Active RIS-Assisted Multi-User Multi-Stream Transmit Precoding Relying on Scalable-Complexity Iterations

Y. Chen[1], H. D. Tuan[2], H. Yu[1], H. V. Poor[3], and L. Hanzo

***Abstract*—This is the first investigation focused on delivering multi-stream information to multiple multi-antenna users employing an active reconfigurable intelligent surface (aRIS)-assisted system. We conceive the joint design of the transmit precoders and of the aRIS's power-amplified reconfigurable elements (APRES) to enhance the log-det rate objective functions for all users, which poses large-scale mixed discrete continuous problems. We develop a max-min log-det solver, which iterates quadratic-solvers of cubic complexity to maximize the nonsmooth function representing the minimum of the users' log-det rate functions. To mitigate the computational burden associated with cubically escalating complexity in large-scale scenarios, we introduce a pair of alternative problems aimed at maximizing the smooth functions representing the sum of the users' log-det rate function (sum log-det) and the soft minimum of the users' log-det rate function (soft min log-det). We develop sum log-det and soft max-min solvers, leveraging closed-form expressions of scalable (linear) complexity for efficient computation. This approach ensures practicality in addressing large-scale scenarios. Furthermore, the soft min log-det enables us to enhance the log-det rates for all users and their sum, ultimately improving the quality of delivering multi-user multi-stream information.**

***Index Terms*—Active reconfigurable intelligent surface (aRIS), multi-user precoding, active power control, log-det matrix function optimization, large-scale computation, mixed discrete continuous optimization, log-det enhancement,**

I. Introduction

At the time of this writing, there is intense interest in utilizing metasurfaces as massive antenna arrays, owing to their promising potential in next-generation networks and beyond [1]–[10]. Nonetheless, the majority of existing research has been predominantly focused on simple scenarios, considering either a single multi-antenna user or multiple single-antenna users. Unfortunately, these scenarios still fall short of capturing the intricacies inherent in real-world communication networks, where multiple data streams are transmitted to numerous users, each with multiple antennas is commonplace. Interference is absent in a single multi-antenna user scenario, which significantly simplifies the information theoretic analysis of the single-user rate function. For scenarios of multiple single-antenna users, optimizing the users' rate functions can be equivalently translated into optimizing the users' signal-to-interference-plus-noise ratios (SINRs) for fractional programming. For multiple multi-antenna users, the individual rate function is defined as the log determinant (log-det) of a nonlinear matrix expression composed of the covariance matrices of the signal and interference. Optimizing these log-det rate functions poses significant computational challenges, even for moderate-dimensional conventional multiple user systems dispensing with metasurfaces, where fractional programming or semi-definite relaxation is entirely ineffective [11]–[13]. It is unsurprising that optimizing the users' rate functions in large-scale supporting multiple single-antenna users has been extensively studied [14]–[21] but its extension to scenarios of multiple multi-antenna users has remained largely unexplored.

The primary objective of multi-user (MU) networks is to ensure a high quality of information delivery (QoD) for all users (MU QoD) in terms of their individual rates. Metasurfaces are distinguished by their programable reconfigurable elements (PREs), which are subject to discrete constraints due to their low resolution in practical implementations. Consequently, optimizing the users' log-det rate functions in metasurface-aided MU networks involves tackling large-scale mixed discrete-continuous problems, which further exacerbates the computational challenges. The sum rate (SR) maximization problem proves inadequate since it often prioritizes enhancing rates the specific users that have the best channel, leaving others with nearly zero rates. Attempting to overcome this limitation of SR maximization by imposing constraints on the log-det rate functions to ensure MU QoD should be avoided since it results in computationally challenging mixed discrete-continuous constraints that remain unsolved. The most straightforward approach of enhancing MU QoD without constraining it is to maximize the minimum of the log-det rate functions, which, nevertheless, constitutes a nonsmooth mixed discrete-continuous problem.

Active reconfigurable intelligent surfaces (aRIS) have been introduced recently [5], [6]. They enable power amplification to the PREs of a metasurface, compensating for the double path-loss caused by dual-hop signal propagation from the source to the destination assisted by the metasurface. In particular, the authors of [6], [22], [23] have used fractional programming to demonstrate the enhancement of SR in aRIS-assisted networks supporting multiple singer-antenna users.

The work was supported in part by the Australian Research Councils Discovery Projects under Grant DP190102501, in part by the Shanghai Sailing Scholar under Grant 23YF1412700 and the Innovation Program of Shanghai Municipal Science Technology Commission under Grant 22511103202, and in part by the U.S National Science Foundation under Grants CNS-2128448 and ECCS-2335876.

[1]School of Communication and Information Engineering, Shanghai University Shanghai 200444, China (email: cyf119@shu.edu.cn, h-w_yu@shu.edu.cn); [2]School of Electrical and Data Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia; [3]Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA (email: poor@princeton.edu) (Corresponding Author: H. Yu)

Again, solely enhancing the SR may not always lead to equitable MU QoD. This paper presents the first advancement to ensure high MU QoD in aRIS-assisted multi-antenna multi-user systems by the joint design of matrices for precoding information streams (transmit precoders), low-resolution phases, and the power allocation for the vector of aRIS-powered amplified PREs (APREs). More explicitly, the paper offers the following contributions to push forward the field of MU services:

- **Large-scale Max-Min Log-Det Optimization:** This is the first attempt to address the challenge of maximizing the minimum of the users log-det rate functions, leading to max-min log det optimization. It poses a nonsmooth, large-scale, mixed-discrete problem due to the significant dimensions of both the individual transmit precoders (T-PCs) and of the APREs' vector, along with their discrete-valued phases. To tackle this, we introduce a novel universal penalized optimization framework, which allows us to utilize cubic complexity (CC) quadratic solvers (QSs) in the alternating optimization of the TPCs or the APREs' vector, while employing closed-form expressions in alternating optimization either for the APREs' power or for their discrete-valued phases;
- **Large-scale Sum-Log-Det Maximization:** Embedded within our universal penalized optimization framework, it takes advantage of the smoothness of the GM-log-det function to leverage closed-form expressions having scalable (linear) complexity in alternating optimization for the TPCs, the APREs' vector, as well as for the APREs' power or their discrete-valued phases;
- **Large-scale Soft Max-Min Log-Det Optimization:** This is the first endeavor to enhance the MU QoD. We introduce a new optimization objective, referred as soft minimum of log-det rate functions. Maximizing this function (soft max-min log-det optimization) not only helps to ensure high MU QoD but also achieves high sum log-det rates. Its smoothness allows us to utilize scalable-complexity closed-form expressions in the alternating optimization of each of the aforementioned TPCs, APREs' vector, APREs' power, and discrete-valued phases with the aid of the universal penalized optimization framework mentioned above.

The paper is structured as follows. Sections II, III, and IV are dedicated to developing algorithms for the optimization of max-min log-det, sum-log-det, and soft max-min log-det. Simulations are presented in Section V, and Section VI concludes the paper. The Appendix provides the mathematical ingredients for the algorithmic derivations.

*Notation.*Only decision variables are printed in boldface;[1] For a vector a, $\mathsf{diag}[a]$ is the diagonal matrix with the entries of a in its diagonal; $\langle X\rangle$ represents the trace of matrix X, and $||X||^2$ is the squared Frobenius norm; $[X]^2$ meansXX^H, so $[X^H]^2 = X^HX$ and $||X||^2 = \langle [X]^2\rangle = \langle [X^H]^2\rangle$; $X \succeq 0$ ($X \succ 0$, resp.) means X is positive semi-definite (positive definite, resp.); For $X \succeq 0$, $\sqrt{X}$ indicates a positive semi-definite matrix ($\sqrt{X} \succeq 0$) so that $[\sqrt{X}]^2 = X$; Naturally, $X \succ Y$ suggests $X - Y \succ 0$. The reader is referred to [24] for matrix optimization.

The distribution of a circular Gaussian random variable with zero means and variance σ is denoted by $\mathcal{C}(0,\sigma)$. Table I presents a summary of the notations used.

TABLE I: Basic notations

Notation	Description
$N/\mathfrak{N}$	# of BS's transmit antennas/index set $\{1,\ldots,N\}$
$K/\mathfrak{K}$	# of users/index set $\{1,\ldots,K\}$
$N_r/\mathfrak{N}_r$	# of user's antenna/index set $\{1,\ldots,N_r\}$
$M/\mathfrak{M}$	total # of APREs/index set $\{1,\ldots,M\}$
b	number bit resolution
c	penalty parameters
$\mathbf{W}_k \in \mathbb{C}^{N\times N_r}$	user k's precoder
$\mathbf{z} \triangleq (\mathbf{z}_1,\ldots,\mathbf{z}_M)^T$	design vector of APREs
$\boldsymbol{\theta} \triangleq (\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_M)^T$	design vector of APREs' phases
$\boldsymbol{p} \triangleq (\boldsymbol{p}_1,\ldots,\boldsymbol{p}_M)^T$	design vector of APREs' power amplification
$W_k^{(\kappa)}$, $z^{(\kappa)}$, $\theta^{(\kappa)}$, $p^{(\kappa)}$	generated values of $\mathbf{W}_k$, $\mathbf{z}$, $\boldsymbol{\theta}$, $\boldsymbol{p}$ at κ-th iteration
$\mathbf{W}/W^{(\kappa)}$	$\begin{bmatrix}\mathbf{W}_1 & \ldots & \mathbf{W}_K\end{bmatrix}/\begin{bmatrix}W_1^{(\kappa)} & \ldots & W_K^{(\kappa)}\end{bmatrix}$
$r_k(\mathbf{W},\mathbf{z})$	achievable rate for user k

II. Mastering iterative CC large scale quadratic solvers for max-min log-det optimization

We consider the wireless network of Figure 1, consisting of an aRIS assisting an N-antenna base station (BS) in serving K N_r-antenna users, which are indexed by $k \in \mathfrak{K} \triangleq \{1,\ldots,K\}$. The aRIS is equipped with M power-amplified programable reconfigurable elements (APREs) $\mathbf{z}_m$, $m \in \mathfrak{M} \triangleq \{1,\ldots,M\}$. For the practical implementation of the aRIS, we make the assumption that the APREs have phases of low-bit resolution, resulting in the following mixed discrete continuous constraints

$$\mathbf{z}_m = \boldsymbol{p}_m e^{\jmath\boldsymbol{\theta}_m}, m \in \mathfrak{M}, \tag{1}$$

with $\boldsymbol{p}_m \in \mathbb{R}$ representing the power amplification coefficients, and

$$\boldsymbol{\theta}_m \in \mathcal{B} \triangleq \left\{\beta\frac{2\pi}{2^b}, \beta = 0,1,\ldots,2^b-1\right\}, m \in \mathfrak{M}, \tag{2}$$

representing the APREs' phases of b-bit resolution.[2] We thus define the vector of APREs by

$$\mathbf{z} \triangleq (\mathbf{z}_1,\ldots,\mathbf{z}_M)^T \in \mathbb{C}^M. \tag{3}$$

Let

$$\mathbb{C}^{N_r\times M} \ni H_{R,k} = \begin{bmatrix} H_{R,k,1} \\ \ldots \\ H_{R,k,N_r}\end{bmatrix}, H_{R,k,n_r} \in \mathbb{C}^{1\times M}, n_r \in \mathfrak{N}_r \tag{4}$$

$$\& \quad G_{B,R} = \begin{bmatrix} G_{B,R,1} \\ \ldots \\ G_{B,R,M}\end{bmatrix} \in \mathbb{C}^{M\times N},$$

[1] We would like to emphasize the intentional use of boldface font in our paper, exclusively employed to highlight decision variables. This deliberate choice serves to explicitly distinguish between feasible points and variables, ensuring the prominence of decision variables and thus helping to recognize the specific structures such as quadratic forms of the functions concerned. These variables may take the form of scalars, vectors, or matrices.

[2] In practice, $\boldsymbol{p}_m$ may be subject to quantization. However, for the sake of focus in this paper, we defer consideration of quantization to future work, assuming that the resolution of $\boldsymbol{p}_m$ is sufficiently high to treat it as analog.

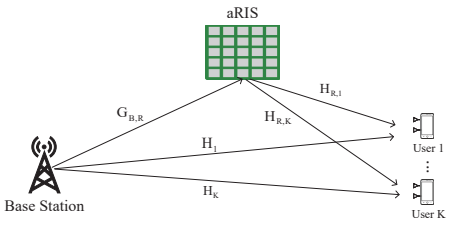


Fig. 1: aRIS-assisted system model

represent the channel spanning from the aRIS to user $k \in \mathfrak{K}$, and that from the BS to the aRIS. $G_{B,R}$ and $H_{R,k}$ are expressed by $\sqrt{\beta_{B,R}}\bar{G}_{B,R}$ and $\sqrt{\beta_{R,k}}\bar{H}_{R,k}$, where $\beta_{B,R}$ and $\beta_{R,k}$ represent the path-loss and large-scale fading of the link spanning from the BS to the aRIS and that from the aRIS to user k. Explicitly, $G_{B,R}$ is the line-of-sight (LoS) channel matrix between the BS and aRIS, while $H_{R,k}$ is modelled by Rician fading between the aRIS and the user k [25], [26]. For $R_{R,k}$ as the spatial correlation matrix of the aRIS elements with respect to user k, the channel from the aRISs to user k is represented by

$$\tilde{H}_{R,k} = \begin{bmatrix} \tilde{H}_{R,k,1} \\ \cdots \\ \tilde{H}_{R,k,N_r} \end{bmatrix} \triangleq H_{R,k}\sqrt{R_{R,k}} = \begin{bmatrix} H_{R,k,1}\sqrt{R_{R,k}} \\ \cdots \\ H_{R,k,N_r}\sqrt{R_{R,k}} \end{bmatrix}.$$

