Appendix A: Additional Proofs

A.1. Additional Proofs in Section 2.1

Proof of Proposition 1. The expressions for Euclidean gradients are obtained via direct computation. For the Riemannian gradient, since $\mathcal{M}_{r+}$ and $\mathcal{M}_r$ are embedded submanifolds of $\mathbb{R}^{p\times p}$ and $\mathbb{R}^{p_1\times p_2}$, respectively and the Euclidean metric is considered, from (Absil et al. 2009, (3.37)), we know the Riemannian gradients are the projections of the Euclidean gradients onto the corresponding tangent spaces. The results follow by observing the projection operator onto $T_{\mathbf{X}}\mathcal{M}_{r+}$ and $T_{\mathbf{X}}\mathcal{M}_r$ given in (9). ■

Proof of Proposition 2. First the expressions for $\nabla^2 g(\mathbf{L},\mathbf{R})[\mathbf{A},\mathbf{A}]$ and $\nabla^2 g_{\text{reg}}(\mathbf{L},\mathbf{R})[\mathbf{A},\mathbf{A}]$ are given in (Ha et al. 2020, Eq. (2.8)) and (Zhu et al. 2018, Section IV-A and Remark 8), respectively. The expressions for $\nabla^2 g(\mathbf{Y})[\mathbf{A}',\mathbf{A}']$ can be obtained by letting $\mathbf{A}=[\mathbf{A}'^\top \quad \mathbf{A}'^\top]^\top$ and $\mathbf{L}=\mathbf{R}=\mathbf{Y}$ in $\nabla^2 g(\mathbf{L},\mathbf{R})[\mathbf{A},\mathbf{A}]$.

Next, we derive the Riemannian Hessian of f. The Riemannian Hessian of an objective function f is usually defined in terms of the Riemannian connection as in (6). But in the case of embedded submanifolds, it can also be defined by means of the so-called second-order retractions.

Given a general smooth manifold $\mathcal{M}$, a *retraction* R is a smooth map from $T\mathcal{M}$ to $\mathcal{M}$ satisfying i) $R(\mathbf{X},0)=\mathbf{X}$ and ii) $\frac{d}{dt}R(\mathbf{X},t\eta)|_{t=0}=\eta$ for all $\mathbf{X}\in\mathcal{M}$ and $\eta\in T_{\mathbf{X}}\mathcal{M}$, where $T\mathcal{M}=\{(\mathbf{X},T_{\mathbf{X}}\mathcal{M}):\mathbf{X}\in\mathcal{M}\}$, is the tangent bundle of $\mathcal{M}$ (Absil et al. 2009, Chapter 4). We also let $R_{\mathbf{X}}$ to be the restriction of R to $T_{\mathbf{X}}\mathcal{M}$ and it satisfies $R_{\mathbf{X}}:T_{\mathbf{X}}\mathcal{M}\to\mathcal{M},\xi\longmapsto R(\mathbf{X},\xi)$. Retraction is in general a first-order approximation of the exponential map (Absil et al. 2009, Chapter 4). A *second-order retraction* is the retraction defined as a second-order approximation of the exponential map (Absil and Malick 2012). As far as convergence of Riemannian optimization methods goes, first-order retraction is sufficient (Absil et al. 2009, Chapter 3), but second-order retraction enjoys the following nice property: the Riemannian Hessian of an objective function f coincides with the Euclidean Hessian of the lifted objective $\hat{f}_{\mathbf{X}}:=f\circ R_{\mathbf{X}}$.

Lemma 1 (Proposition 5.5.5 of Absil et al. (2009)) *Let $R_{\mathbf{X}}$ be a second-order retraction on $\mathcal{M}$. Then* $\text{Hess} f(\mathbf{X})=\nabla^2(f\circ R_{\mathbf{X}})(0)$ *for all* $\mathbf{X}\in\mathcal{M}$.

We present the second-order retractions under both PSD and general low-rank matrix settings in the following Proposition 1.

Proposition 1 (Second-order Retractions in PSD and General Low-rank Matrix Manifolds)

- *PSD case: Suppose $\mathbf{X}\in\mathcal{M}_{r+}$ has eigendecomposition $\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\top$. Then the mapping $R^{(2)}_{\mathbf{X}}:T_{\mathbf{X}}\mathcal{M}_{r+}\to\mathcal{M}_{r+}$ given by*

$$R^{(2)}_{\mathbf{X}}:\xi=[\mathbf{U}\quad\mathbf{U}_\perp]\begin{bmatrix}\mathbf{S} & \mathbf{D}^\top\\ \mathbf{D} & \mathbf{0}\end{bmatrix}[\mathbf{U}\quad\mathbf{U}_\perp]^\top\to\mathbf{W}\mathbf{X}^\dagger\mathbf{W}^\top$$

  is a second-order retraction on $\mathcal{M}_{r+}$, where $\mathbf{W}=\mathbf{X}+\frac{1}{2}\xi^s+\xi^p-\frac{1}{8}\xi^s\mathbf{X}^\dagger\xi^s-\frac{1}{2}\xi^p\mathbf{X}^\dagger\xi^s$, $\xi^s=P_{\mathbf{U}}\xi P_{\mathbf{U}}$ and $\xi^p=P_{\mathbf{U}_\perp}\xi P_{\mathbf{U}}+P_{\mathbf{U}}\xi P_{\mathbf{U}_\perp}$. Furthermore, we have

$$R^{(2)}_{\mathbf{X}}(\xi)=\mathbf{X}+\xi+\mathbf{U}_\perp\mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}^\top\mathbf{U}_\perp^\top+O(\|\xi\|_{\mathrm{F}}^3),\ \ \textit{as}\ \|\xi\|_{\mathrm{F}}\to 0.$$

- *General case: Suppose $\mathbf{X}\in\mathcal{M}_r$ has SVD $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$. Then the mapping $R^{(2)}_{\mathbf{X}}:T_{\mathbf{X}}\mathcal{M}_r\to\mathcal{M}_r$ given by*

$$R^{(2)}_{\mathbf{X}}:\xi=[\mathbf{U}\quad\mathbf{U}_\perp]\begin{bmatrix}\mathbf{S} & \mathbf{D}_2^\top\\ \mathbf{D}_1 & \mathbf{0}\end{bmatrix}[\mathbf{V}\quad\mathbf{V}_\perp]^\top\to\mathbf{W}\mathbf{X}^\dagger\mathbf{W}$$

is a second-order retraction on $\mathcal{M}_r$*, where* $\mathbf{W} = \mathbf{X} + \frac{1}{2}\xi^s + \xi^p - \frac{1}{8}\xi^s\mathbf{X}^\dagger\xi^s - \frac{1}{2}\xi^p\mathbf{X}^\dagger\xi^s - \frac{1}{2}\xi^s\mathbf{X}^\dagger\xi^p$*,* $\xi^s = P_{\mathbf{U}}\xi P_{\mathbf{V}}$ *and* $\xi^p = P_{\mathbf{U}_\perp}\xi P_{\mathbf{V}} + P_{\mathbf{U}}\xi P_{\mathbf{V}_\perp}$*. Furthermore, we have*

$$R_{\mathbf{X}}^{(2)}(\xi) = \mathbf{X} + \xi + \mathbf{U}_\perp\mathbf{D}_1\mathbf{\Sigma}^{-1}\mathbf{D}_2^\top\mathbf{V}_\perp^\top + O(\|\xi\|_{\mathrm{F}}^3), \quad \textit{as } \|\xi\|_{\mathrm{F}} \to 0.$$

Proof of Proposition 1. The results for the PSD case can be found in (Vandereycken and Vandewalle 2010, Proposition 5.10) and the results under the general case can be found in (Vandereycken 2013, Proposition A.1) and (Shalit et al. 2012, Theorem 3). ■

By Lemma 1 and the property of second-order retraction, the sum of the first three dominating terms in the Taylor expansion of $f \circ R_{\mathbf{X}}^{(2)}(\xi)$ w.r.t. ξ are $f(\mathbf{X}) + \langle \mathrm{grad} f(\mathbf{X}), \xi\rangle + \frac{1}{2}\mathrm{Hess} f(\mathbf{X})[\xi,\xi]$. By matching the corresponding terms and the expressions of $R_{\mathbf{X}}^{(2)}$ in Proposition 1, we can get the quadratic expression for $\mathrm{Hess} f(\mathbf{X})[\xi,\xi]$.

Next, we discuss how to obtain $\mathrm{Hess} f(\mathbf{X})[\xi,\xi]$ in PSD and general low-rank matrix manifolds, respectively.

PSD case: Given small enough $\xi = [\mathbf{U} \quad \mathbf{U}_\perp]\begin{bmatrix}\mathbf{S} & \mathbf{D}^\top \\ \mathbf{D} & \mathbf{0}\end{bmatrix}[\mathbf{U} \quad \mathbf{U}_\perp]^\top$, define $\mathbf{U}_p = \mathbf{U}_\perp\mathbf{D}$. By Proposition 1 and Taylor expansion, we have

$$\begin{aligned} f \circ R_{\mathbf{X}}^{(2)}(\xi) &= f(\mathbf{X} + \xi + \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{U}_p^\top + O(\|\xi\|_{\mathrm{F}}^3)) \\ &= f(\mathbf{X} + \xi + \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{U}_p^\top) + O(\|\xi\|_{\mathrm{F}}^3) \\ &= f(\mathbf{X} + \xi) + \langle \nabla f(\mathbf{X}+\xi), \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{U}_p^\top\rangle + O(\|\xi\|_{\mathrm{F}}^3) \\ &= f(\mathbf{X}) + \langle \nabla f(\mathbf{X}), \xi\rangle + \frac{1}{2}\nabla^2 f(\mathbf{X})[\xi,\xi] + \langle \nabla f(\mathbf{X}), \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{U}_p^\top\rangle + O(\|\xi\|_{\mathrm{F}}^3). \end{aligned} \tag{1}$$

Since $\xi^p\mathbf{X}^\dagger\xi^p = \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{U}_p^\top$, where $\xi^p = P_{\mathbf{U}_\perp}\xi P_{\mathbf{U}} + P_{\mathbf{U}}\xi P_{\mathbf{U}_\perp}$, the second order term in (1) is $\frac{1}{2}\nabla^2 f(\mathbf{X})[\xi,\xi] + \langle \nabla f(\mathbf{X}), \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{U}_p^\top\rangle$ and it equals to $\frac{1}{2}\mathrm{Hess} f(\mathbf{X})[\xi,\xi]$.

