Channel Shaping Using Reconfigurable Intelligent Surfaces: From Diagonal Model to Beyond

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

***Abstract*—This paper investigates how a passive Reconfigurable Intelligent Surface (RIS) can manipulate the Multiple-Input Multiple-Output (MIMO) point-to-point channel in terms of singular values and power gain. We depart from the widely-adopted diagonal phase shift model to a Beyond-Diagonal (BD) architecture, which enables cooperative wave scattering thanks to in-group connections between elements. An efficient and universal BD-RIS design framework is proposed over group-wise geodesic Riemannian Conjugate Gradient (RCG) in the Stiefel manifold, which performs multiplicative rotational (instead of add-then-retract) updates for faster convergence. For the singular value problem, we numerically characterize their Pareto frontiers by geodesic RCG and derive analytical bounds for rank-deficient channels and fully-connected (a.k.a. unitary) RIS. For the power gain problem, we provide a closed-form iterative solution by group-wise Singular Value Decomposition (SVD) that features even faster convergence. Those channel shaping designs have two major implications. First, they demystify the performance gain from off-diagonal entries in terms of subchannel rearrangement and subspace alignment. Second, they inspire a two-stage suboptimal solution that decouples joint active and passive beamforming design. As a side product, we tackle the BD-RIS-aided MIMO rate maximization problem by a local-optimal Alternating Optimization (AO) approach and a two-stage shaping-inspired approach. Simulation results indicate that the rate deficit from the latter diminishes as the RIS evolves from diagonal towards unitary, while the relative gain of BD-RIS increase with MIMO dimensions and decrease with spatial correlation.**

***Index Terms*—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 metasurface 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 not only experiences negligible noise and supports full-duplex transmissions, but also features better flexibility than reflectarrays, lighter footprint than various relays, and greater scalability than conventional multi-antenna techniques. The most popular RIS application is *joint beamforming* design with transceivers, which has attracted significant attention in wireless communication [2]–[4], sensing [5]–[7], and power transfer literature [8]–[10]. Since coherent scattering can be decomposed as equal-gain combining and transmission, RIS elements usually offer a squared asymptotic performance (e.g., second-order array gain and fourth-order power scaling law [10]) than transceiver antennas. On the other hand, RIS can also be employed for *backscatter modulation* by periodically switching its reflection pattern. This creates a free-ride message stream (similar to index modulation [11]) that can be either integrated into the legacy transmitter for enhanced channel capacity [12]–[14], or exploited by surrounding low-power devices for energy-efficient uplink communications [15]–[17].

Despite the great potentials and fruitful outcomes, one critical unanswered question is the channel shaping capability: *To what extent can a passive RIS reshape Multiple-Input Multiple-Output (MIMO) channels?* The answer depends heavily on the scattering model and hardware architecture. Most existing works assumed that each scattering element is tuned by a dedicated impedance and acts as an *individual* phase shifter [18]. This ideally translates to a diagonal scattering matrix with unit-magnitude entries, which only controls the phase of the scattered waves but cannot redistribute their amplitude. The concept was soon generalized to Beyond-Diagonal (BD)-RIS [19] which physically groups adjacent elements using passive reconfigurable components.[1] This allows *cooperative* scattering — wave impinging on one element can propagate losslessly within the circuit and depart partially from any element in the same group. It can thus control both the amplitude and phase of the scattered wave without power loss, generalizing the scattering matrix to block-diagonal with unitary blocks. Such a powerful model can be realized at reduced hardware complexity using tree- and forest-connected architectures using graph theory [21]. BD-RIS can also function in hybrid transmitting-and-reflecting mode [22] and multi-sector mode [23] for full-space coverage and multi-user support. Many practical design challenges have been addressed including channel estimation [24], mutual coupling [25], and wideband modelling [26]. Its superiority has been studied extensively in Single-Input Single-Output (SISO) and Multiple-Input Single-Output (MISO) systems in terms of single-user Signal-to-Noise Ratio (SNR) maximization [19], [21], [27], [28] and multi-user Weighted Sum-Rate (WSR)

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).

[1]Those components can be either symmetric (e.g., capacitors and inductors) or asymmetric (e.g., ring hybrids and branch-line hybrids) [20], resulting in symmetric and asymmetric scattering matrices respectively.

maximization [23], [29]–[31]. When it comes to MIMO, most existing works are limited to very specific scenarios. For example, the authors of [32] investigated the rate-optimal joint beamforming design under the assumption of unitary RIS without direct link. Received power maximization problem is extended to MIMO for continuous-valued [27] and discrete-valued [33] scattering matrices, but they only considered single-stream transmission with given precoder and combiner (which is equivalent to SISO). Transmitter-side BD-RIS design over statistical Channel State Information (CSI) has been introduced to massive MIMO networks for improved spectral efficiency [34], but was again constrained to unitary RIS without direct link. A practical frequency-dependent BD-RIS model has been recently proposed for multi-band multi-cell MIMO networks to facilitate practical deployments [35]. Although the results are insightful, it remains unclear how the off-diagonal entries contribute to channel shaping and whether the relative gain of BD-RIS over diagonal RIS scales with MIMO dimensions.

Channel shaping is different from passive beamforming since the RIS is exploited as a stand-alone device to modify the inherent properties of the channel itself, instead of optimizing for a specific performance metric. This not only allows one to decouple the joint RIS-transceiver design, but also unveils the fundamental limits of wave manipulation and provides an all-in-one tool for various wireless applications. Relevant designs can be classified into two categories:

- *Singular value centric:* Focus on the singular values of the equivalent channel matrix, which are sensitive to minor perturbations but precisely related to the capacity and diversity-multiplexing trade-off. It has been mainly studied in terms of minimum singular value [36], effective rank [36], [37], condition number [38], [39], and degree of freedom [40]–[42].
- *Power centric:* Focus on the second-order characteristics of the equivalent channel matrix, which are useful in SISO and MISO but less informative in MIMO. It has been mainly studied in terms of channel power gain [2], [19], [21], [27], [28] and leakage interference [43].

While the above works offer promising glimpses into the channel shaping potential of passive RIS, they neither suggest the achievable singular value region nor address the most general BD-RIS-aided MIMO scenario with direct link and arbitrary group size. On the other hand, the great potential of BD-RIS (especially in MIMO) has not been fully understood and an efficient design framework is still missing. This paper aims for a comprehensive answer to the channel shaping question through theoretical analysis and numerical optimization. The contributions are summarized below.