Let $H_k \in \mathbb{C}^{N_r \times N}$ be the direct channel impinging from the BS to user k. $H_k = \sqrt{\beta_k}\bar{H}_k$ with the path-loss and large-scale fading $\sqrt{\beta_k}$ at the distance $d_{\rm k}$ from the BS to user k. The small-scale fading channel gain $\bar{H}_k$ of the BS to user k obeys the Rayleigh distribution [27]. Then the composite two-hop channel from the BS to user k is given by

$$\tilde{\mathcal{H}}_k(\mathbf{z}) \triangleq H_k + \mathcal{H}_k(\mathbf{z}), \tag{5}$$

where $\mathcal{H}_k(\mathbf{z})$ is the composite channel spanning from the BS to user k via the aRIS defined by

$$\begin{aligned} \mathbb{C}^{N_r \times M} \ni \mathcal{H}_k(\mathbf{z}) &\triangleq \tilde{H}_{R,k}\mathsf{diag}[\mathbf{z}]G_{B,R} && (6)\\ &= \begin{bmatrix} \mathbf{z}^T\mathsf{diag}[\tilde{H}^T_{R,k,1}] \\ \cdots \\ \mathbf{z}^T\mathsf{diag}[\tilde{H}^T_{R,k,N_r}] \end{bmatrix} G_{B,R} && (7)\\ &= \Xi_k(\mathbf{z})G_{B,R} && (8) \end{aligned}$$

for

$$\mathbb{C}^M \ni \mathcal{D}_{k,n_r} \triangleq \mathsf{diag}[\tilde{H}^T_{R,k,n_r}], (k,n_r) \in \mathfrak{K} \times \mathfrak{N}_r, \tag{9}$$

and

$$\mathbb{C}^{N_r \times M} \ni \Xi_k(\mathbf{z}) \triangleq \begin{bmatrix} \mathbf{z}^T\mathcal{D}_{k,1} \\ \cdots \\ \mathbf{z}^T\mathcal{D}_{k,N_r} \end{bmatrix}. \tag{10}$$

Let $s_k \in \mathbb{C}^{N_r}$ obeying $\mathbb{E}([s_k]^2) = I_{N_r}$ represent the N_r number of information streams intended for user k, which is precoded by the TPC of

$$\mathbf{W}_k = \begin{bmatrix} \mathbf{W}_{k,1} & \cdots & \mathbf{W}_{k,N_r} \end{bmatrix} \in \mathbb{C}^{N \times N_r} \tag{11}$$

prior to its transmission from the BS. The signal received at user k can be written as

$$\mathbb{C}^{N_r} \ni y_k = \tilde{\mathcal{H}}_k(\mathbf{z}) \sum_{k \in \mathfrak{K}} \mathbf{W}_k s_k + \Xi_k(\mathbf{z})\nu + n_k, \tag{12}$$

where $\nu \in \mathcal{C}(0, \sigma_\nu I_M)$ is the dynamic noise induced by the aRIS, $n_k \in \mathcal{C}(0, \sigma I_{N_r})$ is the additive white Gaussian noise (AWGN), which encompasses both the background noise and the channel impairments caused by imperfect channel state information (CSI).

For $\mathbf{W} \triangleq \begin{bmatrix} \mathbf{W}_1 & \cdots & \mathbf{W}_K \end{bmatrix}$, the rate in nats for user k is expressed as the following log-det function

$$r_k(\mathbf{W}, \mathbf{z}) = \ln \left| I_{N_r} + [\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]^2 \Phi_k^{-1}(\mathbf{z}, \mathbf{W}) \right|, \tag{13}$$

with $[\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]^2$ representing the covariance of the signal of interest, and

$$\Phi_k(\mathbf{W}, \mathbf{z}) \triangleq \sum_{j \in \mathfrak{K}\setminus\{k\}} [\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_j]^2 + \sigma_\nu[\Xi_k(\mathbf{z})]^2 + \sigma I_{N_r} \tag{14}$$

representing the covariance of the interference plus noise. We also use (8) to represent

$$\begin{aligned} \tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_j &= H_k\mathbf{W}_j + \Xi_k(\mathbf{z})G_{B,R}\mathbf{W}_j \\ &= H_k\mathbf{W}_j + \Delta_k(\mathbf{W}_j, \mathbf{z}), \end{aligned} \tag{15}$$

with

$$\begin{aligned} &\mathbb{C}^{N_r \times N_r} \ni \Delta_k(\mathbf{W}_j, \mathbf{z}) \\ \triangleq\ &\begin{bmatrix} \mathbf{z}^T\mathcal{D}_{k,1}G_{B,R}\mathbf{W}_{j,1} & \cdots & \mathbf{z}^T\mathcal{D}_{k,1}G_{B,R}\mathbf{W}_{j,N_r} \\ \cdots & \cdots & \cdots \\ \mathbf{z}^T\mathcal{D}_{k,N_r}G_{B,R}\mathbf{W}_{j,1} & \cdots & \mathbf{z}^T\mathcal{D}_{k,N_r}G_{B,R}\mathbf{W}_{j,N_r} \end{bmatrix} \\ =\ &\begin{bmatrix} \Delta_{k,1,1}(\mathbf{W}_j)\mathbf{z} & \cdots & \Delta_{k,1,N_r}(\mathbf{W}_j)\mathbf{z} \\ \cdots & \cdots & \cdots \\ \Delta_{k,N_r,1}(\mathbf{W}_j)\mathbf{z} & \cdots & \Delta_{k,N_r,N_r}(\mathbf{W}_j)\mathbf{z} \end{bmatrix}, \end{aligned} \tag{16}$$

where

$$\mathbb{C}^{1\times M} \ni \Delta_{k,\ell,\ell'}(\mathbf{W}_j) \triangleq \mathbf{W}^T_{j,\ell'}G^T_{B,R}\mathcal{D}_{k,\ell}, (\ell,\ell') \in \mathfrak{N}_r \times \mathfrak{N}_r. \tag{17}$$

The TPC power given the budget P is constrained by

$$\sum_{k \in \mathfrak{K}} ||\mathbf{W}_k||^2 \le P, \tag{18}$$

which is a convex quadratic constraint, while the reflected power of the aRISs given the budget P_A is constrained by

$$\begin{aligned} &\sum_{k \in \mathfrak{K}} ||\mathsf{diag}[\mathbf{z}_m]_{m \in \mathfrak{M}}G_{B,R}\mathbf{W}_k||^2 + \sigma_\nu||\mathbf{z}||^2 \le P_A && (19)\\ \Leftrightarrow &\sum_{k \in \mathfrak{K}} \langle \mathcal{Q}_1(\mathbf{z})[\mathbf{W}_k]^2 \rangle \le P_1(\mathbf{z}) && (20)\\ \Leftrightarrow &\mathbf{z}^H\mathcal{Q}_2(\mathbf{W})\mathbf{z} \le P_A, && (21) \end{aligned}$$

where

$$\begin{aligned} \mathcal{Q}_1(\mathbf{z}) &\triangleq (G_{B,R})^H\mathsf{diag}[|\mathbf{z}_m|^2]_{m \in \mathfrak{M}}G_{B,R} \\ P_1(\mathbf{z}) &\triangleq P_A - \sigma_\nu||\mathbf{z}||^2, \\ \mathcal{Q}_2(\mathbf{W}) &\triangleq \sum_{k \in \mathfrak{K}} \sum_{n_r \in \mathfrak{N}_r} \mathsf{diag}[|G_{B,R,m}\mathbf{W}_{k,n_r}|^2]_{m \in \mathfrak{M}} \\ &\quad + \sigma_\nu I_M. \end{aligned} \tag{22}$$

By (20) we represent the biconvex constraint (19) as a convex quadratic constraint in $\mathbf{W}$ with $\mathbf{z}$ held fixed, while by (21) we express the biconvex constraint (19) as a convex quadratic constraint in $\mathbf{z}$ with $\mathbf{W}$ held fixed. To address the MU QoD

offered by aRIS-assisted MU communications, we consider the following problem of max-min log-det optimization:

$$\max_{\mathbf{W},\mathbf{z},\boldsymbol{\theta},\boldsymbol{p}} f(\mathbf{W},\mathbf{z}) \triangleq \min_{k\in\mathfrak{K}} r_k(\mathbf{W},\mathbf{z}) \quad \text{s.t} \quad (1),(2),(18),(19), \tag{23}$$

which is a large-scale problem due to the massive number of its decision variables. For example, under a typical scenarios of $(N, N_r, K, M) = (16, 2, 10, 100)$ it is $KN_rN+3M = 620$. This maximization problem is very computationally challenging, because its objective function (OF) $f(\mathbf{W},\mathbf{z})$ is not only nonconcave and nonconvex but also nonsmooth, representing the pointwise minimum of the log-det functions. Furthermore, the nonlinear equality constraint (1) and the difficult discrete constraint (2) prohibit exact alternating optimization in either $\mathbf{z}$, $\boldsymbol{\theta}$ or $\boldsymbol{p}$. To tackle (23), we adopt the popular penalized optimization framework [28], [29] and integrate the nonlinear equality constraint (1) into the optimization OF, leading to the following penalized optimization problem:

$$\max_{\mathbf{W},\mathbf{z},\boldsymbol{\theta},\boldsymbol{p}} f_\rho(\mathbf{W},\mathbf{z},\boldsymbol{p},\boldsymbol{\theta}) \triangleq \Big[f(\mathbf{W},\mathbf{z}) - \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - \boldsymbol{p}_m e^{\jmath\boldsymbol{\theta}_m}|^2\Big] \quad \text{s.t} \quad (2),(18),(19), \tag{24}$$

where $\rho > 0$ is the penalty parameter. Note that any feasible point for (24) is not feasible for (23), unless we have $\mathbf{z}_m = \boldsymbol{p}_m e^{\jmath\boldsymbol{\theta}_m}$, $m \in \mathfrak{M}$.

This section introduces an alternating ascent algorithm based on cubic-complexity QSs to solve (24). Starting from an initial point $(W^{(0)}, z^{(0)}, p^{(0)}, \theta^{(0)})$ for (24), let $(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})$ be a feasible point for (24) that is found from the $(\kappa-1)$-st iteration. We now describe an alternating ascent in each of $\mathbf{W}$, $\mathbf{z}$, $\boldsymbol{p}$ and $\boldsymbol{\theta}$ to generate $(W^{(\kappa+1)}, z^{(\kappa+1)}, p^{(\kappa+1)}, \theta^{(\kappa+1)})$.

A. Cubic-complexity TPC ascent

To seek $W^{(\kappa+1)}$ such satisfying

$$\begin{aligned} & f_\rho(W^{(\kappa+1)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)}) \geq f_\rho(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)}) \\ \Leftrightarrow \quad & f(W^{(\kappa+1)}, z^{(\kappa)}) >\!\!\geq f(W^{(\kappa)}, z^{(\kappa)}), \end{aligned} \tag{25}$$

we consider the following problem[3]

$$\max_{\mathbf{W}} \ f_1^{(\kappa)}(\mathbf{W}) \triangleq \min_{k\in\mathfrak{K}} r_{1,k}^{(\kappa)}(\mathbf{W}) \quad \text{s.t.} \quad (18), \tag{26a}$$

$$\sum_{k\in\mathfrak{K}} \langle \mathcal{Q}_1^{(\kappa)} [\mathbf{W}_k]^2 \rangle \leq P_1^{(\kappa)}, \tag{26b}$$

with $\mathcal{Q}_1^{(\kappa)} \triangleq \mathcal{Q}_1(z^{(\kappa)})$ and $P_1^{(\kappa)} \triangleq P_1(z^{(\kappa)})$ according to (20) and (22), while according to (13) and (14) we have:

$$\begin{aligned} r_{1,k}^{(\kappa)}(\mathbf{W}) &\triangleq r_k(\mathbf{W}, z^{(\kappa)}) \\ &= \ln \Big| I_{N_r} + [H_{1,k}^{(\kappa)} \mathbf{W}_k]^2 (\Phi_{1,k}^{(\kappa)}(\mathbf{W}))^{-1} \Big|, \end{aligned} \tag{27}$$

for $H_{1,k}^{(\kappa)} \triangleq \tilde{\mathcal{H}}_k(z^{(\kappa)})$, $\Phi_{1,k}^{(\kappa)}(\mathbf{W}) \triangleq \Phi_k(\mathbf{W}, z^{(\kappa)}) = \sum_{j\in\mathfrak{K}\setminus\{k\}} [H_{1,k}^{(\kappa)} \mathbf{W}_j]^2 + \Omega_{1,k}^{(\kappa)}$, and $\Omega_{1,k}^{(\kappa)} \triangleq \sigma_\nu [\Xi_k(z^{(\kappa)})]^2 + \sigma I_{N_r}$.

[3]Subscript 1 in (26) and subsequent equations in Subsection II.A, III.A, and IV.A designates functions associated with TPC ascent.

By applying the inequality (77) in Appendix I for $(\mathbf{V},\mathbf{Y}) = (H_{1,k}^{(\kappa)}\mathbf{W}_k, \Phi_{1,k}^{(\kappa)}(\mathbf{W}))$ and $(\bar{V},\bar{Y}) = (V_{1,k}^{(\kappa)}, Y_{1,k}^{(\kappa)}) \triangleq (H_{1,k}^{(\kappa)} W_k^{(\kappa)}, \Phi_{1,k}^{(\kappa)}(W^{(\kappa)}))$, the following tight concave quadratic minorant of $r_{1,k}^{(\kappa)}(\mathbf{W})$ at $W^{(\kappa)}$ is obtained:

$$\begin{aligned} \tilde{r}_{1,k}^{(\kappa)}(\mathbf{W}) &\triangleq a_{1,k}^{(\kappa)} + 2\Re\{\langle (V_{1,k}^{(\kappa)})^H (Y_{1,k}^{(\kappa)})^{-1} H_{1,k}^{(\kappa)} \mathbf{W}_k \rangle\} \\ &\quad - \langle \Psi_{1,k}^{(\kappa)}, \sum_{j\in\mathfrak{K}} [H_{1,k}^{(\kappa)} \mathbf{W}_j]^2 \rangle \qquad (28) \\ &= a_{1,k}^{(\kappa)} + 2\Re\{\langle \mathcal{B}_{1,k}^{(\kappa)} \mathbf{W}_k \rangle\} - \langle \mathcal{C}_{1,k}^{(\kappa)}, \sum_{j\in\mathfrak{K}} [\mathbf{W}_j]^2 \rangle \\ &= a_{1,k}^{(\kappa)} + 2\Re\{\langle \mathcal{B}_{1,k}^{(\kappa)} \mathbf{W}_k \rangle\} - \sum_{j\in\mathfrak{K}} ||\sqrt{\mathcal{C}_{1,k}^{(\kappa)}} \mathbf{W}_j||^2 \qquad (29) \end{aligned}$$

with $0 \preceq \Psi_{1,k}^{(\kappa)} \triangleq (Y_{1,k}^{(\kappa)})^{-1} - (Y_{1,k}^{(\kappa)} + [V_{1,k}^{(\kappa)}]^2)^{-1}$, $a_{1,k}^{(\kappa)} \triangleq r_{1,k}^{(\kappa)}(W^{(\kappa)}) - \langle [V_{1,k}^{(\kappa)}]^2 (Y_{1,k}^{(\kappa)})^{-1} \rangle - \langle \Psi_{1,k}^{(\kappa)}, \Omega_{1,k}^{(\kappa)} \rangle$, $\mathcal{B}_{1,k}^{(\kappa)} \triangleq (V_{1,k}^{(\kappa)})^H (Y_{1,k}^{(\kappa)})^{-1} H_{1,k}^{(\kappa)}$, $\mathcal{C}_{1,k}^{(\kappa)} \triangleq (H_{1,k}^{(\kappa)})^H \Psi_{1,k}^{(\kappa)} H_{1,k}^{(\kappa)}$.

