On the Absence of Spurious Local Minima in Nonlinear Low-Rank Matrix Recovery Problems: Supplementary Material

Yingjie Bi **Javad Lavaei**
Industrial Engineering and Operations Research, University of California, Berkeley

This supplementary material presents the proofs of Lemma 4 and Lemma 7 in the main paper.

1 PROOF OF LEMMA 4 IN THE PAPER

Without loss of generality, assume that $\|K\|_F = \|L\|_F = 1$. By the δ-RIP$_{2r}$ property of $\mathcal{Q}$, we have

$$(1-\delta)\|K-L\|_F^2 \le [\mathcal{Q}](K-L, K-L) \le (1+\delta)\|K-L\|_F^2,$$
$$(1-\delta)\|K+L\|_F^2 \le [\mathcal{Q}](K+L, K+L) \le (1+\delta)\|K+L\|_F^2.$$

Taking the difference between the above two inequalities, one can obtain

$$4[\mathcal{Q}](K, L) \le (1+\delta)\|K+L\|_F^2 - (1-\delta)\|K-L\|_F^2 = 4\delta + 4\langle K, L\rangle,$$
$$-4[\mathcal{Q}](K, L) \le (1+\delta)\|K-L\|_F^2 - (1-\delta)\|K+L\|_F^2 = 4\delta - 4\langle K, L\rangle,$$

which proves the desired inequality.

2 PROOF OF LEMMA 7 IN THE PAPER

Let $\mathrm{OPT}(X, Z)$ denote the optimal value of the optimization problem

$$\begin{aligned} \min_{\delta,\mathbf{H}} \quad & \delta \\ \text{s.t.} \quad & \|\mathbf{X}^T\mathbf{H}\mathbf{e}\| \le a, \\ & 2I_r \otimes \mathrm{mat}_S(\mathbf{H}\mathbf{e}) + \mathbf{X}^T\mathbf{H}\mathbf{X} \succeq -bI_{nr}, \\ & \mathbf{H} \text{ is symmetric and satisfies } \delta\text{-RIP}_{2r}, \end{aligned} \tag{1}$$

and $\mathrm{LMI}(X, Z)$ denote the optimal value of the optimization problem

$$\begin{aligned} \min_{\delta,\mathbf{H}} \quad & \delta \\ \text{s.t.} \quad & \begin{bmatrix} I_{nr} & \mathbf{X}^T\mathbf{H}\mathbf{e} \\ (\mathbf{X}^T\mathbf{H}\mathbf{e})^T & a^2 \end{bmatrix} \succeq 0, \\ & 2I_r \otimes \mathrm{mat}_S(\mathbf{H}\mathbf{e}) + \mathbf{X}^T\mathbf{H}\mathbf{X} \succeq -bI_{nr}, \\ & (1-\delta)I_{n^2} \preceq \mathbf{H} \preceq (1+\delta)I_{n^2}. \end{aligned} \tag{2}$$

As mentioned in the paper, the first constraint in (1) and the first constraint in (2) are interchangeable. Our goal is to prove that $\mathrm{OPT}(X, Z) = \mathrm{LMI}(X, Z)$ for given $X, Z \in \mathbb{R}^{n\times r}$. Let $(v_1, \dots, v_n)$ be an orthogonal basis of $\mathbb{R}^n$ such that $(v_1, \dots, v_d)$ spans the column spaces of both X and Z. Note that $d \le 2r$. Let $P \in \mathbb{R}^{n\times d}$ be the matrix with the columns $(v_1, \dots, v_d)$ and $P_\perp \in \mathbb{R}^{n\times(n-d)}$ be the matrix with the columns $(v_{d+1}, \dots, v_n)$. Then,

$$P^TP = I_d, \quad P_\perp^TP_\perp = I_{n-d}, \quad P_\perp^TP = 0, \quad P^TP_\perp = 0,$$
$$PP^T + P_\perp P_\perp^T = I_n, \quad PP^TX = X, \quad PP^TZ = Z.$$

Define $\mathbf{P} = P \otimes P$. Consider the auxiliary optimization problem

$$\begin{aligned} \min_{\delta,\mathbf{H}} \quad & \delta \\ \text{s. t.} \quad & \begin{bmatrix} I_{nr} & \mathbf{X}^T\mathbf{H}\mathbf{e} \\ (\mathbf{X}^T\mathbf{H}\mathbf{e})^T & a^2 \end{bmatrix} \succeq 0, \\ & 2I_r \otimes \operatorname{mat}_S(\mathbf{H}\mathbf{e}) + \mathbf{X}^T\mathbf{H}\mathbf{X} \succeq -bI_{nr}, \\ & (1-\delta)I_{d^2} \preceq \mathbf{P}^T\mathbf{H}\mathbf{P} \preceq (1+\delta)I_{d^2}, \end{aligned} \tag{3}$$

