Prof. Wei Yi
Associate Editor
IEEE Transactions on Signal Processing

August 25, 2025

Response to Decision on Manuscript T-SP-33360-2025

Dear Editor and Reviewers,

Thank you for giving us an opportunity to revise the paper "MIMO Channel Shaping and Rate Maximization Using Beyond-Diagonal RIS". Your feedback and suggestions have been invaluable in helping us improve the quality of the manuscript. Below we prepare a point-to-point response and highlight the corresponding changes in text, where labels have been matched to the submission for your convenience. We hope the revisions and clarifications make the manuscript meet the TSP publication standards.

Yours sincerely,

Yang Zhao, Hongyu Li, Bruno Clerckx, and Massimo Franceschetti

Editorial Decision

The paper has been improved after this round of revision. However, reviewers still pointed out several critical issues regarding the novelty, practical merit, performance assessment of the paper. For example, one reviewer criticized that the rate maximization problem is actually extensively studied in existing papers and solved by different methods, by different algorithms, e.g., [36]. The application of the derived bounds on Singular Values (SVs) on Beyond-Diagonal (BD)-Reconfigurable Intelligent Surface (RIS) channel has been well established in early papers, e.g., [44]. I am deciding on the major revision decision. However, in the revised version, the authors need to clarify the above-mentioned very fundamental issues, otherwise the paper can still be rejected.

Response We appreciate your feedback and summary again.

In [36], the authors considered an energy efficiency maximization problem for *multi-user Multiple-Input Single-Output (MISO)* downlink systems. A Penalty Dual Decompistion (PDD) method was proposed for symmetric BD-RIS design, which is a two-layer procedure alternating between the primal variables in the inner layer and the penalty coefficient in the outer layer. The PDD method is often used to tackle coupled constraints (e.g., Signal-to-Interference Noise Ratio (SINR) threshold involving both transmit/receive beamforming and BD-RIS matrices) and is notoriously tricky to implement. Invoking PDD for Multiple-Input Multiple-Output (MIMO) rate maximization problem (36) is theoretically feasible but computationally very inefficient. For example, PDD requires $O(N_s^2)$ and $O(N_s^4)$ flops to obtain a stationary solution for Diagonal (D)-RIS and fully-connected BD-RIS, respectively [36, Table I], whereas our proposed geodesic Riemannian Conjugate Gradient (RCG) Algorithm 1 achieves the same performance with respectively $O(N_s)$ and $O(N_s^3)$ flops. The latter exploits the structure of the block-unitary constraint for acceleration and can accommodate for the symmetric constraint when necessary (as discussed in Section V-D1).

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

We would also like to point out that [44], mentioned by the editor, actually investigated the optimal beamforming design for multi-hop *Amplify-and-Forward (AF) relays* with *no channel SV manipulation bounds* taken into account. Those are active devices that consume power to amplify the signal with noise. The optimal structure of AF relays [44, (17)] comes in the form

$$\mathbf{F} = \mathbf{V}_B \boldsymbol{\Lambda} \mathbf{U}_F^H, \tag{i}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix that models the effect of relay power allocation. The aim of [44] is to show, as quoted from the authors, "The optimal source and relay matrices jointly diagonalize the multi-hop MIMO relay system into a set of parallel scalar channels." On the other hand, our paper aims to characterize the fundamental limits of BD-RIS in MIMO channel SVs shaping. BD-RIS can only manipulate the phase and amplitude of the signal by passive scattering without amplification, which is fundamentally different from AF relays in terms of hardware architecture, power consumption, and noise characteristics. It just happens that *the mathematical modeling of AF relays of unit power coincide with the modeling of fully-connected BD-RIS,* and thus the rate-optimal solutions. The underlying physical constraints and the resulting practical implications are clearly distinct.

> For example, the rate maximization problem [37] has only been tacked in the special case where the direct channel is negligible and the BD-RIS is fully-connected. Under those conditions, the mathematical modeling of BD-RIS coincides with that of AF relay of unit power, although the operation mechanism and noise characteristics are clearly distinct.

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

We also supplied extensive theoretical studies of BD-RIS in Propositions 1 – 3 and the resulting Corollaries, most of which are new results to the topic. The comments from the reviewers have also been addressed to our best efforts as reflected below. Therefore, we would like to politely invite the editor to reconsider the decision for this manuscript.

Reviewer 1

The authors have addressed my concerns.

Response Thank you for your positive feedback and continued support.

