Channel Shaping Using Beyond Diagonal Reconfigurable Intelligent Surfaces

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

***Abstract*—This paper investigates to what extent a passive Reconfigurable Intelligent Surface (RIS) can redistribute the singular values of a Multiple-Input Multiple-Output (MIMO) point-to-point channel. We depart from the widely-adopted diagonal phase shift model to a Beyond-Diagonal (BD)-RIS architecture featuring in-group connections between elements, which enables cooperative wave scattering with amplitude and phase control. Specifically, we first provide unique shaping insights by characterizing the Pareto frontiers of channel singular values via a novel geodesic manifold optimization. The resulting region comprehensively encapsulates most relevant metrics (e.g., spectral norm and condition number). To explore the shaping limits of passive BD-RIS, we also derive individual and collective singular value bounds respectively in rank-deficient channels and for fully-connected BD-RIS via matrix analysis. As a side product, we tackle BD-RIS-aided MIMO achievable rate maximization by a local-optimal Alternating Optimization (AO) approach and a low-complexity shaping-inspired approach. Simulation results highlight that BD-RIS significantly improves the dynamic range of and trade-off between channel singular values. Results also show the power and rate gain of BD-RIS over diagonal RIS increase with MIMO dimensions.**

***Index Terms*—Beyond diagonal reconfigurable intelligent surface, multi-input multi-output, manifold optimization, channel shaping, rate maximization.**

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 the recent advances in Reconfigurable Intelligent Surface (RIS), a programmable passive planar surface that recycles and redistributes ambient electromagnetic waves for improved wireless performance. A typical RIS consists of numerous low-power sub-wavelength non-resonant 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 low-noise, full-duplex operation, and also features better flexibility than reflectarrays, lighter footprint than various relays, and greater scalability than conventional Multiple-Input Multiple-Output (MIMO) systems. The most popular RIS research direction is *joint passive and active beamforming* design with transceivers for a specific performance measure, which has attracted significant attention in wireless communication [2]–[4], sensing [5]–[7], and power transfer literature [8]–[10]. While passive beamforming at RIS suffers severe attenuation from double fading, it offers better asymptotic behaviors than active beamforming at transceivers (e.g., second-order array gain and fourth-order harvested power [10]). Another feature is that RIS can also be used for *backscatter modulation* by periodically switching its reflection pattern within the channel coherence time. This creates a free-ride message stream (similar to index modulation [11]) with dual benefits: integrating with the legacy transmitter for enhanced channel capacity [12]–[14] or serving as a dedicated source for low-power uplink communication [15]–[17]. Different from above directions, *channel shaping* exploits the RIS as a stand-alone device to modify the inherent properties of the wireless environment. To name a few, it can compensate for the Doppler effect [18], transform frequency-selective channels into frequency-flat [19], improve the spatial diversity for MIMO systems [20], and create artificial time diversity for orthogonal [21] and non-orthogonal [22] multiple accesses. Channel shaping also provides a ubiquitous benchmark for different wireless applications and helps to decouple the RIS-transceiver design in a shape-then-transmit manner. At a specific time and frequency, channel metrics can be roughly classified into two categories:

- *Singular value:* Relevant metrics are in general closely related to the performance measures (e.g., achievable rate and harvested power [23]) but sensitive to numerical perturbations of the channel matrix. The impact of RIS has been studied in terms of minimum singular value [24], effective rank [24], [25], condition number [26], [27], and degree of freedom [28]–[30].
- *Power:* Second-order statistics are less informative in MIMO but easier to analyze and optimize. The impact of RIS has been studied in terms of channel power gain [2], [31]–[34] and leakage interference [35].

Although those works offer initial glimpses into the channel shaping potential, one critical question has not been fully addressed: *What is the ultimate channel shaping capability of a passive RIS in terms of singular values?* The answer depends heavily on the hardware architecture of RIS and its scattering model. Most relevant works [2], [24]–[30], [35] considered a diagonal RIS model where each element is connected to a dedicated impedance and is disconnected from others, such that wave impinging on one element is entirely reflected by the same element. This simple architecture mathematically translates to a diagonal scattering matrix with unit-magnitude entries on the main diagonal and zeros elsewhere, applying merely a phase shift to the incoming signal. The concept was

Yang Zhao, Hongyu Li, and Bruno Clerckx are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: {yang.zhao18, c.li21, b.clerckx}@imperial.ac.uk).

Massimo Franceschetti is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla CA 92093, USA (e-mail: massimo@ece.ucsd.edu).

soon generalized to Beyond-Diagonal (BD)-RIS with a group-connected architecture [31] that groups adjacent elements and connects them via passive reconfigurable components.[1] This allows wave impinging on one element to propagate within the circuit and depart partially from any element in the same group. It can thus redistribute both amplitude and phase of the scattered wave with zero power loss, generalizing the scattering matrix to block-diagonal with unitary blocks. Such a powerful model can be realized at reduced hardware cost using tree- and forest-connected architectures inspired by graph theory [33]. BD-RIS can also function in hybrid transmitting-and-reflecting mode [37] and multi-sector mode [38] to provide full-space coverage and multi-user support. Many practical design challenges of BD-RIS have also been addressed including channel estimation [39], mutual coupling [40], and wideband modelling [41]. From the beamforming perspective, the superiority of BD-RIS over diagonal RIS has been studied extensively in Single-Input Single-Output (SISO) and Multiple-Input Single-Output (MISO) systems, where the problems were cast as single-user Signal-to-Noise Ratio (SNR) maximization [31]–[34] and multi-user Weighted Sum-Rate (WSR) maximization [38], [42]–[44]. However, the interplay between BD-RIS and MIMO systems is still in the infancy stage. The authors of [45] investigated the rate-optimal joint beamforming design for a specific BD-RIS-aided MIMO system with blocked direct link and fully-connected RIS. Similar constraints were also adopted in [46], which introduces a transmitter-side BD-RIS to massive MIMO that exploits statistical Channel State Information (CSI) for improved spectral efficiency. Received power maximization problem was studied over continuous-valued [32] and discrete-valued [47] BD-RIS, but the proposed single-stream transceiver is rate-suboptimal and the passive beamforming problem is equivalent to SISO. Although the results are promising, the channel shaping capability of BD-RIS deserves further investigation. Moreover, no previous work has attempted to characterize the singular value region of a MIMO channel manipulated by any type of RIS, and the BD-RIS-aided MIMO rate maximization problem remains unsolved. This paper aims for a comprehensive answer to the channel shaping capability of BD-RIS through theoretical analysis and numerical optimization. The contributions are summarized below.

First, we interpret the channel shaping potential of BD-RIS in terms of channel rearrangement and space alignment. Channel rearrangement refers to rearranging and recombining the forward and backward channel branches (i.e., entries of the channel matrix) according to their strength. Space alignment generalizes phase matching in SISO and MISO to the high-dimensional singular vector space in MIMO. The off-diagonal entries of BD-RIS scattering matrix provides a higher design freedom and exploits channel spatial diversity more effectively than diagonal RIS. This is the first paper to study BD-RIS in general MIMO systems.

Second, we propose an efficient BD-RIS design framework that solves general optimization problems by geodesic[2] Riemannian Conjugate Gradient (RCG). This method modified from [48], [49] exploits the Riemannian geometry of the Stiefel manifold and performs group-wise multiplicative rotational updates along the geodesics, which avoids retraction and facilitates step size selection. It thus converges faster than existing non-geodesic approaches [50], [51]. This is the first work to tailor an efficient optimization framework for BD-RIS.

Third, we adopt the proposed geodesic RCG and provide a numerical answer to the singular value shaping question. The Pareto frontiers of channel singular values are characterized by solving a series of weighted sum maximization problems, where the weight associated with each singular value can be positive, zero, or negative. The enclosed region generalizes most relevant metrics and provides an intuitive channel shaping benchmark. Results show that increasing BD-RIS group size enlarges the singular value region, improving both dynamic range of and trade-off between singular values.

Fourth, we leverage matrix analysis tools and provide an analytical answer to the singular value shaping question. Asymptotic channel singular value bounds shared by diagonal and BD-RIS are derived for rank-deficient forward/backward channel. On the other hand, attainable bounds in the presence of fully-connected RIS are derived for blocked direct channel. Results validate those bounds and emphasizes the shaping advantage of BD-RIS. This is the first work to quantify the singular value redistribution capability from either numerical or analytical perspective.

Fifth, we tackle BD-RIS-aided MIMO rate maximization problem by two beamforming solutions: a local-optimal Alternating Optimization (AO) approach and a low-complexity approach inspired by channel shaping insights. The former updates active beamforming by eigenmode transmission and passive beamforming by geodesic RCG until convergence. The latter suboptimally decouples the problem as channel power gain maximization at RIS and conventional MIMO transmission at transceiver, then solves both in closed form. Results show that BD-RIS provides higher power gain and achievable rate than diagonal RIS, and the rate gap between two approaches diminishes as the RIS evolves from single-connected to fully-connected. It suggests channel shaping offers a promising path to decouple joint RIS-transceiver designs.

Notation: Italic, bold lower-case, and bold upper-case letters indicate scalars, vectors and matrices, respectively. $\jmath$ denotes the imaginary unit. $\mathbb{C}$ represents the set of complex numbers. $\mathbb{H}^{n\times n}$ and $\mathbb{U}^{n\times n}$ denotes the set of $n\times n$ Hermitian and unitary matrices, respectively. $\mathbf{0}$ and $\mathbf{I}$ are the all-zero and identity matrices with appropriate size, respectively. $\Re\{\cdot\}$ takes the real part of a complex number. $\arg(\cdot)$ gives the argument of a complex number. $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ evaluates the trace and determinant of a square matrix, respectively. $\mathrm{diag}(\cdot)$ constructs a square matrix with arguments on the main (block) diagonal and zeros elsewhere. $\mathrm{sv}(\cdot)$ returns the singular value vector. $\sigma_n(\cdot)$ and $\lambda_n(\cdot)$ is the n-th largest singular value and eigenvalue, respectively. $(\cdot)^*$, $(\cdot)^\mathsf{T}$, $(\cdot)^\mathsf{H}$,

[1]Those components can be either symmetric (e.g., capacitors and inductors) or asymmetric (e.g., ring hybrids and branch-line hybrids) [36], resulting in symmetric and asymmetric scattering matrices, respectively.

[2]A geodesic refers to the shortest path between two points in a Riemannian manifold.