First, we demystify the gain from off-diagonal entries of BD-RIS in terms of subchannel rearrangement and subspace alignment. Subchannel rearrangement is a unique feature of cooperative scattering that allows each group to rearrange and recombine the associated forward and backward subchannels. This translates to a higher design flexibility and better shaping capability by exploiting the spatial diversity. On the other hand, phase matching in SISO generalizes to subspace alignment in MIMO and introduces a trade-off between the multiplicative backward-forward channel combination and the additive direct-indirect channel combination. As the MIMO dimensions expand, we show the benefits of BD-RIS in subchannel rearrangement and the limitations of diagonal RIS in subspace alignment become increasingly apparent. This is the first paper to unveil and interpret the potential of BD-RIS in multi-antenna systems.

Second, we exploit the Riemannian geometry of the Stiefel manifold and propose an efficient BD-RIS design framework based on geodesic[2] Riemannian Conjugate Gradient (RCG). This method modified from [44], [45] not only features lower complexity and faster convergence than general non-geodesic approach [46], [47], but also works for arbitrary group size and any smooth optimization problem. Specifically, block-wise multiplicative rotational updates are performed along the geodesics of the Stiefel manifold and compactly evaluated as the exponential map [48]. By exploiting the inherent structure of unitary matrices, this method avoids retractions from the Euclidean space and improves the computational efficiency and stability. This is the first work to tailor an efficient and universal optimization framework for BD-RIS.

Third, we quantify the capability of a BD-RIS to redistribute the singular values of a point-to-point MIMO channel. The Pareto frontiers are characterized by optimizing the weighted sum of singular values, where the weights can be positive, zero, or negative. This problem is solved by the proposed geodesic RCG algorithm. The resulting achievable singular value region generalizes most relevant metrics and provides an intuitive channel shaping benchmark. We also and derive some analytical singular value bounds for rank-deficient MIMO and fully-connected (a.k.a. unitary) RIS. This is the first work to comprehensively answer the channel shaping capability question from a Pareto perspective.

Fourth, we tackle BD-RIS-aided MIMO achievable rate maximization problem with two beamforming solutions: a local-optimal approach via Alternating Optimization (AO) and a low-complexity approach over channel shaping. The former iteratively updates active beamforming by eigenmode transmission and passive beamforming by geodesic RCG until convergence. The latter suboptimally decouples the joint design into a channel power gain maximization subproblem and a conventional MIMO transmission subproblem, then propose a two-stage solution in closed form. Interestingly, the rate deficit from the latter diminishes as the RIS evolves from diagonal to unitary. It suggests that channel shaping offers a promising path towards simplified and practical BD-RIS (and 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

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

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. $(\cdot)^*$, $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$, $(\cdot)^{\dagger}$ $(\cdot)^{(r)}$, $(\cdot)^{\star}$ 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 $(\cdot)_x,(\cdot)_{x+1},\ldots,(\cdot)_y$. $|\cdot|$, $\|\cdot\|$, and $\|\cdot\|_{\mathrm{F}}$ denote the absolute value, Euclidean norm, and Frobenius norm, respectively. $\mathcal{CN}(\mathbf{0},\mathbf{\Sigma})$ 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 [49] that further divides into G individual groups, each containing $L\triangleq N_{\mathrm{S}}/G$ elements interconnected by real-time reconfigurable components [19]. 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 [22], [23], [32]. The overall scattering matrix of the BD-RIS is block-unitary

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

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 $g\in\mathcal{G}\triangleq\{1,\ldots,G\}$. Note that diagonal and unitary RIS can be regarded as its extreme cases with group size $L=1$ and $L=N_{\mathrm{S}}$, respectively. Some potential physical architectures of BD-RIS are illustrated in [19, Fig. 3], [23, Fig. 5], and [21, Fig. 2], where the radiation pattern and circuit topology should 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 subchannels for RIS group g, corresponding to the $(g-1)L$ 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

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

Remark 1. *By analyzing* (4)*, we conclude that the off-diagonal entries of the BD-RIS scattering matrix provide two key potentials for MIMO channel shaping:*

- ***Subchannel rearrangement:*** *This unique feature of BD-RIS allows each group to rearrange and recombine the backward and forward subchannels by 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* $\mathcal{L}\triangleq\{1,\ldots,L\}$*. Note the first summation is over groups and the second summation is over permuted subchannels. By rearrangement inequality, the maximum is attained by pairing the* l*-th strongest backward and forward subchannels within each group. Since the number of subchannels associated with each group is proportional to* $N_{\mathrm{T}}N_{\mathrm{R}}$*, we conclude the advantage of BD-RIS in subchannel rearrangement scales with MIMO dimensions,*
- ***Subspace alignment:*** *Each group can align the singular vectors of the associated backward-forward (intra-group, multiplicative) channels and direct-indirect (inter-group, additive) channels. In SISO, subspace alignment boils down to phase matching and 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 (*$L=1$*, empty null spaces) thus suffices for perfect phase matching in SISO. When it comes to MIMO, each individual scattering element can only apply a common phase shift to the "pinhole" indirect channel* $\mathbf{H}_g\in\mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ *passing through itself. That is, the disadvantage of diagonal RIS in subspace alignment scales with MIMO dimensions. As will be shown later, even if the BD-RIS is fully-connected, there still exists a tradeoff between the alignment of direct-indirect and backward-forward subspaces.*

III. Channel Singular Values Redistribution

A. A Toy Example

We first illustrate the channel shaping capabilities of different RIS 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 unitary BD-RIS has 4 independent angular parameters

$$\mathbf{\Theta}_{\mathrm{U}}=e^{\jmath\phi}\begin{bmatrix}e^{\jmath\alpha}\cos\psi & e^{\jmath\beta}\sin\psi\\ -e^{-\jmath\beta}\sin\psi & e^{-\jmath\alpha}\cos\psi\end{bmatrix}. \tag{6}$$

In particular, ϕ has no impact on the singular value because $\mathrm{sv}(e^{\jmath\phi}\mathbf{A})=\mathrm{sv}(\mathbf{A})$. We also enforce symmetry by $\beta=\pi/2$ such that both architectures have the same number of angular parameters. 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 unitary RIS. It is observed that both singular values can be manipulated up to 9% using

