MIMO Channel Shaping and Rate Maximization Using Beyond-Diagonal RIS

Yang Zhao, *Member, IEEE,* Hongyu Li, *Member, IEEE,*
Bruno Clerckx, *Fellow, IEEE,* and Massimo Franceschetti, *Fellow, IEEE*

Abstract—This paper investigates the limits to which a passive Reconfigurable Intelligent Surface (RIS) can reshape a point-to-point Multiple-Input Multiple-Output (MIMO) channel in terms of singular values and their functions (e.g., achievable rate and harvestable power) for improved wireless performance. We depart from the Diagonal (D) scattering model and adopt a Beyond-Diagonal (BD) model that exploits element-wise connections for passive signal amplitude and phase manipulation. Specifically, analytical tight bounds are derived under typical RIS deployment scenarios to unveil the channel shaping potentials of BD-RIS regarding communication Degrees of Freedom (DoF), singular value spread, power gain, and capacity. An efficient numerical method is then proposed to optimize BD-RIS for any locally Lipschitz function of channel singular values, and showcased to characterize the achievable singular value region. As a side product, we tackle BD-RIS-aided MIMO rate maximization problem by a local-optimal Alternating Optimization (AO) approach and a low-complexity shaping approach. Results show that BD-RIS significantly improves the dynamic range of channel singular values and the tradeoff in manipulating them, thus offering enhanced data rate, harvestable power, and physical-layer security. These advantages become more pronounced when the number of RIS elements, group size, or MIMO dimensions increase. Of particular interest, BD-RIS is shown to activate multi-stream transmission and achieve the asymptotic DoF at much lower transmit power than D-RIS thanks to its proficiency in channel shaping.

Index Terms—MIMO, RIS, channel shaping, rate maximization, singular value analysis, manifold optimization.

I. INTRODUCTION

Today we are witnessing a paradigm shift from connectivity to intelligence, where the wireless environment is no longer a chaotic medium but a conscious agent that can serve on demand. This is empowered by recent developments in Reconfigurable Intelligent Surface (RIS), a programmable surface that recycles and redistributes ambient electromagnetic waves for improved wireless performance. A typical RIS consists of numerous low-power sub-wavelength scattering elements, whose response can be engineered in real-time to manipulate the amplitude, phase, frequency, and polarization of the scattered waves [1]. It enables full-duplex operation while featuring higher flexibility than reflectarrays, lower noise than relays, and greater scalability than multi-antenna transceivers. One popular RIS research topic is *joint beamforming* design with transceivers for a specific performance measure, which has attracted significant interests in wireless communication [2]–[4], sensing [5]–[7], and power transfer [8]–[10]. Although RIS-induced propagation paths suffers attenuation from double fading, passive beamforming at RIS offers better asymptotic behaviors than active beamforming at transceivers (e.g., second-order array gain and fourth-order har-

vested power [10]). Another RIS application is *information modulation* by periodically switching its reflection pattern within the channel coherence time. This creates a free-ride message stream with dual benefits – integrating with legacy transmitter for enhanced channel capacity [11], [12] or serving as individual source for low-power uplink communication [13], [14]. Different from above, *channel shaping* exploits RIS as a stand-alone device to modify the inherent properties of the wireless environment, for example, compensate for the Doppler effect [15], flatten frequency-selective channels [16], improve the channel rank [17], and introduce time diversity for multiple access schemes [18], [19]. This helps decouple joint beamforming problems into a channel shaping stage and a transceiver design stage, offering a modular and versatile solution for diverse wireless applications. At a specific time-frequency resource block, channel shaping metrics can be classified into the two categories below.

- *Singular value:* The impact of RIS has been studied in terms of channel minimum singular value [20], effective rank [21], condition number [22], and Degrees of Freedom (DoF) [23]. Those are closely related to explicit performance measures but sensitive to minor perturbations of the channel matrix;
- *Power:* The impact of RIS has been studied in terms of channel power gain [2], [24]–[27] in point-to-point channels and leakage interference [28] in interference channels. Those second-order metrics are less informative in the Multiple-Input Multiple-Output (MIMO) context but easier to analyze and optimize.

Although above works offered inspiring glimpses into the channel shaping potential of passive RIS, they neither provided in-depth theoretical analysis nor characterized the achievable singular value region. Most works [2], [20]–[23], [28] have also been confined to the conventional Diagonal (D) architecture where each RIS element is connected to a dedicated impedance and behaves independently of the others, namely, the wave impinging on one element is entirely scattered by itself. This architecture is modeled by a diagonal scattering matrix with unit-magnitude diagonal entries that, in the ideal case, applies a phase shift to the incident signal. The idea was soon extended to Beyond-Diagonal (BD)-RIS with group-connected architecture that connects elements within the same group via passive reconfigurable circuit components [24] that can be symmetric (e.g., capacitors and inductors) or asymmetric (e.g, ring hybrids and branch-line hybrids [29]). As such, the wave impinging on one element is able to propagate within the circuit and depart partially from the others of the same group. It can thus manipulate both amplitude and phase of the scattered wave

while remaining globally passive. The main manufacturing complexity of BD-RIS lies in the design and implementation of the circuit network. Fortunately, topologies such as tree- and forest-connected graphs have been proposed to reduce the number of components for a flexible cost-performance tradeoff [26]. Other practical challenges such as channel estimation [30], mutual coupling [31], wideband modelling [32], multi-sector coverage [33], and hardware implementation [34] have also been studied in recent literature. BD-RIS has been proved to achieve higher spectral efficiency than D-RIS and higher energy efficiency than active RIS and Amplify-and-Forward (AF) relay [27], [35], [36]. However, the interplay between BD-RIS and MIMO is still at infancy stage and its channel shaping potential remains largely unexplored. For example, the rate maximization problem [37] has only been tacked in the special case where the direct channel is negligible and the BD-RIS is fully-connected. Under those conditions, the mathematical modeling of BD-RIS coincides with that of AF relay of unit power, although the operation mechanism and noise characteristics are distinct.

When it comes to signal processing, existing works have mainly invoked the quasi-Newton method [24], the Penalty Dual Decompistion (PDD) method [36], and the generic (i.e., non-geodesic) Riemannian Conjugate Gradient (RCG) method [33] for the optimization of BD-RIS. The first solves an unconstrained problem and projects the solution back to the feasible domain without optimality guarantee. The second alternates between the primal variables in the inner layer and the penalty coefficient in the outer layer. It is often used to tackle coupled constraints (e.g., Signal-to-Interference Noise Ratio (SINR) thresholds under active and passive beamforming) and can be computationally expensive (e.g., $\mathcal{O}(N_\mathrm{s}^2)$ for D-RIS and $\mathcal{O}(N_\mathrm{s}^4)$ for fully-connected BD-RIS) [36, Table I]. The third applies the conjugate gradient method on generic Riemannian manifolds. Each iteration consists of an addition on the tangent space and a retraction to the feasible domain, which constitutes a zigzag path departing from and returning to the manifold. However, none of them effectively exploits the special structure of BD-RIS for accelerated convergence.

This paper is motivated by a fundamental question: *What is the channel shaping capability, in terms of singular values and their functions, of a passive RIS in MIMO channels?* Unlike existing works that focus on specific performance metrics or deployment scenarios, we aim for an understanding of the theoretical shaping limits (via analysis) and the achievable shaping results (via optimization) that are broadly applicable across diverse wireless applications. We believe a comprehensive shaping answer can serve as a theoretical support/reference for the vast number of RIS research papers and real-world applications. Without a framework that identifies the fundamental shaping limits of RIS, the design of truly optimal and efficient architectures will remain elusive. The contributions of this paper are summarized below.

First, we pioneer BD-RIS study in general MIMO channels and interpret its shaping ability as branch matching and mode alignment. Branch matching refers to pairing and combining the branches (i.e., entries) of backward and forward channels corresponding to each group of the BD-RIS. Mode alignment refers to aligning and ordering the modes (i.e., singular vectors) of the RIS-induced channels with those of the direct channel. The former arises uniquely from the off-diagonal entries of the BD-RIS scattering matrix while the latter is enabled by its block-unitary transformation.

Second, we provide an analytical answer to the shaping question under typical channel conditions. It is shown that BD-RIS may achieve a larger or smaller communication DoF than D-RIS. When the backward or forward channel is rank-deficient, we derive asymptotic bounds of individual singular values applying to D- and BD-RIS. When the direct channel is negligible, we recast the shaping question for fully-connected BD-RIS as a well-studied linear algebra question and provide tight bounds (with closed-form scattering matrices) on channel singular values, power gain, and capacity. These results help us understand the fundamental limits of channel shaping and serve as a foundation for application-specific designs.

Third, we provide a numerical BD-RIS design framework for any locally Lipschitz function of channel singular values via a geodesic RCG method. It compares favorably to generic manifold optimizers in that the updates are performed along the geodesics, namely the shortest paths on the manifold, for accelerated convergence. The method is then invoked for a Pareto problem to reveal the achievable channel singular value region, which generalizes most relevant metrics and provides an intuitive shaping benchmark.

Fourth, we tackle BD-RIS-aided MIMO rate maximization problem by a local-optimal Alternating Optimization (AO) approach and a low-complexity shaping approach. The former iteratively updates the passive beamforming via geodesic RCG and the active beamforming by eigenmode transmission, until convergence. The latter simply shapes the channel for maximum power gain then performs legacy transmission.

Fifth, we validate the analytical bounds and the numerical methods by extensive simulation. It is concluded that:

- BD-RIS widens the dynamic range of channel singular values for enhanced rate, power, and physical-layer security;
- The shaping benefits of BD-RIS over D-RIS scale with the number of elements, group size, and MIMO dimensions;
- BD-RIS can activate multi-stream transmission at much lower transmit power than D-RIS;
- The performance gap between the optimal joint beamforming and the low-complexity shaping solution diminishes as the RIS evolves from D to fully-connected BD;
- The proposed geodesic RCG method is efficient and the optimization cost of practically-sized BD-RIS is affordable;
- The solutions are robust to channel estimation errors and extendable to symmetric case with minimal degradation.

Notation: Italic, bold lower-case, and bold upper-case letters indicate scalars, vectors and matrices, respectively. $\jmath$ denotes the imaginary unit. $\mathbb{R}$ and $\mathbb{C}$ denote the set of real and complex numbers, respectively. $\mathbb{H}^{n \times n}$, $\mathbb{H}_+^{n \times n}$, $\mathbb{U}^{n \times n}$, and $\mathbb{P}^{n \times n}$ denote the set of $n \times n$ Hermitian, positive semi-definite, unitary, and permutation matrices, respectively. $\mathbf{0}$ and $\mathbf{I}$ are the zero and identity matrices with appropriate size, respectively. $\Re\{\cdot\}$ takes the real part of a complex number. $\mathbb{E}\{\cdot\}$ is the expectation operator. $\mathrm{conv}\{\cdot\}$ returns the convex hull of arguments. $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ evaluate the trace and determinant of a square matrix, respectively. $\mathrm{diagm}(\cdot)$ constructs a square matrix with

arguments on the main (block) diagonal and zeros elsewhere. $\mathrm{sv}(\cdot)$, $\mathrm{ran}(\cdot)$, and $\ker(\cdot)$ evaluate the singular values, range, and kernel of a matrix, respectively. $\mathrm{vec}(\cdot)$ stacks the columns of a matrix as a vector. $|\cdot|$, $\|\cdot\|$, and $\|\cdot\|_{\mathrm{F}}$ denote the absolute value, Euclidean norm, and Frobenius norm, respectively. $\sigma_n(\cdot)$ and $\lambda_n(\cdot)$ are the n-th largest singular value and eigenvalue, respectively. $(\cdot)^*$, $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$, $(\cdot)^{\dagger}$, $(\cdot)^{\star}$ denote the conjugate, transpose, conjugate transpose (Hermitian), Moore-Penrose inverse, and stationary point, respectively. $[N]$ is a shortcut for $\{1,2,...,N\}$. $(\cdot)_{[x:y]}$ is a shortcut for $(\cdot)_x,(\cdot)_{x+1},...,(\cdot)_y$. $\odot$ denotes the Hadamard product. $\mathcal{O}(\cdot)$ is the big-O notation. $\mathcal{N}_{\mathbb{C}}(\mathbf{0},\boldsymbol{\Sigma})$ is the multivariate Circularly Symmetric Complex Gaussian (CSCG) distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$. $\sim$ means "distributed as".

II. SYSTEM MODEL

We model a BD-RIS as an N_{S}-port network that divides into G individual groups, where group $g \in [G]$ contains N_g scattering elements ($N_{\mathrm{S}} = \sum_{g=1}^{G} N_g$) interconnected by real-time reconfigurable components [24]. For the ease of analysis, we assume no mutual coupling and equal group size $N_g = L \triangleq N_{\mathrm{S}}/G$, $\forall g$. The overall scattering matrix is block-diagonal

$$\boldsymbol{\Theta} = \mathrm{diagm}(\boldsymbol{\Theta}_1,...,\boldsymbol{\Theta}_G), \tag{1}$$

where $\boldsymbol{\Theta}_g \in \mathbb{U}^{L \times L}$ is the g-th unitary block modeling the response of group g. D-RIS can be seen an extreme case of (1) with group size $L=1$. Some viable architectures of BD-RIS are illustrated in [24, Fig. 3], [33, Fig. 5], [26, Fig. 2] where the circuit topology have been modeled in the scattering matrix.

Consider a BD-RIS-aided MIMO point-to-point channel with N_{T} and N_{R} transmit and receive antennas, respectively, and N_{S} scattering elements at the BD-RIS. This configuration is denoted as $N_{\mathrm{T}} \times N_{\mathrm{S}} \times N_{\mathrm{R}}$ throughout this paper. Let $\mathbf{H}_{\mathrm{D}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$, $\mathbf{H}_{\mathrm{B}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{S}}}$, $\mathbf{H}_{\mathrm{F}} \in \mathbb{C}^{N_{\mathrm{S}} \times N_{\mathrm{T}}}$ denote the direct (i.e., transmitter-receiver), backward (i.e., RIS-receiver), and forward (i.e., transmitter-RIS) channels, respectively. The equivalent channel is

$$\mathbf{H} = \mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\boldsymbol{\Theta}\mathbf{H}_{\mathrm{F}} = \mathbf{H}_{\mathrm{D}} + \sum_g \mathbf{H}_{\mathrm{B},g}\boldsymbol{\Theta}_g\mathbf{H}_{\mathrm{F},g}, \tag{2}$$

where $\mathbf{H}_{\mathrm{B},g} \in \mathbb{C}^{N_{\mathrm{R}} \times L}$ and $\mathbf{H}_{\mathrm{F},g} \in \mathbb{C}^{L \times N_{\mathrm{T}}}$ are the backward and forward channels associated with group g, corresponding to the $(g-1)L+1$ to gL columns of $\mathbf{H}_{\mathrm{B}}$ and rows of $\mathbf{H}_{\mathrm{F}}$, respectively. Since unitary matrices constitute an algebraic group with respect to multiplication, we can decompose the scattering matrix of group g as

$$\boldsymbol{\Theta}_g = \mathbf{L}_g\mathbf{X}_g\mathbf{R}_g^{\mathsf{H}}, \tag{3}$$

where $\mathbf{L}_g,\mathbf{R}_g \in \mathbb{U}^{L \times L}$ are unitary matrices and $\mathbf{X}_g \in \mathbb{P}^{n \times n}$ is a permutation matrix. Let $\mathbf{H}_g \triangleq \mathbf{H}_{\mathrm{B},g}\boldsymbol{\Theta}_g\mathbf{H}_{\mathrm{F},g}$ be the indirect channel via group g and $\mathbf{H}_{\mathrm{B/F},g} = \mathbf{U}_{\mathrm{B/F},g}\boldsymbol{\Sigma}_{\mathrm{B/F},g}\mathbf{V}_{\mathrm{B/F},g}^{\mathsf{H}}$ be the Singular Value Decomposition (SVD) of the backward and forward channels, respectively. The equivalent channel is

$$\mathbf{H} = \overbrace{\mathbf{H}_{\mathrm{D}}}^{\text{direct-indirect}} + \sum_g \mathbf{U}_{\mathrm{B},g}\boldsymbol{\Sigma}_{\mathrm{B},g}\underbrace{\mathbf{V}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{L}_g\mathbf{X}_g\mathbf{R}_g^{\mathsf{H}}\mathbf{U}_{\mathrm{F},g}}_{\text{backward-forward}}\boldsymbol{\Sigma}_{\mathrm{F},g}\mathbf{V}_{\mathrm{F},g}^{\mathsf{H}}. \tag{4}$$

Remark 1. In (4), the BD-RIS performs a blockwise unitary transformation to combine the backward-forward (intra-group,

multiplicative) channels and direct-indirect (inter-group, additive) channels. These two attributes are refined respectively as:

- *Branch matching:* To pair and combine the branches (i.e., entries) of $\mathbf{H}_{\mathrm{B},g}$ and $\mathbf{H}_{\mathrm{F},g}$ through $\boldsymbol{\Theta}_g$;
- *Mode alignment:* To align and order the modes (i.e., singular vectors) of $\{\mathbf{H}_g\}_{g \in [G]}$ with those of $\mathbf{H}_{\mathrm{D}}$ through $\boldsymbol{\Theta}$.

