SUPPLEMENTARY MATERIAL FOR 'Biconvex Landscape in SDP-based Learning'

A APPENDIX: PROOFS

A.1 Proof of Proposition 1

Proof. Let $\alpha_1, \alpha_2 \geq 0$, $\alpha_1 + \alpha_2 = 1$ and $\forall X, Y_1, Y_2 \in \mathbb{R}^{n\times r}$, according to the definition of f_s in (4) we have

1. if $\tilde{f}(XY^\top) = \tilde{f}(YX^\top)$, then

$$\begin{aligned} f_s(X, \alpha_1 Y_1 + \alpha_2 Y_2) &= \tilde{f}(X(\alpha_1 Y_1 + \alpha_2 Y_2)^\top) \\ &= \tilde{f}(\alpha_1 XY_1^\top + \alpha_2 XY_2^\top) \\ &\leq \alpha_1 \tilde{f}(XY_1^\top) + \alpha_2 \tilde{f}(XY_2^\top) \\ &= \alpha_1 f_s(X, Y_1) + \alpha_2 f_s(X, Y_2), \end{aligned}$$

   where the first and last equalities use the definition of $f_s(\cdot,\cdot)$ in (4), and the inequality uses convexity of $\tilde{f}(\cdot)$.

2. otherwise

$$\begin{aligned} f_s(X, \alpha_1 Y_1 + \alpha_2 Y_2) &= \frac{\tilde{f}(X(\alpha_1 Y_1 + \alpha_2 Y_2)^\top) + \tilde{f}((\alpha_1 Y_1 + \alpha_2 Y_2)X^\top)}{2} \\ &= \frac{\tilde{f}(\alpha_1 XY_1^\top + \alpha_2 XY_2^\top) + \tilde{f}(\alpha_1 Y_1 X^\top + \alpha_2 Y_2 X^\top)}{2} \\ &\leq \frac{\alpha_1 \tilde{f}(XY_1^\top) + \alpha_2 \tilde{f}(XY_2^\top) + \alpha_1 \tilde{f}(Y_1 X^\top) + \alpha_2 \tilde{f}(Y_2 X^\top)}{2} \\ &= \alpha_1 \frac{\tilde{f}(XY_1^\top) + \tilde{f}(Y_1 X^\top)}{2} + \alpha_2 \frac{\tilde{f}(XY_2^\top) + \tilde{f}(Y_2 X^\top)}{2} \\ &= \alpha_1 f_s(X, Y_1) + \alpha_2 f_s(X, Y_2), \end{aligned}$$

   where the first and last equalities use the definition of $f_s(\cdot,\cdot)$ in (4), and the inequality uses convexity of $\tilde{f}(\cdot)$.

Combining 1 and 2, we conclude that $f_s(X,\cdot)$ is convex. Similarly we can conclude that $f_s(\cdot, Y)$ is convex. Hence, $f_s(\cdot,\cdot)$ is biconvex.

As $\frac{\gamma}{2}\|X - Y\|^2$ in (3) is biconvex, $F(\cdot,\cdot;\gamma)$ is biconvex. □

A.2 Proof of Theorem 1

Proof. It is easy to know [Nocedal and Wright, 2006] that Assumption 1 is equivalent to the following inequalities

$$f_s(X_2,\cdot) - f_s(X_1,\cdot) - \text{tr}[(X_2 - X_1)^\top \nabla_X f_s(X_1,\cdot)] \leq \frac{L^X}{2}\|X_2 - X_1\|^2, \tag{18}$$

$$f_s(\cdot, Y_2) - f_s(\cdot, Y_1) - \text{tr}[(Y_2 - Y_1)^\top \nabla_Y f_s(\cdot, Y_1)] \leq \frac{L^Y}{2}\|Y_2 - Y_1\|^2, \tag{19}$$

for $\forall X_1, X_2, Y_1, Y_2 \in \mathbb{R}^{n\times r}$.

From the symmetry of f_s, we have $f_s(X,Y) = f_s(Y,X)$, which implies

$$\nabla_X f_s(X,Y) = \nabla_X f_s(Y,X), \tag{20}$$

$$\nabla_Y f_s(X,Y) = \nabla_Y f_s(Y,X). \tag{21}$$

First, we introduce some Lemmas.