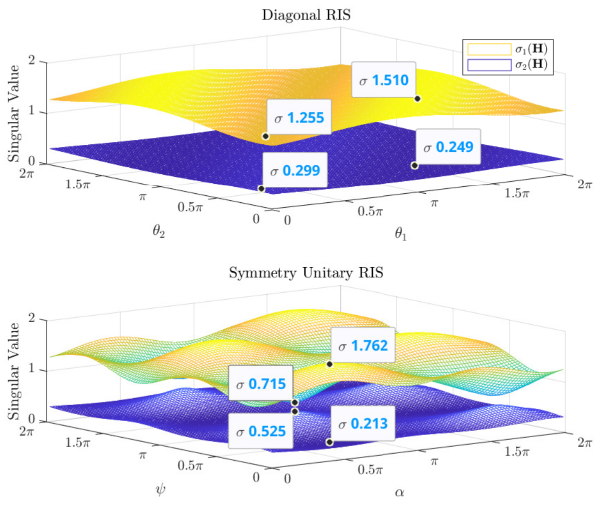


Fig. 1. $2\times2\times2$ (no direct) channel singular value shaping by diagonal and symmetry unitary RIS.

diagonal RIS and 42% using symmetric BD-RIS, despite both architectures have the same number of scattering elements and design parameters. A larger performance gap is expected when asymmetric BD-RIS is available. This example shows BD-RIS can provide a wider dynamic range of channel singular values and motivates further studies on channel shaping.

B. Pareto Frontier Characterization

We then characterize the Pareto frontier of channel singular values by maximizing their weighted sum

$$\max_{\mathbf{\Theta}} \quad \sum_n \rho_n \sigma_n(\mathbf{H}) \tag{7a}$$

$$\text{s.t.} \quad \mathbf{\Theta}_g^{\mathsf{H}}\mathbf{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{7b}$$

where $n \in \{1,\dots,N \triangleq \min(N_\mathrm{T}, N_\mathrm{R})\}$ and ρ_n is the weight of the n-th singular value that can be positive, zero, or negative. Varying $\{\rho_n\}$ unveils the entire achievable singular value region. Thus, the Pareto frontier problem (7) generalizes most relevant metrics and provides a powerful shaping framework. The objective (7a) is smooth in $\mathbf{\Theta}$ and the feasible domain (7b) for group g corresponds to the Stiefel manifold. Next, we zoom out to general smooth maximization problems of asymmetric BD-RIS.

Inspired by [44], [45], we propose a block-wise RCG algorithm along the geodesics on the Lie group of unitary matrices $\mathbb{U}^{L\times L}$. It leverages the fact that unitary matrices are closed under multiplication. At iteration r, the gradient is computed in the Euclidean space and translated to the Riemannian manifold [46]

$$\nabla_{\mathrm{E},g}^{(r)} = \frac{\partial f(\mathbf{\Theta}_g^{(r)})}{\partial \mathbf{\Theta}_g^{*}}, \tag{8}$$

$$\nabla_{\mathrm{R},g}^{(r)} = \nabla_{\mathrm{E},g}^{(r)} \mathbf{\Theta}_g^{(r)\mathsf{H}} - \mathbf{\Theta}_g^{(r)} \nabla_{\mathrm{E},g}^{(r)\mathsf{H}}. \tag{9}$$

The Polak-Ribierre parameter [50] is approximated as [45]

$$\gamma_g^{(r)} = \frac{\mathrm{tr}\big((\nabla_{\mathrm{R},g}^{(r)} - \nabla_{\mathrm{R},g}^{(r-1)})\nabla_{\mathrm{R},g}^{(r)\mathsf{H}}\big)}{\mathrm{tr}\big(\nabla_{\mathrm{R},g}^{(r-1)}\nabla_{\mathrm{R},g}^{(r-1)\mathsf{H}}\big)}, \tag{10}$$

Algorithm 1: Block-wise geodesic RCG for asymmetric BD-RIS

Input: $f(\mathbf{\Theta})$, G
Output: $\mathbf{\Theta}^\star$
Initialize $r \leftarrow 0$, $\mathbf{\Theta}^{(0)}$
Repeat
  For $g \leftarrow 1$ to G
    $\nabla_{\mathrm{E},g}^{(r)} \leftarrow$ (8)
    $\nabla_{\mathrm{R},g}^{(r)} \leftarrow$ (9)
    $\gamma_g^{(r)} \leftarrow$ (10)
    $\mathbf{D}_g^{(r)} \leftarrow$ (11)
    If $\Re\{\mathrm{tr}(\mathbf{D}_g^{(r)\mathsf{H}}\nabla_{\mathrm{R},g}^{(r)})\} < 0$ ▷ not an ascent direction
      $\mathbf{D}_g^{(r)} \leftarrow \nabla_{\mathrm{R},g}^{(r)}$
    End If
    $\mu \leftarrow 1$
    $\mathbf{G}_g^{(r)}(\mu) \leftarrow$ (12)
    While $f\big(\mathbf{G}_g^{(r)}(2\mu)\big) - f(\mathbf{\Theta}_g^{(r)}) \geq \mu \cdot \mathrm{tr}(\mathbf{D}_g^{(r)}\mathbf{D}_g^{(r)\mathsf{H}})/2$
      $\mu \leftarrow 2\mu$
    End While
    While $f\big(\mathbf{G}_g^{(r)}(\mu)\big) - f(\mathbf{\Theta}_g^{(r)}) < \mu/2 \cdot \mathrm{tr}(\mathbf{D}_g^{(r)}\mathbf{D}_g^{(r)\mathsf{H}})/2$
      $\mu \leftarrow \mu/2$
    End While
    $\mathbf{\Theta}_g^{(r+1)} \leftarrow$ (13)
  End For
  $r \leftarrow r+1$
Until $|f(\mathbf{\Theta}^{(r)}) - f(\mathbf{\Theta}^{(r-1)})| / f(\mathbf{\Theta}^{(r-1)}) \leq \epsilon$

and the conjugate direction is

$$\mathbf{D}_g^{(r)} = \nabla_{\mathrm{R},g}^{(r)} + \gamma_g^{(r)} \mathbf{D}_g^{(r-1)}. \tag{11}$$

In the Stiefel manifold, the geodesic emanating from $\mathbf{\Theta}_g^{(r)}$ with velocity $\mathbf{D}_g^{(r)}$ and step size μ is described compactly by the exponential map [48]

$$\mathbf{G}_g^{(r)}(\mu) = \exp(\mu \mathbf{D}_g^{(r)})\mathbf{\Theta}_g^{(r)}. \tag{12}$$

An appropriate $\mu^\star$ can be obtained by the Armijo rule [51].[3] Finally, the scattering matrix is updated along the geodesic as

$$\mathbf{\Theta}_g^{(r+1)} = \mathbf{G}_g^{(r)}(\mu^\star). \tag{13}$$

Algorithm 1 summarizes the proposed block-wise geodesic RCG method for smooth maximization problems of asymmetric BD-RIS. Convergence to stationary points is guaranteed.