$(\cdot)^{\dagger}$ $(\cdot)^{(r)}$, $(\cdot)^{\star}$ denote the conjugate, transpose, conjugate transpose (Hermitian), Moore-Penrose inverse, r-th iterated point, and stationary point, respectively. $(\cdot)_{[x:y]}$ is a shortcut for $(\cdot)_x,(\cdot)_{x+1},...,(\cdot)_y$. $|\cdot|$, $\|\cdot\|$, and $\|\cdot\|_{\mathrm{F}}$ denote the absolute value, Euclidean norm, and Frobenius norm, respectively. $\odot$ represents the element-wise (Hadamard) product. $\mathcal{CN}(\mathbf{0},\mathbf{\Sigma})$ is the multivariate Circularly Symmetric Complex Gaussian (CSCG) distribution with mean $\mathbf{0}$ and covariance $\mathbf{\Sigma}$. $\sim$ means "distributed as".

II. BD-RIS Model

Consider a BD-RIS aided point-to-point MIMO system with N_{T}, N_{S}, N_{R} transmit, scatter, and receive antennas, respectively. This configuration is denoted as $N_{\mathrm{T}}\times N_{\mathrm{S}}\times N_{\mathrm{R}}$ in the following context. The BD-RIS can be modeled as an N_{S}-port network [52] that further divides into G individual groups, each containing $L\triangleq N_{\mathrm{S}}/G$ elements interconnected by real-time reconfigurable components [31]. To simplify the analysis and explore the performance limits, we assume a lossless asymmetric network without mutual coupling between scattering elements, as previously considered in [37], [38], [45]. The overall scattering matrix of the BD-RIS is block-unitary[3]

$$\mathbf{\Theta}=\mathrm{diag}(\mathbf{\Theta}_1,...,\mathbf{\Theta}_G), \tag{1}$$

where $\mathbf{\Theta}_g\in\mathbb{U}^{L\times L}$ is the g-th unitary block (i.e., $\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{\Theta}_g=\mathbf{I}$) that describes the response of group $g\in\mathcal{G}\triangleq\{1,...,G\}$. Note that diagonal RIS can be regarded as its extreme case with group size $L=1$. Some potential physical architectures of BD-RIS are illustrated in [31, Fig. 3], [38, Fig. 5], and [33, Fig. 2], where the circuit topology need to be modelled in the scattering matrix.

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 (transmitter-receiver), backward (RIS-receiver), and forward (transmitter-RIS) channels, respectively. The equivalent channel is a function of the scattering matrix

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

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 for RIS 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. Let $\mathbf{H}_g\triangleq\mathbf{H}_{\mathrm{B},g}\mathbf{\Theta}_g\mathbf{H}_{\mathrm{F},g}$ be the indirect channel via BD-RIS group g. Since unitary matrices constitute an algebraic group with respect to multiplication, the scattering matrix of group g can be decomposed as

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

where $\mathbf{L}_g,\mathbf{R}_g\in\mathbb{U}^{L\times L}$ are two unitary factor matrices. Let $\mathbf{H}_{\mathrm{B},g}=\mathbf{U}_{\mathrm{B},g}\mathbf{\Sigma}_{\mathrm{B},g}\mathbf{V}_{\mathrm{B},g}^{\mathsf{H}}$ and $\mathbf{H}_{\mathrm{F},g}=\mathbf{U}_{\mathrm{F},g}\mathbf{\Sigma}_{\mathrm{F},g}\mathbf{V}_{\mathrm{F},g}^{\mathsf{H}}$ be the compact Singular Value Decomposition (SVD) of the backward and forward channels, respectively. The equivalent channel can thus be rewritten as

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

[3]Following footnote 1, we do not assume the scattering matrix to be symmetric. If required, one can enforce symmetry over the block-unitary result by $\mathbf{\Theta}\leftarrow(\mathbf{\Theta}+\mathbf{\Theta}^{\mathsf{T}})/2$.

By analyzing (4), we conclude that the off-diagonal entries of the BD-RIS scattering matrix provide two key potentials for MIMO channel shaping:

- *Channel rearrangement:* It refers to rearranging and recombining the backward and forward channel branches (i.e., entries of the channel matrix) associated with each group by their strength. In SISO, diagonal RIS with perfect phase matching provides a maximum indirect channel amplitude of $\sum_{n=1}^{N_{\mathrm{S}}}|h_{\mathrm{B},n}||h_{\mathrm{F},n}|$ while BD-RIS can generalize it to $\sum_{g=1}^{G}\sum_{l=1}^{L}|h_{\mathrm{B},\pi_{\mathrm{B},g}(l)}||h_{\mathrm{F},\pi_{\mathrm{F},g}(l)}|$, where $\pi_{\mathrm{B},g}$ and $\pi_{\mathrm{F},g}$ are permutations of $\mathcal{L}\triangleq\{1,...,L\}$. Note the first summation is over groups and the second summation is over permuted channels. We thus conclude that BD-RIS exploits spatial diversity effectively thanks to in-group connections. By rearrangement inequality, the maximum channel gain is attained by pairing the l-th strongest backward and forward branches. Since the number of channels associated with each group is proportional to $N_{\mathrm{T}}N_{\mathrm{R}}$, we conclude the advantage of BD-RIS in channel rearrangement scales with MIMO dimensions.
- *Space alignment:* It refers to aligning the singular vectors of the direct, forward, and backward channels. The BD-RIS needs to strike a balance between the alignment of backward-forward (intra-group, multiplicative) channels and direct-indirect (inter-group, additive) channels. In SISO, singular vectors become scalars and space alignment boils down to phase matching, such that the optimal scattering matrix of group g that maximizes the channel gain is

$$\mathbf{\Theta}_g^{\star}=\exp\big(\jmath\arg(h_{\mathrm{D}})\big)\mathbf{V}_{\mathrm{B},g}\mathbf{U}_{\mathrm{F},g}^{\mathsf{H}}, \tag{5}$$

where $\mathbf{V}_{\mathrm{B},g}=\big[\mathbf{h}_{\mathrm{B},g}/\|\mathbf{h}_{\mathrm{B},g}\|,\mathbf{N}_{\mathrm{B},g}\big]\in\mathbb{U}^{L\times L}$, $\mathbf{U}_{\mathrm{F},g}=\big[\mathbf{h}_{\mathrm{F},g}/\|\mathbf{h}_{\mathrm{F},g}\|,\ \mathbf{N}_{\mathrm{F},g}\big]\in\mathbb{U}^{L\times L}$, and $\mathbf{N}_{\mathrm{B},g},\ \mathbf{N}_{\mathrm{F},g}\in\mathbb{C}^{L\times(L-1)}$ are the orthonormal bases of the null spaces of $\mathbf{h}_{\mathrm{B},g}$ and $\mathbf{h}_{\mathrm{F},g}$, respectively. Diagonal RIS with empty null space thus suffices for perfect phase matching in SISO. When it comes to MIMO, each element of diagonal RIS can only apply a common phase shift to the "pinhole" indirect channel passing through itself by $\mathbf{H}_g=\mathbf{h}_{\mathrm{B},g}\Theta_g\mathbf{h}_{\mathrm{F},g}^{\mathsf{H}}\in\mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$. That is, the disadvantage of diagonal RIS in space alignment scales with MIMO dimensions.

III. BD-RIS Design by Geodesic RCG

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. We notice that the feasible domain of scattering matrix of each BD-RIS group is a L-dimensional Stiefel manifold $\mathbf{\Theta}_g\in\mathbb{U}^{L\times L}$, which is non-convex and non-Euclidean. Therefore, relevant optimization problems are usually solved by relax-then-project methods [42] or universal non-geodesic manifold RCG [37], [38], [45]. The former solves an unconstrained version of original problem by quasi-Newton methods, then projects the solution back to the Stiefel manifold. It has no guarantee of optimality and may suffer from numerical instability. The latter generalizes the conjugate gradient methods to any Riemannian manifold, which updates the feasible point by progressing along the conjugate direction and projecting back to the manifold at each step. In this section, we first discuss the existing non-geodesic RCG method and its drawbacks, then propose a novel group-wise geodesic RCG method that operates directly on the Stiefel manifold for faster convergence.

A. Non-Geodesic RCG

This universal RCG method is applicable to optimization problems over arbitrary manifolds [50], [51]. The main idea is to perform additive updates along the conjugate direction guided by the Riemannian gradient, then project the solution back onto the manifold. For maximization problem with smooth objective f and block-unitary constraint (1), the steps for BD-RIS group g at iteration r are summarized below:

1) *Compute the Euclidean gradient [53]:* The gradient of f with respect to $\mathbf{\Theta}_g^*$ in the Euclidean space is
$$\nabla_{\mathrm{E},g}^{(r)} = \frac{\partial f(\mathbf{\Theta}_g^{(r)})}{\partial \mathbf{\Theta}_g^*}; \tag{6}$$
2) *Translate to the Riemannian gradient [50]:* At point $\mathbf{\Theta}^{(r)}$, the Riemannian gradient lies in the tangent space of the Stiefel manifold $\mathcal{T}_{\mathbf{\Theta}_g^{(r)}}\mathbb{U}^{L\times L} \triangleq \{\mathbf{M}\in\mathbb{C}^{L\times L} \mid \mathbf{M}^{\mathsf{H}}\mathbf{\Theta}_g^{(r)} + \mathbf{\Theta}_g^{(r)\mathsf{H}}\mathbf{M}=\mathbf{0}\}$. It gives the steepest ascent direction of the objective on the manifold can be obtained by projecting the Euclidean gradient onto the tangent space:
$$\nabla_{\mathrm{R},g}^{(r)} = \nabla_{\mathrm{E},g}^{(r)} - \mathbf{\Theta}_g^{(r)}\nabla_{\mathrm{E},g}^{(r)\mathsf{H}}\mathbf{\Theta}_g^{(r)}; \tag{7}$$
3) *Determine the conjugate direction [54]:* The conjugate direction is obtained over the Riemannian gradient and previous direction as
$$\mathbf{D}_g^{(r)} = \nabla_{\mathrm{R},g}^{(r)} + \gamma_g^{(r)}\mathbf{D}_g^{(r-1)}, \tag{8}$$
where $\gamma_g^{(r)}$ is the parameter that deviates the conjugate direction from the tangent space for accelerated convergence. A popular choice is the Polak-Ribière formula
$$\gamma_g^{(r)} = \frac{\mathrm{tr}\big((\nabla_{\mathrm{R},g}^{(r)} - \nabla_{\mathrm{R},g}^{(r-1)})\nabla_{\mathrm{R},g}^{(r)\mathsf{H}}\big)}{\mathrm{tr}\big(\nabla_{\mathrm{R},g}^{(r-1)}\nabla_{\mathrm{R},g}^{(r-1)\mathsf{H}}\big)}; \tag{9}$$
4) *Perform additive update [51]:* The point is updated by moving along a straight path in the conjugate direction
$$\bar{\mathbf{\Theta}}_g^{(r+1)} = \mathbf{\Theta}_g^{(r)} + \mu\mathbf{D}_g^{(r)}, \tag{10}$$
where μ is the step size refinable by the Armijo rule [55];
5) *Retract for feasibility [37], [50]:* The resulting point needs to be projected to the closest point (in terms of Euclidean distance) on the Stiefel manifold by
$$\mathbf{\Theta}_g^{(r+1)} = \bar{\mathbf{\Theta}}_g^{(r+1)}\big(\bar{\mathbf{\Theta}}_g^{(r+1)\mathsf{H}}\bar{\mathbf{\Theta}}_g^{(r+1)}\big)^{-1/2}. \tag{11}$$
One can also combine the addition (10) and retraction (11) in one step
$$\mathbf{\Theta}_g^{(r+1)} = \big(\mathbf{\Theta}_g^{(r)} + \mu\mathbf{D}_g^{(r)}\big)\big(\mathbf{I} + \mu^2\mathbf{D}_g^{(r)\mathsf{H}}\mathbf{D}_g^{(r)}\big)^{-1/2}, \tag{12}$$
and determine the step size therein.