General case: Given small enough $\xi = [\mathbf{U} \quad \mathbf{U}_\perp]\begin{bmatrix}\mathbf{S} & \mathbf{D}_2^\top \\ \mathbf{D}_1 & \mathbf{0}\end{bmatrix}[\mathbf{V} \quad \mathbf{V}_\perp]^\top$, define $\mathbf{U}_p = \mathbf{U}_\perp\mathbf{D}_1$ and $\mathbf{V}_p = \mathbf{V}_\perp\mathbf{D}_2$. By Proposition 1 and Taylor expansion, we have

$$\begin{aligned} f \circ R_{\mathbf{X}}^{(2)}(\xi) &= f(\mathbf{X} + \xi + \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{V}_p^\top + O(\|\xi\|_{\mathrm{F}}^3)) \\ &= f(\mathbf{X} + \xi + \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{V}_p^\top) + O(\|\xi\|_{\mathrm{F}}^3) \\ &= f(\mathbf{X} + \xi) + \langle \nabla f(\mathbf{X}+\xi), \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{V}_p^\top\rangle + O(\|\xi\|_{\mathrm{F}}^3) \\ &= f(\mathbf{X}) + \langle \nabla f(\mathbf{X}), \xi\rangle + \frac{1}{2}\nabla^2 f(\mathbf{X})[\xi,\xi] + \langle \nabla f(\mathbf{X}), \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{V}_p^\top\rangle + O(\|\xi\|_{\mathrm{F}}^3). \end{aligned} \tag{2}$$

Since $\xi^p\mathbf{X}^\dagger\xi^p = \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{V}_p^\top$, where $P_{\mathbf{U}_\perp}\xi P_{\mathbf{V}} + P_{\mathbf{U}}\xi P_{\mathbf{V}_\perp}$, the second order term in (2) is $\frac{1}{2}\nabla^2 f(\mathbf{X})[\xi,\xi] + \langle \nabla f(\mathbf{X}), \mathbf{U}_p\mathbf{\Sigma}^{-1}\mathbf{V}_p^\top\rangle$ and it equals to $\frac{1}{2}\mathrm{Hess} f(\mathbf{X})[\xi,\xi]$. This finishes the proof of this proposition. ■

We note the proof technique for deriving the Riemannian Hessian is analogous to the proof of (Vandereycken 2013, Proposition 2.3). Here we extend it to the setting for a general twice differentiable function f.

A.2. Additional Proofs in Section 3

Proof of Lemma 2. Suppose $\mathbf{X}$ has the eigendecomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top$ and $\mathbf{P} = \mathbf{U}^\top\mathbf{Y}$. Given $\xi = [\mathbf{U} \quad \mathbf{U}_\perp]\begin{bmatrix}\mathbf{S} & \mathbf{D}^\top \\ \mathbf{D} & \mathbf{0}\end{bmatrix}[\mathbf{U} \quad \mathbf{U}_\perp]^\top$. For any $\mathbf{A} \in \mathscr{A}_{\mathbf{Y}}^\xi$, it is easy to check $\mathbf{Y}\mathbf{A}^\top + \mathbf{A}\mathbf{Y}^\top = \xi$, so $\mathscr{A}_{\mathbf{Y}}^\xi \subseteq \{\mathbf{A} : \mathbf{Y}\mathbf{A}^\top + \mathbf{A}\mathbf{Y}^\top = \xi\}$. For any $\mathbf{A}$ such that $\mathbf{Y}\mathbf{A}^\top + \mathbf{A}\mathbf{Y}^\top = \xi$, we have

$$\begin{bmatrix}\mathbf{S} & \mathbf{D}^\top \\ \mathbf{D} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{U}^\top \\ \mathbf{U}_\perp^\top\end{bmatrix}\xi[\mathbf{U} \quad \mathbf{U}_\perp] = \begin{bmatrix}\mathbf{U}^\top \\ \mathbf{U}_\perp^\top\end{bmatrix}(\mathbf{Y}\mathbf{A}^\top + \mathbf{A}\mathbf{Y}^\top)[\mathbf{U} \quad \mathbf{U}_\perp] = \begin{bmatrix}\mathbf{P}\mathbf{A}^\top\mathbf{U} + \mathbf{U}^\top\mathbf{A}\mathbf{P}^\top & \mathbf{P}\mathbf{A}^\top\mathbf{U}_\perp \\ \mathbf{U}_\perp^\top\mathbf{A}\mathbf{P}^\top & \mathbf{0}\end{bmatrix}$$

by observing $\mathbf{Y} = \mathbf{U}\mathbf{P}$. This implies $\mathbf{U}_\perp\mathbf{U}_\perp^\top\mathbf{A} = \mathbf{U}_\perp\mathbf{D}\mathbf{P}^{-\top}$ and $\mathbf{P}\mathbf{A}^\top\mathbf{U} + \mathbf{U}^\top\mathbf{A}\mathbf{P}^\top = \mathbf{S}$. By denoting $\mathbf{S}_1 = \mathbf{U}^\top\mathbf{A}\mathbf{P}^\top$, we have $\mathbf{S}_1 + \mathbf{S}_1^\top = \mathbf{S}$ and $\mathbf{U}\mathbf{U}^\top\mathbf{A} = \mathbf{U}\mathbf{S}_1\mathbf{P}^{-\top}$. Finally, $\mathbf{A} = \mathbf{U}\mathbf{U}^\top\mathbf{A} + \mathbf{U}_\perp\mathbf{U}_\perp^\top\mathbf{A} = (\mathbf{U}\mathbf{S}_1 + \mathbf{U}_\perp\mathbf{D})\mathbf{P}^{-\top} \in \mathscr{A}_{\mathbf{Y}}^\xi$. This proves $\mathscr{A}_{\mathbf{Y}}^\xi \supseteq \{\mathbf{A} : \mathbf{Y}\mathbf{A}^\top + \mathbf{A}\mathbf{Y}^\top = \xi\}$ and finishes the proof. ■

Proof of Lemma 3. First, it is easy to check the dimensions of $\mathscr{A}^{\mathbf{Y}}_{\text{null}}$ and $\mathscr{A}^{\mathbf{Y}}_{\overline{\text{null}}}$ are $(r^2-r)/2$ and $pr-(r^2-r)/2$, respectively. Since $(r^2-r)/2+pr-(r^2-r)/2=pr$, to prove $\mathbb{R}^{p\times r}=\mathscr{A}^{\mathbf{Y}}_{\text{null}}\oplus\mathscr{A}^{\mathbf{Y}}_{\overline{\text{null}}}$, we only need to show $\mathscr{A}^{\mathbf{Y}}_{\text{null}}$ is orthogonal to $\mathscr{A}^{\mathbf{Y}}_{\overline{\text{null}}}$. Suppose $\mathbf{A}=\mathbf{USP}^{-\top}\in\mathscr{A}^{\mathbf{Y}}_{\text{null}}$ and $\mathbf{A}'=(\mathbf{US}'+\mathbf{U}_\perp\mathbf{D}')\mathbf{P}^{-\top}\in\mathscr{A}^{\mathbf{Y}}_{\overline{\text{null}}}$. Then

$$\langle\mathbf{A},\mathbf{A}'\rangle=\langle\mathbf{SP}^{-\top},\mathbf{S}'\mathbf{P}^{-\top}\rangle=\langle\mathbf{S},\mathbf{S}'\mathbf{P}^{-\top}\mathbf{P}^{-1}\rangle\overset{(a)}{=}\langle\mathbf{S},\mathbf{S}'\boldsymbol{\Sigma}^{-1}\rangle\overset{(b)}{=}-\langle\mathbf{S}^\top,(\mathbf{S}'\boldsymbol{\Sigma}^{-1})^\top\rangle=-\langle\mathbf{A},\mathbf{A}'\rangle,$$

where (a) is because $\mathbf{PP}^\top=\boldsymbol{\Sigma}$, (b) is because $\mathbf{S}+\mathbf{S}^\top=\mathbf{0}$, and $\mathbf{S}'\boldsymbol{\Sigma}^{-1}$ is symmetric by the construction of $\mathscr{A}^{\mathbf{Y}}_{\text{null}}$ and $\mathscr{A}^{\mathbf{Y}}_{\overline{\text{null}}}$, respectively. So we have $\langle\mathbf{A},\mathbf{A}'\rangle=0$ and this finishes the proof of this lemma. ■

Proof of Corollary 1. First, by the connection of Riemannian and Euclidean gradients in (18), the connection of FOSPs under two formulations clearly holds.

Suppose $\mathbf{Y}$ is a rank r Euclidean SOSP of (3) and let $\mathbf{X}=\mathbf{YY}^\top$. Given any $\xi\in T_{\mathbf{X}}\mathcal{M}_{r+}$, we have

$$\text{Hess}f(\mathbf{X})[\xi,\xi]\overset{(19)}{=}\nabla^2 g(\mathbf{Y})[\mathcal{L}^{-1}_{\mathbf{Y}}(\xi),\mathcal{L}^{-1}_{\mathbf{Y}}(\xi)]\geqslant 0,$$

where the inequality is by the SOSP assumption on $\mathbf{Y}$. Combining the fact $\mathbf{X}$ is a Riemannian FOSP of (1), this shows $\mathbf{X}=\mathbf{YY}^\top$ is a Riemannian SOSP of (1).