Reviewer 2

This paper explores potentials of a new type of RIS, specifically adopting a BD scattering model, in shaping point-to-point MIMO channels for improved wireless performance. The authors derive analytical bounds under specific scenarios and propose a numerical optimization method for broader shaping problems, both of which are verified by simulation results. The paper tackles an important and timely question regarding the channel shaping capabilities of passive RIS, particularly moving beyond the conventional D-RIS model. The singular value analysis and optimization is a fresh and relevant perspective in the field. However, there are areas which need further improvement:

2.1 *Implementation of BD-RIS: The paper could benefit from more thorough discussions on practical implementation challenges of BD-RIS, including manufacturing complexity, calibration, and hardware imperfections. Addressing these would enhance the paper's practical relevance and applicability.*

Response Thank you very much for the suggestion. The practical implementation of BD-RIS is a profound and challenging issue that deserves further discussion. Some of the authors have a newly released preprint [i] that covers the topic in great depth. It has been discussed lightly in this submission due to the page limitation, and we appreciate your understanding on this point.

> The main manufacturing complexity of BD-RIS lies in the design and implementation of the circuit network. Fortunately, novel topologies such as tree- and forest-connections have been proposed to reduce the number of components for a flexible cost-performance tradeoff [26]. Other practical challenges such as channel estimation [30], mutual coupling [31], wideband modelling [32], multi-sector coverage [33], and hardware implementation [34] have also been studied in recent literature. BD-RIS has been proved to achieve higher spectral efficiency than D-RIS and higher energy efficiency than active RIS and AF relay [27, 35, 36].

2.2 *Comparison between RCG methods: The link could be highlighted by mentioning that the non-geodesic update (additive + retraction) effectively employs a first-order Taylor approximation of the geodesic update (multiplicative via matrix exponential), thus necessitating the retraction step to ensure iterates remain on the manifold.*

Response This is indeed a great point to highlight as it can help the readers to understand the connection and difference between the two RCG methods. We have updated the context as follows.

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

2.3 *Applicability to other BD-RIS models: The focus is strongly on the group-connected BD-RIS versus D-RIS. The authors could clarify how the proposed methods and results might extend to other BD-RIS models, such as multi-sector or multi-layer configurations. This would broaden the applicability of the findings and provide a more comprehensive understanding of the RIS landscape.*

Response We appreciate the reviewer for mentioning this point. While the analysis and optimization methods in this paper have been tailored for the group-connected BD-RIS in shaping point-to-point MIMO channels, we believe they can offer an initial benchmark for other models. We have added a paragraph in the conclusion to clarify this point.

> The analysis and optimization methods in this paper have been tailored for group-connected BD-RIS. Extension to other architectures remains a promising direction for future research. We believe that one straightforward extension to the multi-sector model [33] is to retrieve the optimal scattering matrix for each sector individually by our method and then play with the power splitting factors. Meanwhile, transitioning towards multi-layer RIS models [60] mirrors that from single-hop to multi-hop AF relays; interested readers may be inspired by [44] on this point.

2.4 *Readability of the Technical Parts: The readability of the paper could be improved by having more intuitions and/or explanations after important theorems/lemmas/corollaries.*

Response Apart from Examples 3 and 4, we have now added explanations and typical use cases after those theoretical results to help the readers understanding their meanings and applications. The logic flow and wordings have also been carefully reworked. Please refer to Response **3.1** and the manuscript for details.

Reviewer 3

The reviewer appreciates the authors' efforts to prepare the revision. The reviewer has the following concerns:

3.1 *The authors provide a bunch of bounds (Prop. 2) on SVs of BD-RIS channels. The significance of these bounds is unclear. What can these bounds be used for? One possible solution lies in the achievable maximal channel capacity (Corollary 3.5). However, this result (the well-known forward and backward channel alignment) has long been established, e.g., [44]. Especially, the bounds developed in corollaries 3.1-3.3 are a great number of inequalities coupling with each other. How could these bounds be used?*

Response We appreciate the reviewer for the clinical questions and aim to answer them one-by-one below.

- The SV bounds in Proposition 2 complements the Degrees of Freedom (DoF) result in Proposition 1 by quantifying the dynamic range of extreme singular values in low-multipath scenarios. They reveal a saturation effect of increasing the number of BD-RIS elements and group size in enhancing channel shaping capability. Therefore, the bounds can be used to guide practical RIS configurations, especially in millimeter-wave and terahertz systems under sparse environment, for a balanced performance gain and hardware complexity. Another use case is strategic network planning where the bounds can help compare the shaping capabilities of a single massive RIS versus multiple smaller distributed ones without actually developing them.