It is plausible that $\tilde{f}_1^{(\kappa)}(\mathbf{W}) \triangleq \min_{k\in\mathfrak{K}} \tilde{r}_{1,k}^{(\kappa)}(\mathbf{W})$ is a tight minorant of $f_1^{(\kappa)}(\mathbf{W})$ at $W^{(\kappa)}$ and moreover it is concave as the pointwise minimum of the concave functions $\tilde{r}_{1,k}^{(\kappa)}(\mathbf{W})$ [30]. We thus generate $W^{(\kappa+1)}$ verifying (25) by solving the following large-scale convex quadratic problem (CQP) with CC order of $\mathcal{O}(K^4 N_r^3 N^3)$ [31, p.4] for tight minorant maximization:[4]

$$\max_{\mathbf{W}} \tilde{f}_1^{(\kappa)}(\mathbf{W}) \quad \text{s.t.} \quad (18),(26b). \tag{30}$$

B. Cubic-complexity aRIS ascent

To seek $z^{(\kappa+1)}$ such that

$$\begin{aligned} & f_\rho(W^{(\kappa+1)}, z^{(\kappa+1)}, p^{(\kappa)}, \theta^{(\kappa)}) \quad \geq \\ & f_\rho(W^{(\kappa+1)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)}), \end{aligned} \tag{31}$$

we consider the following problem:[5]

$$\max_{\mathbf{z}} \ f_{\rho,2}^{(\kappa)} \triangleq \left[f_2^{(\kappa)}(\mathbf{z}) - \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - p_m^{(\kappa)} e^{\jmath\theta_m^{(\kappa)}}|^2 \right] \tag{32a}$$

$$\text{s.t.} \quad \mathbf{z}^H \mathcal{Q}_2^{(\kappa)} \mathbf{z} \leq P_A, \tag{32b}$$

with $f_2^{(\kappa)}(\mathbf{z}) \triangleq \min_{k\in\mathfrak{K}} r_{2,k}^{(\kappa)}(\mathbf{z})$, and $\mathcal{Q}_2^{(\kappa)} \triangleq \mathcal{Q}_2(W^{(\kappa+1)})$ according to (21) and (22), while according to (15)-(17):

$$\begin{aligned} r_{2,k}^{(\kappa)}(\mathbf{z}) &\triangleq r_k(W^{(\kappa+1)}, \mathbf{z}) \\ &= \ln \Big| I_{N_r} + [A_{k,k}^{(\kappa)} + \Delta_{k,k}^{(\kappa)}(\mathbf{z})]^2 (\Phi_{2,k}^{(\kappa)}(\mathbf{z}))^{-1} \Big|, \end{aligned} \tag{33}$$

with $A_{k,j}^{(\kappa)} \triangleq H_k W_j^{(\kappa+1)}$, $(k,j) \in \mathfrak{K}\times\mathfrak{K}$, $\Phi_{2,k}^{(\kappa)}(\mathbf{z}) \triangleq \Phi_k(W^{(\kappa+1)}, \mathbf{z}) = \sum_{j\in\mathfrak{K}\setminus\{k\}} [A_{k,j}^{(\kappa)} + \Delta_{k,j}^{(\kappa)}(\mathbf{z})]^2 + \sigma_\nu [\Xi_k(\mathbf{z})]^2 +$

[4]The problem (30) is formulated as the standard convex quadratically constrained problem $\max_{\mathbf{W},t} t \quad \text{s.t.} \quad \tilde{r}_{1,k}^{(\kappa)}(\mathbf{W}) \geq t, k \in \mathfrak{K}; (18),(26b)$.

[5]Subscript 2 in (32) and subsequent equations in Subsections II.B, III.B, and IV.B designates functions related to aRIS ascent.

σI_{N_r},

$$\begin{aligned}\mathbb{C}^{N_r\times N_r} &\ni \Delta_{k,j}^{(\kappa)}(\mathbf{z}) \triangleq \Delta_k(W_j^{(\kappa+1)},\mathbf{z})\\ &= \begin{bmatrix}\Delta_{k,j}^{(\kappa)}(1,1)\mathbf{z} & \dots & \Delta_{k,j}^{(\kappa)}(1,N_r)\mathbf{z}\\ \dots & \dots & \dots\\ \Delta_{k,j}^{(\kappa)}(N_r,1)\mathbf{z} & \dots & \Delta_{k,j}^{(\kappa)}(N_r,N_r)\mathbf{z}\end{bmatrix}\end{aligned}$$

and $\mathbb{C}^{1\times M} \ni \Delta_{k,j}^{(\kappa)}(\ell,\ell') \triangleq (W_{j,\ell'}^{(\kappa+1)})^T G_{B,R}^T \mathcal{D}_{k,\ell}$, $(\ell,\ell') \in \mathfrak{N}_r \times \mathfrak{N}_r$.

Applying the inequality (77) in Appendix I for $(\mathbf{V},\mathbf{Y}) \triangleq (A_{k,k}^{(\kappa)} + \Delta_{k,k}^{(\kappa)}(\mathbf{z}), \Phi_{2,k}^{(\kappa)}(\mathbf{z}))$ and $(\bar{V},\bar{Y}) = (V_{2,k}^{(\kappa)}, Y_{2,k}^{(\kappa)}) \triangleq (A_{k,k}^{(\kappa)} + \Delta_{k,k}^{(\kappa)}(z^{(\kappa)}), \Phi_{2,k}^{(\kappa)}(z^{(\kappa)}))$, the following tight concave quadratic minorant of $r_{2,k}^{(\kappa)}(\mathbf{z})$ is obtained at $z^{(\kappa)}$:

$$\begin{aligned}\tilde{r}_{2,k}^{(\kappa)}(\mathbf{z}) &\triangleq \tilde{a}_{2,k}^{(\kappa)} + 2\Re\{\langle (V_{2,k}^{(\kappa)})^H (Y_{2,k}^{(\kappa)})^{-1}(A_{k,k}^{(\kappa)} + \Delta_{k,k}^{(\kappa)}(\mathbf{z}))\rangle\}\\ &\quad - \langle \Psi_{2,k}^{(\kappa)}, \sum_{j\in\mathfrak{K}}[A_{k,j}^{(\kappa)} + \Delta_{k,j}^{(\kappa)}(\mathbf{z})]^2 + \sigma_\nu[\Xi_k(\mathbf{z})]^2\rangle \qquad (34)\\ &= a_{2,k}^{(\kappa)} + \sum_{j\in\mathfrak{K}} ||\sqrt{\Psi_{2,k}^{(\kappa)}}\Delta_{k,j}^{(\kappa)}(\mathbf{z})||^2 - \sigma_\nu||\sqrt{\Psi_{2,k}^{(\kappa)}}\Xi_k(\mathbf{z})||^2\\ &\quad + 2\sum_{j\in\mathfrak{K}}\Re\{\langle \mathcal{B}_{2,k,j}^{(\kappa)}\Delta_{k,j}^{(\kappa)}(\mathbf{z})\rangle\}\\ &= a_{2,k}^{(\kappa)} + 2\sum_{j\in\mathfrak{K}}\sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r}\Re\{\mathcal{B}_{2,k,j}^{(\kappa)}(\ell,\ell')\Delta_{k,j}^{(\kappa)}(\ell',\ell)\mathbf{z}\}\\ &\quad - \sum_{j\in\mathfrak{K}}\sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r} |\left(\sum_{t=1}^{N_r}\sqrt{\Psi_{2,k}^{(\kappa)}}(\ell,t)\Delta_{k,j}^{(\kappa)}(t,\ell')\right)\mathbf{z}|^2\\ &\quad - \sigma_\nu\sum_{\ell\in\mathfrak{N}_r}||\mathbf{z}^T\sum_{\ell'\in\mathfrak{N}_r}\sqrt{\Psi_{2,k}^{(\kappa)}}(\ell,\ell')\mathcal{D}_{k,\ell'}||^2 \qquad (35)\\ &= a_{2,k}^{(\kappa)} + 2\Re\{b_{2,k}^{(\kappa)}\mathbf{z}\} - \langle \mathcal{C}_{2,k}^{(\kappa)}, [\mathbf{z}]^2\rangle, \qquad (36)\end{aligned}$$

with $0 \preceq \Psi_{2,k}^{(\kappa)} \triangleq (Y_{2,k}^{(\kappa)})^{-1} - (Y_{2,k}^{(\kappa)} + [V_{2,k}^{(\kappa)}]^2)^{-1}$, $\tilde{a}_{2,k}^{(\kappa)} \triangleq r_{2,k}^{(\kappa)}(z^{(\kappa)}) - \langle [V_{2,k}^{(\kappa)}]^2 (Y_{2,k}^{(\kappa)})^{-1}\rangle - \sigma\langle \Psi_{2,k}^{(\kappa)}\rangle$, and $a_{2,k}^{(\kappa)} \triangleq \tilde{a}_{2,k}^{(\kappa)} + 2\Re\{\langle (V_{2,k}^{(\kappa)})^H (Y_{2,k}^{(\kappa)})^{-1} A_{k,k}^{(\kappa)}\rangle\} - \langle \Psi_{2,k}^{(\kappa)}, \sum_{j\in\mathfrak{K}\setminus\{k\}}[A_{k,j}^{(\kappa)}]^2\rangle$,

$$\mathcal{B}_{2,k,j}^{(\kappa)} = \begin{cases} (V_{2,k}^{(\kappa)})^H (Y_{2,k}^{(\kappa)})^{-1} - (A_{k,k}^{(\kappa)})^H \Psi_{2,k}^{(\kappa)} & \text{for} \quad j = k\\ -(A_{k,j}^{(\kappa)})^H \Psi_{2,k}^{(\kappa)} & \text{otherwise}\end{cases}$$

and $b_{2,k}^{(\kappa)} \triangleq \sum_{j\in\mathfrak{K}}\sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r}\mathcal{B}_{2,k,j}^{(\kappa)}(\ell,\ell')\Delta_{k,j}^{(\kappa)}(\ell',\ell)$,

$$\mathcal{C}_{2,k}^{(\kappa)} \triangleq \sum_{j\in\mathfrak{K}}\sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r}\left[(\sum_{t=1}^{N_r}\sqrt{\Psi_{2,k}^{(\kappa)}}(\ell,t)\Delta_{k,j}^{(\kappa)}(t,\ell'))^H\right]^2 + \sigma_\nu\sum_{\ell\in\mathfrak{N}_r}\left[(\sum_{\ell'\in\mathfrak{N}_r}\sqrt{\Psi_{2,k}^{(\kappa)}}(\ell,\ell')\mathcal{D}_{k,\ell'}^T)^H\right]^2.$$

It is plausible that $\tilde{f}_2^{(\kappa)}(\mathbf{z}) \triangleq \min_{k\in\mathfrak{K}}\tilde{r}_{2,k}^{(\kappa)}(\mathbf{z})$ is a tight minorant of $f_2^{(\kappa)}(\mathbf{z})$ at $z^{(\kappa)}$ and moreover it is concave as the pointwise minimum of the concave functions $\tilde{r}_{2,k}^{(\kappa)}(\mathbf{z})$ [30]. We thus generate $z^{(\kappa+1)}$ verifying (31) by solving the following large-scale CQP with CC order of $\mathcal{O}(M^3K)$ [31, p.4] for tight minorant maximization at $z^{(\kappa)}$:[6]

$$\max_{\mathbf{z}}\left[\tilde{f}_2^{(\kappa)}(\mathbf{z}) - \rho\sum_{m\in\mathfrak{M}}|\mathbf{z}_m - p_m^{(\kappa)}e^{\jmath\theta_m^{(\kappa)}}|^2\right] \text{ s.t. } (32b). \quad (37)$$

C. Amplifier and PRE ascent

We generate $p^{(\kappa+1)}$ and $\theta^{(\kappa+1)}$ by

$$\begin{aligned}p_m^{(\kappa+1)} &= \arg\min_{\boldsymbol{p}_m}|z_m^{(\kappa+1)} - \boldsymbol{p}_m e^{\jmath\theta_m^{(\kappa)}}|^2\\ &= |z_m^{(\kappa+1)}|\cos(\angle z^{(\kappa+1)} - \theta_m^{(\kappa)}), m\in\mathfrak{M}, \qquad (38)\end{aligned}$$

and[7]

$$\theta_m^{(\kappa+1)} = \arg\min_{\boldsymbol{\theta}_m\in\mathcal{B}}|z_m^{(\kappa+1)} - p_m^{(\kappa+1)}e^{\jmath\boldsymbol{\theta}_m}|^2 = \lfloor\angle z_m^{(\kappa+1)}\rceil_b, \quad (39)$$

which yield

$$\begin{aligned}f_\rho(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa+1)},\theta^{(\kappa+1)}) &\geq\\ f_\rho(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa+1)},\theta^{(\kappa)}) &\geq\\ f_\rho(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa)},\theta^{(\kappa)}). &\qquad (40)\end{aligned}$$

D. Algorithm and its convergence

Algorithm 1 provides the pseudo code for solving the problem (24). It follows from (25), (31) and (40) that

$$f_\rho(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa+1)},\theta^{(\kappa+1)}) > f_\rho(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)}), \quad (41)$$

as far as $f_\rho(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa+1)},\theta^{(\kappa+1)}) \neq f_\rho(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)})$, so the sequence $\{(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)})\}$ of improved feasible points for (24) converges to $(\bar{W},\bar{z},\bar{p},\bar{\theta})$. Furthermore, by choosing a sufficiently large $\rho > 0$, we can ensure that $\max_{m\in\mathfrak{M}}|z_m^{(\kappa)} - p_m^{(\kappa)}e^{\theta_m^{(\kappa)}}|^2 \to 0$ as $\kappa\to\infty$, which means that $(\bar{z},\bar{p},\bar{\theta})$ satisfies the nonlinear constraint (1). Therefore, $(\bar{W},\bar{z},\bar{p},\bar{\theta})$ is feasible for (23), which turns out to be at least a local solution [28].

