M = 5$ ，每个用户天线数 $K_g = 32(1 \leq g \leq M)$ ，基站天线数 $N = 256$ 。由图可知，随着码长的增加，可达速率呈现对数增长的趋势。在码长较短的情况下，MMR增幅更为明显，表明有限码长带来的系统性能损失更为严重。例如，相较于 $512 \leq n \leq 1024$ 区间，码长在 $64 \leq n \leq 256$ 区间时MMR具有明显更高的速率增益。此外，相比于香农速率最大化以及多用户单天线速率最大化算法，所提算法在不同码长下均具有明显性能优势。这表明所提算法在有限码长通信场景下的鲁棒性更好。特别的，在码长较短的情况下( $64 \leq n \leq 512$ )，所提算法性能优势更为显著。从图6进一步观察到，在码长较短的情况下，香农速率最大化算法的性能不及多用户单天线速率最大化算法。这是因为香农速率最大化算法基于香农域对用户公平性速率进行优化，在用于有限码长域下通信场景时，未考虑到有限长编码对系统性能的影响。而随着码长的增加，有限码长对用户可达速率造成的性能损失逐渐变小，香农速率最大化算法的总吞吐量也随之高于多用户单天线速率最大化算法。

图7展示了MMR与发送端天线平均功率的关系，其中用户数 $M = 5$ ，每个用户天线数 $K_g = 32$

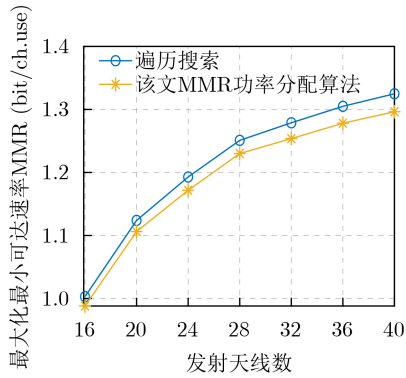


图3 次优解性能评估， $M = 3, K_g = 4(1 \leq g \leq M)$

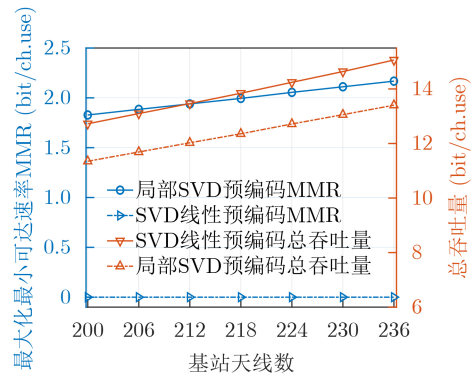
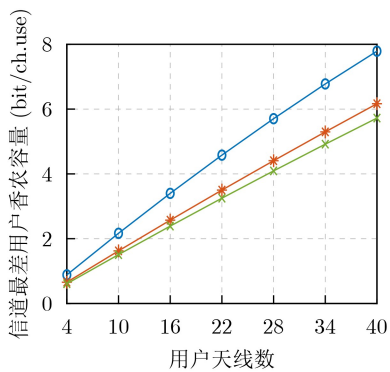
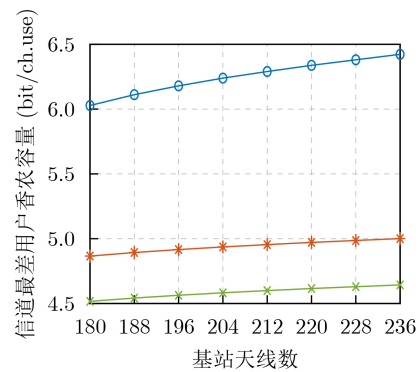