The above method is called non-geodesic since the points are updated in the linear embedding spaces by addition (10) and retraction (11), instead of on the Stiefel manifold itself. It converges to stationary points of the original problem but usually requires a large number of iterations due to inefficient operations in the Euclidean space.

B. Geodesic RCG

Before introducing geodesic RCG, we revisit some basic concepts in differential geometry and Lie algebra.

A Lie group is simultaneously a continuous group and a differentiable manifold. Lie algebra refers to the tangent space of the Lie group at the identity element. The exponential map acts as a bridge between the Lie algebra and Lie group, which allows one to recapture the local group structure using linear algebra techniques. The set of unitary matrices $\mathbb{U}^{L\times L}$ forms a Lie group $U(L)$ under multiplication, and the corresponding Lie algebra $\mathfrak{u}(L) \triangleq \mathcal{T}_{\mathbf{I}}\mathbb{U}^{L\times L} = \{\mathbf{M}\in\mathbb{C}^{L\times L} \mid \mathbf{M}^{\mathsf{H}} + \mathbf{M} = \mathbf{0}\}$ consists of skew-Hermitian matrices. A geodesic emanating from the identity with velocity $\mathbf{D}\in\mathfrak{u}(L)$ can be described by [56]
$$\mathbf{G}_{\mathbf{I}}(\mu) = \exp(\mu\mathbf{D}), \tag{13}$$
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). Note that the right translation is an isometry in $U(L)$. During the optimization of group g, the geodesic evaluated at the identity (13) should be translated to $\mathbf{\Theta}_g^{(r)}$ for successive updates [48]
$$\mathbf{G}_g^{(r)}(\mu) = \mathbf{G}_{\mathbf{I}}(\mu)\mathbf{\Theta}_g^{(r)} = \exp(\mu\mathbf{D}_g^{(r)})\mathbf{\Theta}_g^{(r)}, \tag{14}$$
while the Riemannian gradient evaluated at $\mathbf{\Theta}_g^{(r)}$ (7) should be translated back to the identity for exploiting the Lie algebra [48]
$$\tilde{\nabla}_{\mathrm{R},g}^{(r)} = \nabla_{\mathrm{R},g}^{(r)}\mathbf{\Theta}_g^{(r)\mathsf{H}} = \nabla_{\mathrm{E},g}^{(r)}\mathbf{\Theta}_g^{(r)\mathsf{H}} - \mathbf{\Theta}_g^{(r)}\nabla_{\mathrm{E},g}^{(r)\mathsf{H}}. \tag{15}$$
After gradient translation, the deviation parameter and conjugate direction can be determined similarly to (9) and (8)
$$\tilde{\gamma}_g^{(r)} = \frac{\mathrm{tr}\big((\tilde{\nabla}_{\mathrm{R},g}^{(r)} - \tilde{\nabla}_{\mathrm{R},g}^{(r-1)})\tilde{\nabla}_{\mathrm{R},g}^{(r)\mathsf{H}}\big)}{\mathrm{tr}\big(\tilde{\nabla}_{\mathrm{R},g}^{(r-1)}\tilde{\nabla}_{\mathrm{R},g}^{(r-1)\mathsf{H}}\big)}. \tag{16}$$
$$\mathbf{D}_g^{(r)} = \tilde{\nabla}_{\mathrm{R},g}^{(r)} + \tilde{\gamma}_g^{(r)}\mathbf{D}_g^{(r-1)}, \tag{17}$$
The solution can thus be updated along the geodesic in a multiplicative rotational manner
$$\mathbf{\Theta}_g^{(r+1)} = \mathbf{G}_g^{(r)}(\mu) = \exp(\mu\mathbf{D}_g^{(r)})\mathbf{\Theta}_g^{(r)}, \tag{18}$$
where an appropriate μ may be obtained by the Armijo rule. To double the step size, one can simply square the rotation matrix instead of recomputing the matrix exponential, since $\exp^2(\mu\mathbf{D}_g^{(r)}) = \exp(2\mu\mathbf{D}_g^{(r)})$.

Algorithm 1 summarizes the proposed BD-RIS design framework based on group-wise geodesic RCG. Compared to the non-geodesic approach, it leverages the Lie group properties to replace the add-then-retract update (12) with a multiplicative rotational update (18) along the geodesic. This leads to faster convergence and simplifies the step size tuning thanks to appropriate parameter space. Convergence to a local optimum is still guaranteed if not initialized at a stationary point. Note that the group-wise updates can be performed in parallel to facilitate large-scale BD-RIS optimization problems. Since

Algorithm 1: Group-wise geodesic RCG for BD-RIS design

Input: $f(\mathbf{\Theta})$, G
Output: $\mathbf{\Theta}^\star$
Initialize $r \leftarrow 0$, $\mathbf{\Theta}^{(0)}$
Repeat
  For $g \leftarrow 1$ to G
    $\nabla_{\mathrm{E},g}^{(r)} \leftarrow$ (6)
    $\tilde{\nabla}_{\mathrm{R},g}^{(r)} \leftarrow$ (15)
    $\tilde{\gamma}_g^{(r)} \leftarrow$ (16)
    $\mathbf{D}_g^{(r)} \leftarrow$ (17)
    If $\Re\{\mathrm{tr}(\mathbf{D}_g^{(r)\mathsf{H}} \tilde{\nabla}_{\mathrm{R},g}^{(r)})\} < 0$ ▷ not an ascent direction
      $\mathbf{D}_g^{(r)} \leftarrow \tilde{\nabla}_{\mathrm{R},g}^{(r)}$
    End If
    $\mu \leftarrow 1$
    $\mathbf{G}_g^{(r)}(\mu) \leftarrow$ (14)
    While $f(\mathbf{G}_g^{(r)}(2\mu)) - f(\mathbf{\Theta}_g^{(r)}) \geq \mu \cdot \mathrm{tr}(\mathbf{D}_g^{(r)} \mathbf{D}_g^{(r)\mathsf{H}})/2$
      $\mu \leftarrow 2\mu$
    End While
    While $f(\mathbf{G}_g^{(r)}(\mu)) - f(\mathbf{\Theta}_g^{(r)}) < \mu/2 \cdot \mathrm{tr}(\mathbf{D}_g^{(r)} \mathbf{D}_g^{(r)\mathsf{H}})/2$
      $\mu \leftarrow \mu/2$
    End While
    $\mathbf{\Theta}_g^{(r+1)} \leftarrow$ (18)
  End For
  $r \leftarrow r+1$
Until $|f(\mathbf{\Theta}^{(r)}) - f(\mathbf{\Theta}^{(r-1)})| / f(\mathbf{\Theta}^{(r-1)}) \leq \epsilon$

block-unitary matrices are also closed under multiplication, one can avoid group-wise updates by directly operating on $\mathbf{\Theta}$ and pinching (i.e., keeping the block diagonal and nulling the other entries) the Euclidean gradient (6), with potentially higher computational complexity and slower convergence.

IV. Channel Singular Values Redistribution

In this section, we first provide a toy example to illustrate the channel shaping advantage of BD-RIS architecture. Next, we numerically characterize the channel singular value region based on the proposed Algorithm 1. Finally, we derive some analytical singular value bounds in specific scenarios.

A. Toy Example

We first illustrate the channel shaping capabilities of different RIS models by a toy example. Consider a $2\times2\times2$ setup where the direct link is blocked. The diagonal RIS is modeled by $\mathbf{\Theta}_\mathrm{D} = \mathrm{diag}(e^{\jmath\theta_1}, e^{\jmath\theta_2})$ while the fully-connected RIS has 4 independent angular parameters

$$\mathbf{\Theta}_\mathrm{U} = e^{\jmath\phi} \begin{bmatrix} e^{\jmath\alpha}\cos\psi & e^{\jmath\beta}\sin\psi \\ -e^{-\jmath\beta}\sin\psi & e^{-\jmath\alpha}\cos\psi \end{bmatrix}. \tag{19}$$

It is worth noting that ϕ has no impact on the singular value because $\mathrm{sv}(e^{\jmath\phi}\mathbf{A}) = \mathrm{sv}(\mathbf{A})$. To simplify the analysis, we also enforce symmetry $\mathbf{\Theta}_\mathrm{U} = \mathbf{\Theta}_\mathrm{U}^\mathsf{T}$ by $\beta = \pi/2$ such that both architectures have the same number of optimization variables. Fig. 1 shows the channel singular values achieved by an exhaustive grid search over (θ_1, θ_2) for diagonal RIS and (α, ψ) for symmetric fully-connected RIS. In this particular example, we observe that both singular values can be manipulated up to 9% using diagonal RIS and 42% using symmetric BD-RIS, despite both architectures have the same number of scattering elements and optimization variables.[4] A larger performance gap is expected when the symmetric constraint on (19) can be relaxed. This example shows BD-RIS can provide a wider dynamic range of channel singular values and motivates further studies on channel shaping.

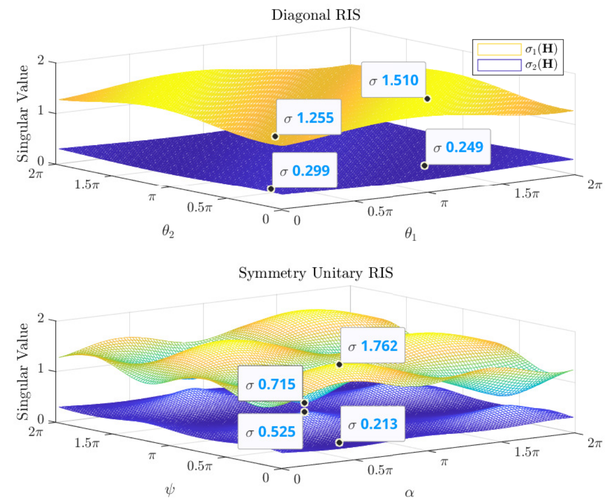


Fig. 1. $2\times2\times2$ (no direct) channel singular value shaping by diagonal and symmetric fully-connected RIS.