Example 1 (Single-Input Single-Output (SISO) channel gain maximization). Denote the direct, backward, forward channels as h_{D}, $\mathbf{h}_{\mathrm{B}} \in \mathbb{C}^{N_{\mathrm{S}} \times 1}$, and $\mathbf{h}_{\mathrm{F}}^{\mathsf{H}} \in \mathbb{C}^{1 \times N_{\mathrm{S}}}$, respectively. In this case, mode alignment boils down to phase matching and the optimal BD-RIS structure is

$$\boldsymbol{\Theta}_{\mathrm{P\text{-}max},g}^{\mathrm{SISO}} = \frac{h_{\mathrm{D}}}{|h_{\mathrm{D}}|}\mathbf{V}_{\mathrm{B},g}\mathbf{U}_{\mathrm{F},g}^{\mathsf{H}}, \quad \forall g, \tag{5}$$

where $\mathbf{V}_{\mathrm{B},g} = [\mathbf{h}_{\mathrm{B},g}/\|\mathbf{h}_{\mathrm{B},g}\|, \ \mathbf{N}_{\mathrm{B},g}] \in \mathbb{U}^{L \times L}$, $\mathbf{U}_{\mathrm{F},g} = [\mathbf{h}_{\mathrm{F},g}/\|\mathbf{h}_{\mathrm{F},g}\|,\mathbf{N}_{\mathrm{F},g}] \in \mathbb{U}^{L \times L}$, and $\mathbf{N}_{\mathrm{B/F},g} \in \mathbb{C}^{L \times (L-1)}$ are the orthonormal bases of kernels of $\mathbf{h}_{\mathrm{B/F},g}$. Evidently, any group size L (including D-RIS $L=1$ with empty kernels) suffices for perfect phase matching. The maximum channel gain still depends on L

$$|h| = |h_{\mathrm{D}}| + \sum_g \sum_l |h_{\mathrm{B},g,\pi_{\mathrm{B},g}(l)}||h_{\mathrm{F},g,\pi_{\mathrm{F},g}(l)}|, \tag{6}$$

where $h_{\mathrm{B/F},g,l}$ are the l-th entries of $\mathbf{h}_{\mathrm{B/F},g}$, and $\pi_{\mathrm{B/F},g}$ are permutations of $[L]$ sorting their magnitude in similar orders. That is, the maximum SISO channel gain is attained when each BD-RIS group, apart from phase shifting, matches the l-th strongest backward and forward channel branches. Increasing L improves the branch matching flexibility and boosts the channel gain.

Example 1 clarifies the difference between branch matching and mode alignment as well as their impacts on channel shaping. When it comes to MIMO, the advantage of BD-RIS in branch matching improves since the number of available branches is proportional to N_{T} and N_{R}. On the other hand, the limitation of D-RIS in mode alignment intensifies since each element can only apply a scalar phase shift to the indirect channel of $\min(N_{\mathrm{T}},N_{\mathrm{S}},N_{\mathrm{R}})$ modes.

III. CHANNEL SHAPING

In this section, we first provide an example demonstrating the MIMO channel shaping advantages of BD-RIS over D-RIS, then derive some analytical bounds on singular values, power gain, and capacity under specific channel conditions. Finally, we propose a numerical method to optimize the BD-RIS for a broad class of singular value functions.

Example 2 ($2 \times 2 \times 2$ shaping). Here D-RIS and fully-connected BD-RIS can be modeled by 2 and 4 independent angular parameters, respectively:

$$\boldsymbol{\Theta}_{\mathrm{D}} = \mathrm{diagm}(e^{j\theta_1},e^{j\theta_2}), \quad \boldsymbol{\Theta}_{\mathrm{BD}} = e^{j\phi}\begin{bmatrix} e^{j\alpha}\cos\psi & e^{j\beta}\sin\psi \\ -e^{-j\beta}\sin\psi & e^{-j\alpha}\cos\psi \end{bmatrix},$$

Consider a special case where the BD-RIS is symmetric (i.e., $\beta = \pi/2$) and the direct channel is negligible such that ϕ has no impact on $\mathrm{sv}(\mathbf{H})$, since $\mathrm{sv}(e^{j\phi}\mathbf{A}) = \mathrm{sv}(\mathbf{A})$. The singular value shaping capabilities of $\boldsymbol{\Theta}_{\mathrm{D}}$ and $\boldsymbol{\Theta}_{\mathrm{BD}}$ can thus be visualized over 2 tunable parameters. With an exhaustive grid search over (θ_1,θ_2) and (α,ψ), Fig. 1
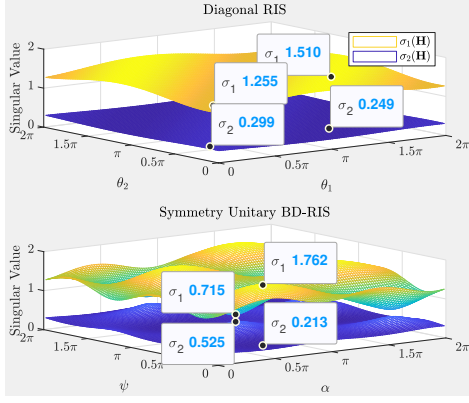
Fig. 1. $2 \times 2 \times 2$ singular value shaping by D-RIS and symmetric fully-connected BD-RIS when the direct channel is negligible. The maximum and minimum of both singular values are marked explicitly on the plot.

shows the achievable singular values of a specific channel instance $\mathbf{H}_\mathrm{B} = \left[\begin{smallmatrix} -0.2059+0.5914\jmath & -0.0909+0.5861\jmath \\ 0.4131+0.2651\jmath & -0.1960+0.4650\jmath \end{smallmatrix}\right]$, $\mathbf{H}_\mathrm{F} = \left[\begin{smallmatrix} -0.6362+0.1332\jmath & -0.1572+1.5538\jmath \\ 0.0196+0.4011\jmath & -0.3170-0.2303\jmath \end{smallmatrix}\right]$. In this example, both singular values can be manipulated up to[1] $\pm 9\%$ by D-RIS (using 2 reconfigurable components) and $\pm 42\%$ by symmetric fully-connected BD-RIS (using 3 reconfigurable components).

Example 2 suggests that the physical interconnection of RIS elements, even if using symmetric circuit components, can create a "cooperation effect" that significantly enhances the dynamic range of channel singular values. This motivates the analytical and numerical shaping studies in Sections III-A and III-B.

A. Analytical Shaping Bounds

Definition 1 (Communication DoF). DoF refers to the maximum number of streams that can be transmitted in parallel over a MIMO channel in the asymptotic high-Signal-to-Noise Ratio (SNR) regime

$$\mathrm{DoF}(\mathbf{H}) = \lim_{\rho \to \infty} \frac{\log\det(\mathbf{I} + \rho\mathbf{H}\mathbf{H}^\mathsf{H})}{\log\rho}, \tag{7}$$

where ρ is the SNR.

Definition 2 (Negligible direct channel). A direct channel is considered negligible when its contribution to the received signal is substantially weaker than that of the RIS-induced indirect channels. Mathematically, this can be defined as

$$\frac{\|\mathbf{H}_\mathrm{D}\|_\mathrm{F}^2}{\|\sum_g \mathbf{H}_{\mathrm{B},g}\mathbf{\Theta}_g\mathbf{H}_{\mathrm{F},g}\|_\mathrm{F}^2} < \epsilon,$$

where ϵ is a small positive threshold. This can result from a very large number of RIS elements or physical obstacles in the propagation path.

The main results of this subsection are summarized in the following Propositions and Corollaries.

Proposition 1 (DoF). *BD-RIS may achieve a larger or smaller MIMO DoF than D-RIS.*

Proof. Please refer to Appendix A. □

[1]The percentage for manipulating $\sigma_n(\mathbf{H})$ is calculated by $\eta_n^+ = \frac{\max\sigma_n(\mathbf{H}) - \mathrm{avg}\sigma_n(\mathbf{H})}{\mathrm{avg}\sigma_n(\mathbf{H})} \times 100\%$ and $\eta_n^- = \frac{\min\sigma_n(\mathbf{H}) - \mathrm{avg}\sigma_n(\mathbf{H})}{\mathrm{avg}\sigma_n(\mathbf{H})} \times 100\%$.

Example 3 (DoF of $4 \times 4 \times 4$ shaping). Consider a $4 \times 4 \times 4$ shaping with $\mathbf{H}_\mathrm{D} = \mathbf{0}$, $\mathbf{H}_\mathrm{B} = \left[\begin{smallmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{smallmatrix}\right]$, and $\mathbf{H}_\mathrm{F} = \mathrm{diagm}(1, 1, 0, 0)$. Evidently, any D-RIS $\mathbf{\Theta}_\mathrm{D} = \mathrm{diagm}(e^{\jmath\theta_1}, e^{\jmath\theta_2}, e^{\jmath\theta_3}, e^{\jmath\theta_4})$ results in

$$\mathbf{H} = \left[\begin{smallmatrix} e^{\jmath\theta_1} & e^{\jmath\theta_2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{smallmatrix}\right]$$

with 1 DoF. On the other hand, a fully-connected BD-RIS can perfectly align or misalign the kernels of $\mathbf{H}_\mathrm{B}$ and $\mathbf{H}_\mathrm{F}$ using the closed-form solutions (47) or (48) in Appendix A. That is,

$$\mathbf{\Theta}_\mathrm{DoF\text{-}max}^\mathrm{MIMO} = \left[\begin{smallmatrix} 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{smallmatrix}\right] \text{ and } \mathbf{\Theta}_\mathrm{DoF\text{-}min}^\mathrm{MIMO} = \left[\begin{smallmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \end{smallmatrix}\right],$$

which correspond to

$$\mathbf{H} = \left[\begin{smallmatrix} 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{smallmatrix}\right], \quad \mathbf{H} = \mathbf{0},$$

and a DoF of 2 and 0, respectively.

Proposition 1 and Example 3 suggest that we can expect more parallel data streams or less crosstalk when shaping the channel with BD-RIS. Increasing the DoF can improve the asymptotic rate performance for point-to-point transmission. Conversely, reducing the DoF can help orthogonalize channels in multi-user networks for the interest of interference alignment and physical layer security. Next, we progress to quantify the limits of singular value redistribution in rank-deficient channels.

Proposition 2 (Rank-deficient channel). *If the minimum rank of backward and forward channels is k ($k \le N \triangleq \min(N_\mathrm{T}, N_\mathrm{R})$), then for D-RIS or BD-RIS of arbitrary number of elements, the n-th singular value of the equivalent channel is bounded above and below respectively by*

$$\sigma_n(\mathbf{H}) \le \sigma_{n-k}(\mathbf{T}), \qquad \text{if } n > k, \tag{8a}$$
$$\sigma_n(\mathbf{H}) \ge \sigma_n(\mathbf{T}), \qquad \text{if } n < N-k+1, \tag{8b}$$

where $\mathbf{T}$ is any auxiliary matrix satisfying

$$\mathbf{T}\mathbf{T}^\mathsf{H} = \begin{cases} \mathbf{H}_\mathrm{D}(\mathbf{I} - \mathbf{V}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H})\mathbf{H}_\mathrm{D}^\mathsf{H}, & \text{if } \mathrm{rank}(\mathbf{H}_\mathrm{F}) = k, \\ \mathbf{H}_\mathrm{D}^\mathsf{H}(\mathbf{I} - \mathbf{U}_\mathrm{B}\mathbf{U}_\mathrm{B}^\mathsf{H})\mathbf{H}_\mathrm{D}, & \text{if } \mathrm{rank}(\mathbf{H}_\mathrm{B}) = k, \end{cases} \tag{9}$$

and $\mathbf{V}_\mathrm{F}$ and $\mathbf{U}_\mathrm{B}$ are any right and left singular matrices of $\mathbf{H}_\mathrm{F}$ and $\mathbf{H}_\mathrm{B}$, respectively.

Proof. Please refer to Appendix B. □

Inequality (8a) states that if $\mathbf{H}_\mathrm{B}$ and $\mathbf{H}_\mathrm{F}$ are at least rank k, then using a D-RIS or BD-RIS of *sufficiently large* N_S, the n-th singular value of $\mathbf{H}$ can be enlarged to the $(n-k)$-th singular value of $\mathbf{T}$, or suppressed to the n-th singular value of $\mathbf{T}$. Moreover, the first k channel singular values are unbounded above[2] while the last k channel singular values can be suppressed to zero. A special case of Line-of-Sight (LoS) channel is presented below[3].

Corollary 2.1 (LoS channel). *If at least one of backward and forward channels is LoS, then a D-RIS or BD-RIS can at most enlarge the n-th ($n \ge 2$) channel singular value to the $(n-1)$-th singular value of $\mathbf{T}$, or suppress the n-th channel*

[2]The energy conservation law $\sum_{n=1}^N \sigma_n^2(\mathbf{H}) \le 1$ still has to be respected in all cases.

[3]A similar eigenvalue result has been derived for D-RIS only [38].

singular value to the n-th singular value of $\mathbf{T}$. That is,

$$\sigma_1(\mathbf{H}) \geq \sigma_1(\mathbf{T}) \geq \sigma_2(\mathbf{H}) \geq ... \geq \sigma_{N-1}(\mathbf{T}) \geq \sigma_N(\mathbf{H}) \geq \sigma_N(\mathbf{T}). \tag{10}$$

Proof. This is a direct result of (8) with $k=1$. $\qquad\square$

We emphasize that Proposition 2 and Corollary 2.1 apply to both D- and BD-RIS configurations regardless of the status of the direct channel. Out of $2N$ bounds in (8) or (10), N of them can be *simultaneously* tight as $N_S \to \infty$, namely when the direct channel becomes negligible. For a finite N_S, the RIS may prioritize a subset of those by aligning the corresponding modes. We will show by simulation that BD-RIS outperforms D-RIS on this purpose. Proposition 2 complements the DoF result in Proposition 1 by quantifying the dynamic range of extreme singular values in low-multipath scenarios. They reveal a diminishing return of increasing the number of BD-RIS elements and group size in enhancing channel shaping capability. Therefore, the bounds can be used to guide practical RIS configurations, especially in millimeter-wave and terahertz systems under sparse propagation environment, for a balanced performance-complexity tradeoff. Next, we progress to quantify the limits of singular value redistribution when the direct channel is negligible.

Proposition 3 (Negligible direct channel). *If the direct channel is negligible, then a fully-connected BD-RIS of arbitrary number of elements can manipulate the channel singular values up to*

$$\mathrm{sv}(\mathbf{H}) = \mathrm{sv}(\mathbf{BF}), \tag{11}$$

where $\mathbf{B}$ and $\mathbf{F}$ are any matrices satisfying $\mathrm{sv}(\mathbf{B}) = \mathrm{sv}(\mathbf{H}_B)$ and $\mathrm{sv}(\mathbf{F}) = \mathrm{sv}(\mathbf{H}_F)$.

Proof. Please refer to Appendix C. $\qquad\square$

Proposition 3 says that if the direct channel is negligible and the BD-RIS is fully-connected, the only singular value bounds on the equivalent channel are those on the product of unitary-transformed backward and forward channels. It is *not necessarily* an asymptotic result and does *not* depend on any relationship between N_T, N_S, and N_R. Its importance lies in that our channel shaping question can be recast as a well-studied linear algebra question: *How the singular values of matrix product are bounded by the singular values of its individual factors?* The question is partially answered in Corollaries 3.1 – 3.3 over definitions $\bar{N} = \max(N_T, N_S, N_R)$ and $\sigma_n(\mathbf{H}) = \sigma_n(\mathbf{H}_F) = \sigma_n(\mathbf{H}_B) = 0$, $\forall n \in [\bar{N}] \setminus [N]$. This is equivalent to padding zero blocks at the end of $\mathbf{H}, \mathbf{H}_B, \mathbf{H}_F$ to make square matrices of dimension $\bar{N}$. The results are by no means complete and interested readers are referred to [39, Chapter 16, 24] and [40, Chapter 3] for more information.

Corollary 3.1 (Product of subset of singular values). *If the direct channel is negligible, then the product of subset of singular values of $\mathbf{H}$ is bounded from above by those of $\mathbf{H}_B$ and $\mathbf{H}_F$, that is,*

$$\prod_{k \in K} \sigma_k(\mathbf{H}) \leq \prod_{i \in I} \sigma_i(\mathbf{H}_B) \prod_{j \in J} \sigma_j(\mathbf{H}_F), \tag{12}$$

for all admissible triples $(I,J,K) \in T_r^{\bar{N}}$ with $r < \bar{N}$, where

$$T_r^{\bar{N}} \triangleq \Big\{ (I,J,K) \in U_r^{\bar{N}} \, \big| \, \forall p < r, \, \forall (F,G,H) \in T_p^r,$$

$$\sum_{f \in F} i_f + \sum_{g \in G} j_g \leq \sum_{h \in H} k_h + \frac{p(p+1)}{2} \Big\},$$

$$U_r^{\bar{N}} \triangleq \Big\{ (I,J,K) \subseteq [\bar{N}]^3 \, \big| \, \sum_{i \in I} i + \sum_{j \in J} j = \sum_{k \in K} k + \frac{r(r+1)}{2} \Big\}.$$

Proof. Please refer to [41, Theorem 8]. $\qquad\square$

Corollary 3.1 applies to arbitrary number of RIS elements as inherited from Proposition 3. The set (12), also recognized as a variation of Horn's inequality [42], provides a comprehensive analytical answer to the shaping question – it can be interpreted as the *outer bounds* of the achievable singular value region of the BD-RIS-aided MIMO channel. An example is given by (45) and their visualization in Fig. 2. Remarkably, the number of inequalities in (12) increases exponentially with N_S[4]. At a first glance the results may seem excessive to be useful; but they are given in this form to be general and one can pick any *subset* of them for specific applications. Below we showcase how to induce some ready-to-use wireless performance bounds with closed-form BD-RIS solutions from Corollary 3.1. The applications mentioned therein are non-exhaustive; we hope readers can discover more results specific to their research.