Next, let us show the other direction: suppose $\mathbf{X}$ is a Riemannian SOSP of (1), then for any $\mathbf{Y}$ such that $\mathbf{YY}^\top=\mathbf{X}$, it is a Euclidean SOSP of (3). To see this, first $\mathbf{Y}$ is of rank r and we have shown $\mathbf{Y}$ is a Euclidean FOSP of (3). Then by (19), we have for any $\mathbf{A}\in\mathbb{R}^{p\times r}$:

$$\nabla^2 g(\mathbf{Y})[\mathbf{A},\mathbf{A}]=\text{Hess}f(\mathbf{X})[\xi^{\mathbf{A}}_{\mathbf{Y}},\xi^{\mathbf{A}}_{\mathbf{Y}}]\geqslant 0.$$

Suppose $\mathbf{Y}$ is a rank r Euclidean strict saddle of (3) and let $\mathbf{X}=\mathbf{YY}^\top$. It implies that there exists $\mathbf{A}\in\mathscr{A}^{\mathbf{Y}}_{\text{null}}$ such that $\nabla^2 g(\mathbf{Y})[\mathbf{A},\mathbf{A}]<0$. Then by (19) $\nabla^2 g(\mathbf{Y})[\mathbf{A},\mathbf{A}]=\text{Hess}f(\mathbf{X})[\mathcal{L}(\mathbf{A}),\mathcal{L}(\mathbf{A})]<0$, and this implies that $\text{Hess}f(\mathbf{X})$ also has at least one eigenvalue. Thus, $\mathbf{X}$ is a Riemannian strict saddle. The proof for the other direction is similar and for simplicity, we omit it here. ■

A.3. Additional Proofs in Section 4

Proof of Lemma 4. Given any tangent vector $\xi=[\mathbf{U}\quad\mathbf{U}_\perp]\begin{bmatrix}\mathbf{S} & \mathbf{D}_2^\top\\ \mathbf{D}_1 & \mathbf{0}\end{bmatrix}[\mathbf{V}\quad\mathbf{V}_\perp]^\top$ in $T_{\mathbf{X}}\mathcal{M}_r$, denote $\mathscr{A}_1=\{\mathbf{A}=[\mathbf{A}_L^\top\quad\mathbf{A}_R^\top]^\top:\mathbf{LA}_R^\top+\mathbf{A}_L\mathbf{R}^\top=\xi\}$ and $\mathscr{A}_2=\{\mathbf{A}=[\mathbf{A}_L^\top\quad\mathbf{A}_R^\top]^\top:\mathbf{LA}_R^\top+\mathbf{A}_L\mathbf{R}^\top=\xi$ and $\mathbf{L}^\top\mathbf{A}_L+\mathbf{A}_L^\top\mathbf{L}-\mathbf{R}^\top\mathbf{A}_R-\mathbf{A}_R^\top\mathbf{R}=\mathbf{0}\}$. The rest of the proof is divided into two steps: in Step 1 we show the results on $\mathscr{A}^{\xi}_{\mathbf{L},\mathbf{R}}$; in Step 2 we show the results on $\widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}}$.

Step 1. It is clear $\dim(\mathscr{A}^{\xi}_{\mathbf{L},\mathbf{R}})=r^2$. For any $\mathbf{A}=[\mathbf{A}_L^\top\quad\mathbf{A}_R^\top]^\top\in\mathscr{A}^{\xi}_{\mathbf{L},\mathbf{R}}$, it is straightforward to check $\mathbf{LA}_R^\top+\mathbf{A}_L\mathbf{R}^\top=\xi$, so $\mathscr{A}^{\xi}_{\mathbf{L},\mathbf{R}}\subseteq\mathscr{A}_1$. For any $\mathbf{A}$ such that $\mathbf{LA}_R^\top+\mathbf{A}_L\mathbf{R}^\top=\xi$, we have

$$\begin{aligned}\begin{bmatrix}\mathbf{S} & \mathbf{D}_2^\top\\ \mathbf{D}_1 & \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{U}^\top\\ \mathbf{U}_\perp^\top\end{bmatrix}\xi[\mathbf{V}\quad\mathbf{V}_\perp]&=\begin{bmatrix}\mathbf{U}^\top\\ \mathbf{U}_\perp^\top\end{bmatrix}(\mathbf{LA}_R^\top+\mathbf{A}_L\mathbf{R}^\top)[\mathbf{V}\quad\mathbf{V}_\perp]\\ &=\begin{bmatrix}\mathbf{P}_1\mathbf{A}_R^\top\mathbf{V}+\mathbf{U}^\top\mathbf{A}_L\mathbf{P}_2^\top & \mathbf{P}_1\mathbf{A}_R^\top\mathbf{V}_\perp\\ \mathbf{U}_\perp^\top\mathbf{A}_L\mathbf{P}_2^\top & \mathbf{0}\end{bmatrix}\end{aligned}\tag{3}$$

by observing $\mathbf{L}=\mathbf{UP}_1,\mathbf{R}=\mathbf{VP}_2$. This implies $P_{\mathbf{U}_\perp}\mathbf{A}_L=\mathbf{U}_\perp\mathbf{D}_1\mathbf{P}_2^{-\top}$, $P_{\mathbf{V}_\perp}\mathbf{A}_R=\mathbf{V}_\perp\mathbf{D}_2\mathbf{P}_1^{-\top}$ and $\mathbf{P}_1\mathbf{A}_R^\top\mathbf{V}+\mathbf{U}^\top\mathbf{A}_L\mathbf{P}_2^\top=\mathbf{S}$. By denoting $\mathbf{S}_1=\mathbf{U}^\top\mathbf{A}_L\mathbf{P}_2^\top$ and $\mathbf{S}_2^\top=\mathbf{V}^\top\mathbf{A}_R\mathbf{P}_1^\top$, we have $\mathbf{S}_1+\mathbf{S}_2=\mathbf{S}$, $P_{\mathbf{U}}\mathbf{A}_L=\mathbf{US}_1\mathbf{P}_2^{-\top}$ and $P_{\mathbf{V}}\mathbf{A}_R=\mathbf{VS}_2^\top\mathbf{P}_1^{-\top}$. Finally, $\mathbf{A}_L=P_{\mathbf{U}}\mathbf{A}_L+P_{\mathbf{U}_\perp}\mathbf{A}_L=(\mathbf{US}_1+\mathbf{U}_\perp\mathbf{D}_1)\mathbf{P}_2^{-\top}$, $\mathbf{A}_R=P_{\mathbf{V}}\mathbf{A}_R+P_{\mathbf{V}_\perp}\mathbf{A}_R=(\mathbf{VS}_2^\top+\mathbf{V}_\perp\mathbf{D}_2)\mathbf{P}_1^{-\top}$. So $\mathbf{A}=[\mathbf{A}_L^\top\quad\mathbf{A}_R^\top]^\top\in\mathscr{A}^{\xi}_{\mathbf{L},\mathbf{R}}$ and $\mathscr{A}^{\xi}_{\mathbf{L},\mathbf{R}}\supseteq\mathscr{A}_1$. This proves the first result.

Step 2. Let us begin by proving $\dim(\widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}}) = (r^2-r)/2$. First, by simple computation, we have $\dim(\widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}}) = \dim(\mathscr{S})$ where

$$\mathscr{S} := \left\{\mathbf{S}_1 \in \mathbb{R}^{r\times r} : \mathbf{P}_1^\top \mathbf{S}_1 \mathbf{P}_2^{-\top} + (\mathbf{P}_1^\top \mathbf{S}_1 \mathbf{P}_2^{-\top})^\top + \mathbf{P}_1^{-1}\mathbf{S}_1\mathbf{P}_2 + (\mathbf{P}_1^{-1}\mathbf{S}_1\mathbf{P}_2)^\top = \mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2 + (\mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2)^\top\right\}.$$

Next, we show $\mathscr{S}$ is of dimension $(r^2-r)/2$. Construct the following linear map $\varphi_{\mathbf{L},\mathbf{R}} : \mathbf{S}' \to \mathbf{P}_1^\top \mathbf{S}'\mathbf{P}_2^{-\top} + \mathbf{P}_1^{-1}\mathbf{S}'\mathbf{P}_2$. We claim $\varphi_{\mathbf{L},\mathbf{R}}$ is a bijective linear map over $\mathbb{R}^{r\times r}$:

- **injective part**: suppose there exists $\mathbf{S}'_1, \mathbf{S}'_2 \in \mathbb{R}^{r\times r}$ such that $\mathbf{S}'_1 \neq \mathbf{S}'_2$ and $\varphi_{\mathbf{L},\mathbf{R}}(\mathbf{S}'_1) = \varphi_{\mathbf{L},\mathbf{R}}(\mathbf{S}'_2)$. Then by definition of $\varphi_{\mathbf{L},\mathbf{R}}$, we have $\mathbf{P}_1^\top(\mathbf{S}'_1-\mathbf{S}'_2)\mathbf{P}_2^{-\top} + \mathbf{P}_1^{-1}(\mathbf{S}'_1-\mathbf{S}'_2)\mathbf{P}_2 = \mathbf{0}$. It further implies $\mathbf{P}_1\mathbf{P}_1^\top(\mathbf{S}'_1-\mathbf{S}'_2) + (\mathbf{S}'_1-\mathbf{S}'_2)\mathbf{P}_2\mathbf{P}_2^\top = \mathbf{0}$. This is a Sylvester equation with respect to $(\mathbf{S}'_1-\mathbf{S}'_2)$ and we know from (Bhatia 2013, Theorem VII.2.1) that it has a unique solution $\mathbf{0}$ due to the fact $\mathbf{P}_1\mathbf{P}_1^\top$ and $-\mathbf{P}_2\mathbf{P}_2^\top$ have disjoint spectra. So we get $\mathbf{S}'_1 = \mathbf{S}'_2$, a contradiction.
- **surjective part**: for any $\widetilde{\mathbf{S}} \in \mathbb{R}^{r\times r}$, we can find a unique $\widetilde{\mathbf{S}}'$ such that $\varphi_{\mathbf{L},\mathbf{R}}(\widetilde{\mathbf{S}}') = \widetilde{\mathbf{S}}$. This follows from the facts: (1) $\{\mathbf{S}' : \mathbf{P}_1^\top\mathbf{S}'\mathbf{P}_2^{-\top} + \mathbf{P}_1^{-1}\mathbf{S}'\mathbf{P}_2 = \widetilde{\mathbf{S}}\} = \{\mathbf{S}' : \mathbf{P}_1\mathbf{P}_1^\top\mathbf{S}' + \mathbf{S}'\mathbf{P}_2\mathbf{P}_2^\top = \mathbf{P}_1\widetilde{\mathbf{S}}\mathbf{P}_2^\top\}$; (2) $\mathbf{P}_1\mathbf{P}_1^\top\mathbf{S}' + \mathbf{S}'\mathbf{P}_2\mathbf{P}_2^\top = \mathbf{P}_1\widetilde{\mathbf{S}}\mathbf{P}_2^\top$ is a Sylvester equation with respect to $\mathbf{S}'$ which has a unique solution again by (Bhatia 2013, Theorem VII.2.1).