Remark 2. *Compared with universal manifold optimization [46], [47], Algorithm 1 inherits a trifold benefit from [44], [45]:*

1) *No retraction thanks to rotational update* (12), (13)*;*
2) *Lower computational complexity per iteration;*
3) *Faster convergence thanks to proper parameter space.*

Lemma 1. *The Euclidean gradient of* (7a) *w.r.t. BD-RIS group g is*

$$\frac{\partial \sum_n \rho_n \sigma_n(\mathbf{H})}{\partial \mathbf{\Theta}_g^{*}} = \mathbf{H}_{\mathrm{B},g}^{\mathsf{H}} \mathbf{U} \mathrm{diag}(\rho_1,\dots,\rho_N) \mathbf{V}^{\mathsf{H}} \mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}, \tag{14}$$

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

Proof. Please refer to Appendix A. □

[3] To double the step size, one only need to square the rotation matrix instead of recomputing the matrix exponential, i.e., $\exp(2\mu\mathbf{D}_g^{(r)}) = \exp^2(\mu\mathbf{D}_g^{(r)})$.

Algorithm 1 can thus be invoked for the Pareto singular value problem (7) where line 4 uses (14) explicitly.

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, then regardless of the RIS size and architecture, the n-th singular value of the equivalent channel is bounded by*

$$\sigma_n(\mathbf{H}) \leq \sigma_{n-k}(\mathbf{T}), \qquad \textit{if } n > k, \tag{15}$$

$$\sigma_n(\mathbf{H}) \geq \sigma_n(\mathbf{T}), \qquad \textit{if } n < N-k+1, \tag{16}$$

where

$$\mathbf{T}\mathbf{T}^{\mathsf{H}} = \begin{cases} \mathbf{H}_{\mathrm{D}}(\mathbf{I}-\mathbf{V}_{\mathrm{F}}\mathbf{V}_{\mathrm{F}}^{\mathsf{H}})\mathbf{H}_{\mathrm{D}}^{\mathsf{H}}, & \textit{if } \mathrm{rank}(\mathbf{H}_{\mathrm{F}}) = k, \\ \mathbf{H}_{\mathrm{D}}^{\mathsf{H}}(\mathbf{I}-\mathbf{U}_{\mathrm{B}}\mathbf{U}_{\mathrm{B}}^{\mathsf{H}})\mathbf{H}_{\mathrm{D}}, & \textit{if } \mathrm{rank}(\mathbf{H}_{\mathrm{B}}) = k, \end{cases} \tag{17}$$

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 RIS, the first k channel singular values are unbounded above while the last k channel singular values can be suppressed to zero.*

Corollary 2.2 (Line-of-Sight (LoS) channel[4])**.** *If the forward or backward channel is LoS, then a RIS can at most enlarge (resp. suppress) the n-th ($n \geq 2$) channel singular value to the ($n-1$)-th (resp. n-th) singular value of $\mathbf{T}$, that is,*

$$\sigma_1(\mathbf{H}) \geq \sigma_1(\mathbf{T}) \geq \sigma_2(\mathbf{H}) \geq \ldots \geq \sigma_{N-1}(\mathbf{T}) \geq \sigma_N(\mathbf{H}) \geq \sigma_N(\mathbf{T}). \tag{18}$$

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

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

$$\mathrm{sv}(\mathbf{H}) = \mathrm{sv}(\mathbf{B}\mathbf{F}), \tag{19}$$

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 $N' = \max(N_{\mathrm{T}}, N_{\mathrm{S}}, N_{\mathrm{R}})$ and $\sigma_n(\mathbf{H}) = \sigma_n(\mathbf{H}_{\mathrm{F}}) = \sigma_n(\mathbf{H}_{\mathrm{B}}) = 0$ for $N < n \leq N'$. We have the following corollaries.

[4]A similar result has been derived for diagonal RIS in [52].

Corollary 3.1 (generic singular value bounds [53])**.**

$$\prod_{k\in K}\sigma_k(\mathbf{H}) \leq \prod_{i\in I}\sigma_i(\mathbf{H}_{\mathrm{B}})\prod_{j\in J}\sigma_j(\mathbf{H}_{\mathrm{F}}), \tag{20}$$

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

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

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

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

$$\sigma_1(\mathbf{H}) \leq \sigma_1(\mathbf{H}_{\mathrm{B}})\sigma_1(\mathbf{H}_{\mathrm{F}}). \tag{21}$$

Corollary 3.3 (upper bound on the product of first k singular values)**.**

$$\prod_{n=1}^{k}\sigma_n(\mathbf{H}) \leq \prod_{n=1}^{k}\sigma_n(\mathbf{H}_{\mathrm{B}})\prod_{n=1}^{k}\sigma_n(\mathbf{H}_{\mathrm{F}}). \tag{22}$$

Corollary 3.4 (upper bound on the sum of first k singular values to the power of p)**.**

$$\sum_{n=1}^{k}\sigma_n^p(\mathbf{H}) \leq \sum_{n=1}^{k}\sigma_n^p(\mathbf{H}_{\mathrm{B}})\sigma_n^p(\mathbf{H}_{\mathrm{F}}), \quad p > 0. \tag{23}$$

When $k = N'$ and $p = 2$, (23) *suggests the channel power is upper bounded by the sum of (sorted) element-wise power product of backward and forward subchannels.*

Tight bounds are inapplicable when a MIMO direct link is present, as the RIS needs to balance the direct-indirect (additive) and backward-forward (multiplicative) subspace alignments. Such a balance often involves optimization approaches and another example will be discussed in Section IV-B.