and denote its optimal value as the function $\overline{\text{LMI}}(X,Z)$. Given an arbitrary symmetric matrix $\mathbf{H} \in \mathbb{R}^{n^2\times n^2}$, if $\mathbf{H}$ satisfies the last constraint in (2), then it obviously satisfies δ-RIP$_{2r}$ and subsequently the last constraint in (1). On the other hand, if H satisfies the last constraint in (1), for every matrix $Y \in \mathbb{R}^{d\times d}$ with $\mathbf{Y} = \operatorname{vec} Y$, since $\operatorname{rank}(PYP^T) \le d \le 2r$ and $\operatorname{vec}(PYP^T) = \mathbf{P}\mathbf{Y}$, by δ-RIP$_{2r}$ property, one arrives at

$$(1-\delta)\|\mathbf{Y}\|^2 = (1-\delta)\|\mathbf{P}\mathbf{Y}\|^2 \le (\mathbf{P}\mathbf{Y})^T\mathbf{H}\mathbf{P}\mathbf{Y} \le (1+\delta)\|\mathbf{P}\mathbf{Y}\|^2 = (1+\delta)\|\mathbf{Y}\|^2,$$

which implies that $\mathbf{H}$ satisfies the last constraint in (3). The above discussion implies that

$$\text{LMI}(X,Z) \ge \text{OPT}(X,Z) \ge \overline{\text{LMI}}(X,Z).$$

Let

$$\hat{X} = P^TX, \quad \hat{Z} = P^TZ.$$

Lemma 2 and Lemma 3 to be stated later will show that

$$\text{LMI}(X,Z) \le \text{LMI}(\hat{X},\hat{Z}) \le \overline{\text{LMI}}(X,Z),$$

which completes the proof.

Before stating Lemma 2 and Lemma 3 that were needed in the above proof, we should first state a preliminary result below.

Lemma 1. *Define $\hat{\mathbf{e}}$ and $\hat{\mathbf{X}}$ in the same way as $\mathbf{e}$ and $\mathbf{X}$, except that X and Z are replaced by $\hat{X}$ and $\hat{Z}$, respectively. Then, it holds that*

$$\begin{aligned} \mathbf{e} &= \mathbf{P}\hat{\mathbf{e}}, \\ \mathbf{X}(I_r \otimes P) &= \mathbf{P}\hat{\mathbf{X}}, \\ \mathbf{P}^T\mathbf{X} &= \hat{\mathbf{X}}(I_r \otimes P)^T. \end{aligned}$$

Proof. Observe that

$$\begin{aligned} \mathbf{e} = \operatorname{vec}(XX^T - ZZ^T) &= \operatorname{vec}(P(\hat{X}\hat{X}^T - \hat{Z}\hat{Z}^T)P^T) = \mathbf{P}\hat{\mathbf{e}}, \\ \mathbf{X}(I_r \otimes P)\operatorname{vec}\hat{U} &= \mathbf{X}\operatorname{vec}(P\hat{U}) = \operatorname{vec}(X\hat{U}^TP^T + P\hat{U}X^T) \\ &= \operatorname{vec}(P(\hat{X}\hat{U}^T + \hat{U}\hat{X}^T)P^T) = \mathbf{P}\hat{\mathbf{X}}\operatorname{vec}\hat{U}, \\ \hat{\mathbf{X}}(I_r \otimes P)^T\operatorname{vec}U &= \hat{\mathbf{X}}\operatorname{vec}(P^TU) = \operatorname{vec}(\hat{X}U^TP + P^TU\hat{X}^T) \\ &= \operatorname{vec}(P^T(XU^T + UX^T)P) = \mathbf{P}^T\mathbf{X}\operatorname{vec}U, \end{aligned}$$

where $U \in \mathbb{R}^{n\times r}$ and $\hat{U} \in \mathbb{R}^{d\times r}$ are arbitrary matrices. □

Lemma 2. *The inequality* $\text{LMI}(\hat{X},\hat{Z}) \ge \text{LMI}(X,Z)$ *holds.*