Then we have $\mathscr{S} = \{\varphi^{-1}_{\mathbf{L},\mathbf{R}}(\mathbf{S}') : \mathbf{S}' + \mathbf{S}'^\top = \mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2 + (\mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2)^\top\}$ and

$$\begin{aligned}\dim(\mathscr{S}) &= \dim(\{\varphi^{-1}_{\mathbf{L},\mathbf{R}}(\mathbf{S}') : \mathbf{S}' + \mathbf{S}'^\top = \mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2 + (\mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2)^\top\})\\ &= \dim(\{\mathbf{S}' : \mathbf{S}' + \mathbf{S}'^\top = \mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2 + (\mathbf{P}_1^{-1}\mathbf{S}\mathbf{P}_2)^\top\}) = (r^2-r)/2.\end{aligned}$$

Finally, we show the second result. For any $\mathbf{A} = [\mathbf{A}_L^\top \quad \mathbf{A}_R^\top]^\top \in \widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}}$, it is straightforward to check $\mathbf{L}\mathbf{A}_R^\top + \mathbf{A}_L\mathbf{R}^\top = \xi$ and $\mathbf{L}^\top\mathbf{A}_L + \mathbf{A}_L^\top\mathbf{L} - \mathbf{R}^\top\mathbf{A}_R - \mathbf{A}_R^\top\mathbf{R} = \mathbf{0}$. So $\widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}} \subseteq \mathscr{A}_2$. For any $\mathbf{A} \in \mathscr{A}_2$, following the same proof of (3) we have $\mathbf{A}_L = (\mathbf{U}\mathbf{S}_1 + \mathbf{U}_\perp\mathbf{D}_1)\mathbf{P}_2^{-\top}, \mathbf{A}_R = (\mathbf{V}\mathbf{S}_2^\top + \mathbf{V}_\perp\mathbf{D}_2)\mathbf{P}_1^{-\top}$ where $\mathbf{S}_1 = \mathbf{U}^\top\mathbf{A}_L\mathbf{P}_2^\top$, $\mathbf{S}_2^\top = \mathbf{V}^\top\mathbf{A}_R\mathbf{P}_1^\top$ and they satisfy $\mathbf{S}_1 + \mathbf{S}_2 = \mathbf{S}$. $\mathbf{L}^\top\mathbf{A}_L + \mathbf{A}_L^\top\mathbf{L} - \mathbf{R}^\top\mathbf{A}_R - \mathbf{A}_R^\top\mathbf{R} = \mathbf{0}$ further requires $\mathbf{S}_1, \mathbf{S}_2$ to satisfy $\mathbf{P}_1^\top\mathbf{S}_1\mathbf{P}_2^{-\top} + \mathbf{P}_2^{-1}\mathbf{S}_1^\top\mathbf{P}_1 - \mathbf{P}_2^\top\mathbf{S}_2^\top\mathbf{P}_1^{-\top} - \mathbf{P}_1^{-1}\mathbf{S}_2\mathbf{P}_2 = \mathbf{0}$. So $\mathbf{A} = [\mathbf{A}_L^\top \quad \mathbf{A}_R^\top]^\top \in \widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}}$ and $\widetilde{\mathscr{A}}^{\xi}_{\mathbf{L},\mathbf{R}} \supseteq \mathscr{A}_2$. This finishes the proof of this lemma. ■

Proof of Lemma 5. We first consider the result of $\mathscr{A}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ and $\mathscr{A}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$. It is easy to check $\mathscr{A}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ and $\mathscr{A}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$ are of dimensions r^2 and $(p_1+p_2-r)r$, respectively. Since $r^2 + (p_1+p_2-r)r = (p_1+p_2)r$, to prove $\mathbb{R}^{(p_1+p_2)\times r} = \mathscr{A}^{\mathbf{L},\mathbf{R}}_{\text{null}} \oplus \mathscr{A}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$, we only need to show $\mathscr{A}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ is orthogonal to $\mathscr{A}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$. Indeed, for any $\mathbf{A} = \begin{bmatrix}\mathbf{U}\mathbf{S}\mathbf{P}_2^{-\top}\\ -\mathbf{V}\mathbf{S}^\top\mathbf{P}_1^{-\top}\end{bmatrix} \in \mathscr{A}^{\mathbf{L},\mathbf{R}}_{\text{null}}$, and $\mathbf{A}' = \begin{bmatrix}(\mathbf{U}\mathbf{S}'\mathbf{P}_2\mathbf{P}_2^\top + \mathbf{U}_\perp\mathbf{D}'_1)\mathbf{P}_2^{-\top}\\ (\mathbf{V}\mathbf{S}'^\top\mathbf{P}_1\mathbf{P}_1^\top + \mathbf{V}_\perp\mathbf{D}'_2)\mathbf{P}_1^{-\top}\end{bmatrix} \in \mathscr{A}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$, by simple calculations, we have $\langle\mathbf{A},\mathbf{A}'\rangle = \langle\mathbf{S},\mathbf{S}'\rangle - \langle\mathbf{S},\mathbf{S}'\rangle = 0$.

Next, we prove the result of $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ and $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$. From the dimension of $\mathscr{S}$ in Step 2 of the proof of Lemma 4, we have $\dim(\mathscr{S}_{\mathbf{L},\mathbf{R}}) = (r^2-r)/2$. As a result of this, we have $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ is of dimension $(r^2-r)/2$. Thus, $(\mathbf{S}_1 - \mathbf{S}_2) \perp \mathscr{S}_{\mathbf{L},\mathbf{R}}$ in the definition of $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$ adds $(r^2-r)/2$ constraints and $\dim(\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}) = (p_1+p_2)r - (r^2-r)/2$. Now, to prove $\mathbb{R}^{(p_1+p_2)\times r} = \widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\text{null}} \oplus \widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$, we only need to show $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ is orthogonal to $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$. In fact, for any $\mathbf{A} = \begin{bmatrix}\mathbf{U}\mathbf{S}\mathbf{P}_2^{-\top}\\ -\mathbf{V}\mathbf{S}^\top\mathbf{P}_1^{-\top}\end{bmatrix} \in \widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ and $\mathbf{A}' = \begin{bmatrix}(\mathbf{U}\mathbf{S}'_1\mathbf{P}_2\mathbf{P}_2^\top + \mathbf{U}_\perp\mathbf{D}'_1)\mathbf{P}_2^{-\top}\\ (\mathbf{V}\mathbf{S}_2'^\top\mathbf{P}_1\mathbf{P}_1^\top + \mathbf{V}_\perp\mathbf{D}'_2)\mathbf{P}_1^{-\top}\end{bmatrix} \in \widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$, we have $\langle\mathbf{A},\mathbf{A}'\rangle = \langle\mathbf{S},\mathbf{S}'_1\rangle - \langle\mathbf{S},\mathbf{S}'_2\rangle = 0$, where the second equality is because $\mathbf{S} \in \mathscr{S}_{\mathbf{L},\mathbf{R}}$ and $(\mathbf{S}'_1 - \mathbf{S}'_2) \perp \mathscr{S}_{\mathbf{L},\mathbf{R}}$ by the construction of $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\text{null}}$ and $\widetilde{\mathscr{A}}^{\mathbf{L},\mathbf{R}}_{\overline{\text{null}}}$, respectively. This finishes the proof of this lemma. ■

Proof of Corollary 2. First, for any Euclidean FOSP $(\mathbf{L},\mathbf{R})$ of (5) or $(\mathbf{L},\mathbf{R})$ such that $\mathbf{L}^\top\mathbf{L} = \mathbf{R}^\top\mathbf{R}$, we have $\nabla g_{\text{reg}}(\mathbf{L},\mathbf{R}) = \nabla g(\mathbf{L},\mathbf{R})$ by (45) and Proposition 1, respectively. The connection on FOSPs of different formulations can be easily obtained by the connection of Riemannian and Euclidean gradients given in (40). Next, we show the equivalence on SOSPs of different formulations.