B. Pareto Frontier Characterization

We then characterize the Pareto frontier of singular values of a general $N_\mathrm{T}\times N_\mathrm{S}\times N_\mathrm{R}$ channel (2) by maximizing their weighted sum

$$\max_{\mathbf{\Theta}} \quad \sum_n \rho_n \sigma_n(\mathbf{H}) \tag{20a}$$

$$\text{s.t.} \quad \mathbf{\Theta}_g^\mathsf{H}\mathbf{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{20b}$$

where $n \in \mathcal{N} \triangleq \{1,\dots,N\}$, $N \triangleq \min(N_\mathrm{T}, N_\mathrm{R})$ is the maximum channel rank, and ρ_n is the weight of the n-th singular value that can be positive, zero, or negative. Varying $\{\rho_n\}_{n\in\mathcal{N}}$ characterizes the Pareto frontier that encloses the entire singular value region. Thus, we claim problem (20) generalizes most singular value shaping problems. It can be solved optimally by Algorithm 1 with the Euclidean gradient given by Lemma 1.

Lemma 1. *The Euclidean gradient of* (20a) *with respect to BD-RIS group g is*

$$\frac{\partial \sum_n \rho_n \sigma_n(\mathbf{H})}{\partial \mathbf{\Theta}_g^*} = \mathbf{H}_{\mathrm{B},g}^\mathsf{H} \mathbf{U} \mathrm{diag}(\rho_1,\dots,\rho_N) \mathbf{V}^\mathsf{H} \mathbf{H}_{\mathrm{F},g}^\mathsf{H}, \tag{21}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left and right compact singular matrices of $\mathbf{H}$, respectively.

Proof. Please refer to Appendix A. □

We then analyze the computational complexity of solving Pareto singular value problem (20) by Algorithm 1. To update each BD-RIS group, compact SVD of $\mathbf{H}$ requires $\mathcal{O}(NN_\mathrm{T}N_\mathrm{R})$,

[4] However, BD-RIS does require more reconfigurable components that interconnect the scattering elements.

Euclidean gradient (21) requires $\mathcal{O}\big(LN(N_\mathrm{T}+N_\mathrm{R}+L+1)\big)$, Riemannian gradient translation (15) requires $\mathcal{O}(L^3)$, deviation parameter (16) and conjugate direction (17) together require $\mathcal{O}(L^2)$, and matrix exponential (18) requires $\mathcal{O}(L^3)$ operations [57]. The overall complexity is thus $\mathcal{O}\big(I_\mathrm{RCG}G\big(NN_\mathrm{T}N_\mathrm{R}+LN(N_\mathrm{T}+N_\mathrm{R}+L+1)+I_\mathrm{BLS}L^3\big)\big)$, where I_RCG and I_BLS are the number of iterations for geodesic RCG and backtracking line search (line 13–18 of Algorithm 1), respectively.

C. Some Analytical Bounds

We then discuss some analytical bounds related to channel singular values.

Proposition 1 (Degree of freedom)**.** *In point-to-point MIMO, BD-RIS cannot achieve a higher Degree of Freedom (DoF) than diagonal RIS.*

Proof. Please refer to Appendix B. □

Proposition 2 (Rank-deficient channel)**.** *If the forward or backward channel is rank-k ($k \le N$), then regardless of the passive RIS size and architecture, the n-th singular value of the equivalent channel is bounded by*

$$\sigma_n(\mathbf{H}) \le \sigma_{n-k}(\mathbf{T}), \qquad \text{if } n > k, \tag{22a}$$

$$\sigma_n(\mathbf{H}) \ge \sigma_n(\mathbf{T}), \qquad \text{if } n < N-k+1, \tag{22b}$$

where

$$\mathbf{T}\mathbf{T}^\mathsf{H} = \begin{cases} \mathbf{H}_\mathrm{D}(\mathbf{I}-\mathbf{V}_\mathrm{F}\mathbf{V}_\mathrm{F}^\mathsf{H})\mathbf{H}_\mathrm{D}^\mathsf{H}, & \text{if } \operatorname{rank}(\mathbf{H}_\mathrm{F})=k, \\ \mathbf{H}_\mathrm{D}^\mathsf{H}(\mathbf{I}-\mathbf{U}_\mathrm{B}\mathbf{U}_\mathrm{B}^\mathsf{H})\mathbf{H}_\mathrm{D}, & \text{if } \operatorname{rank}(\mathbf{H}_\mathrm{B})=k, \end{cases} \tag{23}$$

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

Proof. Please refer to Appendix C. □

Corollary 2.1 (Extreme singular values)**.** *With a sufficiently large passive RIS of arbitrary architecture, the first k channel singular values are unbounded above*[5] *while the last k channel singular values can be suppressed to zero.*

Proof. This is a direct result of (22). □

Corollary 2.2 (Line-of-Sight (LoS) channel[6])**.** *If the forward or backward channel is LoS, then a passive RIS of arbitrary architecture 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 singular value to the n-th singular value of $\mathbf{T}$. That is,*

$$\sigma_1(\mathbf{H}) \ge \sigma_1(\mathbf{T}) \ge \sigma_2(\mathbf{H}) \ge \dots \ge \sigma_{N-1}(\mathbf{T}) \ge \sigma_N(\mathbf{H}) \ge \sigma_N(\mathbf{T}). \tag{24}$$

Proof. This is a direct result of (22) with $k=1$. □

In Section VI, we will show by simulation that a finite-size BD-RIS can better approach those bounds than diagonal RIS.

[5]The energy conservation law $\sum_n \sigma_n^2(\mathbf{H}) \le 1$ still has to be respected. This constraint is omitted in the following context for brevity.

[6]A similar result has been derived for diagonal RIS [58].

Proposition 3 (Fully-connected RIS without direct link)**.** *If the BD-RIS is unitary and the direct link is absent, then the channel singular values can be manipulated up to*

$$\operatorname{sv}(\mathbf{H}) = \operatorname{sv}(\mathbf{B}\mathbf{F}), \tag{25}$$

where $\mathbf{B}$ and $\mathbf{F}$ are arbitrary matrices with the same singular values as $\mathbf{H}_\mathrm{B}$ and $\mathbf{H}_\mathrm{F}$, respectively,

Proof. Please refer to Appendix D. □

The problem now becomes, how the singular values of matrix product are bounded by the singular values of its individual factors. Let $\bar{N} = \max(N_\mathrm{T}, N_\mathrm{S}, N_\mathrm{R})$ and $\sigma_n(\mathbf{H}) = \sigma_n(\mathbf{H}_\mathrm{F}) = \sigma_n(\mathbf{H}_\mathrm{B}) = 0$ for $N < n \le \bar{N}$. We have the following corollaries.

Corollary 3.1 (Generic singular value bounds)**.**

$$\prod_{k\in K}\sigma_k(\mathbf{H}) \le \prod_{i\in I}\sigma_i(\mathbf{H}_\mathrm{B})\prod_{j\in J}\sigma_j(\mathbf{H}_\mathrm{F}), \tag{26}$$

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

$$T_r^{\bar{N}} \triangleq \Big\{(I,J,K)\in U_r^{\bar{N}} \,|\, \forall p<r, (F,G,H)\in T_p^r, \sum_{f\in F} i_f + \sum_{g\in G} j_g \le \sum_{h\in H} k_h + p(p+1)/2\Big\},$$

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

Proof. Please refer to [59, Theorem 8]. □

Corollary (3.1) is by far the most comprehensive singular value bound over Proposition 3, which is also recognized as a variation of Horn's inequality [60]. It is worth mentioning that the number of admissible triples (and bounds) grows exponentially with $\bar{N}$. For example, the number of inequalities described by (26) grows from 12 to 2062 when $\bar{N}$ increases from 3 to 7. This renders it computationally expensive for applications in large-scale MIMO systems. Next, we showcase some useful inequalities enclosed by (26). Readers are referred to [61, Chapter 16, 24] for more examples.

Corollary 3.2 (Upper bound on the largest singular value)**.**

$$\sigma_1(\mathbf{H}) \le \sigma_1(\mathbf{H}_\mathrm{B})\sigma_1(\mathbf{H}_\mathrm{F}). \tag{27}$$

Proof. This is a direct result of (26) with $r=1$. □

Corollary 3.3 (Lower bound on the smallest singular value)**.**

$$\sigma_{\bar{N}}(\mathbf{H}) \ge \sigma_{\bar{N}}(\mathbf{H}_\mathrm{B})\sigma_{\bar{N}}(\mathbf{H}_\mathrm{F}). \tag{28}$$

Proof. This can be deducted from (26) with $r_1 = \bar{N}-1$ and $r_2 = \bar{N}$. □

Corollary 3.4 (Upper bound on the product of first k singular values)**.**

$$\prod_{n=1}^{k}\sigma_n(\mathbf{H}) \le \prod_{n=1}^{k}\sigma_n(\mathbf{H}_\mathrm{B})\prod_{n=1}^{k}\sigma_n(\mathbf{H}_\mathrm{F}). \tag{29}$$

Proof. This is a direct result of (26) with $r=k$. □

Corollary 3.5 (Lower bound on the product of last k singular values)**.**

$$\prod_{n=\bar{N}}^{\bar{N}-k+1}\sigma_n(\mathbf{H})\geq\prod_{n=\bar{N}}^{\bar{N}-k+1}\sigma_n(\mathbf{H}_\mathrm{B})\prod_{n=\bar{N}}^{\bar{N}-k+1}\sigma_n(\mathbf{H}_\mathrm{F}). \tag{30}$$

Proof. This can be deducted from (26) with $r_1=\bar{N}-k$ and $r_2=\bar{N}$. □

Corollaries 3.3 and 3.5 are less informative when $\bar{N}\neq N$ (i.e., unequal number of transmit, scatter, and receive antennas) as the lower bounds would coincide at zero.