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

- The bounds in Corollary 3.1 sketch the outer bounds of the entire achievable channel SV region under the assumption of negligible direct channel and fully-connected BD-RIS. While direct application of every inequality within is not straightforward, they indeed provide a rich theoretical foundation and ultimate performance limit for MIMO channel shaping using BD-RIS. One may choose any subset of those bounds for specific wireless applications and metrics, and some examples are given in Corollaries 3.2 and 3.3. In the revised manuscript, we have also added simulation results comparing the outer bounds in Corollary 3.1 and the achievable region obtained by solving the optimization problem (34) in Fig. 2.
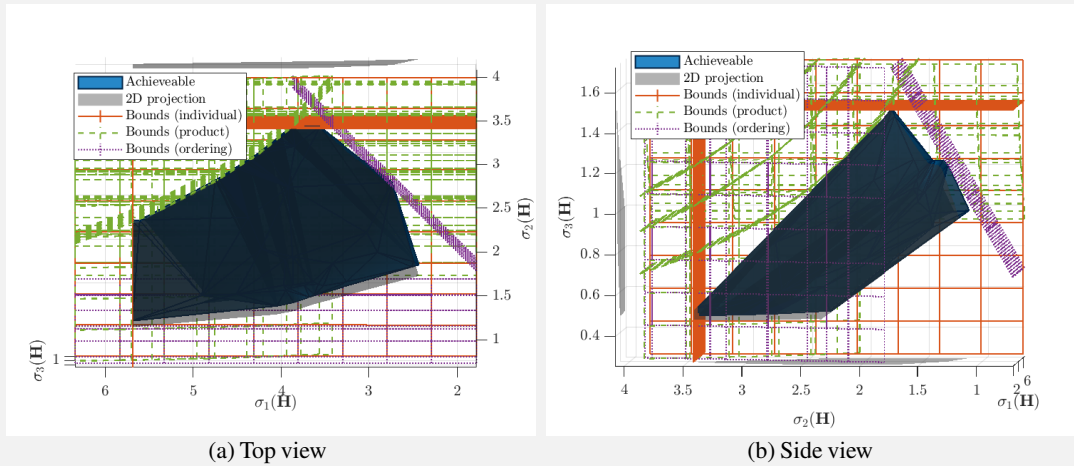

(a) Top view

(b) Side view

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

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

$$\begin{aligned}
&\sigma_1(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_1(\mathbf{H_F}), \quad \sigma_2(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_2(\mathbf{H_F}), \\
&\sigma_2(\mathbf{H}) \le \sigma_2(\mathbf{H_B})\sigma_1(\mathbf{H_F}), \quad \sigma_3(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_3(\mathbf{H_F}), \\
&\sigma_3(\mathbf{H}) \le \sigma_2(\mathbf{H_B})\sigma_2(\mathbf{H_F}), \quad \sigma_3(\mathbf{H}) \le \sigma_3(\mathbf{H_B})\sigma_1(\mathbf{H_F}),
\end{aligned} \tag{iia}$$

$$\begin{aligned}
&\sigma_1(\mathbf{H})\sigma_2(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_2(\mathbf{H_B})\sigma_1(\mathbf{H_F})\sigma_2(\mathbf{H_F}), \\
&\sigma_1(\mathbf{H})\sigma_3(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_3(\mathbf{H_B})\sigma_1(\mathbf{H_F})\sigma_2(\mathbf{H_F}), \\
&\sigma_1(\mathbf{H})\sigma_3(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_2(\mathbf{H_B})\sigma_1(\mathbf{H_F})\sigma_3(\mathbf{H_F}), \\
&\sigma_2(\mathbf{H})\sigma_3(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_2(\mathbf{H_B})\sigma_2(\mathbf{H_F})\sigma_3(\mathbf{H_F}), \\
&\sigma_2(\mathbf{H})\sigma_3(\mathbf{H}) \le \sigma_2(\mathbf{H_B})\sigma_3(\mathbf{H_B})\sigma_1(\mathbf{H_F})\sigma_2(\mathbf{H_F}), \\
&\sigma_2(\mathbf{H})\sigma_3(\mathbf{H}) \le \sigma_1(\mathbf{H_B})\sigma_3(\mathbf{H_B})\sigma_1(\mathbf{H_F})\sigma_3(\mathbf{H_F}),
\end{aligned} \tag{iib}$$

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

where (iia), (iib) are explicit results of (12) while (iic) denotes the ordering of singular values. Those are labeled respectively as 'Bounds (individual)', 'Bounds (product)', and 'Bounds (ordering)' in Fig. 2. The two views confirm that the theoretical outer bounds are not everywhere tight with many entries being redundant, but they provide a conservative estimate of the achievable singular value region. Importantly, the vertices of the region are always on the boundary, which implies that those can be achieved in the closed form without performing optimization.