Suppose $\mathbf{X}$ is a Riemannian SOSP of (2), we claim any $(\mathbf{L},\mathbf{R})$ such that $\mathbf{L}\mathbf{R}^\top = \mathbf{X}$ is a Euclidean SOSP of (4) and any $(\mathbf{L},\mathbf{R})$ such that $\mathbf{L}\mathbf{R}^\top = \mathbf{X}$ and $\mathbf{L}^\top\mathbf{L} = \mathbf{R}^\top\mathbf{R}$ is a Euclidean SOSP of (5). To see it, first $(\mathbf{L},\mathbf{R})$ in both cases are Euclidean FOSP of (4) and (5) as we mentioned before. For any $\mathbf{A} = [\mathbf{A}_L^\top \quad \mathbf{A}_R^\top]^\top \in \mathbb{R}^{(p_1+p_2)\times r}$, by Theorems 2 and 3 we have

$$\nabla^2 g(\mathbf{L},\mathbf{R})[\mathbf{A},\mathbf{A}] \overset{(41)}{=} \mathrm{Hess} f(\mathbf{X})[\xi^{\mathbf{A}}_{\mathbf{L},\mathbf{R}}, \xi^{\mathbf{A}}_{\mathbf{L},\mathbf{R}}] \geqslant 0;$$
$$\nabla^2 g_{\mathrm{reg}}(\mathbf{L},\mathbf{R})[\mathbf{A},\mathbf{A}] \overset{(47)}{\geqslant} \mathrm{Hess} f(\mathbf{X})[\xi^{\mathbf{A}}_{\mathbf{L},\mathbf{R}}, \xi^{\mathbf{A}}_{\mathbf{L},\mathbf{R}}] \geqslant 0.$$

Next we show the reverse direction: suppose $(\mathbf{L},\mathbf{R})$ is a rank r Euclidean SOSP of (4) or (5), then $\mathbf{X} = \mathbf{L}\mathbf{R}^\top$ is a Riemannian SOSP of (2). To see this, for any $\xi \in T_{\mathbf{X}}\mathcal{M}_r$,

$$\mathrm{Hess} f(\mathbf{L}\mathbf{R}^\top)[\xi,\xi] \overset{(41)}{=} \nabla^2 g(\mathbf{L},\mathbf{R})[\mathcal{L}^{-1}_{\mathbf{L},\mathbf{R}}(\xi), \mathcal{L}^{-1}_{\mathbf{L},\mathbf{R}}(\xi)] \geqslant 0,$$
$$\mathrm{Hess} f(\mathbf{L}\mathbf{R}^\top)[\xi,\xi] \overset{(47)}{=} \nabla^2 g_{\mathrm{reg}}(\mathbf{L},\mathbf{R})[\mathcal{L}^{-1}_{\mathbf{L},\mathbf{R}}(\xi), \mathcal{L}^{-1}_{\mathbf{L},\mathbf{R}}(\xi)] \geqslant 0.$$

This shows $\mathbf{X}$ is a Riemannian SOSP of (2).

Suppose $(\mathbf{L},\mathbf{R})$ is a rank r Euclidean strict saddle of (4) or (5), and let $\mathbf{X} = \mathbf{L}\mathbf{R}^\top$. Then by definition there exists $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{(p_1+p_2)\times r}$ such that $\nabla^2 g(\mathbf{L},\mathbf{R})[\mathbf{A}_1,\mathbf{A}_1] < 0$ and $\nabla^2 g_{\mathrm{reg}}(\mathbf{L},\mathbf{R})[\mathbf{A}_2,\mathbf{A}_2] < 0$. Then

$$\mathrm{Hess} f(\mathbf{X})[\xi^{\mathbf{A}_1}_{\mathbf{L},\mathbf{R}}, \xi^{\mathbf{A}_1}_{\mathbf{L},\mathbf{R}}] \overset{(41)}{=} \nabla^2 g(\mathbf{L},\mathbf{R})[\mathbf{A}_1,\mathbf{A}_1] < 0;$$
$$\mathrm{Hess} f(\mathbf{X})[\xi^{\mathbf{A}_2}_{\mathbf{L},\mathbf{R}}, \xi^{\mathbf{A}_2}_{\mathbf{L},\mathbf{R}}] \overset{(47)}{\leqslant} \nabla^2 g_{\mathrm{reg}}(\mathbf{L},\mathbf{R})[\mathbf{A}_2,\mathbf{A}_2] < 0.$$

This implies that $\mathrm{Hess} f(\mathbf{X})$ has negative eigenvalues in both cases, i.e., $\mathbf{X}$ is a Riemannian strict saddle. The proof for the reverse direction is similar and for simplicity, we omit it here. ■

Proof of Theorem 4. This proof is divided into two steps. In Step 1, we show (54); in Step 2, we give the spectrum bounds for the bijective map $\mathcal{J}$ and the spectrum connection between $\nabla^2 g_{\mathrm{reg}}(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})$ and $\nabla^2 g(\mathbf{L},\mathbf{R})$.

Step 1. First, since $\mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top = \mathbf{L}\mathbf{R}^\top$, $\mathbf{L}_{\mathrm{reg}}$ and $\mathbf{L}$ share the same left singular subspace. Thus $\mathbf{L}\boldsymbol{\Delta} = \mathbf{L}\mathbf{L}^\dagger\mathbf{L}_{\mathrm{reg}} = \mathbf{L}_{\mathrm{reg}}$ and $\boldsymbol{\Delta}$ is of rank r. Meanwhile, by $\mathbf{L}\mathbf{R}^\top = \mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top$, we have $\boldsymbol{\Delta}\mathbf{R}_{\mathrm{reg}}^\top = \mathbf{L}^\dagger\mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top = \mathbf{L}^\dagger\mathbf{L}\mathbf{R}^\top = \mathbf{R}^\top$. Moreover, as $(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})$ is a Euclidean FOSP of (5), by (46) we have for any $\mathbf{A} = [\mathbf{A}_L^\top \quad \mathbf{A}_R^\top]^\top \in \mathbb{R}^{(p_1+p_2)\times r}$:

$$\nabla^2 g_{\mathrm{reg}}(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})[\mathbf{A},\mathbf{A}] - \mu\|\mathbf{L}_{\mathrm{reg}}^\top\mathbf{A}_L + \mathbf{A}_L^\top\mathbf{L}_{\mathrm{reg}} - \mathbf{R}_{\mathrm{reg}}^\top\mathbf{A}_R - \mathbf{A}_R^\top\mathbf{R}_{\mathrm{reg}}\|_{\mathrm{F}}^2 = \nabla g^2(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})[\mathbf{A},\mathbf{A}].$$

Next, we show $\nabla g^2(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})[\mathbf{A},\mathbf{A}] = \nabla g^2(\mathbf{L},\mathbf{R})[\mathcal{J}(\mathbf{A}),\mathcal{J}(\mathbf{A})]$. By Proposition 2 we have

$$\begin{aligned}
&\nabla^2 g(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})[\mathbf{A},\mathbf{A}]\\
&= \nabla^2 f(\mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top)[\mathbf{L}_{\mathrm{reg}}\mathbf{A}_R^\top + \mathbf{A}_L\mathbf{R}_{\mathrm{reg}}^\top, \mathbf{L}_{\mathrm{reg}}\mathbf{A}_R^\top + \mathbf{A}_L\mathbf{R}_{\mathrm{reg}}^\top] + 2\langle\nabla f(\mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top), \mathbf{A}_L\mathbf{A}_R^\top\rangle\\
&= \nabla^2 f(\mathbf{L}\mathbf{R}^\top)[\mathbf{L}\boldsymbol{\Delta}\mathbf{A}_R^\top + \mathbf{A}_L\boldsymbol{\Delta}^{-1}\mathbf{R}^\top, \mathbf{L}\boldsymbol{\Delta}\mathbf{A}_R^\top + \mathbf{A}_L\boldsymbol{\Delta}^{-1}\mathbf{R}^\top]\\
&\quad + 2\langle\nabla f(\mathbf{L}\mathbf{R}^\top), \mathbf{A}_L\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta}\mathbf{A}_R^\top\rangle\\
&= \nabla^2 g(\mathbf{L},\mathbf{R})[\mathcal{J}(\mathbf{A}),\mathcal{J}(\mathbf{A})].
\end{aligned}$$

This finishes the proof for the first part.

Step 2. Next, we provide the spectrum bounds for the bijection operator. Suppose $\mathbf{A} = [\mathbf{A}_L^\top \quad \mathbf{A}_R^\top]^\top$ and $\mathcal{J}(\mathbf{A}) = [\mathbf{A}_L'^\top \quad \mathbf{A}_R'^\top]^\top$. Then

$$\|\mathcal{J}(\mathbf{A})\|_{\mathrm{F}}^2 = \|\mathbf{A}'_L\|_{\mathrm{F}}^2 + \|\mathbf{A}'_R\|_{\mathrm{F}}^2 = \|\mathbf{A}_L\boldsymbol{\Delta}^{-1}\|_{\mathrm{F}}^2 + \|\mathbf{A}_R\boldsymbol{\Delta}^\top\|_{\mathrm{F}}^2 \leqslant (\sigma_1(\boldsymbol{\Delta}) \vee (1/\sigma_r(\boldsymbol{\Delta})))^2\,\|\mathbf{A}\|_{\mathrm{F}}^2,$$
$$\|\mathbf{A}\|_{\mathrm{F}}^2 = \|\mathbf{A}_L\|_{\mathrm{F}}^2 + \|\mathbf{A}_R\|_{\mathrm{F}}^2 = \|\mathbf{A}'_L\boldsymbol{\Delta}\|_{\mathrm{F}}^2 + \|\mathbf{A}'_R\boldsymbol{\Delta}^{-\top}\|_{\mathrm{F}}^2 \leqslant (\sigma_1(\boldsymbol{\Delta}) \vee (1/\sigma_r(\boldsymbol{\Delta})))^2\,\|\mathcal{J}(\mathbf{A})\|_{\mathrm{F}}^2.$$

Finally, we provide a spectrum connection of two Euclidean Hessians at FOSPs. By (54), we have $\nabla^2 g_{\mathrm{reg}}(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}}) \succcurlyeq \mathcal{J}^*\nabla g^2(\mathbf{L},\mathbf{R})\mathcal{J}$. So the first inequality of (56) follows from Lemma 3(ii) in the Appendix and (55). Also by (45), (54) and Lemma 5, we have $\nabla^2 g_{\mathrm{reg}}(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}}) - 8\mu\sigma_1(\mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top)\mathcal{I} \preccurlyeq \mathcal{J}^*\nabla g^2(\mathbf{L},\mathbf{R})\mathcal{J}$ and the second inequality in (56) follows from Lemma 3(i) and (55). This finishes the proof. ■

A.4. Additional Proofs in Section 5

Proof of Theorem 5. By Theorem I.1 and Theorem II.2 of Li et al. (2019), we have with probability at least $1-\exp(-C'n)$, the factorization formulation $g(\mathbf{x})$ in (59) has the following geometric landscape properties: (1) $\mathbf{x}^*$ is the unique Euclidean SOSP of $g(\mathbf{x})$; (2) for any other non-zero Euclidean FOSP $\mathbf{x}$ of $g(\mathbf{x})$, it satisfies $\lambda_{\min}(\nabla^2 g(\mathbf{x})) \leqslant -3\|\mathbf{x}^*\|_2^2 = -3\sigma_1(\mathbf{X}^*)$ under the assumptions of Theorem 5.