Corollary 3.6 (Upper bound on the channel power gain)**.** *The channel power gain is upper bounded by the sum of sorted element-wise product of squared singular values of backward and forward channels*

$$\|\mathbf{H}\|_\mathrm{F}^2=\sum_{n=1}^{N}\sigma_n^2(\mathbf{H})\leq\sum_{n=1}^{N}\sigma_n^2(\mathbf{H}_\mathrm{B})\sigma_n^2(\mathbf{H}_\mathrm{F}). \tag{31}$$

Proof. Please refer to [61, Inequality 24.4.7]. □

To achieve the equalities in Corollaries (3.2) – (3.6), the RIS needs to completely align the spaces of $\mathbf{H}_\mathrm{B}$ and $\mathbf{H}_\mathrm{F}$. The resulting scattering matrix is generally required to be unitary

$$\mathbf{\Theta}^\star=\mathbf{V}_\mathrm{B}\mathbf{U}_\mathrm{F}^\mathsf{H}, \tag{32}$$

which can be concluded from (50) and (51) in Appendix D. Interestingly, diagonal RIS can attain those equalities if and only if $\mathbf{H}_\mathrm{B}$ and $\mathbf{H}_\mathrm{F}$ are both rank-1. In such case, the equivalent channel reduces to $\mathbf{H}=\sigma_\mathrm{B}\sigma_\mathrm{F}\mathbf{u}_\mathrm{B}\mathbf{v}_\mathrm{B}^\mathsf{H}\mathbf{\Theta}\mathbf{u}_\mathrm{F}\mathbf{v}_\mathrm{F}^\mathsf{H}$ and the RIS only needs to align $\mathbf{v}_\mathrm{B}^\mathsf{H}$ and $\mathbf{u}_\mathrm{F}$ by

$$\mathbf{\Theta}^\star=\mathbf{v}_\mathrm{B}\mathbf{u}_\mathrm{F}^\mathsf{H}\odot\mathbf{I}, \tag{33}$$

which becomes a special case of (5). On the other hand, when $\mathbf{H}_\mathrm{B}$ and $\mathbf{H}_\mathrm{F}$ are both in Rayleigh fading, the expected maximum channel power gain $\mathbb{E}\big\{\|\mathbf{H}\|_\mathrm{F}\big\}_{\max}$ can be evaluated as

$$\sum_{n=1}^{N}\int_0^\infty f_{\lambda_n^{\min(N_\mathrm{R},N_\mathrm{S})}}(x_n)dx_n\int_0^\infty f_{\lambda_n^{\min(N_\mathrm{S},N_\mathrm{T})}}(x_n)dx_n, \tag{34}$$

where λ_n^K is the n-th eigenvalue of the complex $K\times K$ Wishart matrix with probability density function $f_{\lambda_n^K}(x_n)$ given by [62, Equation 51]. We notice (34) is a generalization of [31, Equation 58] to MIMO.

Tighter bounds are generally inapplicable when the direct link is present or the BD-RIS is not unitary, since the direct-indirect channels and backward-forward channels cannot be completely aligned at the same time. In such case, we can exploit optimization approaches from a singular value perspective (Section IV-B) or a power gain perspective (Section V-A).

V. Power Gain and Achievable Rate Maximization

A. Channel Power Gain

The MIMO channel power gain maximization problem is formulated with respect to the BD-RIS scattering matrix

$$\max_{\mathbf{\Theta}}\quad\|\mathbf{H}_\mathrm{D}+\mathbf{H}_\mathrm{B}\mathbf{\Theta}\mathbf{H}_\mathrm{F}\|_\mathrm{F}^2 \tag{35a}$$

$$\text{s.t.}\quad\mathbf{\Theta}_g^\mathsf{H}\mathbf{\Theta}_g=\mathbf{I},\quad\forall g, \tag{35b}$$

which generalizes the case of SISO [31], MISO [34], [42], single-stream MIMO [32], [47], and direct link-blocked MIMO with fully-connected RIS (32). The key of solving (35) is to balance the additive and multiplicative space alignments. Interestingly, in terms of maximizing the inner product $\langle\mathbf{H}_\mathrm{D},\mathbf{H}_\mathrm{B}\mathbf{\Theta}\mathbf{H}_\mathrm{F}\rangle$, (35) is reminiscent of the weighted orthogonal Procrustes problem [63]

$$\min_{\mathbf{\Theta}}\quad\|\mathbf{H}_\mathrm{D}-\mathbf{H}_\mathrm{B}\mathbf{\Theta}\mathbf{H}_\mathrm{F}\|_\mathrm{F}^2 \tag{36a}$$

$$\text{s.t.}\quad\mathbf{\Theta}^\mathsf{H}\mathbf{\Theta}=\mathbf{I}, \tag{36b}$$

which relaxes the block-unitary constraint (36b) to unitary but still has no trivial solution. One lossy transformation exploits the Moore-Penrose inverse and moves $\mathbf{\Theta}$ to one side of the product [64], formulating two standard orthogonal Procrustes problems

$$\min_{\mathbf{\Theta}}\quad\|\mathbf{H}_\mathrm{B}^\dagger\mathbf{H}_\mathrm{D}-\mathbf{\Theta}\mathbf{H}_\mathrm{F}\|_\mathrm{F}^2 \text{ or } \|\mathbf{H}_\mathrm{D}\mathbf{H}_\mathrm{F}^\dagger-\mathbf{H}_\mathrm{B}\mathbf{\Theta}\|_\mathrm{F}^2 \tag{37a}$$

$$\text{s.t.}\quad\mathbf{\Theta}^\mathsf{H}\mathbf{\Theta}=\mathbf{I}, \tag{37b}$$

which have global optimal solutions

$$\mathbf{\Theta}=\mathbf{U}\mathbf{V}^\mathsf{H}, \tag{38}$$

where $\mathbf{U}$ and $\mathbf{V}$ are respectively the left and right compact singular matrices of $\mathbf{H}_\mathrm{B}^\dagger\mathbf{H}_\mathrm{D}\mathbf{H}_\mathrm{F}^\mathsf{H}$ or $\mathbf{H}_\mathrm{B}^\mathsf{H}\mathbf{H}_\mathrm{D}\mathbf{H}_\mathrm{F}^\dagger$ [65]. We emphasize that (32) and (38) are valid fully-connected RIS solutions to (35) when the direct link is absent and present, but the latter is neither optimal nor a generalization of the former due to the lossy transformation.

Inspired by [66], we propose an optimal solution to problem (35) with arbitrary group size. The idea is to successively approximate the quadratic objective (35a) by local Taylor expansions and solve each step in closed form by group-wise SVD.

Proposition 4. *Starting from any feasible $\mathbf{\Theta}^{(0)}$, the sequence*

$$\mathbf{\Theta}_g^{(r+1)}=\mathbf{U}_g^{(r)}\mathbf{V}_g^{(r)},\quad\forall g. \tag{39}$$

converges to a stationary point of (35)*, where $\mathbf{U}_g^{(r)}$ and $\mathbf{V}_g^{(r)}$ are the left and right compact singular matrices of*

$$\mathbf{M}_g^{(r)}=\mathbf{H}_{\mathrm{B},g}^\mathsf{H}\Big(\mathbf{H}_\mathrm{D}+\mathbf{H}_\mathrm{B}\mathrm{diag}\big(\mathbf{\Theta}_{[1:g-1]}^{(r+1)},\mathbf{\Theta}_{[g:G]}^{(r)}\big)\mathbf{H}_\mathrm{F}\Big)\mathbf{H}_{\mathrm{F},g}^\mathsf{H} \tag{40}$$

Proof. Please refer to Appendix F. □

We then analyze the computational complexity of solving channel gain maximization problem (35) by Proposition 4. To update each BD-RIS group, matrix multiplication (40) requires $\mathcal{O}\big(N_\mathrm{T}N_\mathrm{R}+(G+1)(NL^2+N_\mathrm{T}N_\mathrm{R}L)\big)$ operations and its compact SVD requires $\mathcal{O}(L^3)$ operations. The overall complexity is thus $\mathcal{O}\big(I_\mathrm{SAA}G\big(N_\mathrm{T}N_\mathrm{R}+(G+1)(NL^2+N_\mathrm{T}N_\mathrm{R}L)+L^3\big)\big)$, where I_SAA is the number iterations for successive affine approximation.

B. Achievable Rate Maximization

We aim to maximize the achievable rate of the BD-RIS-aided MIMO system by jointly optimizing the active and passive beamforming

$$\max_{\mathbf{W},\boldsymbol{\Theta}} \quad R = \log\det\left(\mathbf{I} + \frac{\mathbf{W}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{W}}{\eta}\right) \tag{41a}$$

$$\text{s.t.} \quad \|\mathbf{W}\|_{\mathrm{F}}^2 \leq P, \tag{41b}$$

$$\boldsymbol{\Theta}_g^{\mathsf{H}}\boldsymbol{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{41c}$$

where $\mathbf{W}$ is the transmit precoder, R is the achievable rate, η is the average noise power, and P is the transmit power constraint. Problem (41) is non-convex due to the block-unitary constraint (41c) and the coupling between variables. We propose a local-optimal approach via AO and a low-complexity approach based on channel shaping.

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

$$\max_{\boldsymbol{\Theta}} \quad \log\det\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}}{\eta}\right) \tag{42a}$$

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

where $\mathbf{Q} \triangleq \mathbf{W}\mathbf{W}^{\mathsf{H}}$ is the transmit covariance matrix. Problem (42) can be solved optimally by Algorithm 1 with the Euclidean gradient given by Lemma 2.

Lemma 2. *The Euclidean gradient of* (42a) *with respect to BD-RIS block g is*

$$\frac{\partial R}{\partial \boldsymbol{\Theta}_g^*} = \frac{1}{\eta}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}}{\eta}\right)^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}. \tag{43}$$

Proof. Please refer to Appendix E. □

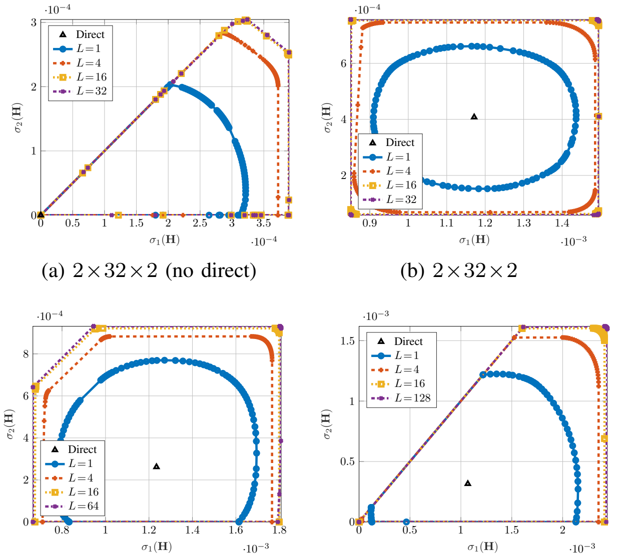
For a given $\boldsymbol{\Theta}$, the global optimal transmit precoder is given by eigenmode transmission [67]

$$\mathbf{W}^\star = \mathbf{V}\operatorname{diag}(\mathbf{s}^\star)^{1/2}, \tag{44}$$

where $\mathbf{V}$ is the right singular matrix of the equivalent channel and $\mathbf{s}^\star$ is the optimal water-filling power allocation obtainable by the iterative method [68].