图4 MMR/总吞吐量变化， $M = 6, K_g = 32(1 \leq g \leq M)$



(a) 香农容量随用户天线数的变化



(b) 香农容量随基站端天线的变化

图5 最大最小香农容量随收发端天线数变化

( $1 \leq g \leq M$ ), 基站天线数  $N = 256$ 。注意到所提优化方案充分考虑了用户公平性, 从而保证信道质量较差的用户也可被分配一定功率资源, 因而该部分用户的可达速率依然可从总功率增加中获得增益, 致使图7中曲线呈现出对数增长趋势。另一方面, 在总功率增加的同时各用户受到的干扰也相应增加, 这导致MMR无法随平均功率线性增加。观察对比方案的性能可知, 相比于香农速率最大化及多用户单天线速率最大化算法, 所提算法具有最好的性能, 且在平均功率增加时性能提升增幅更大, 再次验证了次优解性的有效性, 并表明本文算法能更有效地利用系统功率资源。

图8(a), 图8(b)分别研究了基站天线数和用户天线数对MMR的影响, 其中  $M = 5K_g = 32$  ( $1 \leq g \leq M$ )。由图8(a)可观察到, 最大最小可达速率与发射端天线数几乎呈近似线性增长, 印证了发射端天线数增加对系统空间分集增益的提升。同时, 图8(a)也表明所提算法与两种对比方案相比具有更佳的性能, 且这一性能优势随基站天线数增加而逐渐增加, 表明所提算法在大规模天线场景下的适用性。图8(b)给出了最大最小可达速率与单个用户接收端天

线数的关系。可观察到最大最小可达速率随用户天线数的增加而增加, 再次验证了空间分集所产生的速率增益。同时, 以图8(b)为例, 计算曲线斜率得到本文高效算法、香农速率最大化以及多用户单天线速率最大化算法达到的分集增益分别为每个天线  $0.201 \text{ bit/ch.use}$ ,  $0.138 \text{ bit/ch.use}$ ,  $0.11 \text{ bit/ch.use}$ , 表明所提算法相比于两种对比方案能带来更高的分集增益。

图9研究了系统接入用户数量对MMR的影响, 其中基站天线数  $N = 256$ , 每个用户天线数  $K_g = 16$  ( $1 \leq g \leq M$ )。由图可知, 在总功率不变的情况下, MMR随着接入用户数的增加而减小。这是因为用户数增加, 一方面导致分配给单个用户的功率逐渐减少, 另一方面产生了更多的用户间干扰, 综合作用使得系统MMR明显下降。此外, 所提MMR功率分配算法与两种对比方案相比, 在不同用户数下均有一定的性能优势。