By Corollary 1, we have $\mathbf{X}^* = \mathbf{x}^*\mathbf{x}^{*\top}$ is the unique Riemannian SOSP of (61). In addition, by Theorem 1, for any other Riemannian FOSP $\mathbf{X}$ of (61), we have

$$\lambda_{\min}(\mathrm{Hess} f(\mathbf{X})) \leqslant \frac{1}{4\sigma_1(\mathbf{X})}\lambda_{\min}(\nabla^2 g(\mathbf{x})) \leqslant -\frac{3\sigma_1(\mathbf{X}^*)}{4\sigma_1(\mathbf{X})},$$

where $\mathbf{x}$ is any Euclidean FOSP satisfying $\mathbf{x}\mathbf{x}^\top = \mathbf{X}$. ■

Proof of Theorem 6. First, Zhu et al. (2018) considered the geometric landscape of (5) when f satisfies the $(2r,4r)$-restricted strong convexity and smoothness property. Under the assumptions of Theorem 6, Theorem 3 of Zhu et al. (2018) shows any Euclidean SOSP $(\mathbf{L},\mathbf{R})$ of the regularized factorization formulation satisfies $\mathbf{L}\mathbf{R}^\top = \mathbf{X}^*$. By Corollary 2 of this paper, we further conclude if the input rank $r = r^*$ in (2), then $\mathbf{X}^*$ is the unique Riemannian SOSP of (2) and if $r > r^*$, there is no Riemannian SOSP of (2).

At the same time, by Theorem 3 of Zhu et al. (2018), any Euclidean FOSP $(\mathbf{L},\mathbf{R})$ of (5) that is not a SOSP must be a strict saddle and satisfy

$$\lambda_{\min}(\nabla^2 g_{\mathrm{reg}}(\mathbf{L},\mathbf{R})) \leqslant \begin{cases} -0.08\alpha_1\sigma_r(\mathbf{X}^*), & \text{if } r = r^*; \\ -0.05\alpha_1\cdot(\sigma_{r^c}^2(\mathbf{W}) \wedge 2\sigma_{r^*}(\mathbf{X}^*)), & \text{if } r > r^*, \end{cases}$$

where $\mathbf{W} = [\mathbf{L}^\top \quad \mathbf{R}^\top]^\top$ and r^c is the rank of $\mathbf{W}$. Under the manifold formulation (2), by Theorem 3, any Riemannian FOSP $\mathbf{X}$ that is not a Riemannian SOSP must satisfy

$$\begin{aligned}\lambda_{\min}(\mathrm{Hess} f(\mathbf{X})) &\leqslant \lambda_{\min}(\nabla^2 g_{\mathrm{reg}}(\mathbf{L}',\mathbf{R}'))/2\sigma_1(\mathbf{X}) \\ &\leqslant \begin{cases} -0.08\alpha_1\sigma_r(\mathbf{X}^*)/(2\sigma_1(\mathbf{X})), & \text{if } r = r^*; \\ -0.05\alpha_1\cdot(\sigma_r^2(\mathbf{W}') \wedge 2\sigma_{r^*}(\mathbf{X}^*))/(2\sigma_1(\mathbf{X})), & \text{if } r > r^*, \end{cases}\end{aligned}$$

where $\mathbf{W}' = [\mathbf{L}'^\top \quad \mathbf{R}'^\top]^\top$ and $(\mathbf{L}',\mathbf{R}')$ is a rank r Euclidean FOSP of (5) satisfying $\mathbf{L}'\mathbf{R}'^\top = \mathbf{X}$. Finally, we only need to compute $\sigma_r^2(\mathbf{W}')$. By Lemma 6 we have $\mathbf{L}' = \mathbf{U}\mathbf{P}$ and $\mathbf{R}' = \mathbf{V}\mathbf{P}$ for some invertible $\mathbf{P} \in \mathbb{R}^{r\times r}$, where $\mathbf{U},\mathbf{V}$ are the left and right singular subspaces of $\mathbf{X}$. So $\sigma_r(\mathbf{W}') = \sigma_r([\mathbf{L}'^\top \quad \mathbf{R}'^\top]^\top) = \sqrt{2}\sigma_r(\mathbf{P}) = \sqrt{2\sigma_r(\mathbf{X})}$. This finishes the proof of this theorem. ■

Proof of Theorem 7. Under the assumptions of Theorem 7, by Theorem 3 of Zhu et al. (2018) we have for a rank r Euclidean FOSP $(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})$ of the regularized formulation (5), it is either a Euclidean SOSP satisfying $\mathbf{L}_{\mathrm{reg}}\mathbf{R}_{\mathrm{reg}}^\top = \mathbf{X}^*$ or a strict saddle with $\lambda_{\min}(\nabla^2 g_{\mathrm{reg}}(\mathbf{L}_{\mathrm{reg}},\mathbf{R}_{\mathrm{reg}})) \leqslant -0.08\alpha_1\sigma_r(\mathbf{X}^*)$.

By Corollary 2 and Theorem 4, we have for any rank r Euclidean FOSP $(\mathbf{L},\mathbf{R})$ of (4), it is either a Euclidean SOSP satisfying $\mathbf{L}\mathbf{R}^\top = \mathbf{X}^*$ or a strict saddle with

$$\lambda_{\min}(\nabla^2 g(\mathbf{L},\mathbf{R})) \leqslant \theta_{\mathbf{\Delta}}^2\lambda_{\min}(\nabla^2 g_{\mathrm{reg}}(\mathbf{L}'_{\mathrm{reg}},\mathbf{R}'_{\mathrm{reg}})) \leqslant -0.08\theta_{\mathbf{\Delta}}^2\alpha_1\sigma_r(\mathbf{X}^*),$$

where $\theta_{\mathbf{\Delta}} := (1/\sigma_1(\mathbf{\Delta})) \wedge \sigma_r(\mathbf{\Delta})$, $\mathbf{\Delta} = \mathbf{L}^\dagger\mathbf{L}'_{\mathrm{reg}}$ and $(\mathbf{L}'_{\mathrm{reg}},\mathbf{R}'_{\mathrm{reg}})$ is a rank r Euclidean FOSP of (5) satisfying $\mathbf{L}'_{\mathrm{reg}}\mathbf{R}'^\top_{\mathrm{reg}} = \mathbf{L}\mathbf{R}^\top =: \mathbf{X}$.

Finally, we give a lower bound for $\theta_{\mathbf{\Delta}}$. Notice $\mathbf{L}\mathbf{\Delta} = \mathbf{L}'_{\mathrm{reg}}$, and

$$\begin{aligned}&\sigma_1(\mathbf{\Delta}) = \sigma_1(\mathbf{L}^\dagger\mathbf{L}'_{\mathrm{reg}}) \leqslant \sigma_1(\mathbf{L}^\dagger)\sigma_1(\mathbf{L}'_{\mathrm{reg}}) \overset{(45),\text{Lemma 6}}{=} \sigma_1^{1/2}(\mathbf{X})/\sigma_r(\mathbf{L}), \\ &\sigma_r^{1/2}(\mathbf{X}) \overset{(45),\text{Lemma 6}}{=} \sigma_r(\mathbf{L}'_{\mathrm{reg}}) = \sigma_r(\mathbf{L}\mathbf{\Delta}) = \inf_{\mathbf{x}:\|\mathbf{x}\|_2=1}\|\mathbf{L}\mathbf{\Delta}\mathbf{x}\|_2 \leqslant \sigma_1(\mathbf{L})\inf_{\mathbf{x}:\|\mathbf{x}\|_2=1}\|\mathbf{\Delta}\mathbf{x}\|_2 = \sigma_1(\mathbf{L})\sigma_r(\mathbf{\Delta}).\end{aligned}$$

We have $\theta_{\mathbf{\Delta}} := (1/\sigma_1(\mathbf{\Delta})) \wedge \sigma_r(\mathbf{\Delta}) \geqslant (\sigma_r(\mathbf{L})/\sigma_1^{1/2}(\mathbf{X})) \wedge (\sigma_r^{1/2}(\mathbf{X})/\sigma_1(\mathbf{L}))$. This finishes the proof of this theorem. ■

Appendix B: Additional Lemmas

Recall $\lambda_k(\cdot)$ and $\sigma_k(\cdot)$ are the kth largest eigenvalue and kth largest singular value of matrix $(\cdot)$. Also $\lambda_{\max}(\cdot)$, $\lambda_{\min}(\cdot)$ denote the largest and least eigenvalue of matrix $(\cdot)$.