IV. Achievable Rate Maximization

The MIMO achievable rate maximization problem is formulated w.r.t. joint active and passive beamforming

$$\max_{\mathbf{W},\mathbf{\Theta}} \quad R = \log\det\left(\mathbf{I} + \frac{\mathbf{W}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{W}}{\eta}\right) \tag{24a}$$

$$\text{s.t.} \quad \|\mathbf{W}\|_{\mathrm{F}}^2 \leq P, \tag{24b}$$

$$\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{\Theta}_g = \mathbf{I}, \quad \forall g, \tag{24c}$$

where $\mathbf{W}$ is the transmit precoder, R is the achievable rate, η is the noise power, and P is the transmit power budget. Two methods are proposed below to solve problem (24).

A. Alternating Optimization

Consider an AO approach that updates $\mathbf{\Theta}$ and $\mathbf{W}$ iteratively. For a given $\mathbf{W}$, the passive beamforming subproblem is

$$\max_{\mathbf{\Theta}} \quad \log\det\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}}{\eta}\right) \tag{25a}$$

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

where $\mathbf{Q} \triangleq \mathbf{W}\mathbf{W}^{\mathsf{H}}$ is the transmit covariance matrix.

Lemma 2. *The Euclidean gradient of* (25a) *w.r.t. BD-RIS block g is*

$$\frac{\partial R}{\partial \mathbf{\Theta}_g^*} = \frac{1}{\eta}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}}{\eta}\right)^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}. \tag{26}$$

Proof. Please refer to Appendix E. □

Algorithm 1 is then invoked to solve problem (24) where line 4 uses (26) explicitly. Since (25a) is a concave function of $\mathbf{\Theta}$, convergence to local-optimal points is guaranteed. On the other hand, the global optimal transmit precoder for a fixed $\mathbf{\Theta}$ is given by the eigenmode transmission [54]

$$\mathbf{W}^\star = \mathbf{V}\mathbf{S}^{\star 1/2}, \tag{27}$$

where $\mathbf{V}$ is the right channel singular matrix and $\mathbf{S}^\star$ is the optimal water-filling power allocation matrix. The overall AO algorithm converges to local-optimal points of problem (24) since each subproblem is solved optimally and the objective is bounded above.

B. Low-Complexity Solution

We then propose a low-complexity solution to problem (24) based on channel shaping. The passive beamforming subproblem (25) involves transmit covariance matrix $\mathbf{Q}$ and thus requires iterative RCG update. Instead, we decouple the joint RIS-transceiver design by recasting (25) as channel power maximization

$$\max_{\mathbf{\Theta}} \quad \|\mathbf{H}_\mathrm{D} + \mathbf{H}_\mathrm{B}\mathbf{\Theta}\mathbf{H}_\mathrm{F}\|_\mathrm{F}^2 \tag{28a}$$

$$\text{s.t.} \quad \mathbf{\Theta}_g^{\mathsf{H}}\mathbf{\Theta}_g = \mathbf{I}, \quad \forall g. \tag{28b}$$

Remark 3. *As mentioned in Section III-C, the key of solving* (28) *is to balance the additive and multiplicative subspace alignments. Problem* (28) *is very similar (in terms of maximizing the inner product of $\mathbf{H}_\mathrm{D}$ and $\mathbf{H}_\mathrm{B}\mathbf{\Theta}\mathbf{H}_\mathrm{F}$) to the weighted orthogonal Procrustes problem [55]*

$$\min_{\mathbf{\Theta}} \quad \|\mathbf{H}_\mathrm{D} - \mathbf{H}_\mathrm{B}\mathbf{\Theta}\mathbf{H}_\mathrm{F}\|_\mathrm{F}^2 \tag{29a}$$

$$\text{s.t.} \quad \mathbf{\Theta}^{\mathsf{H}}\mathbf{\Theta} = \mathbf{I}, \tag{29b}$$

which has no trivial solution. One lossy transformation, by moving $\mathbf{\Theta}$ to one side [56], formulates standard orthogonal Procrustes problems

$$\min_{\mathbf{\Theta}} \quad \|\mathbf{H}_\mathrm{B}^{\dagger}\mathbf{H}_\mathrm{D} - \mathbf{\Theta}\mathbf{H}_\mathrm{F}\|_\mathrm{F}^2 \text{ or } \|\mathbf{H}_\mathrm{D}\mathbf{H}_\mathrm{F}^{\dagger} - \mathbf{H}_\mathrm{B}\mathbf{\Theta}\|_\mathrm{F}^2 \tag{30a}$$

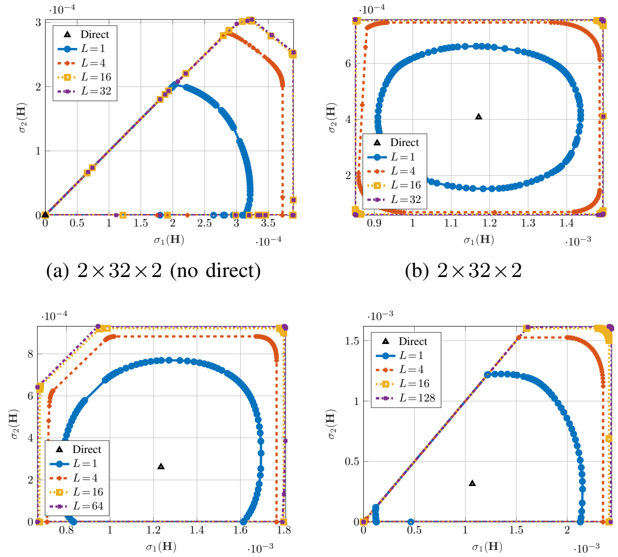
$$\text{s.t.} \quad \mathbf{\Theta}^{\mathsf{H}}\mathbf{\Theta} = \mathbf{I}, \tag{30b}$$

which has global optimal solutions

$$\mathbf{\Theta}^\star = \mathbf{U}\mathbf{V}^{\mathsf{H}} \tag{31}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular matrices of $\mathbf{H}_\mathrm{B}^{\dagger}\mathbf{H}_\mathrm{D}\mathbf{H}_\mathrm{F}^{\mathsf{H}}$ or $\mathbf{H}_\mathrm{B}^{\mathsf{H}}\mathbf{H}_\mathrm{D}\mathbf{H}_\mathrm{F}^{\dagger}$ [57]. This suboptimal solution only applies to fully-connected BD-RIS.