Algorithm 1 CC QS-based algorithm for the large-scale max-min log-det optimization (24)

1: **Initialization:** Randomly generate $(W^{(0)},z^{(0)},p^{(0)},\theta^{(0)})$ feasible for (24). Set $\kappa = 0$.
2: **Repeat until convergence:** Generate $W^{(\kappa+1)}$ by solving the large-scale CQP (30) of the CC $\mathcal{O}(K^4N_r^3N^3)$, and $z^{(\kappa+1)}$ by solving the large-scale CQP (37) of the CC $\mathcal{O}(M^3K)$. Generate $(p^{(\kappa+1)},z^{(\kappa+1)})$ by (38)-(39). Reset $\kappa\leftarrow\kappa+1$.
3: **Output** $(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},z^{(\kappa)})$ and $r_k(W^{(\kappa)},z^{(\kappa)})$, $k\in\mathfrak{K}$.

[6]The problem (37) is formulated as the standard convex quadratically constrained problem $\max_{\mathbf{z},t}\ t$ s.t. $\tilde{r}_{2,k}^{(\kappa)}(\mathbf{z}) - \rho\sum_{m\in\mathfrak{M}}|\mathbf{z}_m - p_m^{(\kappa)}e^{\jmath\theta_m^{(\kappa)}}|^2 \geq t, k\in\mathfrak{K}$; (32$b$).

[7]$\lfloor.\rceil_b$ is the b-bit rounding operation

III. Large-scale Sum-Log-Det maximization

The computational complexities $\mathcal{O}(K^3N_r^3N^3)$ and $\mathcal{O}(M^3)$ of the large-scale CQPs (30) and (37), invoked in Algorithm 1 are massive because both KN and M typically exceed 100. Consequently, the computational time required is quite substantial. In this section, we aim for addressing this computational challenge by considering the following large-scale Sum-Log-Det maximization:

$$\max_{\mathbf{W},\mathbf{z},\boldsymbol{p},\boldsymbol{\theta}} f_{SR}(\mathbf{W},\mathbf{z}) \triangleq \sum_{k\in\mathfrak{K}} r_k(\mathbf{W},\mathbf{z}) \quad \text{s.t.} \quad (1),(2),(18),(19). \tag{42}$$

Similar to (23), we tackle (42) by formulating the following penalized optimization problem:

$$\max_{\mathbf{W},\mathbf{z},\boldsymbol{p},\boldsymbol{\theta}} f_{\rho,SR}(\mathbf{W},\mathbf{z},\boldsymbol{p},\boldsymbol{\theta}) \triangleq [f_{SR}(\mathbf{W},\mathbf{z}) - \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - \boldsymbol{p}_m e^{\jmath\boldsymbol{\theta}_m}|^2] \text{ s.t. } (2),(18),(19). \tag{43}$$

Initialized by a feasible point $(W^{(0)},z^{(0)},p^{(0)},\theta^{(0)})$ for (43), let $(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)})$ represent a feasible point for (43) obtained from the $(\kappa-1)$-st iteration.

A. Scalable-complexity bisection for TPC ascent

To seek $W^{(\kappa+1)}$ so that

$$\begin{aligned} & f_{\rho,SR}(W^{(\kappa+1)},z^{(\kappa)}) \geq f_{\rho,SR}(W^{(\kappa)},z^{(\kappa)}) \\ \Leftrightarrow \quad & f_{SR}(W^{(\kappa+1)},z^{(\kappa)}) \geq f_{SR}(W^{(\kappa)},z^{(\kappa)}), \end{aligned} \tag{44}$$

we consider the following problem:

$$\max_{\mathbf{W}} \ f_{SR,1}^{(\kappa)}(\mathbf{W}) \triangleq \sum_{k\in\mathfrak{K}} r_{1,k}^{(\kappa)}(\mathbf{W}) \quad \text{s.t.} \quad (18),(26b), \tag{45}$$

with $r_{1,k}^{(\kappa)}(\mathbf{W})$ defined in (27). By utilizing (29), we can derive the following concave quadratic minorant of the OF $f_{SR,1}^{(\kappa)}(\mathbf{W})$ in (45) at $W^{(\kappa)}$:

$$\begin{aligned} \tilde{f}_{SR,1}^{(\kappa)}(\mathbf{W}) &\triangleq \sum_{k\in\mathfrak{K}} [a_{1,k}^{(\kappa)} + 2\Re\{\langle \mathcal{B}_{1,k}^{(\kappa)}\mathbf{W}_k\rangle\} \\ &\quad -\langle \mathcal{C}_{1,k}^{(\kappa)}, \sum_{j\in\mathfrak{K}} [\mathbf{W}_j]^2\rangle] \\ &= \sum_{k\in\mathfrak{K}} a_{1,k}^{(\kappa)} + 2\sum_{k\in\mathfrak{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\kappa)}\mathbf{W}_k\rangle\} \\ &\quad -\langle \mathcal{C}_1^{(\kappa)}, \sum_{k\in\mathfrak{K}} [\mathbf{W}_j]^2\rangle, \end{aligned} \tag{46}$$

with $0 \preceq \mathcal{C}_1^{(\kappa)} \triangleq \sum_{k\in\mathfrak{K}} \mathcal{C}_{1,k}^{(\kappa)}$. We thus generate $W^{(\kappa+1)}$ verifying (44) by solving the following CQP of tight minorant maximization at $W^{(\kappa)}$:

$$\max_{\mathbf{W}} \ \tilde{f}_{GM,1}^{(\kappa)}(\mathbf{W}) \quad \text{s.t.} \quad (18),(26b). \tag{47}$$

This matrix optimization problem, which is solvable by a convex solver with CC order of $\mathcal{O}(K^3N_r^3N^3)$, also falls under the category defined by (78). Therefore, it is solvable by the bisection method of scalable complexity order of $\mathcal{O}(KN_rN\log_2 N)$ described in Appendix II, bypassing a convex solver of CC order of $\mathcal{O}((KN_rN)^3)$.

B. Scalable-complex closed-form for aRIS alternating ascent

To seek $z^{(\kappa+1)}$ so that

$$f_{\rho,SR}(W^{(\kappa+1)},z^{(\kappa+1)}) \geq f_{\rho,SR}(W^{(\kappa+1)},z^{(\kappa)}) \tag{48}$$

we consider the following problem:

$$\max_{\mathbf{z}} \ f_{\rho,SR,2}^{(\kappa)}(\mathbf{z}) \triangleq \Big[f_{SR,2}^{(\kappa)}(\mathbf{z}) - \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - p_m^{(\kappa)} e^{\jmath\theta_m^{(\kappa)}}|^2\Big] \quad \text{s.t.} \quad (32b), \tag{49}$$

where $f_{SR,2}^{(\kappa)}(\mathbf{z}) \triangleq \sum_{k\in\mathfrak{K}} r_{2,k}^{(\kappa)}(\mathbf{z})$ with $r_{2,k}^{(\kappa)}(\mathbf{z})$ defined from (33).

By utilizing (36), we can derive the following concave quadratic minorant of $f_{SR,2}^{(\kappa)}(\mathbf{z})$ in (49) at $z^{(\kappa)}$:

$$\begin{aligned} \tilde{f}_{SR,2}^{(\kappa)}(\mathbf{z}) &\triangleq \sum_{k\in\mathfrak{K}} \Big[a_{2,k}^{(\kappa)} + 2\Re\{b_{2,k}^{(\kappa)}\mathbf{z}\} - \langle \mathcal{C}_{2,k}^{(\kappa)}, [\mathbf{z}]^2\rangle\Big] \\ &= a_2^{(\kappa)} + 2\Re\{b_2^{(\kappa)}\mathbf{z}\} + \mathbf{z}^H\mathcal{C}_2^{(\kappa)}\mathbf{z}, \end{aligned} \tag{50}$$

for $a_2^{(\kappa)} \triangleq \sum_{k\in\mathfrak{K}} a_{2,k}^{(\kappa)}$, $b_2^{(\kappa)} \triangleq \sum_{k\in\mathfrak{K}} b_{2,k}^{(\kappa)}$, and $\mathcal{C}_2^{(\kappa)} \triangleq \sum_{k\in\mathfrak{K}} \mathcal{C}_{2,k}^{(\kappa)}$.

We thus generate $z^{(\kappa+1)}$ verifying (48) by solving the following convex problem of tight minorant maximization:

$$\max_{\mathbf{z}} \ \Big[\tilde{f}_{SR,2}^{(\kappa)}(\mathbf{z}) - \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - p_m^{(\kappa)} e^{\jmath\theta_m^{(\kappa)}}|^2\Big] \quad \text{s.t.} \quad (32b), \tag{51}$$

which admits the closed form solution of

$$z^{(\kappa+1)} = \begin{cases} (\mathcal{C}_2^{(\kappa)} + \rho I_M)^{-1}\xi^{(\kappa)} \\ \text{if } \ ||\sqrt{\mathcal{Q}_2^{(\kappa)}}(\mathcal{C}_2^{(\kappa)} + \rho I_M)^{-1}\xi^{(\kappa)}||^2 \leq P_A \\ (\mathcal{C}_2^{(\kappa)} + \rho I_M + \alpha\mathcal{Q}_2^{(\kappa)})^{-1}\xi^{(\kappa)} \quad \text{otherwise,} \end{cases}$$

where $\xi^{(\kappa)} \triangleq (b_2^{(\kappa)})^H + \rho(p_1^{(\kappa)}e^{\jmath\theta_1^{(\kappa)}},\ldots,p_M^{(\kappa)}e^{\jmath\theta_M^{(\kappa)}})^T$ and $\alpha > 0$ is found by bisection so that $||\sqrt{\mathcal{Q}_2^{(\kappa)}}(\mathcal{C}_2^{(\kappa)} + \rho I_M + \alpha\mathcal{Q}_2^{(\kappa)})^{-1}\xi^{(\kappa)}||^2 = P_A$. The computational complexity of (52) is $\mathcal{O}(M\log_2 M)$.

C. Algorithm and its convergence

Algorithm 2 presents the pseudo code for solving the problem (43) by iteratively evaluating the closed forms (47), (52), (38), and (39). The algorithm's convergence follows from the fact that

$$\begin{aligned} & f_{\rho,SM}(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa+1)},\theta^{(\kappa+1)}) \ > \\ & \qquad f_{\rho,SM}(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)}), \end{aligned} \tag{52}$$

as far as $f_{\rho,SM}(W^{(\kappa+1)},z^{(\kappa+1)},p^{(\kappa+1)},\theta^{(\kappa+1)}) \neq f_{\rho,SM}(W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)})$, i.e. Algorithm 2 generates a sequence $\{W^{(\kappa)},z^{(\kappa)},p^{(\kappa)},\theta^{(\kappa)}\}$ of infeasible points for (42), which converges to its local solution [28].

Remark. While this large-scale mixed discrete continuous problem (43) can be solved efficiently, it often exhibits zero or negligible log-det rates for certain users. In essence, optimizing the sum log-det rate without taking into account realistic MU QoD constraints is not meaningful.

Algorithm 2 Scalable-complex SR algorithm for the large-scale Sum-Log-Det maximization (43)

1: **Initialization:** Randomly generate a feasible $(W^{(0)}, z^{(0)}, p^{(0)}, \theta^{(0)})$ for (43). Set $\kappa = 0$.
2: **Repeat until convergence of the objective function in** (43)**:** Generate $W^{(\kappa+1)}$ by solving (47) of the CC $\mathcal{O}(KN_rN\log_2 N)$, $z^{(\kappa+1)}$ by (52) of the CC $\mathcal{O}(M\log_2 M)$. Generate $(p^{(\kappa+1)}, \theta^{(\kappa+1)})$ by (38)-(39). Reset $\kappa \leftarrow \kappa + 1$.
3: **Output** $(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})$ and $r_k(W^{(\kappa)}, z^{(\kappa)})$, $k \in \mathfrak{K}$.

IV. Unleashing scalable-complex iterations for soft max-min log-det optimization and elevating MU QoD

One can see that one of the main reasons for solving the max-min log-det problem (23) by iterative CC QSs instead of scalable-complexity closed forms is the non-smooth nature of its minimum log-det rate OF $f(\mathbf{W}, \mathbf{z}) \triangleq \min_{k\in\mathfrak{K}} r_k(\mathbf{W}, \mathbf{z})$, for which there is no tight and smooth minorant. Now, we embark on exploiting the following scalable version: $f_\mu(\mathbf{W}, \mathbf{z}) \triangleq \min_{k\in\mathfrak{K}} r_{\mu,k}(\mathbf{W}, \mathbf{z}) \triangleq \min_{k\in\mathfrak{K}} \ln \left| I_{N_r} + \frac{1}{\mu}[\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]^H \Phi_k^{-1}(\mathbf{z}, \mathbf{W})[\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k] \right|$, for small $\mu > 0$. Observe that, $f_\mu(\mathbf{W}, \mathbf{z}) > f_1(\mathbf{W}, \mathbf{z}) = f(\mathbf{W}, \mathbf{z})$ for $\mu < 1$. Moreover, $\max_{\mathbf{W},\mathbf{z}} f(\mathbf{W}, \mathbf{z}) \Leftrightarrow \max_{\mathbf{W},\mathbf{z}} f_\mu(\mathbf{W}, \mathbf{z})$ for $N_r = 1$, and maximizing $f_\mu(\mathbf{W}, \mathbf{z})$ enhances $f(\mathbf{W}, \mathbf{z})$ for $N_r > 1$. Furthermore, we can readily show the validity of the following two-sided inequality

$$f_\mu(\mathbf{W}, \mathbf{z}) \geq -f_{SM}(\mathbf{W}, \mathbf{z}) \geq f_\mu(\mathbf{W}, \mathbf{z}) - \ln K, \tag{53}$$

where we have

$$f_{SM}(\mathbf{W}, \mathbf{z}) \triangleq \ln |\sum_{k\in\mathfrak{K}} \Pi_k(\mathbf{W}, \mathbf{z})| \tag{54}$$

with $\Pi_k(\mathbf{W}, \mathbf{z}) \triangleq (I_{N_r} + \frac{1}{\mu}[\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]^H \Phi_k^{-1}(\mathbf{z}, \mathbf{W})[\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k])^{-1} = I_{N_r} - [\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]^H([\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]^2 + \mu\Phi_k(\mathbf{z}, \mathbf{W}))^{-1}[\tilde{\mathcal{H}}_k(\mathbf{z})\mathbf{W}_k]$.