Lemma 2 *Suppose $\mathbf{A}\in\mathbb{S}^{p\times p}$ is symmetric and $\mathbf{P}\in\mathbb{R}^{p\times p}$ is invertible. Then $\lambda_k(\mathbf{P}^\top\mathbf{A}\mathbf{P})$ is sandwiched between $\sigma_p^2(\mathbf{P})\lambda_k(\mathbf{A})$ and $\sigma_1^2(\mathbf{P})\lambda_k(\mathbf{A})$ for $k=1,\ldots,p$.*

Proof. Suppose $\mathbf{u}_1,\ldots,\mathbf{u}_p$ are eigenvectors corresponding to $\lambda_1(\mathbf{A}),\ldots,\lambda_p(\mathbf{A})$ and $\mathbf{v}_1,\ldots,\mathbf{v}_p$ are eigenvectors corresponding to $\lambda_1(\mathbf{P}^\top\mathbf{A}\mathbf{P}),\ldots,\lambda_p(\mathbf{P}^\top\mathbf{A}\mathbf{P})$. For $k=1,\ldots,p$, define

$$\mathcal{U}_k=\text{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_k\},\quad \mathcal{U}'_k=\text{span}\{\mathbf{P}^{-1}\mathbf{u}_1,\ldots,\mathbf{P}^{-1}\mathbf{u}_k\},$$
$$\mathcal{V}_k=\text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_k\},\quad \mathcal{V}'_k=\text{span}\{\mathbf{P}\mathbf{v}_1,\ldots,\mathbf{P}\mathbf{v}_k\}.$$

Let us first consider the case that $\lambda_k(\mathbf{A})\geqslant 0$. By Lemma 4, we have

$$\lambda_k(\mathbf{P}^\top\mathbf{A}\mathbf{P})\geqslant\min_{\mathbf{u}\in\mathcal{U}'_k,\mathbf{u}\neq\mathbf{0}}\frac{\mathbf{u}^\top\mathbf{P}^\top\mathbf{A}\mathbf{P}\mathbf{u}}{\|\mathbf{u}\|_2^2}=\min_{\mathbf{u}\in\mathcal{U}_k,\mathbf{u}\neq\mathbf{0}}\frac{\mathbf{u}^\top\mathbf{A}\mathbf{u}}{\|\mathbf{P}^{-1}\mathbf{u}\|_2^2}\geqslant\min_{\mathbf{u}\in\mathcal{U}_k,\mathbf{u}\neq\mathbf{0}}\frac{\lambda_k(\mathbf{A})\|\mathbf{u}\|_2^2}{\|\mathbf{P}^{-1}\mathbf{u}\|_2^2}\geqslant\lambda_k(\mathbf{A})\sigma_p^2(\mathbf{P})\geqslant 0. \tag{4}$$

On the other hand, we have

$$\begin{aligned}\lambda_k(\mathbf{A})&\overset{\text{Lemma 4}}{\geqslant}\min_{\mathbf{u}\in\mathcal{V}'_k,\mathbf{u}\neq\mathbf{0}}\frac{\mathbf{u}^\top\mathbf{P}^{-\top}\mathbf{P}^\top\mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{u}}{\|\mathbf{u}\|_2^2}=\min_{\mathbf{v}\in\mathcal{V}_k,\mathbf{v}\neq\mathbf{0}}\frac{\mathbf{v}^\top\mathbf{P}^\top\mathbf{A}\mathbf{P}\mathbf{v}}{\|\mathbf{P}\mathbf{v}\|_2^2}\geqslant\min_{\mathbf{v}\in\mathcal{V}_k,\mathbf{v}\neq\mathbf{0}}\frac{\lambda_k(\mathbf{P}^\top\mathbf{A}\mathbf{P})\|\mathbf{v}\|_2^2}{\|\mathbf{P}\mathbf{v}\|_2^2}\\&\overset{(4)}{\geqslant}\frac{\lambda_k(\mathbf{P}^\top\mathbf{A}\mathbf{P})}{\sigma_1^2(\mathbf{P})}.\end{aligned}\tag{5}$$

So we have proved the result for the case that $\lambda_k(\mathbf{A})\geqslant 0$. When $\lambda_k(\mathbf{A})<0$, we have $\lambda_{p+1-k}(-\mathbf{A})=-\lambda_k(\mathbf{A})>0$. Following the same proof of (4) and (5), we have

$$-\lambda_k(\mathbf{P}^\top\mathbf{A}\mathbf{P})=\lambda_{p+1-k}(-\mathbf{P}^\top\mathbf{A}\mathbf{P})\geqslant\sigma_p^2(\mathbf{P})\lambda_{p+1-k}(-\mathbf{A})=-\sigma_p^2(\mathbf{P})\lambda_k(\mathbf{A})>0,$$
$$-\lambda_k(\mathbf{A})=\lambda_{p+1-k}(-\mathbf{A})\geqslant\lambda_{p+1-k}(-\mathbf{P}^\top\mathbf{A}\mathbf{P})/\sigma_1^2(\mathbf{P})=-\lambda_k(\mathbf{P}^\top\mathbf{A}\mathbf{P})/\sigma_1^2(\mathbf{P}).$$

This finishes the proof of this lemma. ■

Lemma 3 *Suppose $\mathbf{A}\in\mathbb{S}^{p\times p},\mathbf{B}\in\mathbb{S}^{q\times q}$ are symmetric matrices with $q\geqslant p$ and $\mathbf{P}\in\mathbb{R}^{q\times p},\mathbf{Q}\in\mathbb{R}^{p\times q}$.*

(i) If $\mathbf{P}^\top\mathbf{B}\mathbf{P}\succcurlyeq\mathbf{A}$, then $\lambda_k(\mathbf{B})\sigma_1^2(\mathbf{P})\vee\lambda_k(\mathbf{B})\sigma_p^2(\mathbf{P})\geqslant\lambda_k(\mathbf{A})$ holds for $k=1,\ldots,p$.

(ii) If $\mathbf{P}^\top\mathbf{B}\mathbf{P}\preccurlyeq\mathbf{A}$, then $\lambda_{k+q-p}(\mathbf{B})\sigma_1^2(\mathbf{P})\wedge\lambda_{k+q-p}(\mathbf{B})\sigma_p^2(\mathbf{P})\leqslant\lambda_k(\mathbf{A})$ holds for $k=1,\ldots,p$.

(iii) If $\mathbf{Q}^\top\mathbf{A}\mathbf{Q}\preccurlyeq\mathbf{B}$, then $\lambda_{\min}(\mathbf{B})\geqslant\sigma_1^2(\mathbf{Q})\lambda_{\min}(\mathbf{A})\wedge 0$.

(iv) If $\mathbf{Q}^\top\mathbf{A}\mathbf{Q}\succcurlyeq\mathbf{B}$, then $\lambda_1(\mathbf{B})\leqslant\sigma_1^2(\mathbf{Q})\lambda_{\max}(\mathbf{A})\vee 0$.

Proof. We first prove the first and the second claims under the assumption that $\sigma_p(\mathbf{P})>0$, i.e., all p columns of $\mathbf{P}$ are linearly independent.

Suppose $\mathbf{u}_1,\ldots,\mathbf{u}_p$ are eigenvectors corresponding to $\lambda_1(\mathbf{A}),\ldots,\lambda_p(\mathbf{A})$, respectively and let $\mathcal{U}_k=\text{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$. Then

$$\lambda_k(\mathbf{B})\overset{(a)}{\geqslant}\inf_{\mathbf{u}\in\mathcal{U}_k}\frac{\mathbf{u}^\top\mathbf{P}^\top\mathbf{B}\mathbf{P}\mathbf{u}}{\|\mathbf{P}\mathbf{u}\|_2^2}\geqslant\inf_{\mathbf{u}\in\mathcal{U}_k}\frac{\mathbf{u}^\top\mathbf{A}\mathbf{u}}{\|\mathbf{P}\mathbf{u}\|_2^2}\geqslant\inf_{\mathbf{u}\in\mathcal{U}_k}\frac{\lambda_k(\mathbf{A})\|\mathbf{u}\|_2^2}{\|\mathbf{P}\mathbf{u}\|_2^2}\geqslant\begin{cases}\lambda_k(\mathbf{A})/\sigma_1^2(\mathbf{P}), & \text{if } \lambda_k(\mathbf{A})\geqslant 0;\\ \lambda_k(\mathbf{A})/\sigma_p^2(\mathbf{P}), & \text{if } \lambda_k(\mathbf{A})<0.\end{cases}$$

Here (a) is because $\{\mathbf{P}\mathbf{u}_1,\ldots,\mathbf{P}\mathbf{u}_k\}$ forms a k dimensional subspace in $\mathbb{R}^q$ and Lemma 4.

To see the second claim under $\sigma_p(\mathbf{P})>0$, suppose $\mathbf{v}_1,\ldots,\mathbf{v}_q$ are eigenvectors corresponding to $\lambda_1(\mathbf{B}),\ldots,\lambda_q(\mathbf{B})$ and let $\mathcal{V}_{k+q-p}=\text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_{k+q-p}\}$.

$$\begin{aligned}\lambda_k(\mathbf{A})&\overset{(a)}{\geqslant}\inf_{\mathbf{v}:\mathbf{P}\mathbf{v}\in\mathcal{V}_{k+q-p}}\frac{\mathbf{v}^\top\mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}\geqslant\inf_{\mathbf{v}:\mathbf{P}\mathbf{v}\in\mathcal{V}_{k+q-p}}\frac{\mathbf{v}^\top\mathbf{P}^\top\mathbf{B}\mathbf{P}\mathbf{v}}{\|\mathbf{v}\|_2^2}\geqslant\inf_{\mathbf{v}:\mathbf{P}\mathbf{v}\in\mathcal{V}_{k+q-p}}\frac{\lambda_{k+q-p}(\mathbf{B})\|\mathbf{P}\mathbf{v}\|_2^2}{\|\mathbf{v}\|_2^2}\\&\geqslant\begin{cases}\sigma_p^2(\mathbf{P})\lambda_{k+q-p}(\mathbf{B}), & \text{if } \lambda_{k+q-p}(\mathbf{B})\geqslant 0\\ \sigma_1^2(\mathbf{P})\lambda_{k+q-p}(\mathbf{B}), & \text{if } \lambda_{k+q-p}(\mathbf{B})<0\end{cases}\end{aligned}\tag{6}$$

Here (a) is because of Lemma 4 and the fact $\{\mathbf{v}:\mathbf{P}\mathbf{v}\in\mathcal{V}_{k+q-p}\}$ has dimension at least k.