The AO algorithm is guaranteed to converge to local-optimal points of problem (41) since each subproblem is solved optimally and the objective is bounded above. Similar to the analysis in Section IV-B, the computational complexity of solving subproblem (42) by geodesic RCG is $\mathcal{O}\big(I_{\mathrm{RCG}}G(NL^2 + LN_{\mathrm{T}}N_{\mathrm{R}} + N_{\mathrm{T}}^2N_{\mathrm{R}} + N_{\mathrm{T}}N_{\mathrm{R}}^2 + N_{\mathrm{R}}^3 + I_{\mathrm{BLS}}L^3)\big)$. On the other hand, the complexity of solving active beamforming subproblem by (44) is $\mathcal{O}\big(NN_{\mathrm{T}}N_{\mathrm{R}}\big)$. The overall complexity is thus $\mathcal{O}\big(I_{\mathrm{AO}}\big(I_{\mathrm{RCG}}G(NL^2 + LN_{\mathrm{T}}N_{\mathrm{R}} + N_{\mathrm{T}}^2N_{\mathrm{R}} + N_{\mathrm{T}}N_{\mathrm{R}}^2 + N_{\mathrm{R}}^3 + I_{\mathrm{BLS}}L^3) + NN_{\mathrm{T}}N_{\mathrm{R}}\big)\big)$, where I_{AO} is the number of iterations for AO.

2) Low-Complexity Solution: We then propose a suboptimal two-stage solution to problem (41) that decouples the joint RIS-transceiver design. The idea is to first consider channel shaping and replace the rate maximization subproblem (42) by channel power gain maximization problem (35), then proceed to conventional eigenmode transmission (44). Both steps are solved in closed form and the computational complexity is $\mathcal{O}\big(I_{\mathrm{SAA}}G\big(N_{\mathrm{T}}N_{\mathrm{R}} + (G+1)(NL^2 + N_{\mathrm{T}}N_{\mathrm{R}}L) + L^3\big) + NN_{\mathrm{T}}N_{\mathrm{R}}\big)$. While suboptimal, this shaping-inspired solution avoids outer iterations and implements inner iterations more efficiently.



(a) $2\times32\times2$ (no direct) (b) $2\times32\times2$

(c) $2\times64\times2$ (d) $2\times128\times2$

Fig. 2. Pareto frontiers of singular values of a 2T2R channel reshaped by a RIS.

VI. Simulation Results

In this section, we provide numerical results to evaluate the proposed BD-RIS designs. Consider a distance-dependent path loss model $\Lambda(d) = \Lambda_0 d^{-\gamma}$ where Λ_0 is the reference path loss at distance $1\,\mathrm{m}$, d is the propagation distance, and γ is the path loss exponent. The small-scale fading model is $\mathbf{H} = \sqrt{\kappa/(1+\kappa)}\mathbf{H}_{\mathrm{LoS}} + \sqrt{1/(1+\kappa)}\mathbf{H}_{\mathrm{NLoS}}$, where κ is the Rician K-factor, $\mathbf{H}_{\mathrm{LoS}}$ is the deterministic LoS component, and $\mathbf{H}_{\mathrm{NLoS}} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ is the Rayleigh component. We set $\Lambda_0 = -30\mathrm{dB}$, $d_{\mathrm{D}} = 14.7\mathrm{m}$, $d_{\mathrm{F}} = 10\mathrm{m}$, $d_{\mathrm{B}} = 6.3\mathrm{m}$, $\gamma_{\mathrm{D}} = 3$, $\gamma_{\mathrm{F}} = 2.4$ and $\gamma_{\mathrm{B}} = 2$ for reference, which corresponds to a typical indoor environment with $\Lambda_{\mathrm{D}} = -65\mathrm{dB}$, $\Lambda_{\mathrm{F}} = -54\mathrm{dB}$, $\Lambda_{\mathrm{B}} = -46\mathrm{dB}$. The indirect path via RIS is thus $35\,\mathrm{dB}$ weaker than the direct path. Rayleigh fading (i.e., $\kappa = 0$) is assumed for all channels unless otherwise specified.

A. Algorithm Evaluation

We first compare the geodesic and non-geodesic RCG algorithm on problem (20) in a 4T4R system with BD-RIS group size $L = 4$. The statistics are averaged over 100 independent runs. It is observed that the geodesic RCG method achieves a slightly higher objective value with significantly (up to $3\times$) lower number of iterations and shorter (up to $4\times$) computational time than the non-geodesic method. The results demonstrate the efficiency of the proposed geodesic RCG algorithm especially in large-scale BD-RIS design problems.

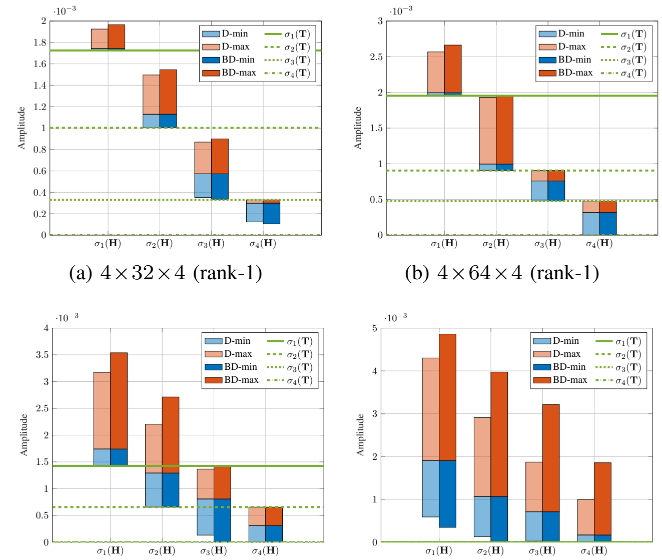
TABLE I
AVERAGE PERFORMANCE OF GEODESIC AND NON-GEODESIC RCG ALGORITHMS ON PROBLEM (20)

RCG path	$N_S = 16$			$N_S = 256$		
	Objective	Iterations	Time [s]	Objective	Iterations	Time [s]
Geodesic	4.359×10^{-3}	11.59	1.839×10^{-2}	1.163×10^{-2}	25.58	3.461
Non-geodesic	4.329×10^{-3}	30.92	5.743×10^{-2}	1.116×10^{-2}	61.40	13.50

B. Channel Singular Values Redistribution

1) Pareto Frontier: Fig. 2 shows the Pareto singular values of a 2T2R MIMO reshaped by a RIS. When the direct link is absent, the achievable regions in Fig. 2(a) are shaped like pizza slices. This is because $\sigma_1(\mathbf{H}) \geq \sigma_2(\mathbf{H}) \geq 0$ and there exists a trade-off between the alignment of two spaces. We observe that the smallest singular value can be enhanced up to 2×10^{-4} by single-connected RIS and 3×10^{-4} by fully-connected RIS, corresponding to a $50\,\%$ gain. When the direct link is present, the shape of the singular value region depends heavily on the relative strength of the indirect link. In Fig. 2(b), a 32-element RIS is insufficient to compensate the $35\,\mathrm{dB}$ path loss 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 $38\,\%$, but also provides a better trade-off in manipulating both singular values. It suggests the design freedom from larger group size allows better alignment of multiple spaces simultaneously. The singular value region also enlarges as the number of scattering elements N_S increases. In particular, Fig. 2(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. Those results demonstrate the superior channel shaping capability of BD-RIS and emphasizes the importance of cooperative wave scattering.

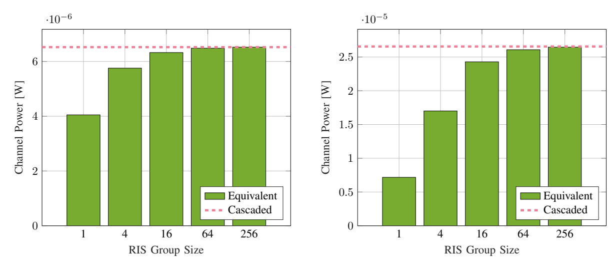
2) Analytical Bounds and Numerical Results: Fig. 3 illustrates the analytical singular value bounds in Proposition 2 and the numerical results obtained by solving problem (20) with $\rho_n = \pm 1$ and $\rho_{n'} = 0$, $\forall n' \neq n$. Here we assme a rank-k forward channel without loss of generality. When the RIS is in the vicinity of the transmitter, Figs. 3(a) and 3(b) show that the achievable channel singular values indeed satisfy Corollary 2.2, 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 diagonal RIS especially for a small N_S. Another example is given in Fig. 3(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 $\mathbf{T}$. Those observations align with Proposition 2 and Corollary 2.1. Finally, Fig. 3(d) confirms there are no extra singular value bounds when both forward and backward channels are full-rank. This can be predicted from (23) where the compact singular matrix $\mathbf{V}_F$ 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 diagonal RIS scales with forward and backward channel ranks.



(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)

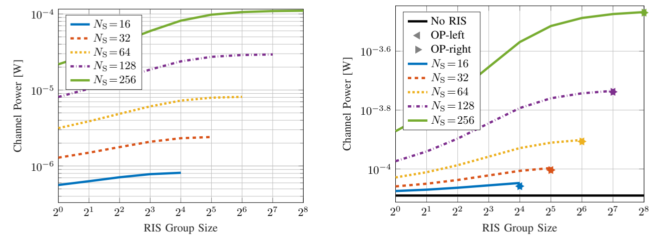
Fig. 3. Achievable channel singular values: analytical bounds (green lines) and numerical optimization results (blue and red bars). The intersections of the blue and red bars denote the singular values of the direct channel. The blue (resp. red) bars are obtained by solving problem (20) with $\rho_n = -1$ (resp. $+1$) and $\rho_{n'} = 0$, $\forall n' \neq n$. 'D' means diagonal 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. 4. Average maximum channel power versus BD-RIS group size and MIMO dimensions. 'Cascaded' refers to the available power of the cascaded channel, i.e., the sum of (sorted) element-wise power product of backward and forward channels.

Fig. 4 compares the analytical channel power bound in Corollary 3.6 and the numerical results obtained by solving problem (35) when the direct link is absent. Here, a fully-connected BD-RIS can attain the upper bound either in closed form (32) or via optimization approach (39). For the SISO case in Fig. 4(a), the maximum channel power is approximately 4×10^{-6} by diagonal RIS and 6.5×10^{-6} by fully-connected BD-RIS, corresponding to a $62.5\,\%$ gain. This aligns with the asymptotic BD-RIS scaling law derived for SISO in [31].