Inspired by [58], we propose a general solution to problem (28) with arbitrary group size. The idea is to successively approximate the quadratic objective (28a) by local Taylor expansions and solve each step in closed form.



(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.

Proposition 4. *Starting from any $\mathbf{\Theta}^{(0)} \in \mathbb{U}^{N_\mathrm{S}\times N_\mathrm{S}}$, the sequence*

$$\mathbf{\Theta}_g^{(r+1)} = \mathbf{U}_g^{(r)}\mathbf{V}_g^{(r)}, \quad \forall g. \tag{32}$$

converges to a stationary point of (28)*, where $\mathbf{U}_g^{(r)}$ and $\mathbf{V}_g^{(r)}$ are the left and right compact singular matrix of*

$$\mathbf{M}_g^{(r)} = \mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\Big(\mathbf{H}_\mathrm{D} + \mathbf{H}_\mathrm{B}\mathrm{diag}\big(\mathbf{\Theta}_{[1:g-1]}^{(r+1)}, \mathbf{\Theta}_{[g:G]}^{(r)}\big)\mathbf{H}_\mathrm{F}\Big)\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} \tag{33}$$

Proof. Please refer to Appendix F. □

Once the channel shaping problem (28) is solved, the transmit precoder can be obtained by (27). This two-stage approach decouples both blocks and is computationally efficient.

V. Simulation Results

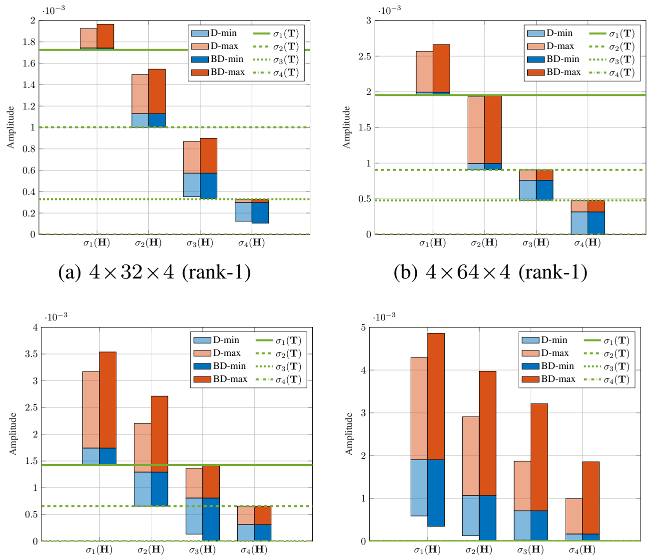
In this section, we provide numerical results to evaluate the proposed BD-RIS designs. Consider a distance-dependent path loss model $\Lambda(d) = \Lambda_0 d^{-\gamma}$ where Λ_0 is the reference path loss at distance $1\,\mathrm{m}$, d is the propagation distance, and γ is the path loss exponent. The 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. $\kappa \to \infty$ is assumed for all channels unless otherwise specified.

A. 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

TABLE I
AVERAGE PERFORMANCE OF BD-RIS DESIGNS

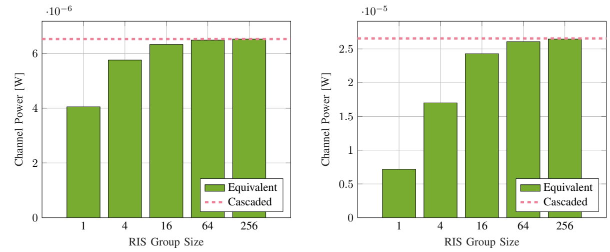
RCG path	$N_\mathrm{S}=16$			$N_\mathrm{S}=256$		
	Objective	Iterations	Time [s]	Objective	Iterations	Time [s]
Geodesic	4.355×10^{-3}	11.61	2.038×10^{-2}	1.164×10^{-2}	25.78	3.216
Non-geodesic	4.168×10^{-3}	169.5	1.420×10^{-1}	8.873×10^{-3}	278.1	27.81



(a) $4\times32\times4$ (rank-1) (b) $4\times64\times4$ (rank-1)

(c) $4\times128\times4$ (rank-2) (d) $4\times256\times4$ (rank-4)

Fig. 3. Achievable channel singular values: analytical bounds (green lines) and numerical optimization results (blue and red bars). 'D' means diagonal RIS and 'BD' means fully-connected BD-RIS. 'rank-k' refers to the forward channel.



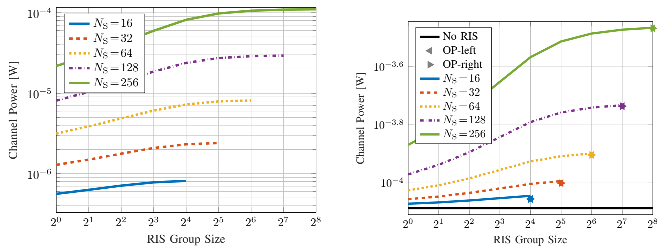
(a) $1\times256\times1$ (no direct) (b) $4\times256\times4$ (no direct)

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 subchannels.

a tradeoff between aligning the two subspaces. We observe that the smallest singular value is enhanced up to 2×10^{-4} by diagonal RIS and 3×10^{-4} by fully-connected BD-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 provides a better tradeoff in subchannel manipulation but also improves the dynamic range of $\sigma_1(\mathbf{H})$ and $\sigma_2(\mathbf{H})$ by $22\,\%$ and $38\,\%$, respectively. The achievable 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 by a 128-element BD-RIS but not by a diagonal one. Those results demonstrate the superior channel shaping capability of BD-RIS for better signal enhancement and interference suppression.