When $\sigma_p(\mathbf{P})=0$, we construct a series of matrices $\mathbf{P}_l$ such that $\lim_{l\to\infty}\mathbf{P}_l=\mathbf{P}$ and $\sigma_p(\mathbf{P}_l)>0$. According to the previous proofs,

$$\lambda_k(\mathbf{B})\sigma_1^2(\mathbf{P}_l)\vee\lambda_k(\mathbf{B})\sigma_p^2(\mathbf{P}_l)\geqslant\lambda_k(\mathbf{P}_l^\top\mathbf{B}\mathbf{P}_l),$$

$$\lambda_{k+q-p}(\mathbf{B})\sigma_1^2(\mathbf{P}_l)\wedge\lambda_{k+q-p}(\mathbf{B})\sigma_p^2(\mathbf{P}_l)\leqslant\lambda_k(\mathbf{P}_l^\top\mathbf{B}\mathbf{P}_l).$$

Since $\sigma_k(\cdot)$ and $\lambda_k(\cdot)$ are continuous functions of the input matrix, by taking $l\to\infty$, we have

$$\lambda_k(\mathbf{B})\sigma_1^2(\mathbf{P})\vee\lambda_k(\mathbf{B})\sigma_p^2(\mathbf{P})\geqslant\lambda_k(\mathbf{P}^\top\mathbf{B}\mathbf{P})\overset{(a)}{\geqslant}\lambda_k(\mathbf{A}),\ \text{under the assumption of Claim 1;}$$

$$\lambda_{k+q-p}(\mathbf{B})\sigma_1^2(\mathbf{P})\wedge\lambda_{k+q-p}(\mathbf{B})\sigma_p^2(\mathbf{P})\leqslant\lambda_k(\mathbf{P}^\top\mathbf{B}\mathbf{P})\overset{(a)}{\leqslant}\lambda_k(\mathbf{A}),\ \text{under the assumption of Claim 2.}$$

Here in (a) we use the fact for any two p_1-by-p_1 symmetric matrices $\mathbf{W}_1,\mathbf{W}_2$, $\mathbf{W}_1\succcurlyeq\mathbf{W}_2$ implies $\lambda_k(\mathbf{W}_1)\geqslant\lambda_k(\mathbf{W}_2)$ for any $k\in[p_1]$. This finishes the proof for the first two claims.

To prove the third claim, suppose $\mathbf{v}_{\min}$ is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{B}$, then

$$\lambda_{\min}(\mathbf{B})=\mathbf{v}_{\min}^\top\mathbf{B}\mathbf{v}_{\min}\geqslant\mathbf{v}_{\min}^\top\mathbf{Q}^\top\mathbf{A}\mathbf{Q}\mathbf{v}_{\min}\geqslant\lambda_{\min}(\mathbf{A})\|\mathbf{Q}\mathbf{v}_{\min}\|_2^2\geqslant\begin{cases}0, & \text{if } \lambda_{\min}(\mathbf{A})\geqslant 0;\\ \sigma_1^2(\mathbf{Q})\lambda_{\min}(\mathbf{A}), & \text{if } \lambda_{\min}(\mathbf{A})<0.\end{cases}$$

To prove the last claim, suppose $\mathbf{v}_{\max}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{B}$, then

$$\lambda_1(\mathbf{B})=\mathbf{v}_{\max}^\top\mathbf{B}\mathbf{v}_{\max}\leqslant\mathbf{v}_{\max}^\top\mathbf{Q}^\top\mathbf{A}\mathbf{Q}\mathbf{v}_{\max}\leqslant\lambda_{\max}(\mathbf{A})\|\mathbf{Q}\mathbf{v}_{\max}\|_2^2\leqslant\begin{cases}0, & \text{if } \lambda_{\max}(\mathbf{A})<0;\\ \sigma_1^2(\mathbf{Q})\lambda_{\max}(\mathbf{A}), & \text{if } \lambda_{\max}(\mathbf{A})\geqslant 0.\end{cases}$$

This finishes the proof of this lemma. ■

Lemma 4 (Max-min Theorem for Eigenvalues (Bhatia 2013, Corollary III.1.2)) *For any p-by-p real symmetric matrix $\mathbf{A}$ with eigenvalues $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$. If $\mathscr{C}_k$ denotes the set of subspaces of $\mathbb{R}^p$ of dimension k, then $\lambda_k=\max_{C\in\mathscr{C}_k}\min_{\mathbf{u}\in C,\mathbf{u}\neq\mathbf{0}}\mathbf{u}^\top\mathbf{A}\mathbf{u}/\|\mathbf{u}\|_2^2$.*

Lemma 5 *Suppose $\mathbf{L}\in\mathbb{R}^{p_1\times r}$ and $\mathbf{R}\in\mathbb{R}^{p_2\times r}$. Then for any $[\mathbf{A}_L^\top\quad\mathbf{A}_R^\top]^\top\in\mathbb{R}^{(p_1+p_2)\times r}$,*

$$\|\mathbf{L}^\top\mathbf{A}_L+\mathbf{A}_L^\top\mathbf{L}-\mathbf{R}^\top\mathbf{A}_R-\mathbf{A}_R^\top\mathbf{L}\|_{\mathrm{F}}^2\leqslant 8(\sigma_1(\mathbf{L})\vee\sigma_1(\mathbf{R}))^2(\|\mathbf{A}_R\|_{\mathrm{F}}^2+\|\mathbf{A}_L\|_{\mathrm{F}}^2).$$

Proof.

$$\begin{aligned}\|\mathbf{L}^\top\mathbf{A}_L+\mathbf{A}_L^\top\mathbf{L}-\mathbf{R}^\top\mathbf{A}_R-\mathbf{A}_R^\top\mathbf{L}\|_{\mathrm{F}}^2&\leqslant 2(\|\mathbf{L}^\top\mathbf{A}_L+\mathbf{A}_L^\top\mathbf{L}\|_{\mathrm{F}}^2+\|\mathbf{R}^\top\mathbf{A}_R+\mathbf{A}_R^\top\mathbf{L}\|_{\mathrm{F}}^2)\\&\leqslant 2(4\|\mathbf{L}^\top\mathbf{A}_L\|_{\mathrm{F}}^2+4\|\mathbf{R}^\top\mathbf{A}_R\|_{\mathrm{F}}^2)\\&\leqslant 8(\sigma_1(\mathbf{L})\vee\sigma_1(\mathbf{R}))^2(\|\mathbf{A}_R\|_{\mathrm{F}}^2+\|\mathbf{A}_L\|_{\mathrm{F}}^2).\end{aligned}$$

This finishes the proof. ■

Lemma 6 *Suppose $\mathbf{L}\in\mathbb{R}^{p_1\times r},\mathbf{R}\in\mathbb{R}^{p_2\times r}$ are two rank r matrices and $\mathbf{L}^\top\mathbf{L}=\mathbf{R}^\top\mathbf{R}$. Let $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ be a SVD of $\mathbf{L}\mathbf{R}^\top$. Then we have $\mathbf{L}=\mathbf{U}\mathbf{P},\mathbf{R}=\mathbf{V}\mathbf{P}$ for some r-by-r full rank matrix $\mathbf{P}$ satisfying $\mathbf{P}\mathbf{P}^\top=\boldsymbol{\Sigma}$.*

Proof. First since $\mathbf{L}\mathbf{R}^\top$ has SVD $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$, we have $\mathbf{L}=\mathbf{U}\mathbf{P}_1$ and $\mathbf{R}=\mathbf{V}\mathbf{P}_2$. Next we show $\mathbf{P}_1=\mathbf{P}_2$. Since $\mathbf{P}_1\mathbf{P}_2^\top=\boldsymbol{\Sigma}$, we have

$$\boldsymbol{\Sigma}^2=\mathbf{P}_1\mathbf{P}_2^\top\mathbf{P}_2\mathbf{P}_1^\top\overset{(a)}{=}\mathbf{P}_1\mathbf{P}_1^\top\mathbf{P}_1\mathbf{P}_1^\top\overset{(b)}{\Longrightarrow}\boldsymbol{\Sigma}=\mathbf{P}_1\mathbf{P}_1^\top.$$

Here (a) is because $\mathbf{L}^\top\mathbf{L}=\mathbf{R}^\top\mathbf{R}$ implies $\mathbf{P}_1^\top\mathbf{P}_1=\mathbf{P}_2^\top\mathbf{P}_2$; and (b) is because a PSD matrix has a unique principal square root (Johnson et al. 2001). This finishes the proof of this lemma. ■

References

Absil PA, Mahony R, Sepulchre R (2009) *Optimization algorithms on matrix manifolds* (Princeton University Press).

Absil PA, Malick J (2012) Projection-like retractions on matrix manifolds. *SIAM Journal on Optimization* 22(1):135–158.

Bhatia R (2013) *Matrix analysis*, volume 169 (Springer Science & Business Media).

Ha W, Liu H, Barber RF (2020) An equivalence between critical points for rank constraints versus low-rank factorizations. *SIAM Journal on Optimization* 30(4):2927–2955.

Johnson CR, Okubo K, Reams R (2001) Uniqueness of matrix square roots and an application. *Linear Algebra and its applications* 323(1-3):51–60.

Li Z, Cai JF, Wei K (2019) Toward the optimal construction of a loss function without spurious local minima for solving quadratic equations. *IEEE Transactions on Information Theory* 66(5):3242–3260.

Shalit U, Weinshall D, Chechik G (2012) Online learning in the embedded manifold of low-rank matrices. *Journal of Machine Learning Research* 13(2).

Vandereycken B (2013) Low-rank matrix completion by Riemannian optimization. *SIAM Journal on Optimization* 23(2):1214–1236.

Vandereycken B, Vandewalle S (2010) A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. *SIAM Journal on Matrix Analysis and Applications* 31(5):2553–2579.

Zhu Z, Li Q, Tang G, Wakin MB (2018) Global optimality in low-rank matrix optimization. *IEEE Transactions on Signal Processing* 66(13):3614–3628.