## 5 结束语

该文面向MIMO辅助的URLLC下行通信系统, 对用户通信速率公平性的优化展开研究。基于此场

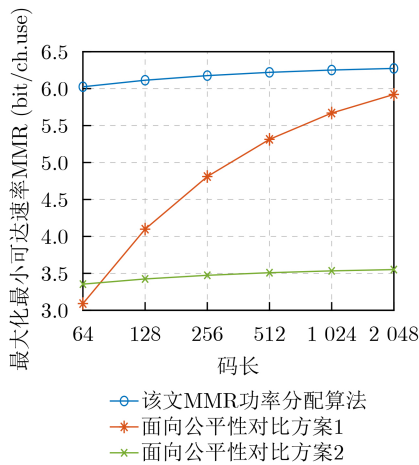


图6 最大化最小可达速率随码长的变化

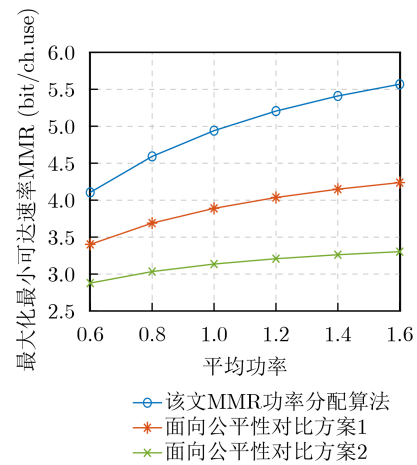
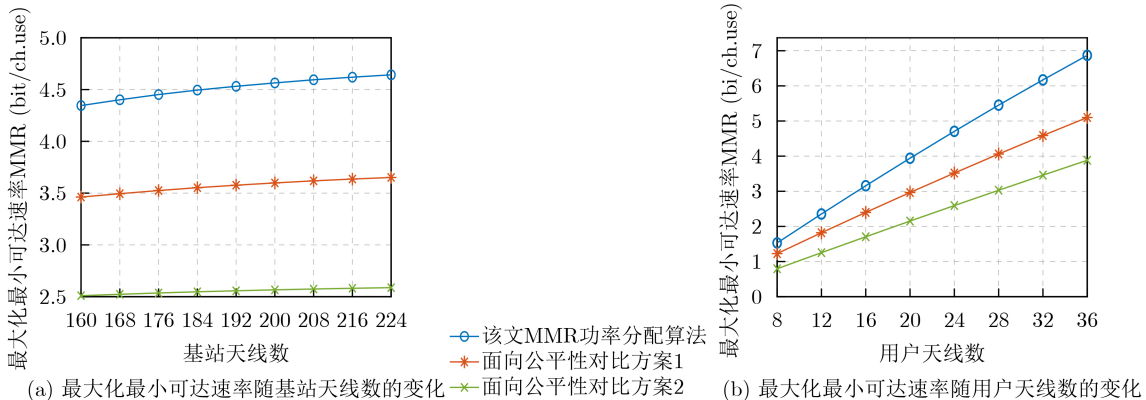


图7 最大化最小可达速率随平均功率的变化,  $N = 256$



(a) 最大化最小可达速率随基站天线数的变化

(b) 最大化最小可达速率随用户天线数的变化

图8 最大最小可达速率随收发天线数变化

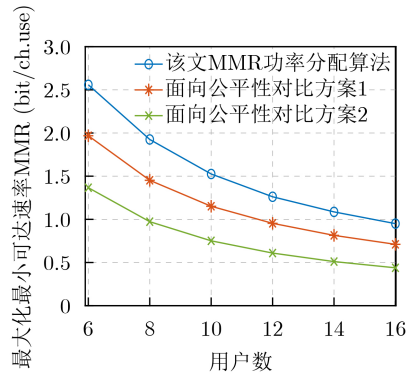


图9 最大化最小可达速率随用户数的变化

景，该文在FBL域下对用户可达速率进行建模，而在发送端总功率资源受限的条件下，建模了最大化所有用户最小可达速率的问题。为了解决该非凸问题，该文首先设计基于局部SVD的低复杂度预编码方案，在消除用户内干扰的同时减少用户间干扰。其次，为了解决原MMR优化问题的非凸性，通过引入辅助变量及PME对原问题进行凸松弛重构。随后对原优化目标函数中的信道色散项进行精准凸近似，得到原问题在可行局部点的凸子问题。进而，该文基于SCA算法对凸子问题迭代求解以获得原问题的次优解。最后，该文通过仿真结果验证了所提预编码设计和优化方案的优势，并探究了码长、基站天线数、用户天线数等参量对系统性能的影响。所提出的预编码方案和面向用户公平性的功率分配设计方法具有一定的可扩展性，如可以拓展到多小区MIMO系统的能效优化、用户间最小速率提升、总吞吐量提升等场景。此外，所提出针对FBL吞吐量优化非凸问题解决思路，包括将问题解耦、辅助变量凸松弛以及基于连续凸近似的求解方案，可为类似多变量非凸FBL性能优化问题提供一定借鉴。

