## Appendix

### A. Proof of Proposition 1

It suffices to consider the rank of the indirect channel. Denote the Singular Value Decomposition (SVD) of the backward and forward channels as

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

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

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

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

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

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

$$\mathbf{\Theta}_{\text{DoF-max}}^{\text{MIMO}} = \mathbf{Q}_{\mathrm{B},2}\mathbf{Q}_{\mathrm{F},2}^{\mathsf{H}}, \tag{ii}$$

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

$$\mathbf{\Theta}_{\text{DoF-min}}^{\text{MIMO}} = \mathbf{Q}_{\mathrm{B},1}\mathbf{Q}_{\mathrm{F},2}^{\mathsf{H}}, \tag{iii}$$

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

### B. Proof of Proposition 2

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

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

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

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

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

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

### C. 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 BD-RIS can be decomposed as

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

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 $\mathbf{X}$

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

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

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

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

### D. Proof of Corollary 3.2

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

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

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

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

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

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

where the first equality follows from (viii) and $\hat{\mathbf{\Sigma}}_{\mathrm{B}}, \hat{\mathbf{\Sigma}}_{\mathrm{F}}$ truncate $\mathbf{\Sigma}_{\mathrm{B}}, \mathbf{\Sigma}_{\mathrm{F}}$ to square matrices of dimension $N_{\mathrm{S}}$, respectively. It is evident that (x) implies (ix) and thus (13b).

### E. Proof of Corollary 3.3

In (14), the set of upper bounds

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

is a special case of (12) with $(I, J, K) \in [N_{\mathrm{S}}]^3$. The minimum[1] of (xi) is selected as the tightest upper bound in (14). On the other hand, the set of lower bounds

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

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

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

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

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

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

### F. Proof of Corollary 3.4

From (vi) and (vii) we have

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

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

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

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

### G. Proof of Corollary 3.6

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

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

where the approximation is $\log(1+x)\approx x$ for small $x$ and the inequality follows from (13a) with $k=1$. In the high-SNR regime, the capacity is achieved by multiple eigenmode transmission with uniform power location [v, (5.27)]

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

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

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

### H. Proof of Proposition 4

The sub-differential of a symmetric gauge function of singular values of a matrix with respect to the matrix itself is given by [vi, Theorem 2]

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

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 $\mathbf{H}$. It implies

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

and therefore $\mathbf{H}_{\mathrm{B},g}^\mathsf{H}\mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{H}\mathbf{H}_{\mathrm{F},g}^\mathsf{H}$ constitutes a sub-gradient of $f\big(\mathrm{sv}(\mathbf{H})\big)$ with respect to $\mathbf{\Theta}_g$. The convex hull of those sub-gradients is the sub-differential (26).

### I. Proof of Lemma 1

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

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

and the corresponding complex derivative is (37).

### J. Proof of Proposition 5

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

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

and the corresponding complex derivative is

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

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

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

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

whose optimal solution is

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

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

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

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

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

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

Adding (xviii) and (xix), we have

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

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

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

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

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

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

## Acknowledgement

The authors would like to thank the anonymous reviewers for their insightful criticisms and suggestions that helped us correct several technical errors.

## References

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

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

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

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

[v] 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.

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

[vii] 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.