Corollary 3.2 (Product of some largest or smallest singular values). *If the direct channel is negligible, then the product of the first (resp. last) k singular values of $\mathbf{H}$ is bounded from above (resp. below) by those of $\mathbf{H}_B$ and $\mathbf{H}_F$, that is,*

$$\prod_{n=1}^{k} \sigma_n(\mathbf{H}) \leq \prod_{n=1}^{k} \sigma_n(\mathbf{H}_B) \sigma_n(\mathbf{H}_F), \tag{13a}$$

$$\prod_{n=\bar{N}}^{\bar{N}-k+1} \sigma_n(\mathbf{H}) \geq \prod_{n=\bar{N}}^{\bar{N}-k+1} \sigma_n(\mathbf{H}_B) \sigma_n(\mathbf{H}_F). \tag{13b}$$

Proof. Please refer to Appendix D. $\qquad\square$

Corollary 3.2 reveals the shaping limits on the product of some extreme channel singular values. The lower bounds (13b) coincide at zero when $\bar{N} \neq N$ (i.e., $N_T = N_S = N_R$ being false). These bounds can be applied, for instance, as a shortcut to establish the capacity of BD-RIS-aided MIMO channels at extreme SNR, as shown in Corollary 3.6. In the special case $k=1$, we arrive at the upper bound on the largest channel singular value $\sigma_1(\mathbf{H}) \leq \sigma_1(\mathbf{H}_B) \sigma_1(\mathbf{H}_F)$. This is particularly useful for MIMO wireless power transfer with Radio Frequency (RF) combining where the harvested power depends merely on, and is a quartic function of, the largest channel singular value [43]. A closed-form BD-RIS solution to attain this upper bound can be found below in (15a).

Corollary 3.3 (Individual singular value). *If the direct channel is negligible, then the n-th channel singular value can be manipulated within the range of*

$$\max_{i+j=n+N_S} \sigma_i(\mathbf{H}_B) \sigma_j(\mathbf{H}_F) \leq \sigma_n(\mathbf{H}) \leq \min_{i+j=n+1} \sigma_i(\mathbf{H}_B) \sigma_j(\mathbf{H}_F), \tag{14}$$

---

[4]For example, the number of inequalities described by (12) grows from 12 to 2062 when N_S increases from 3 to 7.

where $(i,j) \in [N_S]^2$. *The upper and lower bounds are attained respectively at*

$$\Theta_{\text{sv-}n\text{-max}}^{\text{MIMO-ND}} = \mathbf{V}_B \mathbf{P} \mathbf{U}_F^H, \tag{15a}$$

$$\Theta_{\text{sv-}n\text{-min}}^{\text{MIMO-ND}} = \mathbf{V}_B \mathbf{Q} \mathbf{U}_F^H, \tag{15b}$$

where $\mathbf{V}_B, \mathbf{U}_F \in \mathbb{U}^{N_S \times N_S}$ *are any right and left singular matrices of* $\mathbf{H}_B$ *and* $\mathbf{H}_F$, *respectively, and* $\mathbf{P}, \mathbf{Q} \in \mathbb{P}^{n \times n}$ *are any permutation matrices of dimension* N_S *satisfying:*

- *The* (i,j)-*th entry is 1, where*

$$(i,j) = \begin{cases} \underset{i+j=n+1}{\arg\min}\ \sigma_i(\mathbf{H}_B)\sigma_j(\mathbf{H}_F) & \textit{for } \mathbf{P}, \quad (16a) \\ \underset{i+j=n+N_S}{\arg\max}\ \sigma_i(\mathbf{H}_B)\sigma_j(\mathbf{H}_F) & \textit{for } \mathbf{Q}, \quad (16b) \end{cases}$$

  and ties may be broken arbitrarily;

- *After deleting the* i-*th row and* j-*th column, the resulting submatrix* $\mathbf{Y} \in \mathbb{P}^{(N_S-1) \times (N_S-1)}$ *is any permutation matrix satisfying*

$$\sigma_{n-1}(\hat{\mathbf{\Sigma}}_B \mathbf{Y} \hat{\mathbf{\Sigma}}_F) \geq \min_{i+j=n+1} \sigma_i(\mathbf{H}_B)\sigma_j(\mathbf{H}_F) \textit{ for } \mathbf{P}, \tag{17a}$$

$$\sigma_{n+1}(\hat{\mathbf{\Sigma}}_B \mathbf{Y} \hat{\mathbf{\Sigma}}_F) \leq \max_{i+j=n+N_S} \sigma_i(\mathbf{H}_B)\sigma_j(\mathbf{H}_F) \textit{ for } \mathbf{Q}, \tag{17b}$$

  where $\hat{\mathbf{\Sigma}}_B$ *and* $\hat{\mathbf{\Sigma}}_F$ *are diagonal singular value matrices of* $\mathbf{H}_B$ *and* $\mathbf{H}_F$ *with both* i-*th row and* j-*th column deleted, respectively.*

Proof. Please refer to Appendix E. $\qquad\square$

Remark 2. We emphasize that the singular matrices in the SVD are not uniquely defined. When a singular value has multiplicity k, the corresponding singular vectors can be any orthonormal basis of the k-dimensional subspace. Even if all singular values are distinct, the singular vectors of each can be scaled by a phase factor of choice. Consequently, all SVD-based scattering matrices in this paper are inherently non-unique.

Corollary 3.3 and Proposition 2 both reveal the shaping limits of the n-th largest channel singular value. The two results are derived under different assumptions are not special cases of each other. Importantly, Corollary 3.3 establishes upper and lower bounds for *each* channel singular value (c.f. first and last k in Proposition 2) and provides general solutions for fully-connected BD-RIS of arbitrary (c.f. sufficiently large) size to attain the equalities. These bounds enable closed-form passive beamforming, and hence fixed channel and closed-form active beamforming, for spatial multiplexing with a limited number n of RF chains. We emphasize that in (15) the mode alignment is realized by $\mathbf{V}_B$ and $\mathbf{U}_F$ while the ordering is enabled by permutation matrices $\mathbf{P}$ and $\mathbf{Q}$, which are special cases of $\mathbf{X}$ defined in (3). Specially, the extreme channel singular values can be manipulated within the range of

$$\max_{i+j=N_S+1} \sigma_i(\mathbf{H}_B)\sigma_j(\mathbf{H}_F) \leq \sigma_1(\mathbf{H}) \leq \sigma_1(\mathbf{H}_B)\sigma_1(\mathbf{H}_F), \tag{18a}$$

$$\min_{i+j=\bar{N}+1} \sigma_i(\mathbf{H}_B)\sigma_j(\mathbf{H}_F) \geq \sigma_{\bar{N}}(\mathbf{H}) \geq \sigma_{\bar{N}}(\mathbf{H}_B)\sigma_{\bar{N}}(\mathbf{H}_F). \tag{18b}$$

We notice that the right halves in (18a) and (18b) are also special cases of (13a) and (13b) with $k=1$.

Example 4 (Bounds on $3 \times 3 \times 3$ shaping). Consider a $3 \times 3 \times 3$ setup with $\mathbf{H}_D = \mathbf{0}$, $\mathbf{H}_B = \text{diagm}(3,2,1)$, and $\mathbf{H}_F = \text{diagm}(4,0,5)$.

- D-RIS: It is evident that any D-RIS can only achieve $\text{sv}(\mathbf{H}) = [12,5,0]^\mathsf{T}$ due to limited branch matching and mode alignment capabilities;
- BD-RIS: According to (14), a fully-connected BD-RIS can manipulate the singular values within the range of

$$8 \leq \sigma_1(\mathbf{H}) \leq 15, \quad 4 \leq \sigma_2(\mathbf{H}) \leq 10, \quad 0 \leq \sigma_3(\mathbf{H}) \leq 0.$$

  To attain the upper and lower bounds, (i,j) in (15a) and (15b) takes $(1,1)$ and $(2,2)$ when $n=1$, and $(2,1)$ and $(3,2)$ when $n=2$, respectively.

We conclude from Example 4 that a fully-connected BD-RIS can widen the dynamic range of channel singular values by properly aligning and ordering the modes of $\mathbf{H}_B$ and $\mathbf{H}_F$. However, the individual bounds (14) may not be simultaneously tight when the problem of interest is a function of multiple singular values. Some case studies are presented below.

Corollary 3.4 (Channel power gain). *If the direct channel is negligible, then the channel power gain is bounded from above (resp. below) by the inner product of squared singular values of* $\mathbf{H}_B$ *and* $\mathbf{H}_F$ *when they are sorted similarly (resp. oppositely), that is,*

$$\sum_{n=1}^{N} \sigma_n^2(\mathbf{H}_B)\sigma_{N_S-n+1}^2(\mathbf{H}_F) \leq \|\mathbf{H}\|_F^2 \leq \sum_{n=1}^{N} \sigma_n^2(\mathbf{H}_B)\sigma_n^2(\mathbf{H}_F), \tag{19}$$

whose upper and lower bounds are attained respectively at

$$\Theta_{\text{P-max}}^{\text{MIMO-ND}} = \mathbf{V}_B \mathbf{U}_F^H, \tag{20a}$$

$$\Theta_{\text{P-min}}^{\text{MIMO-ND}} = \mathbf{V}_B \mathbf{J} \mathbf{U}_F^H, \tag{20b}$$

where $\mathbf{J}$ *is the exchange (a.k.a. backward identity) matrix of dimension* N_S.

Proof. Please refer to Appendix F. $\qquad\square$

We notice that (20a) and (20b) are special cases of (15a) and (15b) with $\mathbf{P} = \mathbf{I}$ and $\mathbf{Q} = \mathbf{J}$, which also attain the right and left halves of (18), respectively. That is to say, there exists a closed-form BD-RIS solution (20a) maximizing the channel power gain that is also optimal for wireless power transfer. We will shortly see that this solution also achieves the channel capacity. The upper bound (20a) is also reminiscent of the optimal AF relay beamforming design [44, (16), (17)] where the diagonal power allocation matrices boil down to $\mathbf{I}$ due to the passive nature of RIS. As a side note, when both $\mathbf{H}_B$ and $\mathbf{H}_F$ follow Rayleigh fading, the expectation of maximum channel power gain can be numerically evaluated as

$$\mathbb{E}\{\|\mathbf{H}\|_F^2\} = \sum_{n=1}^{N} \iint_0^\infty xy f_{\lambda_n^{\min(N_R,N_S)}}(x) f_{\lambda_n^{\min(N_S,N_T)}}(y) dx dy, \tag{21}$$

where λ_n^K is the n-th eigenvalue of the complex $K \times K$ Wishart matrix with probability density function $f_{\lambda_n^K}(\cdot)$ given by [45, (51)]. (21) generalizes the SISO channel power gain aided by BD-RIS [24, (58)] to MIMO but a closed-form expression is non-trivial. The next corollary has been derived in [37] independently of Proposition 3; we include it here for the completeness of results.

Corollary 3.5 (Channel capacity at general SNR). *If the direct channel is negligible, then the BD-RIS-aided MIMO*

channel capacity is

$$C^{\text{MIMO-ND}} = \sum_{n=1}^{N} \log\left(1 + \frac{s_n \sigma_n^2(\mathbf{H}_{\text{B}})\sigma_n^2(\mathbf{H}_{\text{F}})}{\eta}\right), \quad (22)$$

where η is the average noise power, $s_n = \mu - \frac{\eta}{\sigma_n^2(\mathbf{H}_{\text{B}})\sigma_n^2(\mathbf{H}_{\text{F}})}$ is the power allocated to the n-th mode obtainable by the water-filling algorithm [46]. The capacity-achieving BD-RIS scattering matrix is

$$\mathbf{\Theta}_{\text{R-max}}^{\text{MIMO-ND}} = \mathbf{V}_{\text{B}}\mathbf{U}_{\text{F}}^{\mathsf{H}}. \quad (23)$$

Proof. Please refer to [37, Appendix A]. □

One can observe from (20a) and (23) that the optimal channel shaping solution for channel power gain maximization, wireless power transfer, and wireless communication coincide with each other when the direct channel is negligible and the BD-RIS is fully-connected. If either condition is false, the active and passive beamforming would be coupled and the rate-optimal solution involves numerical optimization. In such case, the power gain-optimal RIS can still provide a low-complexity decoupled solution and the details will be discussed in Section IV.

Corollary 3.6 (Channel capacity at extreme SNR). *If the direct channel is negligible, then the channel capacity at extremely low and high SNR ρ are approximately bounded from above by*

$$C_{\rho\downarrow} \lesssim \rho\sigma_1^2(\mathbf{H}_{\text{B}})\sigma_1^2(\mathbf{H}_{\text{F}}), \quad (24a)$$

$$C_{\rho\uparrow} \lesssim N\log\frac{\rho}{N} + 2\log\prod_{n=1}^{N}\sigma_n(\mathbf{H}_{\text{B}})\sigma_n(\mathbf{H}_{\text{F}}). \quad (24b)$$

Proof. Please refer to Appendix G. □

The ergodic capacity (22) and (24) when both $\mathbf{H}_{\text{B}}$ and $\mathbf{H}_{\text{F}}$ follow Rayleigh fading can be evaluated similarly to (21) and the details are omitted here.

Proposition 1 – 3 and the resulting Corollaries provide a partial answer to the channel shaping question. These analyses provide a theoretical foundation for understanding exactly *to what extent* and *how* a BD-RIS can shape the MIMO channel regarding typical singular value functions under specific channel conditions. Extending the results to more general setups is challenging due to the limited branch matching capability and the inherent tradeoff in mode alignment.

B. Numerical Shaping Solution

Below we propose a numerical method to optimize BD-RIS for a broad class of singular value functions.

Definition 3 (Locally Lipschitz). *A function $f : \mathbb{R}^N \to \mathbb{R}$ is locally Lipschitz if for any compact set $\mathcal{S} \subset \mathbb{R}^N$, there exists a constant $L > 0$ such that $|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}$.*

Proposition 4. *Consider channel shaping problems of the form*

$$\max_{\mathbf{\Theta}} \quad f(\text{sv}(\mathbf{H})) \quad (25a)$$

$$\text{s.t.} \quad \mathbf{\Theta}_g^{\mathsf{H}}\mathbf{\Theta}_g = \mathbf{I}, \quad \forall g, \quad (25b)$$

where $f : \mathbb{R}^N \to \mathbb{R}$ is arbitrary locally Lipschitz function of channel singular values. The Clarke subdifferential of (25a) with respect to BD-RIS block g is

$$\partial_{\mathbf{\Theta}_g^*} f(\text{sv}(\mathbf{H})) = \text{conv}\{\mathbf{H}_{\text{B},g}^{\mathsf{H}}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}\mathbf{H}_{\text{F},g}^{\mathsf{H}}\}, \quad (26)$$

where $\mathbf{D} \in \mathbb{C}^{N_{\text{R}} \times N_{\text{T}}}$ is a rectangular diagonal matrix with $[\mathbf{D}]_{n,n} \in \partial_{\sigma_n(\mathbf{H})} f(\text{sv}(\mathbf{H})), \forall n \in [N]$, and $\mathbf{U}$, $\mathbf{V}$ are any left and right singular matrices of $\mathbf{H}$.

Proof. Please refer to Appendix H. □

Proposition 4 enables subgradient-based optimization for any locally Lipschitz function of channel singular values (e.g., Pareto frontier, power gain, capacity, condition number) via Clarke subdifferential (26). Next, we introduce a *geodesic*[5] *RCG* method modified from [47], [48] for the optimization of BD-RIS. Our contribution is an extension to the block-unitary case with sequential, parallel, or unified updates for accelerated convergence. The steps for updating $\mathbf{\Theta}_g$ at iteration r are summarized below, where the gradients are replaced by Clarke subgradients for non-smooth f.

(i) *Compute the Euclidean gradient at $\mathbf{\Theta}_g^{(r)}$:* The gradient of f with respect to $\mathbf{\Theta}_g$ in the Euclidean space is

$$\nabla_{\text{E},g}^{(r)} = 2\frac{\partial f(\mathbf{\Theta}_g^{(r)})}{\partial \mathbf{\Theta}_g^*}; \quad (27)$$

(ii) *Translate to the Riemannian gradient at $\mathbf{\Theta}_g^{(r)}$:* At point $\mathbf{\Theta}_g^{(r)}$, the Riemannian gradient gives the steepest ascent direction on the manifold. It lies in the tangent space of the manifold $\mathcal{T}_{\mathbf{\Theta}_g^{(r)}}\mathbb{U}^{L\times L} \triangleq \{\mathbf{M} \in \mathbb{C}^{L\times L} \mid \mathbf{M}^{\mathsf{H}}\mathbf{\Theta}_g^{(r)} + \mathbf{\Theta}_g^{(r)\mathsf{H}}\mathbf{M} = \mathbf{0}\}$ and is obtainable by projection:

$$\nabla_{\text{R},g}^{(r)} = \nabla_{\text{E},g}^{(r)} - \mathbf{\Theta}_g^{(r)}\nabla_{\text{E},g}^{(r)\mathsf{H}}\mathbf{\Theta}_g^{(r)}; \quad (28)$$

(iii) *Translate to the Riemannian gradient at the identity:* The Riemannian gradient should be translated back to the identity for exploiting the Lie algebra[6]:

$$\tilde{\nabla}_{\text{R},g}^{(r)} = \nabla_{\text{R},g}^{(r)}\mathbf{\Theta}_g^{(r)\mathsf{H}} = \nabla_{\text{E},g}^{(r)}\mathbf{\Theta}_g^{(r)\mathsf{H}} - \mathbf{\Theta}_g^{(r)}\nabla_{\text{E},g}^{(r)\mathsf{H}}. \quad (29)$$

(iv) *Determine the conjugate direction:* The conjugate direction is obtained over the Riemannian gradient and the previous direction as

$$\mathbf{D}_g^{(r)} = \tilde{\nabla}_{\text{R},g}^{(r)} + \gamma_g^{(r)}\mathbf{D}_g^{(r-1)}, \quad (30)$$

where $\gamma_g^{(r)}$ deviates the conjugate direction from the tangent space for accelerated convergence. A popular choice is the Polak-Ribière formula [49]

$$\gamma_g^{(r)} = \frac{\text{tr}\left((\tilde{\nabla}_{\text{R},g}^{(r)} - \tilde{\nabla}_{\text{R},g}^{(r-1)})\tilde{\nabla}_{\text{R},g}^{(r)\mathsf{H}}\right)}{\text{tr}\left(\tilde{\nabla}_{\text{R},g}^{(r-1)}\tilde{\nabla}_{\text{R},g}^{(r-1)\mathsf{H}}\right)}. \quad (31)$$

(v) *Evaluate the geodesic at the identity:* The geodesic emanating from the identity with velocity $\mathbf{D} \in \mathfrak{u}(L)$ is described by

$$\mathbf{G}_{\mathbf{I}}(\mu) = \exp(\mu\mathbf{D}), \quad (32)$$

where $\exp(\mathbf{A}) = \sum_{k=0}^{\infty}(\mathbf{A}^k/k!)$ is the matrix exponential and μ is the step size (i.e., magnitude of the tangent vector).