### 参考文献

- [1] HAMIDI-SEPEHR F, SAJADIEH M, PANTELEEV S, et al. 5G URLLC: Evolution of high-performance wireless networking for industrial automation[J]. *IEEE Communications Standards Magazine*, 2021, 5(2): 132–140. doi: [10.1109/MCOMSTD.001.2000035](https://doi.org/10.1109/MCOMSTD.001.2000035).
- [2] International Telecommunication Union. IMT Vision—Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond[M]. Geneva: Electronic Publication, 2015.
- [3] ZHANG Zhengquan, XIAO Yue, MA Zheng, et al. 6G wireless networks: Vision, requirements, architecture, and key technologies[J]. *IEEE Vehicular Technology Magazine*, 2019, 14(3): 28–41. doi: [10.1109/MVT.2019.2921208](https://doi.org/10.1109/MVT.2019.2921208).
- [4] 袁俊杰, 李双洋, 种若汐, 等. 面向6G物联网的分布式译码技术[J]. *电子与信息学报*, 2021, 43(1): 21–27. doi: [10.11999/JEIT200343](https://doi.org/10.11999/JEIT200343).
- [5] YUAN Weijie, LI Shuangyang, CHONG Ruoxi, et al. A distributed decoding algorithm for 6G internet-of-things networks[J]. *Journal of Electronics & Information Technology*, 2021, 43(1): 21–27. doi: [10.11999/JEIT200343](https://doi.org/10.11999/JEIT200343).
- [6] 黄崇文, 季然, 魏丽, 等. 面向全息MIMO 6G通信的电磁信道建模理论与方法[J]. *电子与信息学报*, 2024, 46(5): 1940–1950. doi: [10.11999/JEIT231219](https://doi.org/10.11999/JEIT231219).
- [7] HUANG Chongwen, JI Ran, WEI Li, et al. Electromagnetic channel modeling theory and approaches for holographic MIMO wireless communications[J]. *Journal of Electronics & Information Technology*, 2024, 46(5): 1940–1950. doi: [10.11999/JEIT231219](https://doi.org/10.11999/JEIT231219).
- [8] MA Wenyan, ZHU Lipeng, and ZHANG Rui. MIMO capacity characterization for movable antenna systems[J]. *IEEE Transactions on Wireless Communications*, 2024, 23(4): 3392–3407. doi: [10.1109/TWC.2023.3307696](https://doi.org/10.1109/TWC.2023.3307696).
- [9] CHI Yuhao, LIU Lei, SONG Lei, et al. Constrained capacity optimal generalized multi-user MIMO: A theoretical and practical framework[J]. *IEEE Transactions on Communications*, 2022, 70(12): 8086–8104. doi: [10.1109/TCOMM.2022.3207813](https://doi.org/10.1109/TCOMM.2022.3207813).
- [10] LV Zhihan, QIAO Liang, and YOU I. 6G-enabled network in box for internet of connected vehicles[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(8): 5275–5282. doi: [10.1109/TITS.2020.3034817](https://doi.org/10.1109/TITS.2020.3034817).
- [11] WANG Bin, XU Ke, ZHENG Shilian, et al. A deep learning-based intelligent receiver for improving the reliability of the MIMO wireless communication system[J]. *IEEE Transactions on Reliability*, 2022, 71(2): 1104–1115. doi: [10.1109/TR.2022.3148114](https://doi.org/10.1109/TR.2022.3148114).
- [12] POLYANSKIY Y, POOR H V, and VERDU S. Channel coding rate in the finite blocklength regime[J]. *IEEE Transactions on Information Theory*, 2010, 56(5): 2307–2359. doi: [10.1109/TIT.2010.2043769](https://doi.org/10.1109/TIT.2010.2043769).
- [13] YANG Wei, DURISI G, KOCH T, et al. Quasi-static multiple-antenna fading channels at finite blocklength[J]. *IEEE Transactions on Information Theory*, 2014, 60(7): 4232–4265. doi: [10.1109/TIT.2014.2318726](https://doi.org/10.1109/TIT.2014.2318726).
- [14] ZHU Yao, HU Yulin, YUAN Xiaopeng, et al. Joint convexity of error probability in blocklength and transmit power in the finite blocklength regime[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(4): 2409–2423. doi: [10.1109/TWC.2022.3211454](https://doi.org/10.1109/TWC.2022.3211454).
- [15] ZHAO Linlin, YANG Shaoshi, CHI Xuefen, et al. Achieving energy-efficient uplink URLLC with MIMO-aided grant-free access[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(2): 1407–1420. doi: [10.1109/TWC.2021.3104043](https://doi.org/10.1109/TWC.2021.3104043).
- [16] PENG Qihao, REN Hong, PAN Cunhua, et al. Resource allocation for uplink cell-free massive MIMO enabled URLLC in a smart factory[J]. *IEEE Transactions on Communications*, 2023, 71(1): 553–568. doi: [10.1109/TCOMM.2022.3224502](https://doi.org/10.1109/TCOMM.2022.3224502).
- [17] HE Shiwen, AN Zhenyu, ZHU Jianyue, et al. Beamforming design for multiuser uRLLC with finite blocklength transmission[J]. *IEEE Transactions on Wireless*