(a) $16 \times N_{\mathrm{S}} \times 16$ (no direct) (b) $16 \times N_{\mathrm{S}} \times 16$

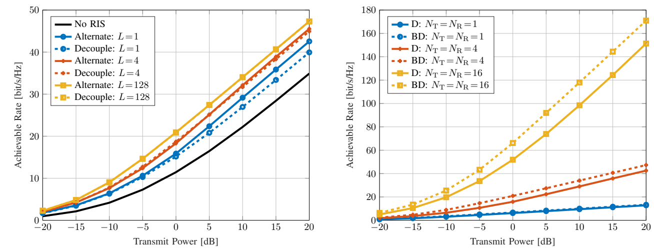
Fig. 5. Average maximum channel power versus RIS configuration. 'OP-left' and 'OP-right' refer to the suboptimal solutions to problem (35) by lossy transformation (37) where $\mathbf{\Theta}$ is to the left and right of the product, respectively.

Interestingly, the gain surges to $270\,\%$ in 4T4R MIMO as shown in Fig. 4(b). This is because space alignment boils down to phase matching in SISO such that both triangular and Cauchy-Schwarz inequalities in [31, (50)] can be simultaneously tight regardless of the group size. That is, diagonal RIS is sufficient for space alignment in SISO while the $62.5\,\%$ gain from BD-RIS comes purely from channel rearrangement. Now consider a diagonal RIS in MIMO. Each element can only apply a common phase shift to the associated rank-1 $N_{\mathrm{R}} \times N_{\mathrm{T}}$ indirect channel. Therefore, perfect space alignment of indirect channels through different elements is generally impossible. It means the disadvantage of diagonal RIS in space alignment and channel rearrangement scales with MIMO dimensions. We thus conclude that the power gain of BD-RIS scales with group size and MIMO dimensions.

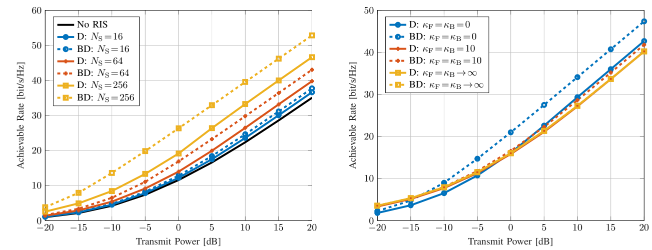
C. Power Gain and Achievable Rate Maximization

We first focus on channel power gain maximization problem (35). Fig. 5 shows the achievable channel power under different RIS configurations. An interesting observation is that the relative power gain of BD-RIS over diagonal RIS is even larger with direct link. For example, a 64-element fully BD-RIS can almost provide the same channel power as a 256-element diagonal RIS in Fig. 5b, but not in Fig. 5a. This is because the RIS needs to balance the multiplicative forward-backward combining and the additive direct-indirect combining, such that the space alignment advantage of BD-RIS becomes more pronounced. We also notice that the suboptimal solutions (38) for fully-connected BD-RIS by lossy transformation (37) are very close to optimal especially for a large N_{S}.

Fig. 6 presents the achievable rate under different MIMO and RIS configurations. At a transmit power of $10\,\mathrm{dB}$, Fig. 6(a) shows that introducing a 128-element diagonal RIS to 4T4R MIMO can improve the achievable rate from $22.2\,\mathrm{bps/Hz}$ to $29.2\,\mathrm{bps/Hz}$ $(+31.5\%)$. In contrast, a BD-RIS of group size 4 and 128 can further improve the rate to $32.1\,\mathrm{bps/Hz}$ $(+44.6\%)$ and $34\,\mathrm{bps/Hz}$ $(+53.2\%)$, respectively. Interestingly, the gap between the optimal AO approach (42)–(44) and the low-complexity solution (39) and (44) narrows as the group size increases, and completely vanishes for a fully-connected BD-RIS. This implies that the RIS-transceiver design can often be decoupled via channel shaping with marginal performance loss. Figs. 6(b) and 6(c) also confirm the advantage of BD-RIS grows with the number of transmit, scatter, and receive antennas. In the low power regime (-20 to $-10\,\mathrm{dB}$), the slope of the achievable rate is significantly larger with BD-RIS, suggesting that multiple streams can be activated at a much lower SNR. This is because BD-RIS not only spreads the channel singular values to a wider range, but also provides a better trade-off between channels (c.f. Fig. 2). Finally, Fig. 6(d) shows that the gap between diagonal and BD-RIS narrows as the Rician K-factor increases and becomes indistinguishable in LoS environment. The observation is expected from previous studies [31], [32], [37] and aligns with Corollary 2.2, which suggests that the BD-RIS should be deployed in rich-scattering environments to exploit its channel shaping potential.



(a) $4 \times 128 \times 4$ (b) $N_{\mathrm{T}} \times 128 \times N_{\mathrm{R}}$



(c) $4 \times N_{\mathrm{S}} \times 4$ (d) $4 \times 128 \times 4$

Fig. 6. Average achievable rate versus MIMO and RIS configurations. The noise power is $\eta = -75\mathrm{dB}$, corresponding to a direct SNR of -10 to $30\,\mathrm{dB}$. 'Alternate' refers to the alternating optimization and 'Decouple' refers to the low-complexity design. 'D' means diagonal RIS and 'BD' refers to fully-connected BD-RIS.

VII. Conclusion

This paper analyzes the channel shaping capability of RIS in terms of singular values redistribution. We consider a general BD architecture that allows elements within the same group to interact, enabling more sophisticated manipulation than diagonal RIS. This translates to a wider dynamic range (with better tradeoff) of singular values and significant power and rate gains, especially in large-scale MIMO systems. We characterize the Pareto frontiers of channel singular values via optimization approach and provide analytical bounds in rank-deficient and unitary scenarios. Specifically, the former is done by proposing an efficient RCG algorithm for smooth BD-RIS optimization problems, which offers better objective value and faster convergence than existing methods. We also present two beamforming designs for rate maximization problem, one based on alternating optimization for optimal performance and the other decouples the RIS-transceiver design for lower complexity. Extensive simulations show that the advantage of BD-RIS stems from its superior space alignment and channel rearrangement capability, which scales with the number of elements, group size, MIMO dimensions, and channel diversity.

One future direction is introducing BD-RIS to MIMO interference channel for interference alignment or cancellation. Another open issue is to exploit different groups of BD-RIS to enhance the channel response (and possibly ride extra information) at different frequencies. Incorporating a RIS at both transmitter and receiver sides provides even stronger manipulation that potentially align both direct-indirect and forward-backward spaces simultaneously.

Appendix

A. Proof of Lemma 1

Let $\mathbf{H}=\sum_n \mathbf{u}_n\sigma_n\mathbf{v}_n^{\mathsf{H}}$ be the compact SVD of the equivalent channel. Since the singular vectors are orthonormal, the n-th singular value can be expressed as

$$\sigma_n=\mathbf{u}_n^{\mathsf{H}}\mathbf{H}\mathbf{v}_n=\mathbf{u}_n^{\mathsf{T}}\mathbf{H}^*\mathbf{v}_n^*, \tag{45}$$

whose differential with respect to $\mathbf{\Theta}_g^*$ is

$$\begin{aligned}\partial\sigma_n&=\partial\mathbf{u}_n^{\mathsf{T}}\underbrace{\mathbf{H}^*\mathbf{v}_n^*}_{\sum_m \mathbf{u}_m^*\sigma_m\mathbf{v}_m^{\mathsf{T}}\mathbf{v}_n}+\mathbf{u}_n^{\mathsf{T}}\cdot\partial\mathbf{H}^*\cdot\mathbf{v}_n^*+\underbrace{\mathbf{u}_n^{\mathsf{T}}\mathbf{H}^*}_{\mathbf{u}_n^{\mathsf{T}}\sum_m \mathbf{u}_m^*\sigma_m\mathbf{v}_m^{\mathsf{T}}}\partial\mathbf{v}_n^*\\&=\underbrace{\partial\mathbf{u}_n^{\mathsf{T}}\mathbf{u}_n^*}_{\partial 1=0}\cdot\sigma_n+\mathbf{u}_n^{\mathsf{T}}\cdot\partial\mathbf{H}^*\cdot\mathbf{v}_n^*+\sigma_n\cdot\underbrace{\mathbf{v}_n^{\mathsf{T}}\partial\mathbf{v}_n^*}_{\partial 1=0}\\&=\mathbf{u}_n^{\mathsf{T}}\mathbf{H}_{\mathrm{B},g}^*\cdot\partial\mathbf{\Theta}_g^*\cdot\mathbf{H}_{\mathrm{F},g}^*\mathbf{v}_n^*\\&=\mathrm{tr}(\mathbf{H}_{\mathrm{F},g}^*\mathbf{v}_n^*\mathbf{u}_n^{\mathsf{T}}\mathbf{H}_{\mathrm{B},g}^*\cdot\partial\mathbf{\Theta}_g^*).\end{aligned}$$

According to [53], the corresponding complex derivative is

$$\frac{\partial\sigma_n}{\partial\mathbf{\Theta}_g^*}=\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{u}_n\mathbf{v}_n^{\mathsf{H}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}. \tag{46}$$

A linear combination of (46) yields (21).

B. Proof of Proposition 1

The scattering matrix of BD-RIS can be decomposed as

$$\mathbf{\Theta}=\mathbf{L}\mathbf{\Theta}_{\mathrm{D}}\mathbf{R}^{\mathsf{H}}, \tag{47}$$

where $\mathbf{\Theta}_{\mathrm{D}}\in\mathbb{U}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$ corresponds to diagonal RIS and $\mathbf{L},\mathbf{R}\in\mathbb{U}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$ are block-diagonal matrices of $L\times L$ unitary blocks. Manipulating $\mathbf{L}$ and $\mathbf{R}$ rotates the linear spans of $\bar{\mathbf{H}}_{\mathrm{B}}\triangleq\mathbf{H}_{\mathrm{B}}\mathbf{L}$ and $\bar{\mathbf{H}}_{\mathrm{F}}\triangleq\mathbf{R}^{\mathsf{H}}\mathbf{H}_{\mathrm{F}}$ and maintains their rank. On the other hand, there exists a $\mathbf{\Theta}_{\mathrm{D}}$ such that

$$\begin{aligned}\mathrm{rank}(\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}})&=\min\big(\mathrm{rank}(\mathbf{H}_{\mathrm{B}}),\mathrm{rank}(\mathbf{\Theta}_{\mathrm{D}}),\mathrm{rank}(\mathbf{H}_{\mathrm{F}})\big)\\&=\min\big(\mathrm{rank}(\bar{\mathbf{H}}_{\mathrm{B}}),N_{\mathrm{S}},\mathrm{rank}(\bar{\mathbf{H}}_{\mathrm{F}})\big)\\&=\max_{\mathbf{\Theta}}\ \mathrm{rank}(\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}})\end{aligned}$$

The same result holds if the direct link is present.