Proof. Let $(\delta,\hat{\mathbf{H}})$ be an arbitrary feasible solution to the optimization problem defining $\text{LMI}(\hat{X},\hat{Z})$ with $\delta \le 1$. It is desirable to show that $(\delta,\mathbf{H})$ with

$$\mathbf{H} = \mathbf{P}\hat{\mathbf{H}}\mathbf{P}^T + (I_{n^2} - \mathbf{P}\mathbf{P}^T)$$

is a feasible solution to the optimization problem defining $\text{LMI}(X, Z)$, which directly proves the lemma. To this end, notice that

$$\mathbf{H} - (1-\delta)I_{n^2} = \mathbf{P}(\hat{\mathbf{H}} - (1-\delta)I_{d^2})\mathbf{P}^T + \delta(I_{n^2} - \mathbf{P}\mathbf{P}^T),$$

which is positive semidefinite because

$$\begin{aligned} I_{n^2} - \mathbf{P}\mathbf{P}^T &= (PP^T + P_\perp P_\perp^T) \otimes (PP^T + P_\perp P_\perp^T) - (PP^T) \otimes (PP^T) \\ &= (PP^T) \otimes (P_\perp P_\perp^T) + (P_\perp P_\perp^T) \otimes (PP^T) + (P_\perp P_\perp^T) \otimes (P_\perp P_\perp^T) \succeq 0. \end{aligned}$$

Similarly,

$$\mathbf{H} - (1+\delta)I_{n^2} \preceq 0,$$

and therefore the last constraint in (2) is satisfied and $\mathbf{H}$ is always positive semidefinite. Next, since

$$\mathbf{X}^T\mathbf{H}\mathbf{e} = \mathbf{X}^T\mathbf{H}\mathbf{P}\hat{\mathbf{e}} = \mathbf{X}^T\mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{e}} = (I_r \otimes P)\hat{\mathbf{X}}^T\hat{\mathbf{H}}\hat{\mathbf{e}},$$

we have

$$\|\mathbf{X}^T\mathbf{H}\mathbf{e}\|^2 = (\hat{\mathbf{X}}^T\hat{\mathbf{H}}\hat{\mathbf{e}})^T(I_r \otimes P^T)(I_r \otimes P)(\hat{\mathbf{X}}^T\hat{\mathbf{H}}\hat{\mathbf{e}}) = \|\hat{\mathbf{X}}^T\hat{\mathbf{H}}\hat{\mathbf{e}}\|^2,$$

and thus the first constraint in (2) is satisfied. Finally, by letting $W \in \mathbb{R}^{d\times d}$ be the vector satisfying $\operatorname{vec} W = \hat{\mathbf{H}}\hat{\mathbf{e}}$, one can write

$$\operatorname{vec}(PWP^T) = \mathbf{P}\operatorname{vec} W = \mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{e}}.$$

Hence,

$$\begin{aligned} 2I_r \otimes \operatorname{mat}_S(\mathbf{H}\mathbf{e}) &= 2I_r \otimes \operatorname{mat}_S(\mathbf{H}\mathbf{P}\hat{\mathbf{e}}) = 2I_r \otimes \operatorname{mat}_S(\mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{e}}) = I_r \otimes (P(W + W^T)P^T) \\ &= 2I_r \otimes (P\operatorname{mat}_S(\hat{\mathbf{H}}\hat{\mathbf{e}})P^T) = 2(I_r \otimes P)(I_r \otimes \operatorname{mat}_S(\hat{\mathbf{H}}\hat{\mathbf{e}}))(I_r \otimes P)^T. \end{aligned}$$

In addition,

$$\mathbf{X}^T\mathbf{H}\mathbf{X}(I_r \otimes P) = \mathbf{X}^T\mathbf{H}\mathbf{P}\hat{\mathbf{X}} = \mathbf{X}^T\mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{X}} = (I_r \otimes P)\hat{\mathbf{X}}^T\hat{\mathbf{H}}\hat{\mathbf{X}}.$$

Therefore, by defining

$$\mathbf{S} := 2I_r \otimes \operatorname{mat}_S(\mathbf{H}\mathbf{e}) + \mathbf{X}^T\mathbf{H}\mathbf{X} + bI_{nr},$$

we have

$$\begin{aligned} (I_r \otimes P)^T\mathbf{S}(I_r \otimes P) &= 2I_r \otimes \operatorname{mat}_S(\hat{\mathbf{H}}\hat{\mathbf{e}}) + \hat{\mathbf{X}}^T\hat{\mathbf{H}}\hat{\mathbf{X}} + bI_{dr} \succeq 0, \\ (I_r \otimes P_\perp)^T\mathbf{S}(I_r \otimes P_\perp) &= (I_r \otimes P_\perp)^T\mathbf{X}^T\mathbf{H}\mathbf{X}(I_r \otimes P_\perp) + bI_{(n-d)r} \succeq 0, \\ (I_r \otimes P_\perp)^T\mathbf{S}(I_r \otimes P) &= 0. \end{aligned}$$

Since the columns of $I_r \otimes P$ and $I_r \otimes P_\perp$ form a basis for $\mathbb{R}^{nr}$, the above inequalities imply that $\mathbf{S}$ is positive semidefinite, and thus the second constraint in (2) is satisfied. □

Lemma 3. *The inequality* $\overline{\text{LMI}}(X, Z) \geq \text{LMI}(\hat{X}, \hat{Z})$ *holds.*