- Communications*, 2021, 20(12): 8096–8109. doi: [10.1109/TWC.2021.3090197](https://doi.org/10.1109/TWC.2021.3090197).
- [16] FANG Hao, HU Han, ZHANG Yao, *et al.* Achievable rate analysis and power optimization for cell-free massive MIMO URLLC systems over aging and correlated channels[J]. *IEEE Internet of Things Journal*, 2024, 11(14): 25239–25250. doi: [10.1109/JIOT.2024.3392298](https://doi.org/10.1109/JIOT.2024.3392298).
- [17] SHENG Zhichao, TUAN H D, NASIR A, *et al.* Power allocation for energy efficiency and secrecy of wireless interference networks[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(6): 3737–3751. doi: [10.1109/TWC.2018.2815626](https://doi.org/10.1109/TWC.2018.2815626).
- [18] JIANG Hao, ZHANG Zaichen, WU Liang, *et al.* A non-stationary geometry-based scattering vehicle-to-vehicle MIMO channel model[J]. *IEEE Communications Letters*, 2018, 22(7): 1510–1513. doi: [10.1109/LCOMM.2018.2834366](https://doi.org/10.1109/LCOMM.2018.2834366).
- [19] HUANG Jie, WANG Chengxiang, CHANG Hengtai, *et al.* Multi-frequency multi-scenario millimeter wave MIMO channel measurements and modeling for B5G wireless communication systems[J]. *IEEE Journal on Selected Areas in Communications*, 2020, 38(9): 2010–2025. doi: [10.1109/JSAC.2020.3000839](https://doi.org/10.1109/JSAC.2020.3000839).
- [20] WANG Jun, WANG Chengxiang, HUANG Jie, *et al.* A general 3D space-time-frequency non-stationary THz channel model for 6G ultra-massive MIMO wireless communication systems[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(6): 1576–1589. doi: [10.1109/JSAC.2021.3071850](https://doi.org/10.1109/JSAC.2021.3071850).
- [21] YANG Guanghua, ZHANG Huan, SHI Zheng, *et al.* Asymptotic outage analysis of spatially correlated rayleigh MIMO channels[J]. *IEEE Transactions on Broadcasting*, 2021, 67(1): 263–278. doi: [10.1109/TBC.2020.3028346](https://doi.org/10.1109/TBC.2020.3028346).
- [22] JOUNG H, JO H S, MUN C, *et al.* Capacity loss due to polarization-mismatch and space-correlation on MISO channel[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(4): 2124–2136. doi: [10.1109/TWC.2014.031314.131079](https://doi.org/10.1109/TWC.2014.031314.131079).
- [23] LIU An, LAU V K N, and KANANIAN B. Stochastic successive convex approximation for non-convex constrained stochastic optimization[J]. *IEEE Transactions on Signal Processing*, 2019, 67(16): 4189–4203. doi: [10.1109/TSP.2019.2925601](https://doi.org/10.1109/TSP.2019.2925601).
- [24] ZHAN Jinlong and DONG Xiaodai. Interference cancellation aided hybrid beamforming for mmwave multi-user massive MIMO systems[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(3): 2322–2336. doi: [10.1109/TVT.2021.3057547](https://doi.org/10.1109/TVT.2021.3057547).
- [25] LONG Wenxuan, CHEN Rui, MORETTI M, *et al.* Joint spatial division and coaxial multiplexing for downlink multi-user OAM wireless backhaul[J]. *IEEE Transactions on Broadcasting*, 2021, 67(4): 879–893. doi: [10.1109/TBC.2021.3081869](https://doi.org/10.1109/TBC.2021.3081869).
- [26] FONTENLA-ROMERO O, PÉREZ-SÁNCHEZ B, and GUIJARRO-BERDIÑAS B. LANN-SVD: A non-iterative SVD-based learning algorithm for one-layer neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(8): 3900–3905. doi: [10.1109/TNNLS.2017.2738118](https://doi.org/10.1109/TNNLS.2017.2738118).
- [27] KOŁODZIEJ S, CASTRO P M, and GROSSMANN I E. Global optimization of bilinear programs with a multiparametric disaggregation technique[J]. *Journal of Global Optimization*, 2013, 57(4): 1039–1063. doi: [10.1007/s10898-012-0022-1](https://doi.org/10.1007/s10898-012-0022-1).
- [28] DE ASSIS L S, CAMPONOGARA E, ZIMBERG B, *et al.* A piecewise McCormick relaxation-based strategy for scheduling operations in a crude oil terminal[J]. *Computers & Chemical Engineering*, 2017, 106: 309–321. doi: [10.1016/j.compchemeng.2017.06.012](https://doi.org/10.1016/j.compchemeng.2017.06.012).
- [29] BOYD S and VANDENBERGHE L. *Convex Optimization*[M]. Cambridge: Cambridge University Press, 2004: 67–95.
- [30] NASIR A A, TUAN H D, NGO H Q, *et al.* Cell-free massive MIMO in the short blocklength regime for URLLC[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(9): 5861–5871. doi: [10.1109/TWC.2021.3070836](https://doi.org/10.1109/TWC.2021.3070836).