- The bounds in Corollary 3.2 reveal the upper (resp. lower) bound on the product of some largest (resp. smallest) channel SVs. These bounds can be applied, for instance, as a shortcut to establish the upper bound of BD-RIS-aided MIMO channel capacity at extreme Signal-to-Noise Ratios (SNRs).

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

- The bounds in Corollary 3.3 reveal the shaping limits of the n-th largest channel SV. These bounds enable closed-form passive beamforming, and hence fixed channel and closed-form active beamforming, for spatial multiplexing with a limited number n of RF chains. Another use case is to enhance the harvested power for MIMO wireless power transfer with RF combining by maximizing the dominant channel SV. One may also improve the channel condition number for numerically stable signal processing by maximizing the smallest channel SV.

Corollary 3.3 and Proposition 2 both reveal the shaping limits of the n-th largest channel singular value. They are derived under different assumptions are not special cases of each other. Importantly, Corollary 3.3 establishes upper and lower bounds for *each* channel singular value (c.f. first and last k in Proposition 2) and provides general solutions for fully-connected BD-RIS of arbitrary (c.f. sufficiently large) size to attain the equalities. These bounds enable closed-form passive beamforming, and hence fixed channel and closed-form active beamforming, for spatial multiplexing with a limited number n of RF chains. We emphasize that in (15) the mode alignment is realized by $\mathbf{V_B}$ and $\mathbf{U_F}$ while the ordering is enabled by permutation matrices

P and Q, which are special cases of $\mathbf{X}$ defined in (3).

- While [44] discussed the optimal beamforming structure for multi-hop AF relays, its focuses of channel diagonalization and relay power allocation are fundamentally different from this paper. The literature provides neither channel SV bounds nor channel capacity bounds. Please refer to our response to the editor on this point.

3.2 *Prop. 4 imposes the assumption that $f(\cdot)$ must be symmetric gauge function in SVs of $\mathbf{H}$. Note gauge function is strong assumption since it should be homogeneous and convex (e.g., [62, Sec 3-I]) in SVs of $\mathbf{H}$, which indeed severely restricts the range of its application. When $f(\cdot)$ is nonconvex, subdifferential ($\mathbf{D}$ in eq. (iv)) generally does not exist. In fact, any commonly adopted performance metrics, e.g., spectral efficiency, energy efficiency, receiving power are not gauge functions (they are inhomogeneous and nonconvex). On the other hand, to obtain partial differential with respect to (w.r.t.) RIS, (i.e., the aim of Prop. 4), the assumption of gauge function is unnecessary. Directly taking the derivative of $f(\cdot)$ w.r.t. RIS is more preferable.*

Response We fully agree with the reviewer that enforcing $f(\mathrm{sv}(\mathbf{H}))$ as a gauge function is too restrictive and unnecessary for the shaping problem. In the revised manuscript, Prop 4 has been extended to locally Lipschitz functions of channel singular values. We sincerely appreciate your insights on this issue.

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

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

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

$$\max_{\boldsymbol{\Theta}} \quad f\big(\mathrm{sv}(\mathbf{H})\big) \tag{iiia}$$

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

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

$$\partial_{\boldsymbol{\Theta}_g^*} f\big(\mathrm{sv}(\mathbf{H})\big) = \mathrm{conv}\big\{\mathbf{H}_{\mathrm{B},g}^{\mathsf{H}}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}\mathbf{H}_{\mathrm{F},g}^{\mathsf{H}}\big\}, \tag{iv}$$

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

Proof. Please refer to Appendix H. □

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

3.3 *The authors propose Riemannian Conjugate Gradient (RCG) algorithm (Algorithm 1) to optimize RIS configuration. Note CG descent on manifold is indeed well established and is widely used for optimization on manifolds, e.g., [ii]. It can be implemented conveniently by off-the-shelf solvers, e.g., Manopt [iii]. Then, what is the novelty or significance of Sec. III-B?*