Proof. The dual problem of the optimization problem defining $\text{LMI}(\hat{X}, \hat{Z})$ can be expressed as

$$\begin{aligned} \max_{\hat{U}_1, \hat{U}_2, \hat{V}, \hat{G}, \hat{\lambda}, \hat{y}} \quad & \operatorname{tr}(\hat{U}_1 - \hat{U}_2) - \operatorname{tr}(\hat{G}) - a^2\hat{\lambda} - b\operatorname{tr}(\hat{V}) \\ \text{s.t.} \quad & \operatorname{tr}(\hat{U}_1 + \hat{U}_2) = 1, \\ & \sum_{j=1}^{r}(\hat{\mathbf{X}}\hat{y} - \operatorname{vec}\hat{V}_{j,j})\hat{\mathbf{e}}^T + \sum_{j=1}^{r}\hat{\mathbf{e}}(\hat{\mathbf{X}}\hat{y} - \operatorname{vec}\hat{V}_{j,j})^T - \hat{\mathbf{X}}\hat{V}\hat{\mathbf{X}}^T = \hat{U}_1 - \hat{U}_2, \\ & \begin{bmatrix} \hat{G} & -\hat{y} \\ -\hat{y}^T & \hat{\lambda} \end{bmatrix} \succeq 0, \\ & \hat{U}_1 \succeq 0, \quad \hat{U}_2 \succeq 0, \quad \hat{V} = \begin{bmatrix} \hat{V}_{1,1} & \cdots & \hat{V}_{r,1} \\ \vdots & \ddots & \vdots \\ \hat{V}_{r,1}^T & \cdots & \hat{V}_{r,r} \end{bmatrix} \succeq 0. \end{aligned} \tag{4}$$

Since

$$\hat{U}_1 = \frac{1}{2d^2} I_{d^2} - \frac{\mu}{2} M, \quad \hat{U}_2 = \frac{1}{2d^2} I_{d^2} + \frac{\mu}{2} M, \quad \hat{V} = \mu I_{dr}, \quad \hat{G} = I_{dr}, \quad \hat{\lambda} = 1, \quad \hat{y} = 0,$$

where

$$M = r((\operatorname{vec} I_d)\hat{\mathbf{e}}^T + \hat{\mathbf{e}}(\operatorname{vec} I_d)^T) + \hat{\mathbf{X}}\hat{\mathbf{X}}^T,$$

is a strict feasible solution to the above dual problem (4) as long as $\mu > 0$ is sufficiently small, Slater's condition implies that strong duality holds for the optimization problem defining $\operatorname{LMI}(\hat{X}, \hat{Z})$. Therefore, we only need to prove that the optimal value of (4) is smaller than or equal to the optimal value of the dual of the optimization problem defining $\overline{\operatorname{LMI}}(X, Z)$ given by:

$$\begin{aligned} \max_{U_1,U_2,V,G,\lambda,y} \quad & \operatorname{tr}(U_1 - U_2) - \operatorname{tr}(G) - a^2\lambda - b\operatorname{tr}(V) \\ \text{s.t.} \quad & \operatorname{tr}(U_1 + U_2) = 1, \\ & \sum_{j=1}^{r} (\mathbf{X}y - \operatorname{vec} V_{j,j})\mathbf{e}^T + \sum_{j=1}^{r} \mathbf{e}(\mathbf{X}y - \operatorname{vec} V_{j,j})^T - \mathbf{X}V\mathbf{X}^T = \mathbf{P}(U_1 - U_2)\mathbf{P}^T, \\ & \begin{bmatrix} G & -y \\ -y^T & \lambda \end{bmatrix} \succeq 0, \\ & U_1 \succeq 0, \quad U_2 \succeq 0, \quad V = \begin{bmatrix} V_{1,1} & \cdots & V_{r,1} \\ \vdots & \ddots & \vdots \\ V_{r,1}^T & \cdots & V_{r,r} \end{bmatrix} \succeq 0. \end{aligned} \tag{5}$$

The above claim can be verified by noting that given any feasible solution

$$(\hat{U}_1, \hat{U}_2, \hat{V}, \hat{G}, \hat{\lambda}, \hat{y})$$

to (4), the matrices

$$U_1 = \hat{U}_1, \quad U_2 = \hat{U}_2, \quad V = (I_r \otimes P)\hat{V}(I_r \otimes P)^T,$$
$$\begin{bmatrix} G & -y \\ -y^T & \lambda \end{bmatrix} = \begin{bmatrix} I_r \otimes P & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{G} & -\hat{y} \\ -\hat{y}^T & \hat{\lambda} \end{bmatrix} \begin{bmatrix} (I_r \otimes P)^T & 0 \\ 0 & 1 \end{bmatrix}$$

form a feasible solution to (5), and both solutions have the same optimal value. □