胡钰林：男，教授，博士生导师，研究方向为工业物联网、高可靠低时延通信、无人机通信、移动边缘计算等。

肖志成：男，硕士生，研究方向为高可靠低时延通信、多用户MIMO。

徐浩：男，硕士生，研究方向为高可靠低时延通信、移动边缘计算。

责任编辑：陈倩

## Efficient Power Allocation Algorithm for Throughput Optimization of Multi-User Massive MIMO Systems in Finite Blocklength Regime

HU Yulin    XIAO Zhicheng    XU Hao

(Electronic Information School, Wuhan University, Wuhan 430000, China)

### Abstract:

**Objective** The 6th Generation (6G) mobile communication network aims to provide Ultra-Reliable and Low Latency Communication (URLLC) services to a large number of nodes. To support URLLC for massive users, Multiple-In-Multiple-Out (MIMO) technology has become a key enabler for improving system performance in



6G. However, URLLC systems typically operate with Finite BlockLength (FBL) codes, which pose unique challenges for resource allocation design due to their deviation from traditional methods in the infinite blocklength regime. Although prior studies have explored resource allocation strategies for MIMO-assisted URLLC, power allocation design that considers user fairness remains unresolved. This paper proposes an efficient power allocation algorithm for multi-user MIMO systems in the FBL regime, addressing the issue of user fairness.