Response Manopt is indeed a canonical tool for manifold optimization. Upon inspecting the source code, we found it invokes a generic non-geodesic solver for feasible domain `stiefelcomplexfactory(n, p,`

k) that does not exploit the block-unitary structure of our problem. This is also confirmed by MATLAB profiler that 60% of its running time was spent on projection (i.e. retraction from the Euclidean space to the manifold, due to non-geodesic update). This is the main bottleneck of the algorithm.
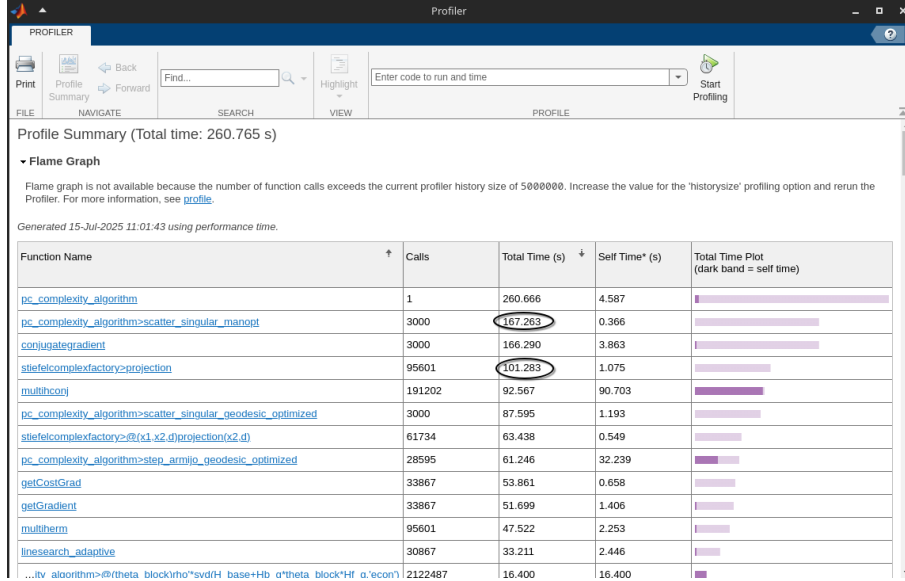

Figure 3: MATLAB profiler report of Manopt vs proposed RCG on 1000 independent channel realizations. The simulation parameters are inline with Table 1.

In contrast, our proposed geodesic RCG algorithm is tailored to group-connected BD-RIS designs, offering significant computational advantages as discussed in context. The novelty of Section III-B lies in:

- The formulation of the channel shaping problem as an optimization problem on the product of unitary manifolds, which is a natural and appropriate setting for BD-RIS design.

- The derivation of the translated Riemannian gradient at the identity of the manifold, which offers a compact expression for the geodesic update.

- The development of a specialized geodesic RCG algorithm that leverages the block-unitary structure of BD-RIS for accelerated convergence and reduced computational complexity compared to general-purpose manifold optimization solvers.

- The application of this algorithm to reveal the entire achievable channel SV region, which is a novel contribution to the understanding of BD-RIS shaping capabilities.

In the revised manuscript, we have added Table 1 to benchmark the performance of two RCG algorithms on the Pareto channel SV problem (34)

Table 1: Performance of RCG Algorithms on (34) with $N_\text{T} = N_\text{R} = 4$, $L = 16$

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

Table 1 benchmarks two RCG algorithms on the Pareto singular value problem (34) with $N_\text{T} = N_\text{R} = 4$ and $L = 16$. The geodesic RCG is implemented with pinched gradients w.r.t. Θ and unified step size selection; please refer to the discussion below (33) for details. The non-geodesic RCG is implemented by `Manopt` toolbox at commit `a879a0d` [iii]. Both algorithms employ a stopping criterion of relative change in the objective function with a tolerance of $\epsilon = 1 \times 10^{-4}$, such that the final values are identical

within reasonable precision. The statistics are averaged over 1000 independent channel realizations. We observe that the non-geodesic RCG typically requires 1 to 2 more iterations than its geodesic counterpart. This is because the addition is in the tangent space of the manifold and is less effective than manifold-native updates. When it comes to elapsed time, the geodesic RCG is 1333% faster than the non-geodesic counterpart when $N_S = 16$. The main reason is that the geodesic RCG avoids the retraction step from the Euclidean space to the manifold. According to the profiler report, around 60% of the non-geodesic RCG runtime is spent on retraction, which becomes the main bottleneck of the algorithm. The advantage narrows down to 391.6% and 33.35% when $N_S = 64$ and 256, respectively. This is because accurately evaluating matrix exponential can be time-consuming for large N_S.

References

[i] H. Li, M. Nerini, S. Shen, and B. Clerckx, "A tutorial on beyond-Diagonal Reconfigurable Intelligent Surfaces: Modeling, architectures, system design and optimization, and applications," *arXiv [eess.SP]*, 22 May 2025.

[ii] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*. Cambridge, UK: Cambridge University Press, Mar 2023.

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