Lemma 1. *If $(\bar{X},\bar{Y})$ is a stationary point of* (3)*, then $\nabla_X f_s(\bar{X},\bar{Y}) = -\nabla_Y f_s(\bar{X},\bar{Y})$.*

Proof. Since $(\bar{X},\bar{Y})$ is a stationary point of (3), $\nabla_X F(\bar{X},\bar{Y};\gamma) = 0, \nabla_Y F(\bar{X},\bar{Y};\gamma) = 0$, and so

$$\nabla_X f_s(\bar{X},\bar{Y}) = \gamma(\bar{Y} - \bar{X}), \tag{22}$$

$$\nabla_Y f_s(\bar{X},\bar{Y}) = \gamma(\bar{X} - \bar{Y}). \tag{23}$$

Hence, $\nabla_X f_s(\bar{X},\bar{Y}) = -\nabla_Y f_s(\bar{X},\bar{Y})$, and result follows. □

Lemma 2. *Assume that $(\bar{X},\bar{Y})$ is a stationary point of* (3) *and $f_s(\bar{X},\bar{Y}) = f_s(\bar{Y},\bar{X})$. Then, $\bar{X} = \bar{Y}$ if $\gamma > \frac{1}{4}L^X$.*

Proof. By contradiction, assume $\bar{X} \neq \bar{Y}$ when $\gamma > \frac{1}{4}L^X$. From (23) we have

$$\gamma \left\|\bar{X} - \bar{Y}\right\|^2 = \mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_Y f_s(\bar{X}, \bar{Y}),$$

such that

$$\begin{aligned}
\gamma &= \frac{2\mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_Y f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{X} - \bar{Y}\right\|^2} \\
&= \frac{\mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_Y f_s(\bar{X}, \bar{Y})] + \mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_Y f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{X} - \bar{Y}\right\|^2} \\
&= \frac{\mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_Y f_s(\bar{X}, \bar{Y})] - \mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_X f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{X} - \bar{Y}\right\|^2} \\
&\leq \frac{f_s(\bar{X}, \bar{X}) - f_s(\bar{X}, \bar{Y}) - \mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_X f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{X} - \bar{Y}\right\|^2} \\
&= \frac{f_s(\bar{X}, \bar{X}) - f_s(\bar{Y}, \bar{X}) - \mathrm{tr}[(\bar{X} - \bar{Y})^\top \nabla_X f_s(\bar{Y}, \bar{X})]}{2\left\|\bar{X} - \bar{Y}\right\|^2} \\
&\leq \frac{\frac{1}{2}L^X \left\|\bar{X} - \bar{Y}\right\|^2}{2\left\|\bar{X} - \bar{Y}\right\|^2} \\
&= \frac{1}{4}L^X. \qquad (24)
\end{aligned}$$

The third equality uses Lemma 1, the first inequality is obtained from the first-order condition of the convexity of $f_s(\bar{X}, \cdot)$, the fourth equality uses (20), the last inequality is obtained from (18).

From (24) we have $\gamma \leq \frac{1}{4}L^X$, which contradicts with the starting assumption $\gamma > \frac{1}{4}L^X$. □

Lemma 3. *Assume that* $(\bar{X}, \bar{Y})$ *is a stationary point of* (3) *and* $f_s(\bar{X}, \bar{Y}) = f_s(\bar{Y}, \bar{X})$. *Then,* $\bar{X} = \bar{Y}$ *if* $\gamma > \frac{1}{4}L^Y$.

Proof. By contradiction, assume $\bar{X} \neq \bar{Y}$ when $\gamma > \frac{1}{4}L^Y$. From (22) we have

$$\gamma \left\|\bar{Y} - \bar{X}\right\|^2 = \mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_X f_s(\bar{X}, \bar{Y}),$$

such that

$$\begin{aligned}
\gamma &= \frac{2\mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_X f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{Y} - \bar{X}\right\|^2} \\
&= \frac{\mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_X f_s(\bar{X}, \bar{Y})] + \mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_X f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{Y} - \bar{X}\right\|^2} \\
&= \frac{\mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_X f_s(\bar{X}, \bar{Y})] - \mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_Y f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{Y} - \bar{X}\right\|^2} \\
&\leq \frac{f_s(\bar{Y}, \bar{Y}) - f_s(\bar{X}, \bar{Y}) - \mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_Y f_s(\bar{X}, \bar{Y})]}{2\left\|\bar{Y} - \bar{X}\right\|^2} \\
&= \frac{f_s(\bar{Y}, \bar{Y}) - f_s(\bar{Y}, \bar{X}) - \mathrm{tr}[(\bar{Y} - \bar{X})^\top \nabla_Y f_s(\bar{Y}, \bar{X})]}{2\left\|\bar{Y} - \bar{X}\right\|^2} \\
&\leq \frac{\frac{1}{2}L^Y \left\|\bar{Y} - \bar{X}\right\|^2}{2\left\|\bar{Y} - \bar{X}\right\|^2} \\
&= \frac{1}{4}L^Y. \qquad (25)
\end{aligned}$$

The third equality uses Lemma 1, the first inequality is obtained from the first-order condition of the convexity of $f_s(\cdot, \bar{Y})$, the fourth equality uses (21), the last inequality is obtained from (19).