With μ small, the term $\ln K$ in (53) is small compared to other terms. Therefore, the function $-f_{SM}(\mathbf{W}, \mathbf{z})$ can regarded as a soft-min function [32], since it is smooth and approximates the nonsmooth minimum function $f_\mu(\mathbf{W}, \mathbf{z})$. Rather than maximizing the nonsmooth minimum function $f_\mu(\mathbf{W}, \mathbf{z})$ we aim for maximizing its soft-minimum counterpart $-f_{SM}(\mathbf{W}, \mathbf{z})$, or for minimizing $f_{SM}(\mathbf{W}, \mathbf{z})$. This leads us to the following soft max-min log-det problem:

$$\min_{\mathbf{W},\mathbf{z},\boldsymbol{p},\boldsymbol{\theta}} f_{SM}(\mathbf{W}, \mathbf{z}) \quad \text{s.t} \quad (1), (2), (18), (19). \tag{55}$$

Similarly to (23), the penalized optimization formulation for (55) is

$$\min_{\mathbf{W},\mathbf{z}} f_{\rho,SM}(\mathbf{W}, \mathbf{z}, \boldsymbol{p}, \boldsymbol{\theta}) \triangleq [f_{SM}(\mathbf{W}, \mathbf{z}) + \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - \boldsymbol{p}_m e^{\jmath\boldsymbol{\theta}_m}|^2] \quad \text{s.t} \quad (2), (18), (19), \tag{56}$$

Initialized by a feasible point $(W^{(0)}, z^{(0)}, p^{(0)}, \theta^{(0)})$ for (56), let $(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})$ be a feasible point for (56) that is found from the $(\kappa-1)$-st iteration. We now provide an alternating descent in each $\mathbf{W}$ and $\mathbf{z}$, since it is plausible that alternating descent in $(\boldsymbol{p}, \boldsymbol{\theta})$ is still based on the closed forms (38)-(39).

A. Scalable-complexity bisection for TPC descent

To seek $W^{(\kappa+1)}$ so that $f_{\rho,SM}(W^{(\kappa+1)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)}) \leq f_{\rho,SM}(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})$, i.e.

$$f_{SM}(W^{(\kappa+1)}, z^{(\kappa)}) \leq f_{SM}(W^{(\kappa)}, z^{(\kappa)}), \tag{57}$$

we consider the following problem:

$$\min_{\mathbf{W}} f_{SM,1}^{(\kappa)}(\mathbf{W}) \quad \text{s.t.} \quad (18), (26b), \tag{58}$$

where in accordance to (27), $f_{SM,1}^{(\kappa)}(\mathbf{W}) \triangleq f_{SM}(\mathbf{W}, z^{(\kappa)}) \triangleq \ln|\Pi_1^{(\kappa)}(\mathbf{W})|$ for $\Pi_1^{(\kappa)}(\mathbf{W}) \triangleq \sum_{k\in\mathfrak{K}}(I_{N_r} - [H_{1,k}^{(\kappa)}\mathbf{W}_k]^H([H_{1,k}^{(\kappa)}\mathbf{W}_k]^2 + \mu\Phi_{1,k}^{(\kappa)}(\mathbf{W}))^{-1}[H_{1,k}^{(\kappa)}\mathbf{W}_k])$.

Using the inequality (76) in Appendix I for $(\mathbf{V}_k, \mathbf{Y}_k) = (H_{1,k}^{(\kappa)}\mathbf{W}_k, [H_{1,k}^{(\kappa)}\mathbf{W}_k]^2 + \mu\Phi_{1,k}^{(\kappa)}(\mathbf{W}))$, $k \in \mathfrak{K}$, and $(\bar{V}_k, \bar{Y}_k) = (V_{1,k}^{(\kappa)}, Y_{1,k}^{(\kappa)}) \triangleq (H_{1,k}^{(\kappa)}W_k^{(\kappa)}, [H_{1,k}^{(\kappa)}W_k^{(\kappa)}]^2 + \mu\Phi_{1,k}^{(\kappa)}(W^{(\kappa)}))$, yields the following tight majorant of $f_{SM}(\mathbf{W}, z^{(\kappa)})$ at $W^{(\kappa)}$:

$$\begin{aligned} f_{1,SM}^{(\kappa)}(\mathbf{W}) &\triangleq a_1^{(\kappa)} - 2\sum_{k\in\mathfrak{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\kappa)}\mathbf{W}_k\rangle\} + \sum_{k\in\mathfrak{K}} \langle \tilde{\mathcal{C}}_{1,k}^{(\kappa)}, \\ &\quad [H_{1,k}^{(\kappa)}\mathbf{W}_k]^2 + \mu \sum_{j\in\mathfrak{K}\setminus\{k\}} [H_{1,k}^{(\kappa)}\mathbf{W}_j]^2\rangle \quad (59) \\ &= a_1^{(\kappa)} - 2\sum_{k\in\mathfrak{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\kappa)}\mathbf{W}_k\rangle\} \\ &\quad + \sum_{k\in\mathfrak{K}} \langle \mathcal{C}_{1,k}^{(\kappa)}, [\mathbf{W}_k]^2\rangle, \quad (60) \end{aligned}$$

where we have

$$\begin{aligned} a_1^{(\kappa)} &\triangleq f_{SM,1}^{(\kappa)}(W^{(\kappa)}) + \textstyle\sum_{k\in\mathfrak{K}} \langle [\Pi_1^{(\kappa)}(W^{(\kappa)})]^{-1}(V_{1,k}^{(\kappa)})^H \\ &\quad (Y_{1,k}^{(\kappa)})^{-1}V_{1,k}^{(\kappa)}\rangle + \mu \textstyle\sum_{k\in\mathfrak{K}} \langle \mathcal{C}_{1,k}^{(\kappa)}, \Omega_{1,k}^{(\kappa)}\rangle, \\ \mathcal{B}_{1,k}^{(\kappa)} &\triangleq [\Pi_1^{(\kappa)}(W^{(\kappa)})]^{-1}(V_{1,k}^{(\kappa)})^H(Y_{1,k}^{(\kappa)})^{-1}H_{1,k}^{(\kappa)}, \\ \tilde{\mathcal{C}}_{1,k}^{(\kappa)} &\triangleq (Y_{1,k}^{(\kappa)})^{-1}V_{1,k}^{(\kappa)}[\Pi_1^{(\kappa)}(W^{(\kappa)})]^{-1}(V_{1,k}^{(\kappa)})^H(Y_{1,k}^{(\kappa)})^{-1}, \end{aligned}$$

and $\mathcal{C}_{1,k}^{(\kappa)} \triangleq (H_{1,k}^{(\kappa)})^H\tilde{\mathcal{C}}_{1,k}^{(\kappa)}H_{1,k}^{(\kappa)} + \mu\sum_{j\in\mathfrak{K}\setminus\{k\}}(H_{1,j}^{(\kappa)})^H\tilde{\mathcal{C}}_{1,j}^{(\kappa)}H_{1,j}^{(\kappa)}$, $k \in \mathfrak{K}$.

We thus solve the following convex problem of majorant minimization of (58) to generate $W^{(\kappa+1)}$ verifying (57):

$$\min_{\mathbf{W}} f_{1,SM}^{(\kappa)}(\mathbf{W}) \quad \text{s.t.} \quad (18), (26b). \tag{61}$$

This matrix optimization problem still belongs to the class of problems defined by (78), so it can be computed by the bisection procedure of scalable complexity order of $\mathcal{O}(KN_rN\log_2 N)$ described in Appendix II, avoiding using a convex solver of CC order of $\mathcal{O}((KN_r)^3)$.

B. Scalable-complexity closed form for aRIS descent

To seek $z^{(\kappa+1)}$ so that

$$\begin{aligned} & f_{\rho,SM}(W^{(\kappa+1)}, z^{(\kappa+1)}, p^{(\kappa)}, \theta^{(\kappa)}) \\ \leq \; & f_{\rho,SM}(W^{(\kappa+1)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)}), \end{aligned} \tag{62}$$

we consider the following problem:

$$\min_{\mathbf{z}} \; f_{\rho,SM,3} \triangleq \left[f^{(\kappa)}_{SM,2}(\mathbf{z}) + \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - p^{(\kappa)}_m e^{\jmath\theta^{(\kappa)}_m}|^2 \right] \quad \text{s.t.} \quad (32b), \tag{63}$$

where in accordance to (33), $f^{(\kappa)}_{SM,2}(\mathbf{z}) \triangleq f_{SM}(W^{(\kappa+1)}, \mathbf{z}) = \ln|\Pi^{(\kappa)}_2(\mathbf{z})|$, with

$$\begin{aligned} \Pi^{(\kappa)}_2(\mathbf{z}) \triangleq \sum_{k\in\mathfrak{K}} & (I_{N_r} - [A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z})]^H ([A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z})]^2 \\ & + \mu\Phi^{(\kappa)}_{2,k}(\mathbf{z}))^{-1} [A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z})]). \end{aligned} \tag{64}$$

Using the inequality (76) in Appendix I for $(\mathbf{V}_k, \mathbf{Y}_k) = (A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z}), [A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z})]^2 + \mu\Phi^{(\kappa)}_{2,k}(\mathbf{z}))$, and $(\bar{V}_k, \bar{Y}_k) = (V^{(\kappa)}_{2,k}, Y^{(\kappa)}_{2,k}) \triangleq (A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(z^{(\kappa)}), [A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(z^{(\kappa)})]^2 + \mu\Phi^{(\kappa)}_{2,k}(z^{(\kappa)}))$, yields the following tight majorant of $f^{(\kappa)}_{SM,2}(\mathbf{z})$ at $z^{(\kappa)}$:

$$\begin{aligned} \tilde{f}^{(\kappa)}_{2,SM}(\mathbf{z}) \triangleq \; & \tilde{a}^{(\kappa)}_2 - 2\sum_{k\in\mathfrak{K}} \Re\{\langle \tilde{\mathcal{B}}^{(\kappa)}_{2,k} (A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z}))\rangle\} \\ & + \sum_{k\in\mathfrak{K}} \langle \Psi^{(\kappa)}_{2,k}, [A^{(\kappa)}_{k,k} + \Delta^{(\kappa)}_{k,k}(\mathbf{z})]^2 + \mu \Bigg(\sum_{j\in\mathfrak{K}\setminus\{k\}} \\ & [A^{(\kappa)}_{k,j} + \Delta^{(\kappa)}_{k,j}(\mathbf{z})]^2 + \sigma_\nu [\Xi_k(\mathbf{z})]^2 \Bigg) \rangle \qquad (65) \\ = \; & a^{(\kappa)}_2 - 2\sum_{k\in\mathfrak{K}} \sum_{j\in\mathfrak{K}} \Re\{\langle \mathcal{B}^{(\kappa)}_{2,k,j} \Delta^{(\kappa)}_{k,j}(\mathbf{z})\rangle\} + \sum_{k\in\mathfrak{K}} \\ & \Bigg(||\sqrt{\Psi^{(\kappa)}_{2,k}} \Delta^{(\kappa)}_{k,k}(\mathbf{z})||^2 + \mu \sum_{j\in\mathfrak{K}\setminus\{k\}} ||\sqrt{\Psi^{(\kappa)}_{2,k}} \\ & \Delta^{(\kappa)}_{k,j}(\mathbf{z})||^2 + \mu\sigma_\nu ||\sqrt{\Psi^{(\kappa)}_{2,k}} \Xi_k(\mathbf{z})||^2 \Bigg) \qquad (66) \\ = \; & a^{(\kappa)}_2 - 2\sum_{k\in\mathfrak{K}} \Re\{b^{(\kappa)}_{2,k}\mathbf{z}\} + \sum_{k\in\mathfrak{K}} \langle \mathcal{C}^{(\kappa)}_{2,k}, [\mathbf{z}]^2\rangle \qquad (67) \\ = \; & a^{(\kappa)}_2 - 2\Re\{b^{(\kappa)}_2\mathbf{z}\} + \langle \mathcal{C}^{(\kappa)}_2, [\mathbf{z}]^2\rangle, \qquad (68) \end{aligned}$$

where

$$\begin{aligned} \tilde{a}^{(\kappa)}_2 \; & \triangleq f^{(\kappa)}_{SM,z}(z^{(\kappa)}) + \textstyle\sum_{k\in\mathfrak{K}} \langle [\Pi^{(\kappa)}_2(z^{(\kappa)})]^{-1} (V^{(\kappa)}_{2,k})^H \\ & (Y^{(\kappa)}_{2,k})^{-1} V^{(\kappa)}_{2,k}\rangle + \mu\sigma \textstyle\sum_{k\in\mathfrak{K}} \langle \Psi^{(\kappa)}_{2,k}\rangle, \\ \mathcal{B}^{(\kappa)}_{2,k} \; & \triangleq [\Pi^{(\kappa)}_2(z^{(\kappa)})]^{-1} (V^{(\kappa)}_{2,k})^H (Y^{(\kappa)}_{2,k})^{-1}, \\ \Psi^{(\kappa)}_{2,k} \; & \triangleq (Y^{(\kappa)}_{2,k})^{-1} V^{(\kappa)}_{2,k} [\Pi^{(\kappa)}_2(z^{(\kappa)})]^{-1} (V^{(\kappa)}_{2,k})^H (Y^{(\kappa)}_{2,k})^{-1}, \end{aligned}$$

in (65), and

$$\begin{aligned} a^{(\kappa)}_2 \; & \triangleq \tilde{a}^{(\kappa)}_2 - 2\textstyle\sum_{k\in\mathfrak{K}} \Re\{\langle \tilde{\mathcal{B}}^{(\kappa)}_{2,k} A^{(\kappa)}_{k,k}\rangle\} \\ & + \textstyle\sum_{k\in\mathfrak{K}} \langle \Psi^{(\kappa)}_{2,k}, [A^{(\kappa)}_{k,k}]^2 + \mu \textstyle\sum_{j\in\mathfrak{K}\setminus\{k\}} [A^{(\kappa)}_{k,j}]^2\rangle, \\ \mathcal{B}^{(\kappa)}_{2,k,j} \; & = \begin{cases} (\tilde{\mathcal{B}}^{(\kappa)}_{2,k} - (A^{(\kappa)}_{k,k})^H \Psi^{(\kappa)}_{2,k} & \text{for} \quad j = k \\ -\mu (A^{(\kappa)}_{k,j})^H \Psi^{(\kappa)}_{2,k} & \text{otherwise} \end{cases} \end{aligned}$$

in (66), and $b^{(\kappa)}_{2,k} \triangleq \sum_{j\in\mathfrak{K}} \sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r} \mathcal{B}^{(\kappa)}_{2,k,j}(\ell,\ell') \Delta^{(\kappa)}_{k,j}(\ell',\ell)$,

$$\begin{aligned} \mathcal{C}^{(\kappa)}_{2,k} \triangleq \; & \sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r} \left[\left(\sum_{t=1}^{N_r} \sqrt{\Psi^{(\kappa)}_{2,k}}(\ell,t) \Delta^{(\kappa)}_{k,k}(t,\ell') \right)^H \right]^2 \\ & + \mu \sum_{j\in\mathfrak{K}\setminus\{k\}} \sum_{(\ell,\ell')\in\mathfrak{N}_r\times\mathfrak{N}_r} \left[\left(\sum_{t=1}^{N_r} \sqrt{\Psi^{(\kappa)}_{2,k}}(\ell,t) \Delta^{(\kappa)}_{k,j}(t,\ell') \right)^H \right]^2 \\ & + \mu\sigma_\nu \textstyle\sum_{\ell\in\mathfrak{N}_r} \left[\left(\sum_{\ell'\in\mathfrak{N}_r} \sqrt{\Psi^{(\kappa)}_{2,k}}(\ell,\ell') \mathcal{D}^T_{k,\ell'} \right)^H \right]^2 \end{aligned}$$

in (67), and $b^{(\kappa)}_2 \triangleq \sum_{k\in\mathfrak{K}} b^{(\kappa)}_{2,k}$, $\mathcal{C}^{(\kappa)}_2 \triangleq \sum_{k\in\mathfrak{K}} \mathcal{C}^{(\kappa)}_{2,k}$ in (68).