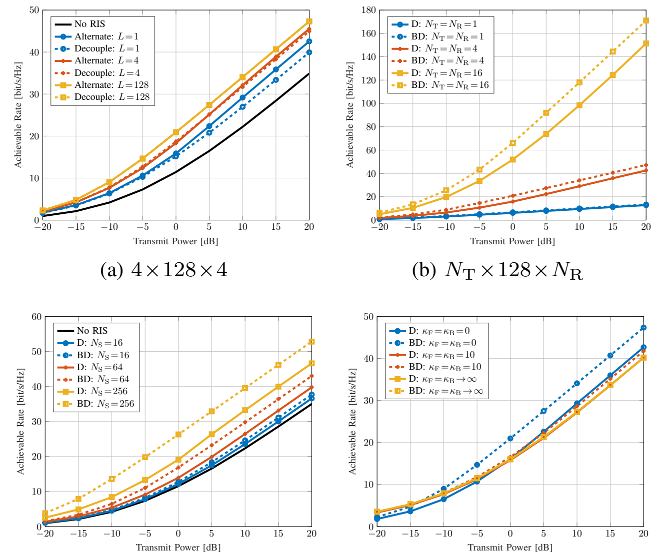
2) Analytical Bounds and Numerical Results: Fig. 3 illustrates the analytical singular value bounds in Proposition 2 and the numerical results obtained by solving problem (7) with $\rho_n=\pm1$ and $\rho_{n'}=0, \forall n'\neq n$. Here we assme a rank-$k$ forward channel without loss of generality. When 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 (17) where the compact singular matrix $\mathbf{V}_\mathrm{F}$ 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.

Fig. 4 compares the analytical channel power bound in Corollary 3.4 with $k=N'$, $p=2$ and the numerical results obtained by solving problem (28) when the direct link is absent. Here, a fully-connected BD-RIS can attain the upper bound either in closed form (**??**) or via optimization approach (32). 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 [19]. Interestingly, the gain surges to $270\,\%$ in 4T4R MIMO as shown in Fig. 4(b). This is because subspace alignment boils down to phase matching in SISO such that both triangular and Cauchy-Schwarz inequalities in [19, (50)] can be simultaneously tight regardless of the group size. That is, diagonal RIS is sufficient for subspace alignment in SISO while the $62.5\,\%$ gain from BD-RIS comes purely from subchannel rearrangement (i.e., pairing the forward and backward channels from strongest to weakest). Now consider a diagonal RIS



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

Fig. 5. Average maximum channel power versus RIS configuration. 'OP-left' and 'OP-right' refer to the suboptimal solutions to problem (28) by lossy transformation (30) where $\mathbf{\Theta}$ 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. 6. Average achievable rate versus MIMO and RIS configurations. The noise power is $\eta = -75\,\text{dB}$, corresponding to a direct SNR of -10 to $30\,\text{dB}$. 'Alternate' refers to the alternating optimization and 'Decouple' refers to the low-complexity design. 'D' means diagonal RIS and 'BD' means fully-connected BD-RIS.

in MIMO. Each element can only apply a common phase shift to the associated rank-1 $N_R \times N_T$ indirect channel. Therefore, perfect subspace alignment of indirect channels through different elements is generally impossible. It means the disadvantage of diagonal RIS in subspace alignment and subchannel rearrangement scales with MIMO dimensions. We thus conclude that the power gain of BD-RIS scales with group size and MIMO dimensions.

B. Achievable Rate Maximization

We first focus on channel shaping subproblem (42). 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 subspace alignment advantage of BD-RIS becomes more pronounced. We also notice that the suboptimal solutions (31) for fully-connected BD-RIS by lossy transformation (30) 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 $10\,\text{dB}$, Fig. 6(a) shows that introducing a 128-element diagonal RIS to 4T4R MIMO can improve the achievable rate from $22.2\,\text{bps/Hz}$ to $29.2\,\text{bps/Hz}$ ($+31.5\%$). In contrast, a BD-RIS of group size 4 and 128 can further improve the rate to $32.1\,\text{bps/Hz}$ ($+44.6\%$) and $34\,\text{bps/Hz}$ ($+53.2\%$), respectively. Interestingly, the gap between the optimal AO approach (25)–(27) and the low-complexity solution (32) and (27) narrows as the group size increases, and completely vanishes for a fully-connected BD-RIS. This implies that the RIS-transceiver design can be completely 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 $-10\,\text{dB}$), 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 tradeoff between subchannels (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 [19], [22], [27] 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.

VI. 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 fully-connected scenarios. An efficient RCG algorithm is proposed for smooth BD-RIS optimization problems, which offers lower computation complexity 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 subspace alignment and subchannel 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 subspaces 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

$$\sigma_n=\mathbf{u}_n^{\mathsf{H}}\mathbf{H}\mathbf{v}_n=\mathbf{u}_n^{\mathsf{T}}\mathbf{H}^*\mathbf{v}_n^*, \tag{34}$$

whose differential w.r.t. $\mathbf{\Theta}_g^*$ 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 [59], the corresponding complex derivative is

$$\frac{\partial\sigma_n}{\partial\mathbf{\Theta}_g^*}=\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{u}_n\mathbf{v}_n^{\mathsf{H}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}. \tag{35}$$

A linear combination of (35) yields (14).

B. Proof of Proposition 1

The scattering matrix of BD-RIS can be decomposed as[5]

$$\mathbf{\Theta}=\mathbf{L}\mathbf{\Theta}_{\mathrm{D}}\mathbf{R}^{\mathsf{H}}, \tag{36}$$

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

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

[5]This is because (block) unitary matrices are closed under multiplication.

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 [57], the n-th eigenvalue of $\mathbf{G}$ is bounded by

$$\lambda_n(\mathbf{G})\le\lambda_{n-k}(\mathbf{Y}),\qquad \text{if } n>k, \tag{37}$$

$$\lambda_n(\mathbf{G})\ge\lambda_n(\mathbf{Y}),\qquad \text{if } n<N-k+1. \tag{38}$$

Since $\mathbf{Y}=\mathbf{T}\mathbf{T}^{\mathsf{H}}$ is positive semi-definite, taking the square roots of (37) and (38) gives (15) and (16).

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

$$\mathbf{\Theta}=\mathbf{V}_{\mathrm{B}}\mathbf{X}\mathbf{U}_{\mathrm{F}}^{\mathsf{H}}, \tag{39}$$

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 $\mathbf{X}$

$$\mathbf{H}=\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}}=\mathbf{U}_{\mathrm{B}}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{X}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{V}_{\mathrm{F}}^{\mathsf{H}}. \tag{40}$$

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 w.r.t. $\mathbf{\Theta}_g^*$ is [59]

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

and the corresponding complex derivative is (26).