C. 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 compact SVD of the forward channel. The channel Gram matrix $\mathbf{G}\triangleq\mathbf{H}\mathbf{H}^{\mathsf{H}}$ can be written as

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

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 $\mathbf{Z}\triangleq\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{U}_{\mathrm{F}}\mathbf{\Sigma}_{\mathrm{F}}+\mathbf{H}_{\mathrm{D}}\mathbf{V}_{\mathrm{F}}\in\mathbb{C}^{N_{\mathrm{R}}\times k}$. That is to say, $\mathbf{G}$ can be expressed as a Hermitian matrix plus k rank-1 perturbations. According to the Cauchy interlacing formula [65], the n-th eigenvalue of $\mathbf{G}$ is bounded by

$$\lambda_n(\mathbf{G})\leq\lambda_{n-k}(\mathbf{Y}),\qquad\text{if } n>k, \tag{48}$$

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

Since $\mathbf{Y}=\mathbf{T}\mathbf{T}^{\mathsf{H}}$ is positive semi-definite, taking the square roots of (48) and (49) gives (22a) and (22b).

D. Proof of Proposition 3

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

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

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 $\mathbf{X}$

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

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

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

where $\bar{\mathbf{U}}_{\mathrm{B/F}}$ and $\bar{\mathbf{V}}_{\mathrm{B/F}}$ are arbitrary unitary matrices.

E. Proof of Lemma 2

The differential of R with respect to $\mathbf{\Theta}_g^*$ is [53]

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

and the corresponding complex derivative is (43).

$$2\Re\Big\{\sum_g \mathrm{tr}(\tilde{\boldsymbol{\Theta}}_g^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}) + \sum_{g_1,g_2} \mathrm{tr}(\tilde{\boldsymbol{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\boldsymbol{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}})\Big\} \geq 2\Re\Big\{\sum_g \mathrm{tr}(\boldsymbol{\Theta}_g^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}) + \sum_{g_1,g_2} \mathrm{tr}(\boldsymbol{\Theta}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\boldsymbol{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}})\Big\} \tag{56}$$

$$\sum_{g_1,g_2} \mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\tilde{\boldsymbol{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\tilde{\boldsymbol{\Theta}}_{g_2}\mathbf{H}_{\mathrm{F},g_2}) - 2\Re\Big\{\sum_{g_1,g_2} \mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\tilde{\boldsymbol{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\boldsymbol{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2})\Big\} + \sum_{g_1,g_2} \mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\boldsymbol{\Theta}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\boldsymbol{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}) \geq 0 \tag{57}$$

F. Proof of Proposition 4

The differential of (35a) with respect to $\boldsymbol{\Theta}_g^*$ is

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

and the corresponding complex derivative is

$$\frac{\partial\|\mathbf{H}\|_{\mathrm{F}}^2}{\partial\boldsymbol{\Theta}_g^*} = \mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\boldsymbol{\Theta}\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} = \mathbf{M}_g. \tag{52}$$

First, we approximate the quadratic objective (35a) by its local Taylor expansion

$$\max_{\boldsymbol{\Theta}} \quad \sum_g 2\Re\big\{\mathrm{tr}(\boldsymbol{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\big\} \tag{53a}$$

$$\text{s.t.} \quad \boldsymbol{\Theta}_g^{\mathsf{H}}\boldsymbol{\Theta}_g = \mathbf{I}, \quad \forall g. \tag{53b}$$

Let $\mathbf{M}_g = \mathbf{U}_g\boldsymbol{\Sigma}_g\mathbf{V}_g^{\mathsf{H}}$ be the compact SVD of $\mathbf{M}_g$. We have

$$\Re\big\{\mathrm{tr}(\boldsymbol{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\big\} = \Re\big\{\mathrm{tr}(\boldsymbol{\Sigma}_g\mathbf{V}_g^{\mathsf{H}}\boldsymbol{\Theta}_g^{\mathsf{H}}\mathbf{U}_g)\big\} \leq \mathrm{tr}(\boldsymbol{\Sigma}_g). \tag{54}$$

The upper bound is tight when $\mathbf{V}_g^{\mathsf{H}}\boldsymbol{\Theta}_g^{\mathsf{H}}\mathbf{U}_g = \mathbf{I}$, which implies the optimal solution of (53) is $\tilde{\boldsymbol{\Theta}}_g = \mathbf{U}_g\mathbf{V}_g^{\mathsf{H}}$, $\forall g$.

Next, we prove that solving (53) successively does not decrease (35a). Since $\tilde{\boldsymbol{\Theta}}$ optimal for problem (53), we have $\sum_g 2\Re\big\{\mathrm{tr}(\tilde{\boldsymbol{\Theta}}_g^{\mathsf{H}}\mathbf{M}_g)\big\} \geq \sum_g 2\Re\big\{\mathrm{tr}(\boldsymbol{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\big\}$ which is explicitly expressed by (56). On the other hand, expanding $\|\sum_g \mathbf{H}_{\mathrm{B},g}\tilde{\boldsymbol{\Theta}}_g\mathbf{H}_{\mathrm{F},g} - \sum_g \mathbf{H}_{\mathrm{B},g}\boldsymbol{\Theta}_g\mathbf{H}_{\mathrm{F},g}\|_{\mathrm{F}}^2 \geq 0$ gives (57). Adding (56) and (57), we have

$$\begin{aligned}&2\Re\Big\{\mathrm{tr}(\tilde{\boldsymbol{\Theta}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}}^{\mathsf{H}})\Big\} + \mathrm{tr}(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\tilde{\boldsymbol{\Theta}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}\tilde{\boldsymbol{\Theta}}\mathbf{H}_{\mathrm{F}})\\ &\geq 2\Re\Big\{\mathrm{tr}(\boldsymbol{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}}^{\mathsf{H}})\Big\} + \mathrm{tr}(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\boldsymbol{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}\boldsymbol{\Theta}\mathbf{H}_{\mathrm{F}}),\end{aligned} \tag{55}$$

which suggests that updating $\tilde{\boldsymbol{\Theta}}$ does not decrease (35a).

Finally, we prove that the converging point of (53), denoted by $\tilde{\boldsymbol{\Theta}}^?$, is a stationary point of (35). The Karush-Kuhn-Tucker (KKT) conditions of (35) and (53) are equivalent in terms of primal/dual feasibility and complementary slackness, while the stationary conditions are respectively, $\forall g$,

$$\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\boldsymbol{\Theta}^\star\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} - \boldsymbol{\Theta}_g^\star\boldsymbol{\Lambda}_g^{\mathsf{H}} = 0, \tag{58}$$

$$\mathbf{M}_g - \boldsymbol{\Theta}_g^\star\boldsymbol{\Lambda}_g^{\mathsf{H}} = 0. \tag{59}$$

On convergence, (59) becomes $\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\boldsymbol{\Theta}^?\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} - \boldsymbol{\Theta}_g^?\boldsymbol{\Lambda}_g^{\mathsf{H}} = 0$ and reduces to (58). 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] E. Basar, M. Wen, R. Mesleh, M. D. Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16 693–16 746, Aug 2017.

[12] 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.

[13] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Transactions on Communications*, vol. 68, pp. 3187–3196, May 2020.

[14] 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.

[15] 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.

[16] 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.

[17] H. Yang, H. Ding, K. Cao, M. Elkashlan, H. Li, and K. Xin, "A RIS-segmented symbiotic ambient backscatter communication system," *IEEE Transactions on Vehicular Technology*, vol. 73, pp. 812–825, Jan 2024.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] Y. Zheng, T. Lin, and Y. Zhu, "Passive beamforming for IRS-assisted MU-MIMO systems with one-bit ADCs: An SER minimization design approach," *IEEE Communications Letters*, vol. 26, pp. 1101–1105, May 2022.
[27] 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.
[28] A. H. Bafghi, V. Jamali, M. Nasiri-Kenari, and R. Schober, "Degrees of freedom of the K-user interference channel assisted by active and passive IRSs," *IEEE Transactions on Communications*, vol. 70, pp. 3063–3080, May 2022.
[29] S. Zheng, B. Lv, T. Zhang, Y. Xu, G. Chen, R. Wang, and P. C. Ching, "On DoF of active RIS-assisted MIMO interference channel with arbitrary antenna configurations: When will RIS help?" *IEEE Transactions on Vehicular Technology*, Dec 2023.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] ——, "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.
[36] H.-R. Ahn, *Asymmetric Passive Components in Microwave Integrated Circuits*. Hoboken, NJ, USA: Wiley, 2006.
[37] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Transactions on Wireless Communications*, vol. 22, pp. 2311–2324, Apr 2023.
[38] ——, "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.
[39] H. Li, S. Shen, Y. Zhang, and B. Clerckx, "Channel estimation and beamforming for beyond diagonal reconfigurable intelligent surfaces," *arXiv:2403.18087*, 2024.
[40] 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.
[41] H. Li, M. Nerini, S. Shen, and B. Clerckx, "Wideband modeling and beamforming for beyond diagonal reconfigurable intelligent surfaces," *arXiv:2403.12893*, 2024.
[42] 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.
[43] 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.
[44] M. Soleymani, I. Santamaria, E. Jorswieck, and B. Clerckx, "Optimization of rate-splitting multiple access in beyond diagonal RIS-assisted URLLC systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, Jul 2024.
[45] 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.
[46] A. Mishra, Y. Mao, C. D'Andrea, S. Buzzi, and B. Clerckx, "Transmitter side beyond-diagonal reconfigurable intelligent surface for massive MIMO networks," *IEEE Wireless Communications Letters*, vol. 13, pp. 352–356, Feb 2024.
[47] M. Nerini, S. Shen, and B. Clerckx, "Discrete-value group and fully connected architectures for beyond diagonal reconfigurable intelligent surfaces," *IEEE Transactions on Vehicular Technology*, vol. 72, pp. 16 354–16 368, Dec 2023.
[48] 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.
[49] 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.
[50] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2009.
[51] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. L. Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, pp. 883–917, Aug 2022.
[52] M. T. Ivrlac and J. A. Nossek, "Toward a circuit theory of communication," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, pp. 1663–1683, Jul 2010.
[53] 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.
[54] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, Sep 2006.
[55] L. Armijo, "Minimization of functions having lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, pp. 1–3, Jan 1966.
[56] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, pp. 303–353, Jan 1998.
[57] 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.
[58] 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.
[59] W. Fulton, "Eigenvalues, invariant factors, highest weights, and schubert calculus," *Bulletin of the American Mathematical Society*, vol. 37, pp. 209–249, Apr 2000.
[60] 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.
[61] L. Hogben, Ed., *Handbook of Linear Algebra*. Boca Raton, FL, USA: CRC press, 2013.
[62] 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.
[63] J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*. Oxford, UK: Oxford University Press, 2004.
[64] 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.
[65] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins University Press, 2013.
[66] 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.
[67] 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.
[68] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, May 2005.