From (25) we have $\gamma \leq \frac{1}{4}L^Y$, which contradicts with the starting assumption $\gamma > \frac{1}{4}L^Y$. □

Theorem 1 follows on combining Lemmas 2 and 3. □

A.3 Proof of Proposition 2

Proof. Specified to (10) and (11), we compute τ_k^X and τ_k^Y in Algorithm 1 respectively.

1. Computing the optimal stepsize τ_k^X
   Let $\hat{d} = \nabla_X F(\hat{X}_{k-1}, Y_{k-1}; \gamma_{\frac{k}{2}})$, we have

$$\begin{aligned} X_k &= \hat{X}_{k-1} - \tau_k^X \nabla_X F(\hat{X}_{k-1}, Y_{k-1}; \gamma_{\frac{k}{2}}) \\ &= \hat{X}_{k-1} - \tau_k^X \hat{d} \end{aligned}$$

   in step 6 of Algorithm 1.
   So, the optimal τ_k^X is the solution of following problem

$$\begin{aligned} \tau_k^X &= \arg\min_{\tau \geq 0} : F(\hat{X}_{k-1} - \tau \hat{d}, Y_{k-1}; \gamma_{\frac{k}{2}}) \\ &= \arg\min_{\tau \geq 0} : \frac{1}{2} \mathrm{vec}(\hat{X}_{k-1} - \tau \hat{d})^\top \mathcal{A}_k^Y \mathrm{vec}(\hat{X}_{k-1} - \tau \hat{d}) - \mathrm{vec}(\hat{X}_{k-1} - \tau \hat{d})^\top b_k^Y + constant \\ &= \arg\min_{\tau \geq 0} : \frac{1}{2} [\mathrm{vec}(\hat{d})^\top \mathcal{A}_k^Y \mathrm{vec}(\hat{d})] \tau^2 - [\mathrm{vec}(\hat{d})^\top (\mathcal{A}_k^Y \mathrm{vec}(\hat{X}_{k-1}) - b_k^Y)] \tau + constant \\ &= \frac{\mathrm{vec}(\hat{d})^\top (\mathcal{A}_k^Y \mathrm{vec}(\hat{X}_{k-1}) - b_k^Y)}{\mathrm{vec}(\hat{d})^\top \mathcal{A}_k^Y \mathrm{vec}(\hat{d})} \\ &= \frac{\mathrm{vec}(\hat{d})^\top \mathrm{vec}(\nabla_X F(\hat{X}_{k-1}, Y_{k-1}; \gamma_{\frac{k}{2}}))}{\mathrm{vec}(\hat{d})^\top \mathcal{A}_k^Y \mathrm{vec}(\hat{d})} \\ &= \frac{\mathrm{vec}(\hat{d})^\top \mathrm{vec}(\hat{d})}{\mathrm{vec}(\hat{d})^\top \mathcal{A}_k^Y \mathrm{vec}(\hat{d})} \\ &= \frac{\left\|\hat{d}\right\|^2}{\mathrm{vec}(\hat{d})^\top \mathcal{A}_k^Y \mathrm{vec}(\hat{d})} \\ &= \frac{\|d_k^X\|^2}{{d_k^X}^\top \mathcal{A}_k^Y d_k^X} \end{aligned}$$

   where the last fourth equality uses $\mathcal{A}_k^Y \mathrm{vec}(\hat{X}_{k-1}) - b_k^Y = \mathrm{vec}(\nabla_X F(\hat{X}_{k-1}, Y_{k-1}; \gamma_{\frac{k}{2}}))$, which can be obtained from (10), the last third equality uses $\hat{d} = \nabla_X F(\hat{X}_{k-1}, Y_{k-1}; \gamma_{\frac{k}{2}})$, the last first equality uses $\mathrm{vec}(\hat{d}) = \mathrm{vec}(\nabla_X F(\hat{X}_{k-1}, Y_{k-1}; \gamma_{\frac{k}{2}})) = d_k^X$ defined in (16).

2. Computing the optimal stepsize τ_k^Y is similar to τ_k^X.

$\square$