**Methods** This study investigates the MIMO-assisted URLLC downlink communication scenario. The system performance is first characterized based on FBL theory, revealing the achievable rate of MIMO downlink users, which introduces significant nonconvexity compared to the infinite blocklength regime. Given the base station's limited power resources, setting system throughput as the optimization objective fails to ensure user fairness. To address this, a Maximum Minimum Rate (MMR) optimization problem is formulated, with power allocation factors as the optimization variables, subject to a total power constraint. The formulated problem is highly nonconvex due to the nonconvex terms in the objective function. To develop an efficient power allocation design for the MMR problem, a low-complexity precoding strategy is first proposed to mitigate both inter-user and intra-user interference. This precoding strategy, based on the local Singular Value Decomposition (SVD) method, reduces complexity compared with the traditional global SVD precoding strategy and effectively suppresses interference. To address the nonconvexity introduced by the Shannon capacity and channel dispersion terms in the objective function, convex relaxation and approximation techniques are introduced. The convex relaxation involves auxiliary variables and piecewise McCormick envelopes to manage the Shannon capacity term, transforming the MMR problem into an optimization problem where only the channel dispersion term remains nonconvex. For the relaxed problem, the channel dispersion term is then approximated by an upper bound function, rigorously shown to be convex through analytical findings. Based on this convex relaxation and approximation, the original MMR problem can be approximated by a convex subproblem at a given feasible point and efficiently solved using the Successive Convex Approximation (SCA) algorithm. Convergence and optimality analyses of the proposed algorithm are provided.

**Results and Discussions** The proposed power allocation design is evaluated and validated through numerical simulations. To demonstrate the superiority of the proposed design, several benchmarks are introduced for comparison. For the proposed precoding design based on local SVD, the global SVD precoding method is included to validate its advantage in terms of complexity. The Shannon-rate-oriented rate maximization methods are also introduced to verify the accuracy of the proposed convex relaxation design. Moreover, the suboptimality of the SCA-based algorithm is validated through comparisons with other benchmarks, including the exhaustive search method. First, the low-complexity advantage of the proposed precoding design is demonstrated (Fig. 2). The proposed precoding strategy exhibits low complexity in both small-scale and large-scale user scenarios. The performance of the suboptimal solutions obtained using the proposed SCA-based algorithm is compared with the globally optimal solutions from the exhaustive search method (Fig. 3). The results confirm the accuracy of the proposed algorithm, with a performance loss of less than 3%. The effect of the base station's antenna number on both the MMR problem and the system throughput maximization problem is illustrated (Fig. 4), further validating the effectiveness and applicability of the local SVD precoding strategy. The tightness of the convex relaxation is examined (Fig. 5), confirming that the proposed design becomes more advantageous as the user antenna number increases. The effect of blocklength on MMR performance is explored (Fig. 6), while the variation trend of MMR with respect to average transmit power is shown (Fig. 7). The influence of antenna number at both base station and user sides on MMR performance is investigated (Fig. 8), while the effect of user number on MMR performance is demonstrated (Fig. 9).

**Conclusions** This paper investigates the optimization of user rate fairness in downlink MIMO-assisted URLLC scenarios. The system performance in the FBL regime is characterized, and based on this modeling, an MMR optimization problem is formulated under a sum power constraint, which is inherently nonconvex. To address this nonconvexity, a local SVD-based precoding design is proposed to reduce precoding complexity while ensuring fairness. Furthermore, convex relaxation is applied by introducing auxiliary variables and piecewise McCormick envelopes. The relaxed objective function is then approximated by an upper bound function, whose convexity is rigorously proven. Building on this relaxation and approximation, an SCA-based algorithm is developed to effectively solve the MMR problem. The proposed design is validated through numerical simulations, where its validity and parameter influence on system performance are discussed. The approach can be extended to other URLLC scenarios, such as multi-cell MIMO, and provides valuable insights for solving nonconvex optimization problems in related fields.

**Key words:** Ultra-Reliable and Low-Latency Communication (URLLC); Multi-user MIMO; Finite BlockLength (FBL); Successive Convex Approximation (SCA)