F. Proof of Proposition 4

The differential of (28a) w.r.t. $\mathbf{\Theta}_g^*$ 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

$$\frac{\partial\|\mathbf{H}\|_{\mathrm{F}}^2}{\partial\mathbf{\Theta}_g^*}=\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}}+\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}=\mathbf{M}_g. \tag{41}$$

First, we approximate the quadratic objective (28a) by its local Taylor expansion

$$\max_{\mathbf{\Theta}}\quad\sum_g 2\Re\big\{\mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\big\} \tag{42a}$$

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

$$2\Re\Big\{\sum_g \mathrm{tr}(\tilde{\mathbf{\Theta}}_g^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}) + \sum_{g_1,g_2} \mathrm{tr}(\tilde{\mathbf{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}})\Big\} \geq 2\Re\Big\{\sum_g \mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}) + \sum_{g_1,g_2} \mathrm{tr}(\mathbf{\Theta}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}})\Big\} \quad (45)$$

$$\sum_{g_1,g_2} \mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\tilde{\mathbf{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\tilde{\mathbf{\Theta}}_{g_2}\mathbf{H}_{\mathrm{F},g_2}) - 2\Re\Big\{\sum_{g_1,g_2} \mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\tilde{\mathbf{\Theta}}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2})\Big\} + \sum_{g_1,g_2} \mathrm{tr}(\mathbf{H}_{\mathrm{F},g_1}^{\mathsf{H}}\mathbf{\Theta}_{g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_1}^{\mathsf{H}}\mathbf{H}_{\mathrm{B},g_2}\mathbf{\Theta}_{g_2}\mathbf{H}_{\mathrm{F},g_2}) \geq 0 \quad (46)$$

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

$$\Re\big\{\mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\big\} = \Re\big\{\mathrm{tr}(\mathbf{\Sigma}_g\mathbf{V}_g^{\mathsf{H}}\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{U}_g)\big\} \leq \mathrm{tr}(\mathbf{\Sigma}_g). \quad (43)$$

The upper bound is tight when $\mathbf{V}_g^{\mathsf{H}}\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{U}_g = \mathbf{I}$, which implies the optimal solution of (42) is $\tilde{\mathbf{\Theta}}_g = \mathbf{U}_g\mathbf{V}_g^{\mathsf{H}}$, $\forall g$.

Next, we prove that solving (42) successively does not decrease (28a). Since $\tilde{\mathbf{\Theta}}$ optimal for problem (42), we have $\sum_g 2\Re\big\{\mathrm{tr}(\tilde{\mathbf{\Theta}}_g^{\mathsf{H}}\mathbf{M}_g)\big\} \geq \sum_g 2\Re\big\{\mathrm{tr}(\mathbf{\Theta}_g^{\mathsf{H}}\mathbf{M}_g)\big\}$ which is explicitly expressed by (45). On the other hand, expanding $\|\sum_g \mathbf{H}_{\mathrm{B},g}\tilde{\mathbf{\Theta}}_g\mathbf{H}_{\mathrm{F},g} - \sum_g \mathbf{H}_{\mathrm{B},g}\mathbf{\Theta}_g\mathbf{H}_{\mathrm{F},g}\|_{\mathrm{F}}^2 \geq 0$ gives (46). Adding (45) and (46), we have

$$\begin{aligned} &2\Re\Big\{\mathrm{tr}(\tilde{\mathbf{\Theta}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}}^{\mathsf{H}})\Big\} + \mathrm{tr}(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\tilde{\mathbf{\Theta}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}\tilde{\mathbf{\Theta}}\mathbf{H}_{\mathrm{F}}) \\ &\geq 2\Re\Big\{\mathrm{tr}(\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{D}}\mathbf{H}_{\mathrm{F}}^{\mathsf{H}})\Big\} + \mathrm{tr}(\mathbf{H}_{\mathrm{F}}^{\mathsf{H}}\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}^{\mathsf{H}}\mathbf{H}_{\mathrm{B}}\mathbf{\Theta}\mathbf{H}_{\mathrm{F}}), \end{aligned} \quad (44)$$

which suggests that updating $\tilde{\mathbf{\Theta}}$ does not decrease (28a).

Finally, we prove that the converging point of (42), denoted by $\tilde{\mathbf{\Theta}}^{?}$, is a stationary point of (28). The Karush-Kuhn-Tucker (KKT) conditions of (28) and (42) are equivalent in terms of primal/dual feasibility and complementary slackness, while the stationary conditions are respectively, $\forall g$,

$$\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\mathbf{\Theta}^{\star}\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} - \mathbf{\Theta}_g^{\star}\mathbf{\Lambda}_g^{\mathsf{H}} = 0, \quad (47)$$

$$\mathbf{M}_g - \mathbf{\Theta}_g^{\star}\mathbf{\Lambda}_g^{\mathsf{H}} = 0. \quad (48)$$

On convergence, (48) becomes $\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}(\mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{B}}\mathbf{\Theta}^{?}\mathbf{H}_{\mathrm{F}})\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}} - \mathbf{\Theta}_g^{?}\mathbf{\Lambda}_g^{\mathsf{H}} = 0$ and reduces to (47). 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," Dec 2022.

[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] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, pp. 106–112, Jan 2020.

[19] 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.

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

[21] 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.

[22] 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.

[23] ——, "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.

[24] H. Li, S. Shen, Y. Zhang, and B. Clerckx, "Channel estimation and beamforming for beyond diagonal reconfigurable intelligent surfaces," *arXiv:2403.18087*, 2024.

[25] 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.

[26] H. Li, M. Nerini, S. Shen, and B. Clerckx, "Wideband modeling and beamforming for beyond diagonal reconfigurable intelligent surfaces," *arXiv:2403.12893*, 2024.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] A. S. de Sena, M. Rasti, N. H. Mahmood, and M. Latva-aho, “Beyond diagonal RIS for multi-band multi-cell MIMO networks: A practical frequency-dependent model and performance analysis,” *arXiv:2401.06475*, 2024.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] I. Santamaria, M. Soleymani, E. Jorswieck, and J. Gutiérrez, “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.

[44] 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.

[45] 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.

[46] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2009.

[47] 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.

[48] 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.

[49] 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.

[50] E. Polak and G. Ribiere, “Note sur la convergence de méthodes de directions conjuguées,” *Revue française d’informatique et de recherche opérationnelle. Série rouge*, vol. 3, pp. 35–43, 1969.

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

[52] 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.

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

[54] 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.

[55] J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*. Oxford, UK: Oxford University Press, 2004.

[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] 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.

[59] 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.