(vi) *Translate to the geodesic at $\mathbf{\Theta}_g^{(r)}$:* The geodesic emanating from $\mathbf{\Theta}_g^{(r)}$ terminates at $\mathbf{\Theta}_g^{(r+1)}$ by multiplicative updates

$$\mathbf{\Theta}_g^{(r+1)} = \mathbf{G}_{\mathbf{\Theta}_g^{(r)}}(\mu) = \mathbf{G}_{\mathbf{I}}(\mu)\mathbf{\Theta}_g^{(r)} = \exp(\mu\mathbf{D}_g^{(r)})\mathbf{\Theta}_g^{(r)}, \quad (33)$$

---

[5]A geodesic is a curve representing the shortest path between two points in a Riemannian manifold, whose tangent vectors remain parallel when transporting along the curve.

[6]Lie algebra refers to the tangent space of the Lie group at the identity element. A Lie group is simultaneously a continuous group and a differentiable manifold. In this example, $\mathbb{U}^{L\times L}$ formulates a Lie group and the corresponding Lie algebra consists of skew-Hermitian matrices $\mathfrak{u}(L) \triangleq \mathcal{T}_{\mathbf{I}}\mathbb{U}^{L\times L} = \{\mathbf{M} \in \mathbb{C}^{L\times L} \mid \mathbf{M}^{\mathsf{H}} + \mathbf{M} = \mathbf{0}\}$.

Algorithm 1 Geodesic RCG for BD-RIS design

Input: $f(\boldsymbol{\Theta}), G$
Output: $\boldsymbol{\Theta}^\star$
1: **Initialize** $r \leftarrow 0$, $\boldsymbol{\Theta}^{(0)}$
2: **Repeat**
3:    **For** $g \leftarrow 1$ to G
4:       $\nabla_{\mathrm{E},g}^{(r)} \leftarrow (27)$, $\tilde{\nabla}_{\mathrm{R},g}^{(r)} \leftarrow (29)$, $\gamma_g^{(r)} \leftarrow (31)$, $\mathbf{D}_g^{(r)} \leftarrow (30)$
5:       **If** $\Re\{\mathrm{tr}(\mathbf{D}_g^{(r)\mathsf{H}}\tilde{\nabla}_{\mathrm{R},g}^{(r)})\} < 0$ ▷ Not ascent
6:          $\mathbf{D}_g^{(r)} \leftarrow \tilde{\nabla}_{\mathrm{R},g}^{(r)}$
7:       **End If**
8:       $\mu \leftarrow 0.1$, $\mathbf{G}\boldsymbol{\Theta}_g^{(r)}(\mu) \leftarrow (33)$
9:       **While** $f(\mathbf{G}\boldsymbol{\Theta}_g^{(r)}(2\mu)) - f(\boldsymbol{\Theta}_g^{(r)}) \geq \mu \cdot \frac{\mathrm{tr}(\mathbf{D}_g^{(r)}\mathbf{D}_g^{(r)\mathsf{H}})}{2}$
10:          $\mu \leftarrow 2\mu$
11:       **End While**
12:       **While** $f(\mathbf{G}\boldsymbol{\Theta}_g^{(r)}(\mu)) - f(\boldsymbol{\Theta}_g^{(r)}) < \frac{\mu}{2} \cdot \frac{\mathrm{tr}(\mathbf{D}_g^{(r)}\mathbf{D}_g^{(r)\mathsf{H}})}{2}$
13:          $\mu \leftarrow \mu/2$
14:       **End While**
15:       $\boldsymbol{\Theta}_g^{(r+1)} \leftarrow (33)$
16:    **End For**
17:    $r \leftarrow r+1$
18: **Until** $|f(\boldsymbol{\Theta}^{(r)}) - f(\boldsymbol{\Theta}^{(r-1)})|/f(\boldsymbol{\Theta}^{(r-1)}) \leq \epsilon$

---

where μ is the step size refinable[7] by the Armijo rule [50].

Algorithm 1 summarizes the proposed geodesic RCG method with sequential group-wise updates. Each iteration leverages Lie algebra to perform a multiplicative update (33) along the geodesics of the Stiefel manifold. This appropriate parameter space leads to faster convergence and easier step size tuning. We remark that the additive update of the non-geodesic RCG can be interpreted a first-order Taylor approximation to the multiplicative update of the proposed geodesic RCG, thus necessitating a retraction step to remain on the manifold. The group-wise updates can be performed in parallel to facilitate large-scale BD-RIS design problems. One may also operate on $\boldsymbol{\Theta}$ and pinching (i.e., keeping the main block diagonal and nulling the rest) (27) to unify the step size selection for further acceleration.

We now analyze the computational complexity of solving singular value shaping problem (25) by Algorithm 1. To update each BD-RIS group, SVD of $\mathbf{H}$ requires $\mathcal{O}(NN_\mathrm{T}N_\mathrm{R})$ flops, Euclidean subgradient (26) requires $\mathcal{O}(LN(N_\mathrm{T}+N_\mathrm{R}+L))$ flops, Riemannian subgradient translation (29) requires $\mathcal{O}(L^3)$ flops, deviation parameter (31) and conjugate direction (30) together require $\mathcal{O}(L^2)$ flops, and matrix exponential (33) requires $\mathcal{O}(L^3)$ flops [51]. The overall complexity is thus $\mathcal{O}(I_\mathrm{RCG}G(NN_\mathrm{T}N_\mathrm{R}+LN(N_\mathrm{T}+N_\mathrm{R}+L)+I_\mathrm{BLS}L^3))$, where I_RCG and I_BLS are the number of iterations for geodesic RCG and backtracking line search (i.e., lines 9 – 14 of Algorithm 1), respectively. That is, $\mathcal{O}(N_\mathrm{S})$ for D-RIS and $\mathcal{O}(N_\mathrm{S}^3)$ for fully-connected BD-RIS.

To validate Algorithm 1 and quantify the shaping capability of BD-RIS, we aim to characterize the achievable singular value region of BD-RIS-aided MIMO channel by considering the Pareto optimization problem

$$\max_{\boldsymbol{\Theta}} \quad \sum_{n=1}^{N} \rho_n \sigma_n(\mathbf{H}) \tag{34a}$$
$$\text{s.t.} \quad \boldsymbol{\Theta}_g^\mathsf{H}\boldsymbol{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{34b}$$

where $\rho_n \geq 0$ is the weight associated with the n-th channel singular value. Varying those weights help to characterize the Pareto frontier that encloses the achievable singular value

---
[7]To double the step size, one can simply square the rotation matrix instead of recomputing the matrix exponential, that is, $\exp^2(\mu\mathbf{D}_g^{(r)}) = \exp(2\mu\mathbf{D}_g^{(r)})$.

region. While the objective (34a) may seem obscure, a larger quantity translates to a stronger singular value redistribution capability and thus better wireless performance (e.g., channel capacity for communication [46], detection probability for sensing [52], and harvested power for power transfer [43]). Problem (34) also generalizes the DoF problem in Proposition 1 and the individual singular value shaping problem in Proposition 2 and Corollary 3.3. It can be solved optimally by Algorithm 1 with $[\mathbf{D}]_{n,n} = \rho_n$ in (26).

IV. RATE MAXIMIZATION

In this section, we first solve the BD-RIS-aided MIMO rate maximization problem optimally by joint beamforming design, and then exploit channel shaping for a low-complexity two-stage solution. The problem is formulated as

$$\max_{\mathbf{W},\boldsymbol{\Theta}} \quad R = \log\det\left(\mathbf{I} + \frac{\mathbf{W}^\mathsf{H}\mathbf{H}^\mathsf{H}\mathbf{HW}}{\eta}\right) \tag{35a}$$
$$\text{s.t.} \quad \|\mathbf{W}\|_\mathrm{F}^2 \leq P, \tag{35b}$$
$$\boldsymbol{\Theta}_g^\mathsf{H}\boldsymbol{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{35c}$$

where $\mathbf{W}$ is the transmit precoder, R is the achievable rate, η is the average noise power, and P is maximum average transmit power. Problem (35) is non-convex due to the block-unitary constraint (35c) and the coupling between variables.

A. Alternating Optimization

This approach updates $\boldsymbol{\Theta}$ and $\mathbf{W}$ iteratively until convergence. For a given $\mathbf{W}$, the passive beamforming subproblem is

$$\max_{\boldsymbol{\Theta}} \quad \log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^\mathsf{H}}{\eta}\right) \tag{36a}$$
$$\text{s.t.} \quad \boldsymbol{\Theta}_g^\mathsf{H}\boldsymbol{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{36b}$$

where $\mathbf{Q} \triangleq \mathbf{WW}^\mathsf{H}$ is the transmit covariance matrix. Problem (36) can be solved optimally by Algorithm 1 with the partial derivative given in Lemma 1.

Lemma 1. *The partial derivative of* (36a) *with respect to BD-RIS block g is*

$$\frac{\partial R}{\partial \boldsymbol{\Theta}_g^*} = \frac{1}{\eta}\mathbf{H}_{\mathrm{B},g}^\mathsf{H}\left(\mathbf{I} + \frac{\mathbf{HQH}^\mathsf{H}}{\eta}\right)^{-1}\mathbf{HQH}_{\mathrm{F},g}^\mathsf{H}, \tag{37}$$

which is a special case of (26).

Proof. Please refer to Appendix I. □

For a given $\boldsymbol{\Theta}$, the optimal transmit precoder is given by eigenmode transmission [46]

$$\mathbf{W}^\star = \mathbf{V}\mathrm{diagm}(\mathbf{s}^\star)^{1/2}, \tag{38}$$

where $\mathbf{V}$ is the right singular matrix of $\mathbf{H}$ and $\mathbf{s}^\star$ can be retrieved by the water-filling algorithm [46]. The AO algorithm is guaranteed to converge to local-optimal points of problem (35) since each subproblem is solved optimally and the objective is bounded above. The computational complexity of solving subproblem (36) by geodesic RCG is $\mathcal{O}(I_\mathrm{RCG}G(NL^2 + LN_\mathrm{T}N_\mathrm{R} + N_\mathrm{T}^2N_\mathrm{R} + N_\mathrm{T}N_\mathrm{R}^2 + N_\mathrm{R}^3 + I_\mathrm{BLS}L^3))$. On the other hand, the complexity of active beamforming (38) is $\mathcal{O}(NN_\mathrm{T}N_\mathrm{R})$. The overall complexity is thus $\mathcal{O}(I_\mathrm{AO}(I_\mathrm{RCG}G(NL^2 + LN_\mathrm{T}N_\mathrm{R} + N_\mathrm{T}^2N_\mathrm{R} + N_\mathrm{T}N_\mathrm{R}^2 + N_\mathrm{R}^3 + I_\mathrm{BLS}L^3) + NN_\mathrm{T}N_\mathrm{R}))$, where I_AO is the number of iterations for AO. That is, $\mathcal{O}(N_\mathrm{S})$ for D-RIS and $\mathcal{O}(N_\mathrm{S}^3)$ for fully-connected BD-RIS.

B. Low-Complexity Solution

To reduce computational complexity, we decouple the joint beamforming design by first shaping the MIMO channel by BD-RIS for maximum power gain and then performing eigenmode transmission. The shaping subproblem is formulated as

$$\max_{\boldsymbol{\Theta}} \quad \|\mathbf{H}_D + \mathbf{H}_B \boldsymbol{\Theta} \mathbf{H}_F\|_F^2 \tag{39a}$$

$$\text{s.t.} \quad \boldsymbol{\Theta}_g^H \boldsymbol{\Theta}_g = \mathbf{I}, \quad \forall g. \tag{39b}$$

While similar problems have been studied in single-mode cases [24], [27], generalizing those methods to MIMO remains non-trivial due to the tradeoff between multi-mode alignments. One can see that the objective function (39a) is equivalent to $\sum_{n=1}^{N} \sigma_n^2(\mathbf{H})$ and thus readily solvable by Algorithm 1. Below we propose a more elegant power iteration method inspired by [53] that iterates in closed-form by orthogonal projection. The idea is to approximate the quadratic objective (39a) by its first-order Taylor expansion and solve each subproblem by group-wise SVD.

Proposition 5. *Starting from any feasible* $\boldsymbol{\Theta}^{(0)}$, *the orthogonal projection of*

$$\mathbf{M}_g^{(r)} = \mathbf{H}_{B,g}^H \left(\mathbf{H}_D + \mathbf{H}_B \operatorname{diagm}\left(\boldsymbol{\Theta}_{[1:g-1]}^{(r+1)}, \boldsymbol{\Theta}_{[g:G]}^{(r)}\right) \mathbf{H}_F \right) \mathbf{H}_{F,g}^H \tag{40}$$

onto the Stiefel manifold, given in the closed-form [54]

$$\boldsymbol{\Theta}_g^{(r+1)} = \operatorname*{argmin}_{\mathbf{X}_g \in \mathbb{U}^{L \times L}} \|\mathbf{M}_g - \mathbf{X}_g\|_F = \mathbf{U}_g^{(r)} \mathbf{V}_g^{(r)H}, \tag{41}$$

monotonically increases the objective function (39a), *where* $\mathbf{U}_g^{(r)}$ *and* $\mathbf{V}_g^{(r)}$ *are any left and right singular matrices of* $\mathbf{M}_g^{(r)}$. *When* (40) *converges,* (41) *leads to a convergence of the objective function* (39a) *towards a stationary point.*

Proof. Please refer to Appendix J. $\square$

Remark 3. Although a rigorous convergence proof remains intricate due to the non-uniqueness of SVD, empirical evidence from extensive simulation suggests that (40) converges reliably from random initializations such that (41) consistently provides an optimal solution to problem (39).

To update each BD-RIS group, the matrix multiplication (40) requires $\mathcal{O}(N_T N_R + N L^2 + N_T N_R L)$ flops and its SVD requires $\mathcal{O}(L^3)$ flops. The overall complexity is thus $\mathcal{O}(I_{SAA} G(N_T N_R + N L^2 + N_T N_R L + L^3))$, where I_{SAA} is the number iterations for successive affine approximation. That is, $\mathcal{O}(N_S)$ for D-RIS and $\mathcal{O}(N_S^3)$ for fully-connected BD-RIS. It is worth mentioning that the computational complexity for fully-connected BD-RIS can be further reduced:

- *Negligible direct channel:* The optimal solution to (39) has been solved in closed form by (20a);
- *Non-negligible direct channel:* In terms of maximizing the inner product $\langle \mathbf{H}_D, \mathbf{H}_B \boldsymbol{\Theta} \mathbf{H}_F \rangle$, (39) is reminiscent of the weighted orthogonal Procrustes problem [55]

$$\min_{\boldsymbol{\Theta}} \quad \|\mathbf{H}_D - \mathbf{H}_B \boldsymbol{\Theta} \mathbf{H}_F\|_F^2 \tag{42a}$$

$$\text{s.t.} \quad \boldsymbol{\Theta}^H \boldsymbol{\Theta} = \mathbf{I}, \tag{42b}$$

which still has no trivial solution. One *lossy* transformation [56] shifts $\boldsymbol{\Theta}$ to sides of the product by Moore-Penrose

TABLE I
PERFORMANCE OF RCG ALGORITHMS ON (34) WITH $N_T = N_R = 4$, $L = 16$

RCG path	$N_S = 16$		$N_S = 64$		$N_S = 256$	
	Iterations	Time [ms]	Iterations	Time [ms]	Iterations	Time [ms]
Geodesic	6.493	1.807	9.003	7.378	12.98	49.41
Non-geodesic (Manopt)	8.601	25.90	11.09	36.27	14.29	65.89

inverse, formulating two standard orthogonal Procrustes problems

$$\min_{\boldsymbol{\Theta}} \quad \|\mathbf{H}_B^\dagger \mathbf{H}_D - \boldsymbol{\Theta} \mathbf{H}_F\|_F^2 \text{ or } \|\mathbf{H}_D \mathbf{H}_F^\dagger - \mathbf{H}_B \boldsymbol{\Theta}\|_F^2 \tag{43a}$$

$$\text{s.t.} \quad \boldsymbol{\Theta}^H \boldsymbol{\Theta} = \mathbf{I}, \tag{43b}$$

with optimal solutions [57, (6.4.1)]

$$\boldsymbol{\Theta}_{\text{P-max-approx}}^{\text{MIMO}} = \mathbf{U} \mathbf{V}^H, \tag{44}$$

where $\mathbf{U}$ and $\mathbf{V}$ are respectively any left and right singular matrices of $\mathbf{H}_B^\dagger \mathbf{H}_D \mathbf{H}_F^H$ or $\mathbf{H}_B^H \mathbf{H}_D \mathbf{H}_F^\dagger$.

Although (20a) and (44) are of similar form, the latter is neither optimal nor a generalization of the former due to the lossy transformation. We will show by simulation that (44) still achieves near-optimal performance on average. Once the channel is shaped in closed form by (41) or (20a) or (44), the active beamforming is retrieved in closed form by (38). This two-stage solution avoids outer iterations and efficiently handles (or avoids) inner iterations.