We thus solve the following convex problem of majorant minimization of (63) to generate $z^{(\kappa+1)}$ verifying (62):

$$\min_{\mathbf{z}} \left[\tilde{f}^{(\kappa)}_{2,SM}(\mathbf{z}) + \rho \sum_{m\in\mathfrak{M}} |\mathbf{z}_m - p^{(\kappa)}_m e^{\jmath\theta^{(\kappa)}_m}|^2 \right] \quad \text{s.t.} \quad (32b), \tag{69}$$

which admits the closed-form solution of

$$z^{(\kappa+1)} = \begin{cases} (\mathcal{C}^{(\kappa)}_2 + \rho I_M)^{-1} \xi^{(\kappa)}_2 \\ \text{if} \quad ||\sqrt{\mathcal{Q}^{(\kappa)}_2} (\mathcal{C}^{(\kappa)}_2 + \rho I_M)^{-1} \xi^{(\kappa)}_2||^2 \leq P_A \\ (\mathcal{C}^{(\kappa)}_2 + \rho I_M + \alpha\mathcal{Q}^{(\kappa)}_2)^{-1} \xi^{(\kappa)}_2 \quad \text{otherwise,} \end{cases}$$

where $\xi^{(\kappa)}_2 \triangleq (b^{(\kappa)}_2)^H + \rho(p^{(\kappa)}_1 e^{\jmath\theta^{(\kappa)}_1}, \ldots, p^{(\kappa)}_M e^{\jmath\theta^{(\kappa)}_M})^T$, and $\alpha > 0$ is found by bisection so that $||\sqrt{\mathcal{Q}^{(\kappa)}_2} (\mathcal{C}^{(\kappa)}_2 + \rho I_M + \alpha\mathcal{Q}^{(\kappa)}_2)^{-1} \xi^{(\kappa)}_2||^2 = P_A$. The computational complexity of (70) is $\mathcal{O}(M \log_2 M)$.

C. Convergence and computational efficiency

Algorithm 3 provides the pseudo code for solving the soft max-min log-det problem (56). It follows from (57), (62) and (40) that

$$\begin{aligned} & f_{\rho,SM}(W^{(\kappa+1)}, z^{(\kappa+1)}, p^{(\kappa+1)}, \theta^{(\kappa+1)}) \quad < \\ & \qquad f_{\rho,SM}(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)}), \end{aligned} \tag{70}$$

as far as $f_{\rho,SM}(W^{(\kappa+1)}, z^{(\kappa+1)}, p^{(\kappa+1)}, \theta^{(\kappa+1)}) \neq f_{\rho,SM}(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})$, so the sequence $\{(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})\}$ of improved feasible points for (56) converges to $(\bar{w}, \bar{z}, \bar{p}, \bar{\theta})$, which is a feasible point for (55).

Given the cubic-time computational complexity of each iteration in the quadratic-solver-based Algorithm 1, it's noteworthy that each iteration in both Algorithm 2 and Algorithm 3 exhibits scalable complexities, rendering them computationally efficient.

V. Simulations

The various components and parameters in (4) are set following [25]–[27]. The large scale fading coefficients, $\beta_{B,R}$ and $\beta_{B,R}$ in (4), β_k in (5), are modeled as [26], [27]

$$\beta_{B,R} = G_{BS} + G_{RIS} - 35.9 - 22\log_{10}(d_{B,R}), \tag{71}$$

$$\beta_{R,k} = G_{RIS} - 33.05 - 30\log_{10}(d_{R,k}), \tag{72}$$

Algorithm 3 Scalable-complex soft max-min log-det algorithm

1: **Initialization:** Randomly generate a feasible $(W^{(0)}, z^{(0)}, p^{(0)}, \theta^{(0)})$ for (56). Set $\kappa = 0$.
2: **Repeat until convergence:** Generate $W^{(\kappa+1)}$ by solving (61) of the CC $\mathcal{O}(KN_rN\log_2 N)$, and $z^{(\kappa+1)}$ by (70) of the CC $\mathcal{O}(M\log_2 M)$. Generate $(p^{(\kappa+1)}, \theta^{(\kappa+1)})$ by (38)-(39). Reset $\kappa \leftarrow \kappa + 1$.
3: **Output** $(W^{(\kappa)}, z^{(\kappa)}, p^{(\kappa)}, \theta^{(\kappa)})$ and $r_k(W^{(\kappa)}, z^{(\kappa)})$, $k \in \mathfrak{K}$.

$$\beta_k = G_{\text{BS}} - 33.05 - 36.7\log_{10}(d_{\text{k}}), \tag{73}$$

where the antenna gains are $G_{RIS} = G_{BS} = 5$ dBi. Furthermore, $\bar{G}_{B,R}$ with $\bar{G}_{B,R}(m,n) = e^{\jmath\pi\left((m_1-1)\sin\bar{\theta}_m \sin\bar{\psi}_m + (n-1)\sin e^{\jmath\theta_m}\sin\psi_m\right)}$ represents the small-scale Rician fading, where $\bar{\theta}_m = \pi - \theta_m$ and $\bar{\phi}_m = \pi + \phi_m$ with $e^{\jmath\theta_m}$ and ψ_m following a uniform distribution within $(0,\pi)$ and $(0,2\pi)$ Lastly, $\bar{H}_{R,k} = \sqrt{\frac{L}{L+1}}\hat{H}_{R,k}^{LoS} + \sqrt{\frac{1}{L+1}}\hat{H}_{R,k}^{NLoS}$, where $\hat{H}_{R,k}^{LoS}$ and $\hat{H}_{R,k}^{NLoS}$ are the LoS and non-LoS (NLoS) components, while $L = 4.7$ dB is the Rician factor.

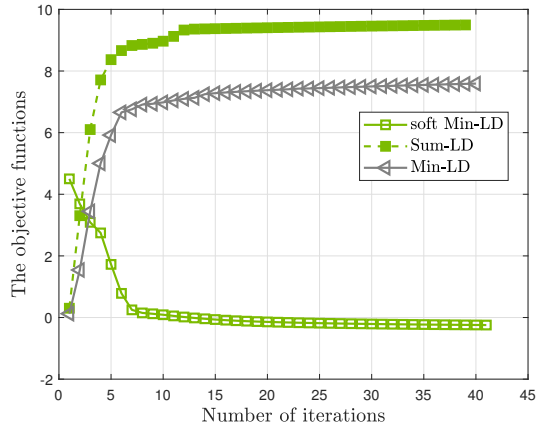
3-D coordinates of the BS and aRIS in Fig. 1 are set to $(0,0,25)$ m and $(50,80,40)$ m, respectively. A total of $K = 10$ users are distributed randomly in a (200m 100m) area right of the BS and aRIS.

Recalling that N is the number of BS antennas, N_r is the number of each user's antennas, P is the transmit power, and M is the number of APREs, we set $(N, N_r, M, P) = (16, 2, 100, 20\text{dBm})$ unless specified otherwise. Furthermore, the power split between the BS and aRIS is $0.99P$ and $0.01P$.

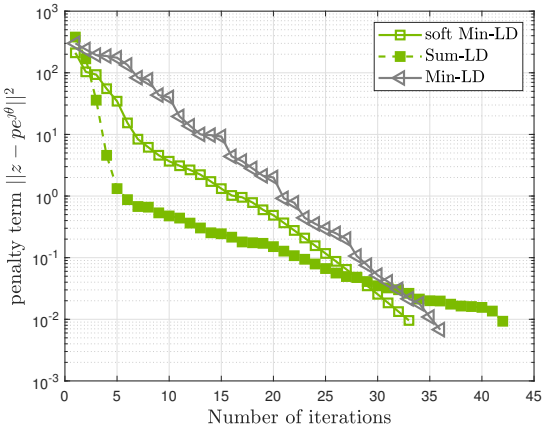
The following legends are used to describe the proposed implementations: (i) Min-LD denotes the performance of the CC QS-based Algorithm 1, designed to solve the max-min log-det problem (24); (ii) Sum-LD (Sum-LD w/o, resp.) represent the performance of the scalable-complexity closed form-based Algorithm 2, designed to solve the Sum-Log-Det problem (43) (without aRIS, resp.); (iii) soft Min-LD (soft Min-LD w/o, resp.) indicates the performance of the scalable-complexity closed form-based Algorithm 3, designed for solving the soft max-min log-det problem (55) (without aRIS, resp.).

A. Algorithmic convergence and efficiency

To guarantee the convergence of the proposed algorithms and usher the penalty terms in (24), (43), and (56) towards zero, the penalty parameter ρ starts at $\rho = 10^{-4}$ and it is incrementally increased by a factor of 1.2 in each iteration. Fig. 2 plots the typical convergence of the OFs and penalty terms. To illustrate the convergence behaviors of the proposed algorithms, we utilize the mean log-det rate value within the objective function for the sum log-det rate problem, as depicted in Fig. 2(a). According to Fig. 2(a), the OFs of the Min-LD and Sum-LD (soft Min-LD, resp.) increase (decrease, resp.) rapidly within 35-45 iterations, confirming (41) and (52) ((70), resp.), and then gradually converge. Furthermore,



(a)



(b)

Fig. 2: The convergence of the proposed algorithms

Fig. 2(b) clearly shows that the penalty terms practically converge to zero within 30-45 iterations.

Table II presents the average computational time of the proposed algorithms, measured using a CPU with a 3.7 GHz Intel Core i9 and 32 GB RAM. Alg. 2 and Alg. 3, which are based on our scalable-complexity iterations, run more than 20 times faster than Alg. 1, which is based on cubic-complexity iterations, confirming the superiority of scalable-complexity iterations.

Selecting a smaller value of μ in the soft max-min log-det problem (56) does not always result in improved minimum log-det performance, as demonstrated by the data seen in Table III, which provides the achieved minimum log-det values for various $\mu \in \{1, 0.5, 0.1\}$. For $N = 8$, soft max-min log-det optimization with $\mu = 1$ yields the lowest minimum rate. The performance of soft max-min log-det problem with $\mu = 0.5$ is better than that with $\mu = 0.1$. In the case of $N \in \{16, 24, 32, 40\}$, soft max-min log-det problem with $\mu = 1$ demonstrates the minimum rate is quite similar to that with $\mu = 0.5$. It is evident that $\mu = 0.5$ consistently delivers better performance and it is therefore our choice for subsequent simulations.

TABLE II: The average computational time of the proposed algorithms.

	N = 8	N = 16	N = 24	N = 32	N = 40
Algorithm 1	11.41 min.	15.46 min.	17.37 min.	22.63 min.	31.37min.
Algorithm 2	0.47 min.	0.74 min.	1.03 min.	1.45 min.	1.79 min.
Algorithm 3	0.35 min.	0.56 min.	0.69 min.	1.10 min.	1.42 min.

TABLE III: The minimum log-det versus μ for different numbers of BS antennas N achieved by soft max-min log-det optimization.

	N = 8	N = 16	N = 24	N = 32	N = 40
μ = 0.1	1.4398	6.3335	11.7243	12.5828	13.1305
μ = 0.5	1.9508	7.6659	12.3407	14.3264	15.7510
μ = 1	1.0443	7.3683	11.7848	14.4856	15.9516

We compare the performance of our Sum-LD algorithm and [6, Algorithm 1], referred to as Zhang et al., which is designed to solve the sum-rate problem (43) with $N_r = 1$ and unquantized PREs. Fig. 3 plots the sum rate versus transmit power budget P, with $(N, M) = (8, 100)$ (8 BS antennas and 100 aRIS elements). It can be observed from Fig. 3 that the performance of Sum-LD is better than that of Zhang et al.

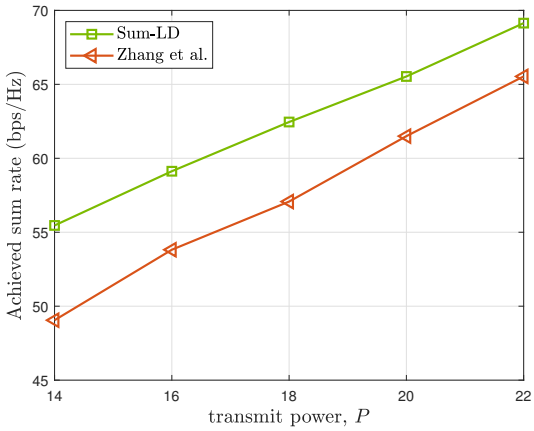


Fig. 3: The sum rate versus the transmit power P under the power split of $0.99P$ and $0.01P$ between the BS and aRIS.

B. MU QoD performance

Fig. 4(a) and Fig. 4(b) plot the minimum rate and sum rate achieved versus the number of BS antennas N, respectively. Fig. 4 clearly illustrates the performance advantage of the proposed algorithms attained by more BS antennas. As depicted in Fig. 4(a), illustrating the minimum rate performances, Min-LD exhibits the best performance, while Sum-LD surpasses its Sum-LD w/o counterpart. Soft Min-LD and Min-LD outperform Sum-LD, with Sum-LD w/o showing the poorest performance. Furthermore, both soft Min-LD and Min-LD benefit more substantially from increasing the number of BS antennas compared to their soft Min-LD w/o and Min-LD w/o counterparts. Regarding the sum rate performance shown in Fig. 4(b), Sum-LD is the top performer, but for $N \geq 16$, soft Min-LD catches up. For $N \in \{32, 40\}$, soft min-LD w/o provides the lowest sum-rate. Additionally, Fig. 4(b) highlights the significant performance enhancement achieved by soft Min-LD compared to its soft Min-LD w/o counterparts.