V. SIMULATION RESULTS

In this section, we provide numerical results to evaluate the proposed solutions and algorithms.[8] Consider a distance-dependent path loss model $\Lambda(d) = \Lambda_0 d^{-\gamma}$ where Λ_0 is the reference path loss at distance 1 m, d is the propagation distance, and γ is the path loss exponent. We set $\Lambda_0 = -30$dB, $\gamma_D = 3$, $\gamma_F = 2.4$, $\gamma_B = 2$, $d_D = 14.7$m, $d_F = 10$m, $d_B = 6.3$m, which corresponds to a typical indoor environment with $\Lambda_D = -65$dB, $\Lambda_F = -54$dB, $\Lambda_B = -46$dB. The small-scale fading model is $\mathbf{H} = \sqrt{\kappa/(1+\kappa)}\mathbf{H}_{\text{LoS}} + \sqrt{1/(1+\kappa)}\mathbf{H}_{\text{NLoS}}$, where κ is the Rician K-factor, $\mathbf{H}_{\text{LoS}}$ is the deterministic LoS component, and $\mathbf{H}_{\text{NLoS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$ is the Rayleigh component. Unless otherwise specified, we assume the direct channel is present, $\kappa = 0$ (i.e., Rayleigh fading) for all channels, and $\eta = -75$dB.

A. Algorithm Evaluation

Here are benchmarks obtained running MATLAB R2023a on an octa-core AMD Ryzen7 5800U processor @ 4.5 GHz with 16 GiB memory. Table I benchmarks two RCG algorithms on the Pareto singular value problem (34) with $N_T = N_R = 4$ and $L = 16$. The geodesic RCG is implemented with pinched gradients w.r.t. $\boldsymbol{\Theta}$ and unified step size selection; please refer to the discussion below (33) for details. The non-geodesic RCG is implemented by Manopt toolbox at commit a879a0d [58]. Both algorithms employ a stopping criterion of relative change in the objective function with a tolerance of $\epsilon = 1 \times 10^{-4}$, such that the final values are identical within reasonable precision. The statistics are averaged over 1000 independent channel realizations. We observe that the non-geodesic RCG typically requires 1 to 2 more iterations than its geodesic counterpart. This is because the addition is in the tangent space of the manifold and is less effective than manifold-native updates. When it

---

[8]Source code is available at https://github.com/snowztail/channel-shaping.

TABLE II
PERFORMANCE OF D-RIS AND BD-RIS ON (35) WITH $N_\mathrm{T} = N_\mathrm{R} = 4$

RIS type	$N_\mathrm{S} = 16$		$N_\mathrm{S} = 64$		$N_\mathrm{S} = 256$	
	Iterations	Time [ms]	Iterations	Time [ms]	Iterations	Time [ms]
Diagonal	2.010	7.848	2.023	36.33	2.141	261.1
Fully-connected BD	2.049	4.878	2.027	15.17	2.030	305.5

comes to elapsed time, the geodesic RCG is 1333% faster than the non-geodesic counterpart when $N_\mathrm{S} = 16$. The main reason is that the geodesic RCG avoids the retraction step from the Euclidean space to the manifold. According to the profiler report, around 60% of the non-geodesic RCG runtime is spent on retraction, which becomes the main bottleneck of the algorithm. The advantage narrows down to 391.6% and 33.35% when $N_\mathrm{S} = 64$ and 256, respectively. This is because accurately evaluating matrix exponential can be time-consuming for large N_S.

Table II compares the performance of D-RIS and fully-connected BD-RIS on rate maximization problem (35) using the AO design in Section IV-A, where $N_\mathrm{T} = N_\mathrm{R} = 4$ and $P = 20$dB. The statistics are averaged over 1000 independent runs. Interestingly, as opposite to the asymptotic complexity analysis, the optimization of fully-connected BD-RIS actually takes shorter elapsed time than D-RIS when N_S is not excessively large. One possible reason is that fully-connected BD-RIS only involves 1 backtracking line search per iteration while D-RIS requires N_S times. Another reason is that the group-wise update of D-RIS leads to slower convergence of inner iterations. These numerical results, together with the closed-form solutions provided in the analysis section, together suggest that designing a practically-sized BD-RIS may be less computational expensive than expected.

B. Channel Singular Value Redistribution

1) Achievable Singular Value Region: Fig. 2 compares the achievable singular value region obtained by solving problem (34) and its outer bounds suggested by Corollary 3.1. Here $\bar{N} = N_\mathrm{T} = N_\mathrm{S} = N_\mathrm{R} = 3$ and the bounds are enumerated as

$$
\begin{aligned}
\sigma_1(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F}), \\
\sigma_2(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_2(\mathbf{H}_\mathrm{F}), \\
\sigma_2(\mathbf{H}) &\le \sigma_2(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F}), \\
\sigma_3(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_3(\mathbf{H}_\mathrm{F}), \\
\sigma_3(\mathbf{H}) &\le \sigma_2(\mathbf{H}_\mathrm{B})\sigma_2(\mathbf{H}_\mathrm{F}), \\
\sigma_3(\mathbf{H}) &\le \sigma_3(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F}),
\end{aligned}
\tag{45a}
$$

$$
\begin{aligned}
\sigma_1(\mathbf{H})\sigma_2(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_2(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F})\sigma_2(\mathbf{H}_\mathrm{F}), \\
\sigma_1(\mathbf{H})\sigma_3(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_3(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F})\sigma_2(\mathbf{H}_\mathrm{F}), \\
\sigma_1(\mathbf{H})\sigma_3(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_2(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F})\sigma_3(\mathbf{H}_\mathrm{F}), \\
\sigma_2(\mathbf{H})\sigma_3(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_2(\mathbf{H}_\mathrm{B})\sigma_2(\mathbf{H}_\mathrm{F})\sigma_3(\mathbf{H}_\mathrm{F}), \\
\sigma_2(\mathbf{H})\sigma_3(\mathbf{H}) &\le \sigma_2(\mathbf{H}_\mathrm{B})\sigma_3(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F})\sigma_2(\mathbf{H}_\mathrm{F}), \\
\sigma_2(\mathbf{H})\sigma_3(\mathbf{H}) &\le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_3(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F})\sigma_3(\mathbf{H}_\mathrm{F}),
\end{aligned}
\tag{45b}
$$

$$
\sigma_1(\mathbf{H}) \ge \sigma_2(\mathbf{H}) \ge \sigma_3(\mathbf{H}),
\tag{45c}
$$

where (45a), (45b) are explicit results of (12) while (45c) denotes the ordering of singular values. Those are labeled respectively as 'Bounds (individual)', 'Bounds (product)', and 'Bounds (ordering)' in Fig. 2. The two views confirm that the theoretical outer bounds are not everywhere tight with many entries being redundant, but they provide a conservative estimate of the achievable singular value region. Importantly,

(a) Top view

(b) Side view
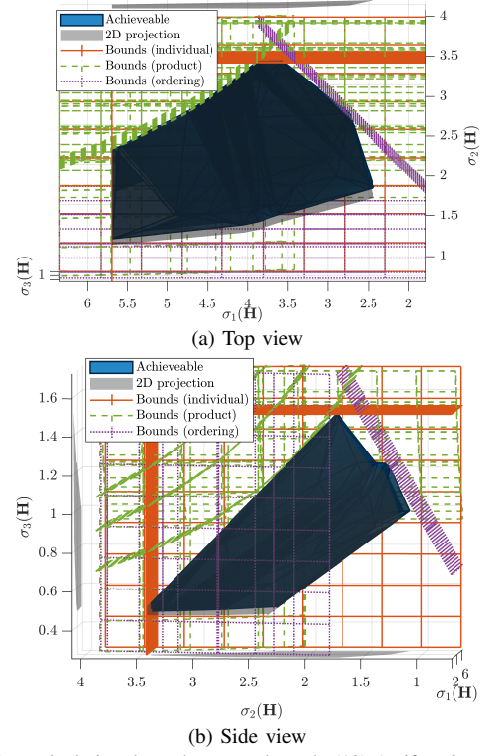
Fig. 2. Theoretical singular value outer bounds (12) (uniformly-spaced mesh grids) vs achievable singular value region by solving (34) (solid dark shape) for one channel realization, where $N_\mathrm{T} = N_\mathrm{S} = N_\mathrm{R} = 3$, the direct channel is negligible, and the BD-RIS is fully-connected. Small offsets are introduced on both views such that the active bounds are highlighted by densely-spaced curves/lines that marginally overlap the region from above. The achievable region lies entirely within the intersection of the bounding surfaces in the 3D space.

(a) $2 \times 32 \times 2$ (no direct)

(b) $2 \times 32 \times 2$

(c) $2 \times 64 \times 2$

(d) $2 \times 128 \times 2$
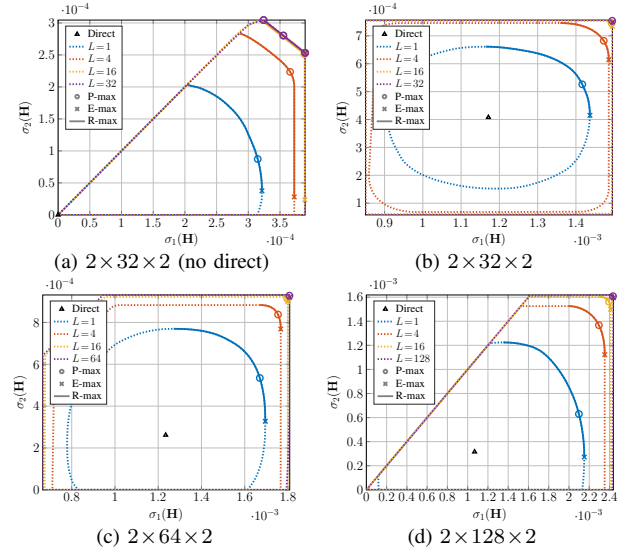
Fig. 3. Achievable singular value regions of an $N_\mathrm{T} = N_\mathrm{R} = 2$ channel shaped by BD-RIS. The singular value pair of the direct channel are marked as baseline. On the Pareto frontiers, 'P-max', 'E-max', and 'R-max' refer to the channel power gain-optimal point, wireless power transfer-optimal point, and rate-optimal arc, respectively.

the vertices of the region lie on the bounding surfaces and can be obtained in closed form without performing optimization.

Fig. 3 illustrates the achievable regions of singular values of an $N_\mathrm{T} = N_\mathrm{R} = 2$ point-to-point MIMO shaped by RIS, where the channel power gain-optimal point, wireless power transfer-optimal point, and rate-optimal arc are highlighted on the Pareto frontiers. The results are obtained by solving the

channel shaping problem (34) merely without any application-specific optimization. As the SNR increases, the rate-optimal point proceeds on the arc from the east (favoring $\sigma_1(\mathbf{H})$) to the north (favoring $\sigma_2(\mathbf{H})$), which aligns with the expected behavior of water filling. When the direct channel is negligible, the achievable regions in Fig. 3(a) are shaped like pizza slices. This is because $\sigma_1(\mathbf{H}) \geq \sigma_2(\mathbf{H})$ and there exists a tradeoff between the alignment of two modes. The smallest singular value can be enhanced up to 2×10^{-4} by D-RIS and 3×10^{-4} by fully-connected BD-RIS, corresponding to a 50% gain. We also see that for fully-connected BD-RIS, there exists a point that is simultaneously optimal for channel power gain, harvested power of wireless power transfer, and achievable rate of wireless communication, as indicated by (20a), (15a), and (23). Interestingly, this observation still holds in Figs. 3(b) – 3(d) where the direct channel is *not* negligible. It is a pity that we could not provide a formal proof on this due to the non-trivial solution structures. The shape of the singular value region depends heavily on the relative strength of the indirect channels, which increases with N_S from the baseline $\Lambda_F \Lambda_B / \Lambda_D = -35\text{dB}$. Fig. 3(b) shows that a 32-element RIS is insufficient to compensate this imbalance and results in a limited singular value region that is symmetric around the direct point. As the group size L increases, the shape of the region evolves from elliptical to square. This transformation not only improves the dynamic range of $\sigma_1(\mathbf{H})$ and $\sigma_2(\mathbf{H})$ by 22% and 38% respectively, but also provides a better tradeoff in manipulating both singular values. The observation verifies that the design flexibility of BD-RIS allows better alignment of multiple modes simultaneously. As a consequence, the optimally shaped channels for power gain, communication, and power transfer coincide, implying that a fully-connected BD-RIS may be designed in closed-form for simultaneous multi-functional optimality. The singular value region also enlarges as the number of scattering elements N_S increases. In particular, Fig. 3(d) shows that the equivalent channel can be completely nulled (corresponding to the origin) by a 128-element BD-RIS but not by a diagonal one. The effect may be leveraged for interference cancellation and covert communication. Those results demonstrate the superior channel shaping capability of BD-RIS and emphasizes the importance of adding reconfigurable components between RIS elements.

2) Analytical Bounds and Numerical Results: We focus on achieving the asymptotic bounds in Proposition 2 by finite N_S, since most results from Proposition 3 are supplied with closed-form RIS solutions. For a rank-k forward channel, Fig. 4 compares the individual singular value bounds in Proposition 2 and the numerical results obtained by solving problem (34) with proper weights. When the RIS is in the LoS of the transmitter, Figs. 4(a) and 4(b) show that the achievable channel singular values indeed satisfy Corollary 2.1, namely $\sigma_1(\mathbf{H}) \geq \sigma_1(\mathbf{T})$, $\sigma_2(\mathbf{T}) \leq \sigma_2(\mathbf{H}) \leq \sigma_1(\mathbf{T})$, etc. It is obvious that BD-RIS can approach those bounds better than D-RIS with a small N_S. Another example is given in Fig. 4(c) with rank-2 forward channel. The first two channel singular values are unbounded above and bounded below by the first two singular values of $\mathbf{T}$, while the last two singular values can be suppressed to zero and bounded above by the first two singular values of $\mathbf{T}$.

(a) $4 \times 32 \times 4$ (rank-1) (b) $4 \times 64 \times 4$ (rank-1)

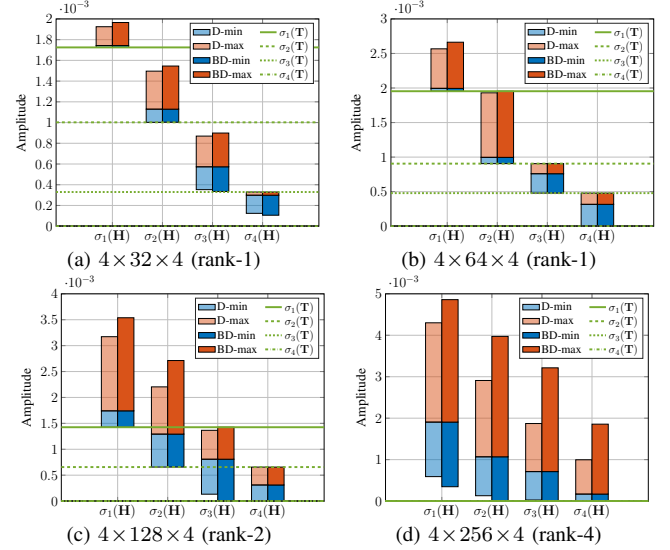(c) $4 \times 128 \times 4$ (rank-2) (d) $4 \times 256 \times 4$ (rank-4)

Fig. 4. Achievable channel singular values: analytical bounds (lines) and numerical results (bars). Baselines of bars denote the singular values of the direct channel. Blue (resp. red) bars denote the lower (resp. upper) dynamic range of singular values obtained by solving (34) with $\rho_n / \rho_{n'} \to 0$ (resp. $\to \infty$), $\forall n, \ n' \neq n$. 'D' means D-RIS and 'BD' refers to fully-connected BD-RIS. 'rank-k' refers to the rank of the forward channel.

(a) $1 \times 256 \times 1$ (no direct) (b) $4 \times 256 \times 4$ (no direct)
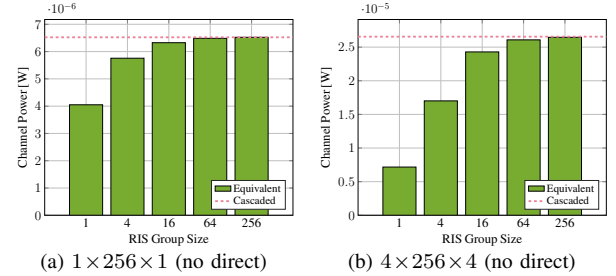
Fig. 5. Average maximum channel power gain versus BD-RIS group size and MIMO dimensions. 'Cascaded' refers to the upper bound in (19).

Those observations align with Proposition 2. Finally, Fig. 4(d) confirms there are no extra singular value bounds when both backward and forward channels are full-rank. This can be predicted from (9) where $\mathbf{V}_F$ becomes unitary and $\mathbf{T} = \mathbf{0}$. The numerical results are consistent with the analytical bounds, and we conclude that the channel shaping advantage of BD-RIS over D-RIS scales with the rank of backward and forward channels.

Fig. 5 compares the analytical bounds on the channel power gain in Corollary 3.4 and the numerical results obtained by solving problem (39) when the direct channel is negligible. Here, a fully-connected BD-RIS can attain the upper bound either in closed form (20a) or via optimization approach (41). For the SISO case in Fig. 5(a), the maximum channel power gain is approximately 4×10^{-6} by D-RIS and 6.5×10^{-6} by fully-connected BD-RIS, corresponding to a 62.5% gain. It comes purely from branch matching as discussed in Example 1 and confirms the asymptotic power scaling law derived in [24, (30)] Interestingly, Fig. 5(b) shows that this relative gain, inferrable from the expectation analysis (21), surges to 270% in $N_T = N_R = 4$ MIMO. We thus conclude that the power gain of BD-RIS scales with the group size and MIMO dimensions.

C. Achievable Rate Maximization

We first focus on the channel power gain problem (39). Fig. 6 shows the maximum channel power gain under different

(a) Power: $2 \times N_S \times 2$ (b) Rate: $4 \times N_S \times 4$

Fig. 8. Impact of RIS symmetry on the MIMO power gain and achievable rate.

(a) $16 \times N_S \times 16$ (no direct)   (b) $16 \times N_S \times 16$

Fig. 6. Average maximum channel power gain versus RIS configuration. 'Explicit' refers to the optimal solution (20a) when the direct channel is negligible. 'OP-left' and 'OP-right' refer to the suboptimal solutions, when the direct channel is significant, by lossy transformation (43) where $\boldsymbol{\Theta}$ is to the left and right of the product, respectively.
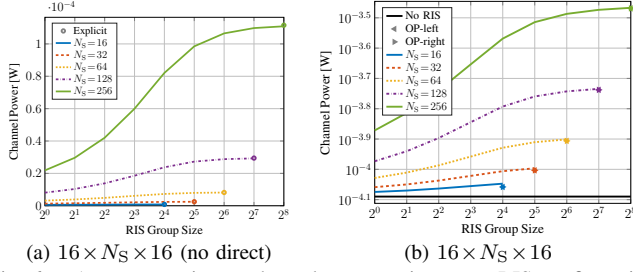

(a) $4 \times 128 \times 4$ (b) $N_T \times 128 \times N_R$

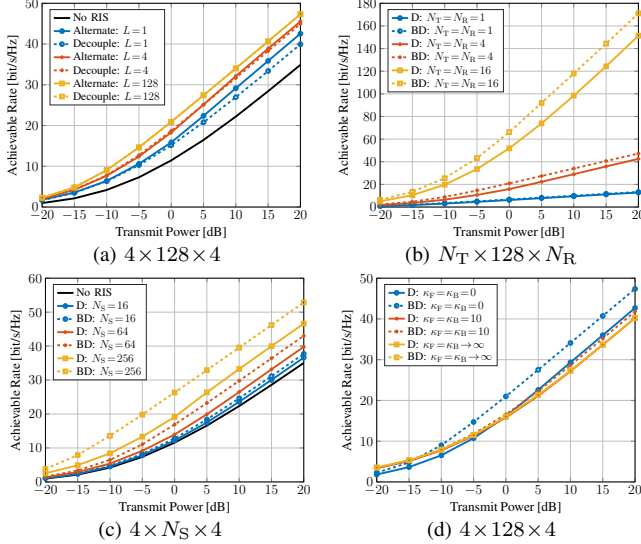(c) $4 \times N_S \times 4$ (d) $4 \times 128 \times 4$

Fig. 7. Average achievable rate versus MIMO and RIS configurations. The transmit power corresponds to a direct SNR of -10 to 30 dB. 'Alternate' refers to the alternating optimization and 'Decouple' refers to the low-complexity design. 'D' means D-RIS and 'BD' refers to fully-connected BD-RIS.

RIS configurations. An interesting observation is that the relative power gain of BD-RIS over D-RIS is even larger when the direct channel is significant. As shown in Figs. 6(a) and 6(b), a 64-element fully BD-RIS can almost provide the same channel power gain as a 256-element D-RIS when the direct channel is significant, but less so when it is negligible. This is because the mode alignment advantage of BD-RIS becomes more pronounced when the modes of direct channel is taken into account. We also notice that the suboptimal solutions (44) for fully-connected BD-RIS by lossy transformation (43) are very close to optimal especially for a large N_S.

Fig. 7 presents the achievable rate under different MIMO and RIS configurations. At a transmit power $P = 10$ dB, Fig. 7(a) shows that introducing a 128-element D-RIS to $N_T = N_R = 4$ MIMO can improve the achievable rate from 22.2 bps/Hz to 29.2 bps/Hz ($+31.5\%$). A BD-RIS of group size 4 and 128 can further elevate those to 32.1 bps/Hz ($+44.6\%$) and 34 bps/Hz ($+53.2\%$), respectively. An interesting observation is that the rate gap between the optimal AO approach in Section IV-A and the low-complexity shaping solution in Section IV-B narrows as group size L increases and completely vanishes for a fully-connected BD-RIS. This implies that joint beamforming designs may be decoupled with minimal performance degradation by first shaping the wireless channel and then optimizing the
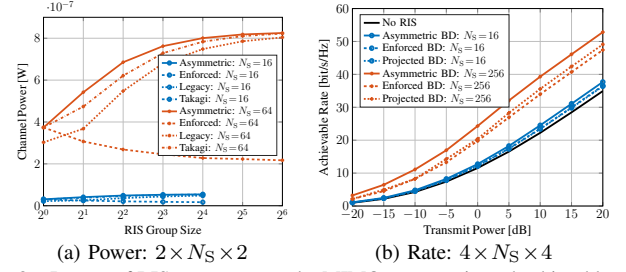
transceiver, which substantially simplifies the design. Figs. 7(b) and 7(c) also show that both *absolute and relative* rate gains of BD-RIS over D-RIS increases with the number of transmit and receive antennas and scattering elements, especially at high SNR. For $N_S = 128$ and $P = 20$ dB, the achievable rate ratio of BD-RIS over D-RIS is 1.04, 1.11, and 1.13 for $N_T = N_R = 1, 4$, and 16, respectively. For $N_T = N_R = 4$ and $P = 20$ dB, this ratio amounts to 1.03, 1.08, and 1.13 for $N_S = 16, 64$, and 256, respectively. Those observations align with the power gain results in Fig. 6 and highlight the rate benefits of BD-RIS over D-RIS in large-scale MIMO systems. In the low power regime (-20 to -10 dB), we also notice that the slope of the achievable rate of BD-RIS is steeper than that of D-RIS. That is, BD-RIS can help to activate more streams and achieve the asymptotic DoF at a low transmit SNR. This is particularly visible in Fig. 7(c) where the topmost curve is almost a linear function of the transmit power. It can be predicted from Fig. 3 that BD-RIS can significantly enlarge all channel singular values for higher receive SNR. Finally, Fig. 7(d) shows that the gap between D- and BD-RIS narrows as the Rician K-factor increases and becomes indistinguishable in LoS environment. The observation is expected from previous studies [24], [25] and aligns with Corollary 2.1, which suggests that the BD-RIS should be deployed in rich-scattering environments to exploit its channel shaping potential.

D. Practical Constraints

1) RIS Symmetry: Symmetric RIS satisfying $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{\mathsf{T}}$ are often considered in the literature due to hardware constraints. This study aim to investigate the impact of RIS symmetry on the system performance.

Remark 4. All proposed asymmetric BD-RIS solutions are readily modifiable for symmetry. In particular,

(i) *SVD-based (e.g., (15), (20), (23), (41), (44)):* Those closed-form asymmetric solutions are constructed from the product of singular matrices. If symmetry is required, one can replace the $\mathbf{U}, \mathbf{V}^{\mathsf{H}}$ in the SVD of $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$ by $\mathbf{Q}, \mathbf{Q}^{\mathsf{T}}$ in the Autonne-Takagi factorization [59] of $\frac{\mathbf{A} + \mathbf{A}^{\mathsf{T}}}{2} = \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^{\mathsf{T}}$ to construct $\boldsymbol{\Theta}$;

(ii) *RCG-based (e.g., (26), (37)):* The symmetry constraint is added to the corresponding optimization problems, and one can project the solution to the nearest symmetric point $\boldsymbol{\Theta} \leftarrow \frac{\boldsymbol{\Theta} + \boldsymbol{\Theta}^{\mathsf{T}}}{2}$ after each iteration.

Figs. 8(a) and 8(b) compare the power gain and achievable rate of MIMO point-to-point channel under asymmetric and various symmetric RIS configurations. Here, 'asymmetric' refers to the benchmark solution by (41) or (37), 'enforced'
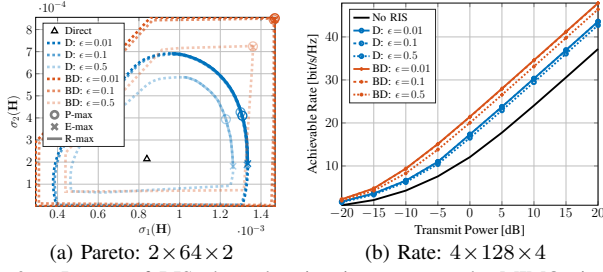
Fig. 9. Impact of RIS channel estimation error on the MIMO singular value region and achievable rate. A higher transparency of the Pareto frontier indicates a larger channel estimation error. 'D' means D-RIS and 'BD' refers to fully-connected BD-RIS.

refers to enforcing symmetry on above, 'legacy' refers to a straightforward extension of the single-mode SNR-optimal solution [27, (6)], 'Takagi' refers to the modification (i), and 'projection' refers to the modification (ii). We observe that the performance gaps between the asymmetric and symmetric RIS configurations are insignificant and tends to widen with the number of scattering elements. The two proposed modifications also outperform other candidates in both problems.

2) Channel Estimation Error: Figs. 9(a) and 9(b) investigates how RIS channel estimation errors affect the system performance in terms of singular value region and achievable rate. We assume the direct channel can be perfectly acquired and the estimated backward and forward channels are modeled by

$$\hat{\mathbf{H}}_{\mathrm{B/F}} = \mathbf{H}_{\mathrm{B/F}} + \tilde{\mathbf{H}}_{\mathrm{B/F}},$$

where the error follows $\mathrm{vec}(\tilde{\mathbf{H}}_{\mathrm{B/F}}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \epsilon\Lambda_{\mathrm{B}}\Lambda_{\mathrm{F}}\mathbf{I})$. The results are evaluated over the ground truth channels. It is observed that the proposed channel shaping and joint beamforming solutions are reasonably robust to channel estimation errors. An interesting observation is that a BD-RIS designed over extremely poorly estimated channels ($\epsilon = 0.5$) may still outperform a D-RIS designed over almost perfectly estimated channels ($\epsilon = 0.01$). We hope those results can motivate further research on the robust shaping design and provide insights for practical BD-RIS deployment.

VI. CONCLUSION

This paper investigated the capability of BD-RIS to shape a MIMO channel in terms of singular values and their functions. We started from a toy example and derived some analytical bounds (with closed-form solutions) on the channel singular values, power gain, and capacity. An efficient framework was then proposed to optimize the BD-RIS for a broader class of singular value functions. We also presented two beamforming designs for the rate maximization problem, one for optimal performance and the other exploits shaping implications for much lower complexity while remaining close-to-optimal. Extensive simulation show that the significant power and rate gains of BD-RIS over D-RIS stems from its superior MIMO branch matching and mode alignment potentials, which scales with the number of elements, group size, and MIMO dimensions.

The analysis and optimization methods in this paper have been tailored for group-connected BD-RIS. Extension to other RIS architectures remains a promising area for future research. One straightforward extension to the multi-sector model [33] is to retrieve the optimal scattering matrix for each sector

individually by methods in this paper and then play with the power splitting factors. Meanwhile, transitioning from single- to multi-layer RIS models [60] mirrors that from single- to multi-hop AF relays; interested readers may be inspired by [44].

Finally, we remark that the principle of channel shaping is not limited to point-to-point MIMO. Algorithm 1 and the two solutions in Section IV are readily extendable to weighted sum-rate maximization and leakage interference minimization in MIMO interference channel.

APPENDIX

A. Proof of Proposition 1

It suffices to consider the rank of the indirect channel. Denote the SVD of the backward and forward channels as

$$\mathbf{H}_{\mathrm{B/F}} = \begin{bmatrix} \mathbf{U}_{\mathrm{B/F},1} & \mathbf{U}_{\mathrm{B/F},2} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{\mathrm{B/F},1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathrm{B/F},1}^{\mathsf{H}} \\ \mathbf{V}_{\mathrm{B/F},2}^{\mathsf{H}} \end{bmatrix},$$

where $\mathbf{U}_{\mathrm{B/F},1}$ and $\mathbf{V}_{\mathrm{B/F},1}$ are any left and right singular matrices of $\mathbf{H}_{\mathrm{B/F}}$ corresponding to non-zero singular values $\mathbf{\Sigma}_{\mathrm{B/F},1}$, and $\mathbf{U}_{\mathrm{B/F},2}$ and $\mathbf{V}_{\mathrm{B/F},2}$ are those corresponding to zero singular values. The rank of the indirect channel is [39, (16.5.10.b)]

$$\mathrm{rank}(\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}}) = \mathrm{rank}(\mathbf{H}_{\mathrm{B}}) - \dim\big(\ker(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\mathbf{\Theta}^{\mathsf{H}}) \cap \mathrm{ran}(\mathbf{H}_{\mathrm{B}}^{\mathsf{H}})\big)$$
$$= \mathrm{rank}(\mathbf{H}_{\mathrm{B}}) - \dim\big(\mathrm{ran}(\mathbf{\Theta}\mathbf{U}_{\mathrm{F},2}) \cap \mathrm{ran}(\mathbf{V}_{\mathrm{B},1})\big)$$
$$\triangleq r_{\mathrm{B}} - r_{\mathrm{L}}(\mathbf{\Theta}),$$

where we define $r_{\mathrm{L}}(\mathbf{\Theta}) \triangleq \dim\big(\mathrm{ran}(\mathbf{\Theta}\mathbf{U}_{\mathrm{F},2}) \cap \mathrm{ran}(\mathbf{V}_{\mathrm{B},1})\big)$ and $r_{\mathrm{B/F}} \triangleq \mathrm{rank}(\mathbf{H}_{\mathrm{B/F}})$. Since $\mathbf{U}_{\mathrm{F},2} \in \mathbb{U}^{N_{\mathrm{S}} \times (N_{\mathrm{S}} - r_{\mathrm{F}})}$ and $\mathbf{V}_{\mathrm{B},1} \in \mathbb{U}^{N_{\mathrm{S}} \times r_{\mathrm{B}}}$, we have $\max(r_{\mathrm{B}} - r_{\mathrm{F}}, 0) \leq r_{\mathrm{L}}(\mathbf{\Theta}) \leq \min(N_{\mathrm{S}} - r_{\mathrm{F}}, r_{\mathrm{B}})$ and thus

$$\max(r_{\mathrm{B}} + r_{\mathrm{F}} - N_{\mathrm{S}}, 0) \leq \mathrm{rank}(\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}}) \leq \min(r_{\mathrm{B}}, r_{\mathrm{F}}). \quad (46)$$

To attain the upper bound in (46), the RIS needs to minimize $r_{\mathrm{L}}(\mathbf{\Theta})$ by aligning the ranges of $\mathbf{\Theta}\mathbf{U}_{\mathrm{F},2}$ and $\mathbf{V}_{\mathrm{B},2}$ as much as possible. This is achieved by

$$\mathbf{\Theta}_{\mathrm{DoF\text{-}max}}^{\mathrm{MIMO}} = \mathbf{Q}_{\mathrm{B},2}\mathbf{Q}_{\mathrm{F},2}^{\mathsf{H}}, \quad (47)$$

where $\mathbf{Q}_{\mathrm{B},2}$ and $\mathbf{Q}_{\mathrm{F},2}$ are the unitary matrices of the QR decomposition of $\mathbf{V}_{\mathrm{B},2}$ and $\mathbf{U}_{\mathrm{F},2}$, respectively. Similarly, the lower bound in (46) is attained at

$$\mathbf{\Theta}_{\mathrm{DoF\text{-}min}}^{\mathrm{MIMO}} = \mathbf{Q}_{\mathrm{B},1}\mathbf{Q}_{\mathrm{F},2}^{\mathsf{H}}, \quad (48)$$

where $\mathbf{Q}_{\mathrm{B},1}$ is the unitary matrix of the QR decomposition of $\mathbf{V}_{\mathrm{B},1}$. While the DoF-optimal structures (47) and (48) are always feasible for fully-connected BD-RIS, they are generally infeasible for D-RIS unless there exist some QR decomposition that diagonalize $\mathbf{Q}_{\mathrm{B},2}\mathbf{Q}_{\mathrm{F},2}^{\mathsf{H}}$ and $\mathbf{Q}_{\mathrm{B},1}\mathbf{Q}_{\mathrm{F},2}^{\mathsf{H}}$ simultaneously. That is, BD-RIS may achieve a larger or smaller number of DoF of indirect channel, and thus equivalent channel, than D-RIS.