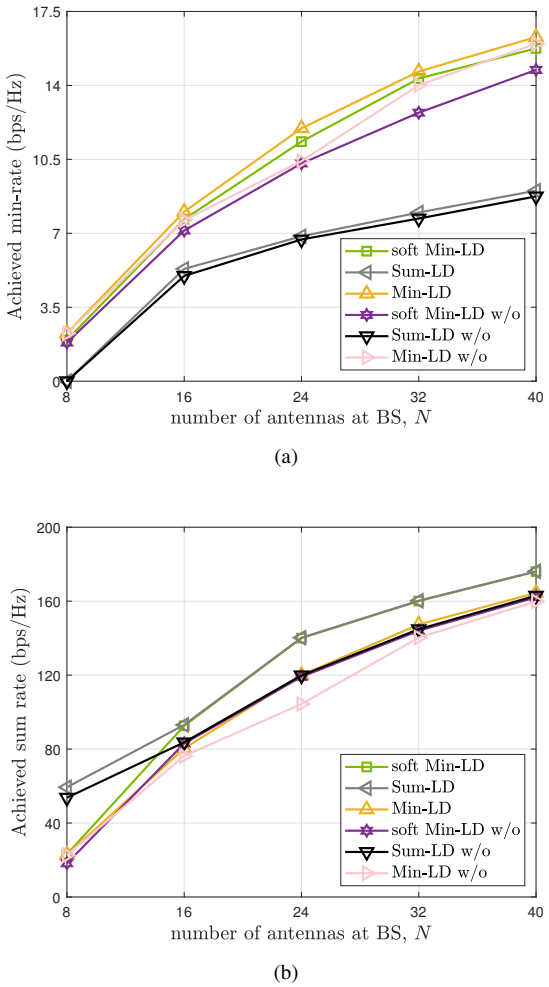
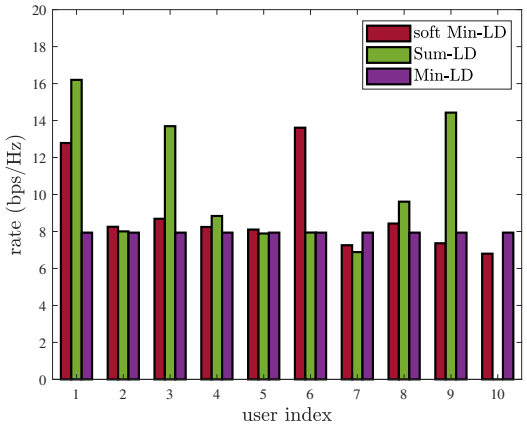


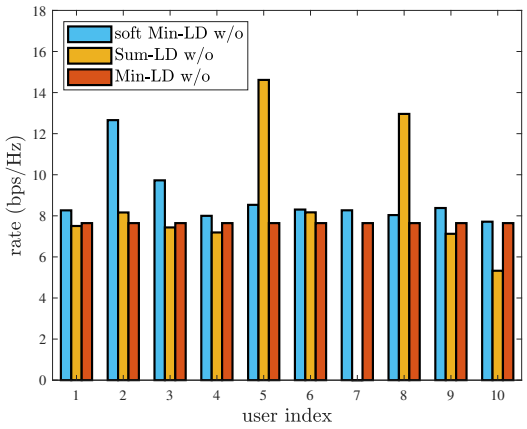
Fig. 4: The min-rate and sum rate achieved versus the number of BS antennas N.

Figures 5(a) and 5(b) illustrate the log-det rate distributions produced by the proposed algorithms. Notably, the log-det rates generated by the soft Min-LD and Min-LD exhibit a commendable uniformity across all users, highlighting their

ability to ensure equitable performance. By contrast, both Sum-LD and Sum-LD w/o allocate zero rates to specific users (UE 10 and UE 7, respectively), underscoring their limitations in achieving fair MU QoD.



(a)



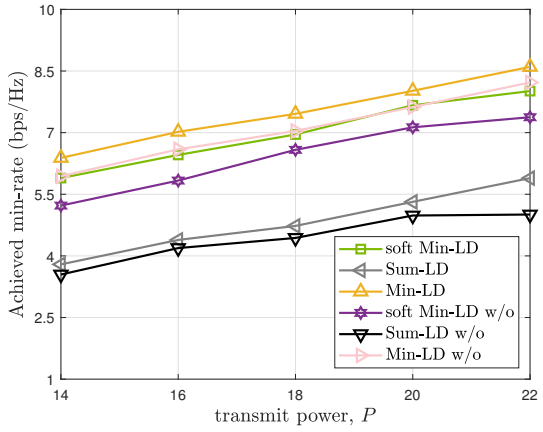
(b)

Fig. 5: Rate distribution versus user index.

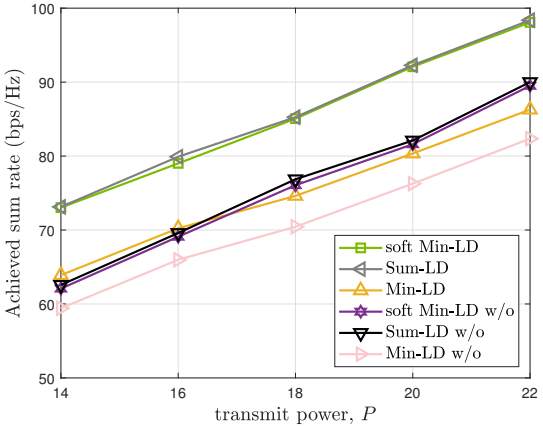
Fig. 6(a) and Fig. 6(b) show the minimum rate and sum rate achieved upon varying the transmit power budget P, respectively. In terms of the minimum rate attained, Fig. 6(a) indicates that Min-LD is the best performer, followed by soft Min-LD, while Min-LD w/o performs better than soft Min-LD w/o. Sum-LD w/o exhibits the worst performance. In terms of the sum rate achieved, Fig. 6(b) shows that both soft Min-LD and Sum-LD perform similarly. The sum-rate of soft Min-LD and Sum-LD is better than that of soft Min-LD w/o and Sum-LD w/o. Min-LD w/o exhibits the lowest sum rate.

Furthermore, Fig.7 shows that the minimum rate achieved of soft Min-LD, Sum-LD and Min-LD improves with an increase in the number of APREs. As expected, Min-LD achieves the best minimum rate. The performance of soft Min-LD and Min-LD is better than that of soft Min-LD w/o and Min-LD w/o.

Fig. 8 demonstrates that soft Min-LD, Sum-LD and Min-LD achieve the highest minimum log det rate, when the power split P_B is set to $0.99P$. It can be observed from Fig. 8 that



(a)



(b)

Fig. 6: The min-rate and sum rate achieved versus the transmit power P.

the minimum log det rate of soft Min-LD w/o, Sum-LD w/o and Min-LD w/o remains constant.

Lastly, Fig. 9(a) and Fig. 9(b) illustrate the advantage of increasing the resolution b. In Fig. 9, it is evident that with the increase in b, there is a slightly rise in both the minimum rate and the sum-rate. However, when the phase of aRIS achieves infinite resolution, a substantial performance improvement is observed.

VI. Conclusions

The paper has considered the join design of the TPCs and of the aRIS's power-amplified reconfigurable elements (APRES) for enhancing the quality of delivering multi-stream information to multiple multi-antenna users. Users' rates have been characterized in terms of their complex log-det functions, leading to challenging large-scale mixed discrete-continuous optimization problems. We have initially shown that the maximization of the minimum of the users' log-det rate functions can be solved based on quadratic problems of cubically increasing complexity. To circumvent this complexity escalation, which is certainly excessive in large-scale

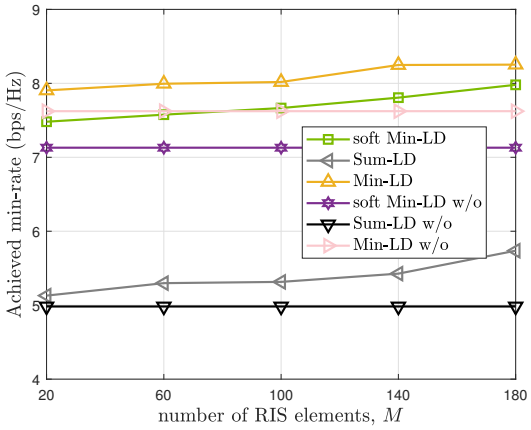


Fig. 7: The achieved minimum rate versus the number of RIS elements M.

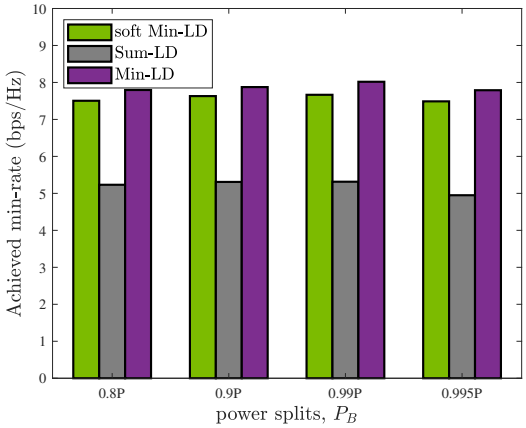
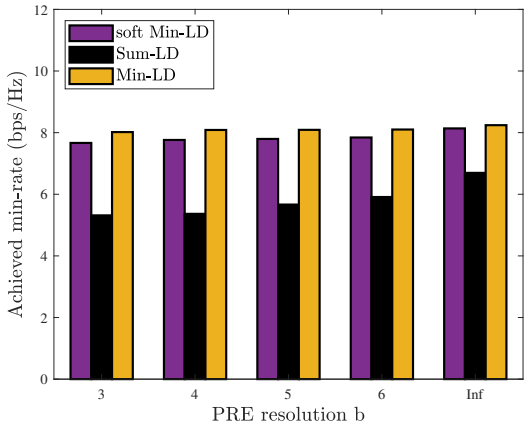
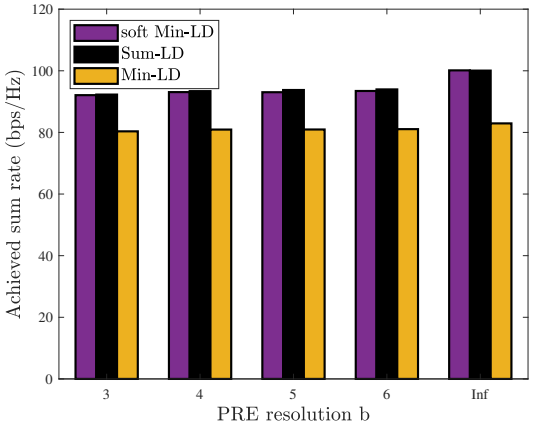


Fig. 8: The min-rate of different power splits, P_B.



(a)



(b)

Fig. 9: The minimum rate and sum rate achieved versus the resolution b.

scenarios, we have developed a pair of alternative algorithms, which rely on closed-form expressions of scalable (linear) complexity in maximizing the sum of the users' log-det rate functions (sum log-det) and of the soft minimum of the users' log-det rate functions (soft min log-det). These algorithms mitigate the computational burden typically associated with large-scale computations. Furthermore, the soft min log-det optimization not only ensures fair and effective log-det rate enhancements for all users but also achieves a high sum log-det, hence improving the delivery quality of multi-user multi-stream information.

Appendix I: Fundamental tight minorants and majorants of log-det rate functions

Recall from [30, p. 366] that a function $\bar{f}$ is considered a tight minorant of a function f at $\bar{x}$ over the domain $\mathrm{dom}(f)$ if $f(x) \geq \bar{f}(x)\ \forall\ x \in \mathrm{dom}(f)$ and $f(\bar{x}) = \bar{f}(\bar{x})$. Then it can be readily shown that $f(\bar{x}_{\max}) > f(\bar{x})$ as far as $\bar{f}(\bar{x}_{\max}) \neq \bar{f}(\bar{x})$ for $\bar{x}_{\max} = \arg\max_{x \in \mathrm{dom}(f)} \bar{f}(x)$, i.e. maximizing a tight minorant at the point $\bar{x}$ helps to obtain a better point.
Analogously, a function $\bar{f}$ is considered a tight majorant of a function f at $\bar{x}$ over the domain $\mathrm{dom}(f)$ if $f(x) \leq \bar{f}(x)\ \forall\ x \in \mathrm{dom}(f)$ and $f(\bar{x}) = \bar{f}(\bar{x})$. Then it can be readily shown that $f(\bar{x}_{\min}) < f(\bar{x})$ as far as $\bar{f}(\bar{x}_{\min}) \neq \bar{f}(\bar{x})$ for $\bar{x}_{\min} = \arg\min_{x \in \mathrm{dom}(f)} \bar{f}(x)$, i.e. minimizing a tight majorant at the point $\bar{x}$ helps to obtain a better point.

For the matrix variables $\mathbf{V} \triangleq (\mathbf{V}_1, \ldots, \mathbf{V}_K)$ and $\mathbf{Y} \triangleq (\mathbf{Y}_1, \ldots, \mathbf{Y}_K)$ with $\mathbf{V}_k \in \mathbb{C}^{N \times M}$, and $0 \prec \mathbf{Y} \in \mathbb{C}^{N \times N}$, $k \in K$, consider the following matrix-valued mapping:

$$\mathbb{C}^{M \times M} \ni \Pi(\mathbf{V}, \mathbf{Y}) \triangleq \sum_{k \in \mathfrak{K}} (I_M - \mathbf{V}_k^H \mathbf{Y}_k^{-1} \mathbf{V}_k), \tag{74}$$

in the domain

$$\mathrm{dom}_\Pi = \left\{ (\mathbf{V}, \mathbf{Y}) : [\mathbf{V}_k]^2 \prec \mathbf{Y}_k, k \in K \right\}. \tag{75}$$

By using the Shur's complement [24], we can show that $\Pi_k(\mathbf{V}_k, \mathbf{Y}_k) \triangleq I_M - \mathbf{V}_k^H \mathbf{Y}_k^{-1} \mathbf{V}_k \succ 0$, $k \in K$, so $\Pi(\mathbf{V}, \mathbf{Y}) \succ 0$, and moreover it satisfies the following concavity matrix inequality [33, Appendix A]: $\Pi(\lambda(\bar{V}, \bar{Y}) + (1-\lambda)(\tilde{V}, \tilde{Y})) \succeq \lambda\Pi(\bar{V}, \bar{Y}) + (1-\lambda)\Pi(\tilde{V}, \tilde{Y}))$, $(\bar{V}, \bar{Y}) \in \mathrm{dom}_\Pi$, $(\tilde{V}, \tilde{Y}) \in \mathrm{dom}_\Pi$, $\lambda \in [0, 1]$, making the composite function $\ln \left| \Pi(\mathbf{V}, \mathbf{Y}) \right|$ concave. Consequently, the following inequality holds for all

$(\mathbf{V},\mathbf{Y})\in \text{dom}_\Pi$ and $(\bar{V},\bar{Y})\in \text{dom}_\Pi$:

$$\begin{aligned}\ln\left|\Pi(\mathbf{V},\mathbf{Y})\right| \;\leq\; & \ln\left|\Pi(\bar{V},\bar{Y})\right| + \sum_{k\in\mathfrak{K}}\langle\Pi^{-1}(\bar{V},\bar{Y})\bar{V}_k^H\bar{Y}_k^{-1}\bar{V}_k\rangle \\ & -2\sum_{k\in\mathfrak{K}}\Re\{\langle\Pi^{-1}(\bar{V},\bar{Y})\bar{V}_k^H\bar{Y}_k^{-1}\mathbf{V}_k\rangle\} \\ & +\sum_{k\in\mathfrak{K}}\langle\bar{Y}_k^{-1}\bar{V}_k\Pi^{-1}(\bar{V},\bar{Y})\bar{V}_k^H\bar{Y}_k^{-1}\mathbf{Y}_k\rangle. \quad (76)\end{aligned}$$

In fact, the right-hand-side (RHS) of (76) represents the linearized function of the concave function $\ln\left|\Pi(\mathbf{V},\mathbf{Y})\right|$ on the left-hand-side (LHS), so the former serves a tight majorant of the latter at $(\bar{V},\bar{Y})\in\text{dom}_\Pi$ [30].

The following inequality has been established in [11], [13] for all matrices $\mathbf{V}$ and $\bar{V}$ of size $n\times m$ and matrices $\mathbf{Y}\succ 0$ and $\bar{Y}\succ 0$ of size $N\times N$:

$$\begin{aligned}\ln\left|I_N+[\mathbf{V}]^2\mathbf{Y}^{-1}\right| \;\geq\; & \ln\left|I_N+[\bar{V}]^2\bar{Y}^{-1}\right| - \langle[\bar{V}]^2\bar{Y}^{-1}\rangle + \\ & 2\Re\{\langle\bar{V}^H\bar{Y}^{-1}\mathbf{V}\rangle\} - \left\langle\bar{Y}^{-1}-(\bar{Y}+[\bar{V}]^2)^{-1},[\mathbf{V}]^2+\mathbf{Y}\right\rangle, \quad (77)\end{aligned}$$

i.e. the function on the RHS, which is a concave quadratic function because $(\bar{Y})^{-1}-(\bar{Y}+[\bar{V}]^2)^{-1}\succeq 0$, serves as a tight minorant of the log-det function on the LHS at the point $(\bar{V},\bar{Y})$.

Appendix II: Bisection of scalable complexity for quadratic constrained matrix optimization

We consider the following quadratic constrained problem of multi-variable matrix optimization:

$$\begin{aligned}\min_{\mathbf{V}_k\in\mathbb{C}^{N\times M},k\in\mathfrak{K}} & -2\sum_{k\in\mathfrak{K}}\Re\{\langle B_k^H\mathbf{V}_k\rangle\}+\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{1,k}}\mathbf{V}_k||^2 \\ \text{s.t.} \quad & \sum_{k\in\mathfrak{K}}||\mathbf{V}_k||^2\leq P_1, \sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}\mathbf{V}_k||^2\leq P_2, \quad (78)\end{aligned}$$

with given $\mathbb{C}^{N\times N}\ni\mathcal{Q}_{1,k}\succeq 0$ and $\mathbb{C}^{N\times N}\ni\mathcal{Q}_{2,k}\succeq 0$ and $B_k\in\mathbb{C}^{N\times M}$, $k\in\mathfrak{K}$, and $P_1>0$ and $P_2>0$.

Using the matrix least squares approach, the solution of the unconstrained problem $-2\sum_{k\in\mathfrak{K}}\Re\{\langle B_k^H\mathbf{V}_k\rangle\}+\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{1,k}}\mathbf{V}_k||^2\to\min$ is given by

$$V_k^{opt}=\mathcal{Q}_{1,k}^{-1}B_k, k\in\mathfrak{K} \quad (79)$$

which is still the solution of (78), provided that it satisfies both constraints in (78), i.e.

$$\sum_{k\in\mathfrak{K}}||\mathcal{Q}_{1,k}^{-1}B||^2\leq P_1 \quad (80)$$

and

$$\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}\mathcal{Q}_{1,k}^{-1}B_k||^2\leq P_2. \quad (81)$$

When (81) is not met, we use bisection to find $\tilde{\lambda}_2>0$, so that

$$\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}(\mathcal{Q}_{1,k}+\tilde{\lambda}_2\mathcal{Q}_{2,k})^{-1}B_k||^2=P_2. \quad (82)$$

The solution of (78) is

$$V_k^{opt}=(\mathcal{Q}_{1,k}+\tilde{\lambda}_2\mathcal{Q}_{2,k})^{-1}B_k, k\in\mathfrak{K}, \quad (83)$$

whenever

$$\sum_{k\in\mathfrak{K}}||(\mathcal{Q}_{1,k}+\tilde{\lambda}_2\mathcal{Q}_{2,k})^{-1}B_k||^2\leq P_1. \quad (84)$$

If (80) is not met, we use bisection to find $\tilde{\lambda}_1>0$ so that

$$\sum_{k\in\mathfrak{K}}||(\mathcal{Q}_{1,k}+\tilde{\lambda}_1 I_N)^{-1}B_k||^2=P_1. \quad (85)$$

The solution of (78) is

$$V_k^{opt}=(\mathcal{Q}_{1,k}+\tilde{\lambda}_1 I_N)^{-1}B_k, k\in\mathfrak{K} \quad (86)$$

whenever

$$\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}(\mathcal{Q}_{1,k}+\lambda_1 I_N)^{-1}B_k||^2\leq P_2. \quad (87)$$

In the remaining case, the optimal solution is represented as $V_k^{opt}=(\mathcal{Q}_{1,k}+\bar{\lambda}_1 I_N+\bar{\lambda}_2\mathcal{Q}_{2,k})^{-1}B_k$, $k\in\mathfrak{K}$, where $\bar{\lambda}_1>0$ and $\bar{\lambda}_2>0$ are Lagrangian multipliers so that $\sum_{k\in\mathfrak{K}}||(\mathcal{Q}_{1,k}+\bar{\lambda}_1 I_N+\bar{\lambda}_2\mathcal{Q}_{2,k})^{-1}B_k||^2=P_1$, and $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}(\mathcal{Q}_{1,k}+\bar{\lambda}_1 I_N+\bar{\lambda}_2\mathcal{Q}_{2,k})^{-1}B_k||^2=P_2$, which constitute nonlinear equations in two unknown and thus are computationally intractable. We now avoid solving them by employing the partial Lagrangian multiplier method, which finds $\lambda_2>$ such that the solution V_k^{opt} of the following problem satisfies $\sum_{k\in\mathfrak{K}}\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k^{opt}||^2=P_2$:

$$\begin{aligned}\min_{\mathbf{V}_k} & -2\sum_{k\in\mathfrak{K}}\Re\{\langle B_k^H\mathbf{V}_k\rangle\}+\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{k,1}}\mathbf{V}_k||^2 \\ & +\lambda_2\left(\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}\mathbf{V}_k||^2-P_2\right) \\ & \text{s.t.} \quad \sum_{k\in\mathfrak{K}}||\mathbf{V}_k||^2\leq P_1. \quad (88)\end{aligned}$$

For a fixed λ_2, the solution of (88) is given by

$$V_k(\lambda_2)=\begin{cases}(\mathcal{Q}_{1,k}+\lambda_2\mathcal{Q}_{2,k})^{-1}B \\ \text{if} \quad \sum_{k\in\mathfrak{K}}||(\mathcal{Q}_{1,k}+\lambda_2\mathcal{Q}_{2,k})^{-1}B_k||^2\leq P_1 \\ \sum_{k\in\mathfrak{K}}(\mathcal{Q}_{1,k}+\lambda_2\mathcal{Q}_{2,k}+\hat{\lambda}_1 I_N)^{-1}B_k \text{ otherwise,}\end{cases}$$

where $\hat{\lambda}_1$ is found by bisection such that $\sum_{k\in\mathfrak{K}}||(\mathcal{Q}_{1,k}+\lambda_2\mathcal{Q}_{2,k}+\hat{\lambda}_1 I_N)^{-1}B_k||^2=P_1$. It follows from (87) that $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k(0)||^2>P_2$, while for λ_2 sufficient large we have $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k(\lambda_2)||^2<P_2$. We thus start from $\lambda_l=0$ and λ_u is chosen to satisfy $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k(\lambda_u)||^2<P_2$ and do the following bisection.

Bisection procedure. Set $\lambda_2=(\lambda_u+\lambda_l)/2$ and compute $V_k(\lambda_2)$ using (88). Terminate the procedure if $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k(\lambda_2)||^2\approx P_2$. Otherwise, update $\lambda_l\leftarrow\lambda_2$ if $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k(\lambda_2)||^2>P_2$ or $\lambda_u\leftarrow\lambda_2$ if $\sum_{k\in\mathfrak{K}}||\sqrt{\mathcal{Q}_{2,k}}V_k(\lambda_2)||^2<P_2$.

References

[1] D. Dardari, "Communicating with large intelligent surfaces: Fundamental limits and models," *IEEE J. Sel. Areas Commun.*, vol. 38, pp. 2526–2537, Nov. 2020.

[2] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, 2021.

[3] H. Zhang, B. Di, K. Bian, Z. Han, H. V. Poor, and L. Song, "Toward ubiquitous sensing and localization with reconfigurable intelligent surfaces," *Proc. IEEE*, vol. 110, no. 9, pp. 1401–1422, 2022.

[4] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. D. Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wirel. Commun.*, vol. 27, no. 5, pp. 118–125, 2020.

[5] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 20, pp. 4962–4975, Aug. 2021.

[6] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?," *IEEE Trans. Commun.*, vol. 71, pp. 1707–1725, Mar. 2023.

[7] L. Wei, C. Huang, G. C. Alexandropoulos, W. E. I. Sha, Z. Zhang, M. Debbah, and C. Yuen, "Multi-user holographic MIMO surfaces: Channel modeling and spectral efficiency analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 16, pp. 1112–1124, May 2022.

[8] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?," *IEEE Commun. Lett.*, vol. 26, pp. 1150–1154, May 2022.

[9] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wirel. Commun.*, vol. 28, pp. 106–113, Feb 2021.

[10] D. Dardari and N. Decarli, "Holographic communication using intelligent surfaces," *IEEE Commun. Mag.*, vol. 59, pp. 35–41, June 2021.

[11] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, pp. 2340–2353, June 2016.

[12] H. H. M. Tam, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "MIMO energy harvesting in full-duplex multi-user networks," *IEEE Trans. Wirel. Commun.*, vol. 16, pp. 3282–3297, May 2017.

[13] H. D. Tuan, H. H. M. Tam, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Superposition signaling in broadcast interference networks," *IEEE Trans. Commun.*, vol. 65, pp. 4646–4656, Nov. 2017.

[14] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge Univ. Press, 2016.

[15] L. D. Nguyen, H. D. Tuan, T. Q. Duong, O. A. Dobre, and H. V. Poor, "Downlink beamforming for energy-efficient heterogeneous networks with massive MIMO and small cells," *IEEE Trans. Wirel. Commun.*, vol. 17, pp. 3386–3400, May 2018.

[16] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Multi-user regularized zero-forcing beamforming," *IEEE Trans. Signal Process.*, vol. 67, pp. 2839–2853, June 2019.

[17] L. D. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Energy-efficient multi-cell massive MIMO subject to minimum user-rate constraints," *IEEE Trans. Commun.*, vol. 69, pp. 914–928, Feb. 2021.

[18] H. D. Tuan, A. A. Nasir, H. Q. Ngo, E. Dutkiewicz, and H. V. Poor, "Scalable user rate and energy-efficiency optimization in cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 70, pp. 6050–6065, Sept. 2022.

[19] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Regularized zero-forcing aided hybrid beamforming for millimeter-wave multiuser MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 22, pp. 3280–3295, May 2023.

[20] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Low-resolution hybrid beamforming in millimeter-wave multi-user systems," *IEEE Trans. Veh. Techn.*, vol. 72, pp. 8941–8955, July 2023.

[21] Y. Chen, H. D. Tuan, Y. Fang, H. Yu, H. V. Poor, and L. Hanzo, "Enhancing the downlink rate fairness of low-resolution active RIS-aided signaling by closed-form expression-based iterative optimization," *IEEE Trans. Veh. Techn. (early access)*.

[22] Y. Ma, M. Li, Y. Liu, Q. Wu, and Q. Liu, "Active reconfigurable intelligent surface for energy efficiency in MU-MISO systems," *IEEE Trans. Veh. Techn.*, vol. 72, pp. 4103–4107, Mar. 2023.

[23] H. Niu *et al.*, "Active RIS assisted rate-splitting multiple access network: Spectral and energy efficiency tradeoff," *IEEE J. Sel. Areas Commun.*, vol. 41, pp. 1452–1467, May 2023.

[24] R. A. Horn and C. R. Johnson, *Matrix analysis (second edition*. Cambridge Univ. Press, 1985.

[25] Q. U. A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 12, pp. 7748–7764, 2020.

[26] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 5, pp. 581–585, 2020.

[27] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2020.

[28] J. F. Bonnans, J. C. Gilbert, C. Lemarechal, and C. Sagastigabal, *Numerical Optimization-Theoretical and Practical Aspects (second edition)*. Springer, 2006.

[29] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo, "Nonsmooth optimization for efficient beamforming in cognitive radio multicast transmission," *IEEE Trans. Signal Process.*, vol. 60, pp. 2941–2951, Jun. 2012.

[30] H. Tuy, *Convex Analysis and Global Optimization (second edition)*. Springer International, 2017.

[31] D. Peaucelle, D. Henrion, Y. Labit, and K. Taitz, "Users guide for SeDuMi interface 1.04," 2002.

[32] A. Epasto, M. Mahdian, V. Mirrokni, and M. Zampetakis, "Optimal approximation-smoothness tradeoffs for soft-max functions," in *Adv. Neural Inf. Process. Syst.*, pp. 1–10, 2020.

[33] U. Rashid, H. D. Tuan, and H. H. Nguyen, "Relay beamforming designs in multi-user wireless relay networks based on throughput maximin optimization," *IEEE Trans. Commun.*, vol. 61, pp. 1739–1749, May 2013.