B. Proof of Proposition 2

We consider rank-k forward channel and the proof follows similarly for rank-k backward channel. Let $\mathbf{H}_{\mathrm{F}} = \mathbf{U}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{V}_{\mathrm{F}}^{\mathsf{H}}$ be the SVD of the forward channel. The channel Gram matrix $\mathbf{G} \triangleq \mathbf{H}\mathbf{H}^{\mathsf{H}}$ can be written as

$$\mathbf{G} = \mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{D}}^{\mathsf{H}} + \mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{U}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}}^{\mathsf{H}}\mathbf{U}_{\mathrm{F}}^{\mathsf{H}}\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}$$
$$+ \mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{U}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{V}_{\mathrm{F}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}^{\mathsf{H}} + \mathbf{H}_{\mathrm{D}}\mathbf{V}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{U}_{\mathrm{F}}^{\mathsf{H}}\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}$$
$$= \mathbf{H}_{\mathrm{D}}(\mathbf{I} - \mathbf{V}_{\mathrm{F}}\mathbf{V}_{\mathrm{F}}^{\mathsf{H}})\mathbf{H}_{\mathrm{D}}^{\mathsf{H}}$$
$$+ (\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{U}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}} + \mathbf{H}_{\mathrm{D}}\mathbf{V}_{\mathrm{F}})(\mathbf{\Sigma}_{\mathrm{F}}\mathbf{U}_{\mathrm{F}}^{\mathsf{H}}\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}} + \mathbf{V}_{\mathrm{F}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}^{\mathsf{H}})$$
$$= \mathbf{Y} + \mathbf{Z}\mathbf{Z}^{\mathsf{H}},$$

where we define $\mathbf{Y} \triangleq \mathbf{H}_\mathrm{D}(\mathbf{I} - \mathbf{V}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H})\mathbf{H}_\mathrm{D}^\mathsf{H} \in \mathbb{H}^{N_\mathrm{R} \times N_\mathrm{R}}$ and $\mathbf{Z} \triangleq \mathbf{H}_\mathrm{B}\boldsymbol{\Theta}\mathbf{U}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F} + \mathbf{H}_\mathrm{D}\mathbf{V}_\mathrm{F} \in \mathbb{C}^{N_\mathrm{R} \times k}$. That is to say, $\mathbf{G}$ can be expressed as a Hermitian matrix plus k rank-1 perturbations. According to the Cauchy interlacing formula [57, Theorem 8.4.3], the n-th eigenvalue of $\mathbf{G}$ is bounded by

$$\lambda_n(\mathbf{G}) \leq \lambda_{n-k}(\mathbf{Y}), \qquad \text{if } n > k, \tag{49a}$$
$$\lambda_n(\mathbf{G}) \geq \lambda_n(\mathbf{Y}), \qquad \text{if } n < N - k + 1. \tag{49b}$$

Since $\mathbf{Y} = \mathbf{T}\mathbf{T}^\mathsf{H}$ is positive semi-definite, taking the square roots of (49a) and (49b) gives (8a) and (8b).

C. Proof of Proposition 3

Let $\mathbf{H}_\mathrm{B} = \mathbf{U}_\mathrm{B}\boldsymbol{\Sigma}_\mathrm{B}\mathbf{V}_\mathrm{B}^\mathsf{H}$ and $\mathbf{H}_\mathrm{F} = \mathbf{U}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H}$ be the SVD of the backward and forward channels, respectively. The scattering matrix of fully-connected BD-RIS can be decomposed as

$$\boldsymbol{\Theta} = \mathbf{V}_\mathrm{B}\mathbf{X}\mathbf{U}_\mathrm{F}^\mathsf{H}, \tag{50}$$

where $\mathbf{X} \in \mathbb{U}^{N_\mathrm{S} \times N_\mathrm{S}}$ is a unitary matrix to be designed. The equivalent channel is thus a function of $\mathbf{X}$

$$\mathbf{H} = \mathbf{H}_\mathrm{B}\boldsymbol{\Theta}\mathbf{H}_\mathrm{F} = \mathbf{U}_\mathrm{B}\boldsymbol{\Sigma}_\mathrm{B}\mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H}. \tag{51}$$

Since $\mathrm{sv}(\mathbf{U}\mathbf{A}\mathbf{V}^\mathsf{H}) = \mathrm{sv}(\mathbf{A})$ for unitary $\mathbf{U}$ and $\mathbf{V}$, we have

$$\begin{aligned}
\mathrm{sv}(\mathbf{H}) &= \mathrm{sv}(\mathbf{U}_\mathrm{B}\boldsymbol{\Sigma}_\mathrm{B}\mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H}) \\
&= \mathrm{sv}(\boldsymbol{\Sigma}_\mathrm{B}\mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}) \\
&= \mathrm{sv}(\bar{\mathbf{U}}_\mathrm{B}\boldsymbol{\Sigma}_\mathrm{B}\bar{\mathbf{V}}_\mathrm{B}^\mathsf{H}\bar{\mathbf{U}}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F}\bar{\mathbf{V}}_\mathrm{F}^\mathsf{H}) \\
&= \mathrm{sv}(\mathbf{B}\mathbf{F}),
\end{aligned} \tag{52}$$

where $\bar{\mathbf{U}}_\mathrm{B} \in \mathbb{U}^{N_\mathrm{R} \times N_\mathrm{R}}$, $\bar{\mathbf{V}}_\mathrm{B}, \bar{\mathbf{U}}_\mathrm{F} \in \mathbb{U}^{N_\mathrm{S} \times N_\mathrm{S}}$, and $\bar{\mathbf{V}}_\mathrm{F} \in \mathbb{U}^{N_\mathrm{T} \times N_\mathrm{T}}$ can be designed arbitrarily.

D. Proof of Corollary 3.2

(13a) follows from (12) when $r = k$. On the other hand, if we can prove

$$\prod_{n=1}^{\bar{N}} \sigma_n(\mathbf{H}) = \prod_{n=1}^{\bar{N}} \sigma_n(\mathbf{H}_\mathrm{B})\sigma_n(\mathbf{H}_\mathrm{F}), \tag{53}$$

then (13b) follows from (13a) and the non-negativity of singular values. To see (53), we start from a stricter result

$$\prod_{n=1}^{N_\mathrm{S}} \sigma_n(\mathbf{H}) = \prod_{n=1}^{N_\mathrm{S}} \sigma_n(\mathbf{H}_\mathrm{B})\sigma_n(\mathbf{H}_\mathrm{F}), \tag{54}$$

which is provable by cases. When $N_\mathrm{S} > N$, both sides of (54) become zero since $\sigma_n(\mathbf{H}) = \sigma_n(\mathbf{H}_\mathrm{B}) = \sigma_n(\mathbf{H}_\mathrm{F}) = 0$ for $n > N$. When $N_\mathrm{S} \leq N$, we have

$$\begin{aligned}
\prod_{n=1}^{N_\mathrm{S}} \sigma_n(\mathbf{H}) &= \prod_{n=1}^{N_\mathrm{S}} \sigma_n(\boldsymbol{\Sigma}_\mathrm{B}\mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}) \\
&= \prod_{n=1}^{N_\mathrm{S}} \sigma_n(\hat{\boldsymbol{\Sigma}}_\mathrm{B}\mathbf{X}\hat{\boldsymbol{\Sigma}}_\mathrm{F}) \\
&= \det(\hat{\boldsymbol{\Sigma}}_\mathrm{B}\mathbf{X}\hat{\boldsymbol{\Sigma}}_\mathrm{F}) \\
&= \det(\hat{\boldsymbol{\Sigma}}_\mathrm{B})\det(\mathbf{X})\det(\hat{\boldsymbol{\Sigma}}_\mathrm{F}) \\
&= \prod_{n=1}^{N_\mathrm{S}} \sigma_n(\boldsymbol{\Sigma}_\mathrm{B})\sigma_n(\boldsymbol{\Sigma}_\mathrm{F}),
\end{aligned}$$

where the first equality follows from (52) and $\hat{\boldsymbol{\Sigma}}_\mathrm{B}, \hat{\boldsymbol{\Sigma}}_\mathrm{F}$ truncate $\boldsymbol{\Sigma}_\mathrm{B}, \boldsymbol{\Sigma}_\mathrm{F}$ to square matrices of dimension N_S, respectively. It is evident that (54) implies (53) and thus (13b).

E. Proof of Corollary 3.3

In (14), the set of upper bounds

$$\{\sigma_n(\mathbf{H}) \leq \sigma_i(\mathbf{H}_\mathrm{B})\sigma_j(\mathbf{H}_\mathrm{F}) \,|\, [i,j,k] \in [N_\mathrm{S}]^3, i+j=n+1\} \tag{55}$$

is a special case of (12) with $(I,J,K) \in [N_\mathrm{S}]^3$. The minimum[9] of (55) is selected as the tightest upper bound in (14). On the other hand, the set of lower bounds

$$\{\sigma_n(\mathbf{H}) \geq \sigma_i(\mathbf{H}_\mathrm{B})\sigma_j(\mathbf{H}_\mathrm{F}) \,|\, [i,j,k] \in [N_\mathrm{S}]^3, i+j=n+N_\mathrm{S}\} \tag{56}$$

can be induced by (55), (54), and the non-negativity of singular values. The maximum of (56) is selected as the tightest lower bound in (14). Interested readers are also referred to [61, (2.0.3)].

To attain the upper bound, the BD-RIS needs to maximize the minimum of the first n channel singular values. It follows from (15a) that

$$\begin{aligned}
\mathrm{sv}(\mathbf{H}) &= \mathrm{sv}(\mathbf{H}_\mathrm{B}\mathbf{V}_\mathrm{B}\mathbf{P}\mathbf{U}_\mathrm{F}^\mathsf{H}\mathbf{H}_\mathrm{F}) \\
&= \mathrm{sv}(\mathbf{U}_\mathrm{B}\boldsymbol{\Sigma}_\mathrm{B}\mathbf{V}_\mathrm{B}^\mathsf{H}\mathbf{V}_\mathrm{B}\mathbf{P}\mathbf{U}_\mathrm{F}^\mathsf{H}\mathbf{U}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F}\mathbf{U}_\mathrm{F}^\mathsf{H}) \\
&= \mathrm{sv}(\boldsymbol{\Sigma}_\mathrm{B}\mathbf{P}\boldsymbol{\Sigma}_\mathrm{F}).
\end{aligned}$$

On the one hand, $\mathbf{P}_{ij} = 1$ with (i,j) satisfying (16a) ensures $\min_{i+j=n+1}\sigma_i(\mathbf{H}_\mathrm{B})\sigma_j(\mathbf{H}_\mathrm{F})$ is a singular value of $\mathbf{H}$. It is actually among the first n since the number of pairs (i',j') not majorized by (i,j) is $n-1$. On the other hand, (17a) ensures the first $(n-1)$-th singular values are no smaller than $\min_{i+j=n+1}\sigma_i(\mathbf{H}_\mathrm{B})\sigma_j(\mathbf{H}_\mathrm{F})$. Combining both facts, we claim the upper bound $\sigma_n(\mathbf{H}) = \min_{i+j=n+1}\sigma_i(\mathbf{H}_\mathrm{B})\sigma_j(\mathbf{H}_\mathrm{F})$ is attainable by (15a). The attainability of the lower bound can be proved similarly and the details are omitted.

F. Proof of Corollary 3.4

From (50) and (51) we have

$$\begin{aligned}
\|\mathbf{H}\|_\mathrm{F}^2 &= \mathrm{tr}\big(\mathbf{V}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F}^\mathsf{H}\mathbf{X}^\mathsf{H}\boldsymbol{\Sigma}_\mathrm{B}^\mathsf{H}\mathbf{U}_\mathrm{B}^\mathsf{H}\mathbf{U}_\mathrm{B}\boldsymbol{\Sigma}_\mathrm{B}\mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H}\big) \\
&= \mathrm{tr}\big(\boldsymbol{\Sigma}_\mathrm{B}^\mathsf{H}\boldsymbol{\Sigma}_\mathrm{B} \cdot \mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F}^\mathsf{H}\mathbf{X}^\mathsf{H}\big) \\
&\triangleq \mathrm{tr}\big(\tilde{\mathbf{B}}\tilde{\mathbf{F}}\big),
\end{aligned} \tag{57}$$

where $\mathbf{X} \triangleq \mathbf{V}_\mathrm{B}^\mathsf{H}\boldsymbol{\Theta}\mathbf{U}_\mathrm{F} \in \mathbb{U}^{N_\mathrm{S} \times N_\mathrm{S}}$, $\tilde{\mathbf{B}} \triangleq \boldsymbol{\Sigma}_\mathrm{B}^\mathsf{H}\boldsymbol{\Sigma}_\mathrm{B} \in \mathbb{H}_+^{N_\mathrm{S} \times N_\mathrm{S}}$, and $\tilde{\mathbf{F}} \triangleq \mathbf{X}\boldsymbol{\Sigma}_\mathrm{F}\boldsymbol{\Sigma}_\mathrm{F}^\mathsf{H}\mathbf{X}^\mathsf{H} \in \mathbb{H}_+^{N_\mathrm{S} \times N_\mathrm{S}}$. By Ruhe's trace inequality for positive semi-definite matrices [62, (H.1.g) and (H.1.h)],

$$\sum_{n=1}^{N} \lambda_n(\tilde{\mathbf{B}})\lambda_{N_\mathrm{S}-n+1}(\tilde{\mathbf{F}}) \leq \mathrm{tr}(\tilde{\mathbf{B}}\tilde{\mathbf{F}}) \leq \sum_{n=1}^{N} \lambda_n(\tilde{\mathbf{B}})\lambda_n(\tilde{\mathbf{F}}),$$

which simplifies to (19). The upper bound is attained when $\mathbf{X}$ is chosen to match the singular values of $\tilde{\mathbf{F}}$ to those of $\tilde{\mathbf{B}}$ in similar order. Apparently this occurs at $\mathbf{X} = \mathbf{I}$ and $\boldsymbol{\Theta} = \mathbf{V}_\mathrm{B}\mathbf{U}_\mathrm{F}^\mathsf{H}$. On the other hand, the lower bound is attained when the singular values of $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{B}}$ are matched in reverse order, namely $\mathbf{X} = \mathbf{J}$ and $\boldsymbol{\Theta} = \mathbf{V}_\mathrm{B}\mathbf{J}\mathbf{U}_\mathrm{F}^\mathsf{H}$.

G. Proof of Corollary 3.6

When perfect Channel State Information (CSI) is available at the transmitter, in the low-SNR regime, the capacity is achieved by dominant eigenmode transmission [46, (5.26)]

$$\begin{aligned}
C_{\rho_\downarrow} &= \log\big(1 + \rho\lambda_1(\mathbf{H}^\mathsf{H}\mathbf{H})\big) \\
&= \log\big(1 + \rho\sigma_1^2(\mathbf{H})\big) \\
&\approx \rho\sigma_1^2(\mathbf{H}) \\
&\leq \rho\sigma_1^2(\mathbf{H}_\mathrm{B})\sigma_1^2(\mathbf{H}_\mathrm{F}),
\end{aligned}$$

where the approximation is $\log(1+x) \approx x$ for small x and the inequality follows from (13a) with $k = 1$. In the high-SNR

---

[9]One may think to take the maximum of those upper bounds as the problem of interest is the attainable dynamic range of n-th singular value. This is infeasible since the singular values will be reordered.

regime, the capacity is achieved by multiple eigenmode transmission with uniform power location [46, (5.27)]

$$
\begin{aligned}
C_{\rho_\uparrow} &= \sum\nolimits_{n=1}^{N} \log\left(1 + \frac{\rho}{N}\lambda_n(\mathbf{H}^{\mathsf{H}}\mathbf{H})\right) \\
&\approx \sum\nolimits_{n=1}^{N} \log\left(\frac{\rho}{N}\sigma_n^2(\mathbf{H})\right) \\
&= N\log\frac{\rho}{N} + \sum\nolimits_{n=1}^{N} \log\sigma_n^2(\mathbf{H}) \\
&= N\log\frac{\rho}{N} + \log\prod\nolimits_{n=1}^{N}\sigma_n^2(\mathbf{H}) \\
&\leq N\log\frac{\rho}{N} + 2\log\prod\nolimits_{n=1}^{N}\sigma_n(\mathbf{H}_{\mathrm{B}})\sigma_n(\mathbf{H}_{\mathrm{F}}),
\end{aligned}
$$

where the approximation is $\log(1+x) \approx \log(x)$ for large x and the inequality follows from (13a) with $k=N$.

We now show (23) can achieve the upper bounds in (24a) and (24b) simultaneously. On the one hand, (23) is a special case of (15a) with $\mathbf{P}=\mathbf{I}$, which satisfies (16a) and (17a) for $n=1$ and thus attain $\sigma_1(\mathbf{H}) = \sigma_1(\mathbf{H}_{\mathrm{B}})\sigma_1(\mathbf{H}_{\mathrm{F}})$. On the other hand, since $\log(\cdot)$ is a monotonic function, we can prove similar to Appendix F that $\sum_{n=1}^{N}\log\sigma_n^2(\mathbf{H}) \leq \sum_{n=1}^{N}\log\sigma_n^2(\mathbf{H}_{\mathrm{B}})\sigma_n^2(\mathbf{H}_{\mathrm{F}})$ and the bound is tight at (23). The proof is complete.

H. Proof of Proposition 4

A straightforward extension to [63, Theorem 2] shows that the Clarke subdifferential of a locally Lipschitz function of singular values of a matrix with respect to the matrix itself is given by

$$
\partial_{\mathbf{H}^*} f\big(\mathrm{sv}(\mathbf{H})\big) = \mathrm{conv}\{\mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}\}, \tag{58}
$$

where $\mathbf{D} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ is a rectangular diagonal matrix with $[\mathbf{D}]_{n,n} \in \partial_{\sigma_n(\mathbf{H})} f\big(\mathrm{sv}(\mathbf{H})\big)$, $\forall n\in[N]$, and $\mathbf{U}$, $\mathbf{V}$ are any left and right singular matrices of $\mathbf{H}$. Therefore,

$$
\begin{aligned}
\partial f\big(\mathrm{sv}(\mathbf{H})\big) \ni \mathrm{tr}\big(\mathbf{V}^*\mathbf{D}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\partial\mathbf{H}^*\big) \\
= \mathrm{tr}\big(\mathbf{V}^*\mathbf{D}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\mathbf{H}_{\mathrm{B},g}^*\partial\mathbf{\Theta}_g^*\mathbf{H}_{\mathrm{F},g}^*\big) \\
= \mathrm{tr}\big(\mathbf{H}_{\mathrm{F},g}^*\mathbf{V}^*\mathbf{D}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\mathbf{H}_{\mathrm{B},g}^*\partial\mathbf{\Theta}_g^*\big),
\end{aligned}
$$

such that $\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}$ constitutes a Clarke subgradient of $f\big(\mathrm{sv}(\mathbf{H})\big)$ with respect to $\mathbf{\Theta}_g$. The convex hull of those subgradients is the subdifferential (26).

I. Proof of Lemma 1

The differential of R with respect to $\mathbf{\Theta}_g^*$ is [64]

$$
\begin{aligned}
\partial R &= \frac{1}{\eta}\mathrm{tr}\left\{\partial\mathbf{H}^*\cdot\mathbf{Q}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\left(\mathbf{I}+\frac{\mathbf{H}^*\mathbf{Q}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}}{\eta}\right)^{-1}\right\} \\
&= \frac{1}{\eta}\mathrm{tr}\left\{\mathbf{H}_{\mathrm{B},g}^*\cdot\partial\mathbf{\Theta}_g^*\cdot\mathbf{H}_{\mathrm{F},g}^*\mathbf{Q}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\left(\mathbf{I}+\frac{\mathbf{H}^*\mathbf{Q}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}}{\eta}\right)^{-1}\right\} \\
&= \frac{1}{\eta}\mathrm{tr}\left\{\mathbf{H}_{\mathrm{F},g}^*\mathbf{Q}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\left(\mathbf{I}+\frac{\mathbf{H}^*\mathbf{Q}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}}{\eta}\right)^{-1}\mathbf{H}_{\mathrm{B},g}^*\cdot\partial\mathbf{\Theta}_g^*\right\},
\end{aligned}
$$

and the corresponding complex derivative is (37).

J. Proof of Proposition 5

The differential of (39a) with respect to $\mathbf{\Theta}_g^*$ is

$$
\begin{aligned}
\partial\|\mathbf{H}\|_{\mathrm{F}}^2 &= \mathrm{tr}\big(\mathbf{H}_{\mathrm{B},g}^*\cdot\partial\mathbf{\Theta}_g^*\cdot\mathbf{H}_{\mathrm{F},g}^*(\mathbf{H}_{\mathrm{D}}^{\mathsf{T}}+\mathbf{H}_{\mathrm{F}}^{\mathsf{T}}\mathbf{\Theta}^{\mathsf{T}}\mathbf{H}_{\mathrm{B}}^{\mathsf{T}})\big) \\
&= \mathrm{tr}\big(\mathbf{H}_{\mathrm{F},g}^*(\mathbf{H}_{\mathrm{D}}^{\mathsf{T}}+\mathbf{H}_{\mathrm{F}}^{\mathsf{T}}\mathbf{\Theta}^{\mathsf{T}}\mathbf{H}_{\mathrm{B}}^{\mathsf{T}})\mathbf{H}_{\mathrm{B},g}^*\cdot\partial\mathbf{\Theta}_g^*\big)
\end{aligned}
$$

and the corresponding complex derivative is

$$
\frac{\partial\|\mathbf{H}\|_{\mathrm{F}}^2}{\partial\mathbf{\Theta}_g^*} = \mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}}+\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} \triangleq \mathbf{M}_g, \tag{59}
$$

whose SVD is denoted as $\mathbf{M}_g = \mathbf{U}_g\mathbf{\Sigma}_g\mathbf{V}_g^{\mathsf{H}}$. The quadratic objective (39a) can be successively approximated by its first-order Taylor expansion, resulting in the subproblem

$$
\max_{\mathbf{\Theta}} \quad \sum_g 2\Re\{\mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\} \tag{60a}
$$

$$
\text{s.t.} \quad \mathbf{\Theta}_g^{\mathsf{H}}\mathbf{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{60b}
$$

whose optimal solution is

$$
\tilde{\mathbf{\Theta}}_g = \mathbf{U}_g\mathbf{V}_g^{\mathsf{H}}, \quad \forall g. \tag{61}
$$

This is because $\Re\{\mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\} = \Re\{\mathrm{tr}(\mathbf{\Sigma}_g\mathbf{V}_g^{\mathsf{H}}\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{U}_g)\} \leq \mathrm{tr}(\mathbf{\Sigma}_g)$ and the bound is tight when $\mathbf{V}_g^{\mathsf{H}}\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{U}_g = \mathbf{I}$.

Next, we prove that solving the affine approximation (60) by (61) does not decrease (39a). Since $\tilde{\mathbf{\Theta}} = \mathrm{diagm}(\tilde{\mathbf{\Theta}}_1,...,\tilde{\mathbf{\Theta}}_G)$ is optimal for (60), we have

$$
\begin{aligned}
2\Re\Big\{&\sum_g \mathrm{tr}(\tilde{\mathbf{\Theta}}_g^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}) \\
&+ \sum_{g_1,g_2} \mathrm{tr}(\tilde{\mathbf{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}})\Big\} \\
\geq 2\Re\Big\{&\sum_g \mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}) \\
&+ \sum_{g_1,g_2} \mathrm{tr}(\mathbf{\Theta}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}})\Big\}.
\end{aligned} \tag{62a}
$$

Besides, $\|\sum_g\mathbf{H}_{\mathrm{B},g}\tilde{\mathbf{\Theta}}_g\mathbf{H}_{\mathrm{F},g} - \sum_g\mathbf{H}_{\mathrm{B},g}\mathbf{\Theta}_g\mathbf{H}_{\mathrm{F},g}\|_{\mathrm{F}}^2 \geq 0$ implies

$$
\begin{aligned}
\sum_{g_1,g_2}&\mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\tilde{\mathbf{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\tilde{\mathbf{\Theta}}_{g_2}\mathbf{H}_{\mathrm{F},g_2}) \\
&+ \sum_{g_1,g_2}\mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\mathbf{\Theta}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}) \\
\geq 2\Re\Big\{&\sum_{g_1,g_2}\mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\tilde{\mathbf{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2})\Big\}.
\end{aligned} \tag{62b}
$$

Adding (62a) and (62b), we have

$$
\begin{aligned}
2\Re\{&\mathrm{tr}(\tilde{\mathbf{\Theta}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}}^{\mathsf{H}})\} + \mathrm{tr}(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\tilde{\mathbf{\Theta}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}\tilde{\mathbf{\Theta}}\mathbf{H}_{\mathrm{F}}) \\
\geq 2\Re\{&\mathrm{tr}(\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}}^{\mathsf{H}})\} + \mathrm{tr}(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}}),
\end{aligned} \tag{63}
$$

which suggests that (39a) is non-decreasing as the solution iterates over (61). Since (39a) is also bounded from above, the sequence of objective value converges.

Finally, we prove that any solution when (40) converges, denoted by $\mathbf{\Theta}'$, is a stationary point of (39). The Karush-Kuhn-Tucker (KKT) conditions of (39) and (60) are equivalent in terms of primal/dual feasibility and complementary slackness, while the stationary conditions are respectively, $\forall g$,

$$
\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}}+\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}^\star\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} - \mathbf{\Theta}_g^\star\mathbf{\Lambda}_g^{\mathsf{H}} = 0, \tag{64a}
$$

$$
\mathbf{M}_g - \mathbf{\Theta}_g^\star\mathbf{\Lambda}_g^{\mathsf{H}} = 0. \tag{64b}
$$

When (40) converges, $\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\mathbf{\Theta}'\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} = \mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}}+\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}^\star\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}$ and (64b) reduces to (64a). The proof is thus completed.

REFERENCES

[1] E. Basar, M. D. Renzo, J. D. Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[2] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 5394–5409, Nov 2019.

[3] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Transactions on Wireless Communications*, vol. 19, pp. 3064–3076, May 2020.

[4] Y. Liu, Y. Zhang, X. Zhao, S. Geng, P. Qin, and Z. Zhou, "Dynamic-controlled RIS assisted multi-user MISO downlink system: Joint beamforming design," *IEEE Transactions on Green Communications and Networking*, vol. 6, pp. 1069–1081, Jun 2022.

[5] Y. He, Y. Cai, H. Mao, and G. Yu, "RIS-assisted communication radar coexistence: Joint beamforming design and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 40, pp. 2131–2145, Jul 2022.

[6] H. Luo, R. Liu, M. Li, Y. Liu, and Q. Liu, "Joint beamforming design for RIS-assisted integrated sensing and communication systems," *IEEE Transactions on Vehicular Technology*, vol. 71, pp. 13 393–13 397, Dec 2022.

[7] M. Hua, Q. Wu, C. He, S. Ma, and W. Chen, "Joint active and passive beamforming design for IRS-aided radar-communication," *IEEE Transactions on Wireless Communications*, vol. 22, pp. 2278–2294, Apr 2023.

[8] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1735–1748, Aug 2020.

[9] Z. Feng, B. Clerckx, and Y. Zhao, "Waveform and beamforming design for intelligent reflecting surface aided wireless power transfer: Single-user and multi-user solutions," *IEEE Transactions on Wireless Communications*, 2022.

[10] Y. Zhao, B. Clerckx, and Z. Feng, "IRS-aided SWIPT: Joint waveform, active and passive beamforming design under nonlinear harvester model," *IEEE Transactions on Communications*, vol. 70, pp. 1345–1359, 2022.

[11] R. Karasik, O. Simeone, M. D. Renzo, and S. S. Shitz, "Beyond max-SNR: Joint encoding for reconfigurable intelligent surfaces," in *2020 IEEE International Symposium on Information Theory (ISIT)*, Jun 2020, pp. 2965–2970.

[12] J. Ye, S. Guo, S. Dang, B. Shihada, and M.-S. Alouini, "On the capacity of reconfigurable intelligent surface assisted MIMO symbiotic communications," *IEEE Transactions on Wireless Communications*, vol. 21, pp. 1943–1959, Mar 2022.

[13] Y.-C. Liang, Q. Zhang, E. G. Larsson, and G. Y. Li, "Symbiotic radio: Cognitive backscattering communications for future wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, pp. 1242–1255, Dec 2020.

[14] Y. Zhao and B. Clerckx, "RIScatter: Unifying backscatter communication and reconfigurable intelligent surface," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, Dec 2024.

[15] E. Basar, "Reconfigurable intelligent surfaces for doppler effect and multipath fading mitigation," *Frontiers in Communications and Networks*, vol. 2, May 2021.

[16] E. Arslan, I. Yildirim, F. Kilinc, and E. Basar, "Over-the-air equalization with reconfigurable intelligent surfaces," *IET Communications*, vol. 16, pp. 1486–1497, Aug 2022.

[17] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Using intelligent reflecting surfaces for rank improvement in MIMO communications," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 9160–9164.

[18] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Transactions on Communications*, vol. 68, pp. 4522–4535, Jul 2020.

[19] G. Chen and Q. Wu, "Fundamental limits of intelligent reflecting surface aided multiuser broadcast channel," *IEEE Transactions on Communications*, vol. 71, pp. 5904–5919, Oct 2023.

[20] M. A. ElMossallamy, H. Zhang, R. Sultan, K. G. Seddik, L. Song, G. Y. Li, and Z. Han, "On spatial multiplexing using reconfigurable intelligent surfaces," *IEEE Wireless Communications Letters*, vol. 10, pp. 226–230, Feb 2021.

[21] S. Meng, W. Tang, W. Chen, J. Lan, Q. Y. Zhou, Y. Han, X. Li, and S. Jin, "Rank optimization for MIMO channel with RIS: Simulation and measurement," *IEEE Wireless Communications Letters*, vol. 13, pp. 437–441, Feb 2024.

[22] W. Huang, B. Lei, S. He, C. Kai, and C. Li, "Condition number improvement of IRS-aided near-field MIMO channels," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2023, pp. 1210–1215.

[23] S. H. Chae and K. Lee, "Cooperative communication for the rank-deficient MIMO interference channel with a reconfigurable intelligent surface," *IEEE Transactions on Wireless Communications*, vol. 22, pp. 2099–2112, Mar 2023.

[24] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Transactions on Wireless Communications*, vol. 21, pp. 1229–1243, Feb 2022.

[25] M. Nerini, S. Shen, and B. Clerckx, "Closed-form global optimization of beyond diagonal reconfigurable intelligent surfaces," *IEEE Transactions on Wireless Communications*, vol. 23, pp. 1037–1051, Feb 2024.

[26] M. Nerini, S. Shen, H. Li, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces utilizing graph theory: Modeling, architecture design, and optimization," *IEEE Transactions on Wireless Communications*, pp. 1–1, May 2024.

[27] I. Santamaria, M. Soleymani, E. Jorswieck, and J. Gutiérrez, "SNR maximization in beyond diagonal RIS-assisted single and multiple antenna links," *IEEE Signal Processing Letters*, vol. 30, pp. 923–926, 2023.

[28] ——, "Interference leakage minimization in RIS-assisted MIMO interference channels," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 39, Jun 2023, pp. 1–5.

[29] H.-R. Ahn, *Asymmetric Passive Components in Microwave Integrated Circuits*. Hoboken, NJ, USA: Wiley, 2006.

[30] H. Li, S. Shen, Y. Zhang, and B. Clerckx, "Channel estimation and beamforming for beyond diagonal reconfigurable intelligent surfaces," *IEEE Transactions on Signal Processing*, vol. 72, pp. 3318–3332, Jan 2024.

[31] H. Li, S. Shen, M. Nerini, M. D. Renzo, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces with mutual coupling: Modeling and optimization," *IEEE Communications Letters*, pp. 1–1, Oct 2024.

[32] H. Li, M. Nerini, S. Shen, and B. Clerckx, "Wideband modeling and beamforming for beyond diagonal reconfigurable intelligent surfaces," in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun 2024, pp. 926–930.

[33] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: A multi-sector mode enabling highly directional full-space wireless coverage," *IEEE Journal on Selected Areas in Communications*, vol. 41, pp. 2446–2460, Aug 2023.

[34] J. Tapie, M. Nerini, B. Clerckx, and P. del Hougne, "Beyond-Diagonal RIS Prototype and Performance Evaluation," *arXiv [eess.SP]*, 19 May 2025.

[35] T. Fang and Y. Mao, "A low-complexity beamforming design for beyond-diagonal RIS aided multi-user networks," *IEEE Communications Letters*, pp. 1–1, Jul 2023.

[36] Y. Zhou, Y. Liu, H. Li, Q. Wu, S. Shen, and B. Clerckx, "Optimizing power consumption, energy efficiency and sum-rate using beyond diagonal RIS — a unified approach," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[37] G. Bartoli, A. Abrardo, N. Decarli, D. Dardari, and M. D. Renzo, "Spatial multiplexing in near field MIMO channels with reconfigurable intelligent surfaces," *IET Signal Processing*, vol. 17, Mar 2023.

[38] D. Semmler, M. Joham, and W. Utschick, "High SNR analysis of RIS-aided MIMO broadcast channels," in *2023 IEEE 24th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Sep 2023, pp. 221–225.

[39] L. Hogben, Ed., *Handbook of Linear Algebra*. Boca Raton, FL, USA: CRC press, 2013.

[40] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, UK: Cambridge University Press, Jun 1994.

[41] W. Fulton, "Eigenvalues, invariant factors, highest weights, and schubert calculus," *Bulletin of the American Mathematical Society*, vol. 37, pp. 209–249, Apr 2000.

[42] R. Bhatia, "Linear algebra to quantum cohomology: The story of alfred horn's inequalities," *The American Mathematical Monthly*, vol. 108, pp. 289–318, Apr 2001.

[43] S. Shen and B. Clerckx, "Beamforming optimization for MIMO wireless power transfer with nonlinear energy harvesting: RF combining versus DC combining," *IEEE Transactions on Wireless Communications*, vol. 20, pp. 199–213, Jan 2021.

[44] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6068–6077, Dec 2009.

[45] A. Zanella, M. Chiani, and M. Win, "On the marginal distribution of the eigenvalues of wishart matrices," *IEEE Transactions on Communications*, vol. 57, pp. 1050–1060, Apr 2009.

[46] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*. Waltham, MA, USA: Academic Press, 2013.

[47] T. E. Abrudan, J. Eriksson, and V. Koivunen, "Steepest descent algorithms for optimization under unitary matrix constraint," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1134–1147, Mar 2008.

[48] T. Abrudan, J. Eriksson, and V. Koivunen, "Conjugate gradient algorithm for optimization under unitary matrix constraint," *Signal Processing*, vol. 89, pp. 1704–1714, Sep 2009.

[49] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific journal of Optimization*, vol. 2, 2006.

[50] L. Armijo, "Minimization of functions having lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, pp. 1–3, Jan 1966.

[51] C. Moler and C. V. Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," *SIAM Review*, vol. 45, pp. 3–49, Jan 2003.

[52] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, pp. 1728–1767, Jun 2022.

[53] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problem on the Stiefel manifold," *Science China Information Sciences*, vol. 60, p. 112101, Nov 2017.

[54] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, pp. 635–650, Mar 2002.

[55] T. Viklands, "Algorithms for the weighted orthogonal procrustes problem and other least squares problems," PhD dissertation, Datavetenskap, Umeå University, 2006.

[56] T. Bell, "Global positioning system-based attitude determination and the orthogonal procrustes problem," *Journal of Guidance, Control, and Dynamics*, vol. 26, pp. 820–822, Sep 2003.

[57] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins University Press, 2013.

[58] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, no. 42, pp. 1455–1459, 2014. [Online]. Available: https://www.manopt.org

[59] K. D. Ikramov, "Takagi's decomposition of a symmetric unitary matrix as a finite algorithm," *Computational Mathematics and Mathematical Physics*, vol. 52, pp. 1–3, Jan 2012.

[60] J. An, C. Xu, D. W. K. Ng, G. C. Alexandropoulos, C. Huang, C. Yuen, and L. Hanzo, "Stacked Intelligent Metasurfaces for Efficient Holographic MIMO Communications in 6G," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2380–2396, Aug. 2023.

[61] F. Zhang, Ed., *The Schur Complement and Its Applications*, ser. Numerical Methods and Algorithms. New York, NY, USA: Springer, Apr 2005.

[62] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., ser. Springer Series in Statistics. New York, NY, USA: Springer, Dec 2010.

[63] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and its Applications*, vol. 170, no. 1, pp. 33–45, 1992.

[64] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2740–2746, Jun 2007.