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 \( \delta \)-RIP\(_{2r} \) property of \( Q \), we have
\[
(1 - \delta)\|K - L\|_F^2 \leq \langle Q(K - L, K - L) \rangle \leq (1 + \delta)\|K - L\|_F^2,
\]
\[
(1 - \delta)\|K + L\|_F^2 \leq \langle Q(K + L, K + L) \rangle \leq (1 + \delta)\|K + L\|_F^2.
\]
Taking the difference between the above two inequalities, one can obtain
\[
4\langle Q, (K, L) \rangle \leq (1 + \delta)\|K + L\|_F^2 - (1 - \delta)\|K - L\|_F^2 = 4\delta + 4\langle K, L \rangle,
\]
\[
-4\langle Q, (K, L) \rangle \leq (1 + \delta)\|K - L\|_F^2 - (1 - \delta)\|K + L\|_F^2 = 4\delta - 4\langle K, L \rangle,
\]
which proves the desired inequality.

2 PROOF OF LEMMA 7 IN THE PAPER

Let OPT\((X, Z)\) denote the optimal value of the optimization problem
\[
\begin{align*}
\min_{\delta, H} & \quad \delta \\
\text{subject to} & \quad \|X^T He\| \leq a, \\
& \quad 2I_r \otimes \text{mat}_S(He) + X^T H X \succeq -bI_{nr}, \\
& \quad H \text{ is symmetric and satisfies } \delta \text{-RIP}_{2r},
\end{align*}
\]
and LMI\((X, Z)\) denote the optimal value of the optimization problem
\[
\begin{align*}
\min_{\delta, H} & \quad \delta \\
\text{subject to} & \quad \begin{bmatrix}
I_{nr} & X^T He \\
(X^T He)^T & a^2
\end{bmatrix} \succeq 0, \\
& \quad 2I_r \otimes \text{mat}_S(He) + X^T H X \succeq -bI_{nr}, \\
& \quad (1 - \delta)I_{n^2} \preceq H \preceq (1 + \delta)I_{n^2}.
\end{align*}
\]
As mentioned in the paper, the first constraint in (1) and the first constraint in (2) are interchangeable. Our goal is to prove that OPT\((X, Z) = \text{LMI}(X, Z)\) for given \(X, Z \in \mathbb{R}^{nr} \). Let \((v_1, \ldots, v_n)\) be an orthogonal basis of \(\mathbb{R}^n\) such that \((v_1, \ldots, v_d)\) spans the column spaces of both \(X\) and \(Z\). Note that \(d \leq 2r\). Let \(P \in \mathbb{R}^{n \times d}\) be the matrix with the columns \((v_1, \ldots, v_d)\) and \(P_\perp \in \mathbb{R}^{n \times (n - d)}\) be the matrix with the columns \((v_{d+1}, \ldots, v_n)\). Then,
\[
P^T P = I_d, \quad P_\perp^T P_\perp = I_{n - d}, \quad P_\perp^T P = 0, \quad P^T P_\perp = 0,
\]
\[
PP^T + P_\perp P_\perp^T = I_n, \quad PP^T X = X, \quad PP^T Z = Z.
\]
Define \( P = P \otimes P \). Consider the auxiliary optimization problem

\[
\begin{align*}
\min_{\delta, H} & \quad \delta \\
\text{s.t.} & \quad \begin{bmatrix} I_{nr} & X^T H e \\ (X^T H e)^T & \frac{1}{a^2} \end{bmatrix} \succeq 0,
\end{align*}
\]

(3)

and denote its optimal value as the function \( \text{LMI}(X, Z) \). Given an arbitrary symmetric matrix \( H \in \mathbb{R}^{n \times n^2} \), if \( H \) satisfies the last constraint in (2), then it obviously satisfies \( \delta\text{-RIP}_{2r} \) 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 \( Y = \text{vec} Y \), since \( \text{rank}(PYP^T) \leq d \leq 2r \) and \( \text{vec}(PYP^T) = PY \), by \( \delta\text{-RIP}_{2r} \) property, one arrives at

\[
(1 - \delta)||Y||^2 = (1 - \delta)||PY||^2 \leq (1 + \delta)||PY||^2 \leq (1 + \delta)||Y||^2,
\]

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

\( \text{LMI}(X, Z) \geq \text{OPT}(X, Z) \geq \text{LMI}(X, Z) \).

Let

\( \hat{X} = P^T X, \quad \hat{Z} = P^T Z \).

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

\( \text{LMI}(X, Z) \leq \text{LMI}(\hat{X}, \hat{Z}) \leq \text{LMI}(X, Z) \),

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 \( \hat{e} \) and \( \hat{X} \) in the same way as \( e \) and \( X \), except that \( X \) and \( Z \) are replaced by \( \hat{X} \) and \( \hat{Z} \), respectively. Then, it holds that

\[
\begin{align*}
e & = \text{P} \hat{e}, \\
x(I_r \otimes I) & = \text{P} \hat{X}, \\
P^T X & = \hat{X}(I_r \otimes P)^T.
\end{align*}
\]

**Proof.** Observe that

\[
\begin{align*}
e & = \text{vec}(X X^T - ZZ^T) = \text{vec}(P(\hat{X} \hat{X}^T - \hat{Z} \hat{Z}^T)P^T) = \text{P} \hat{e}, \\
x(I_r \otimes I) \text{vec} \hat{U} & = \text{X vec}(P \hat{U}) = \text{vec}(X \hat{U}^T P^T + P \hat{U} X^T) \\
& = \text{vec}(P(\hat{X} \hat{U}^T + \hat{U} \hat{X}^T)P^T) = \text{P} \hat{X} \text{vec} \hat{U}, \\
\hat{X}(I_r \otimes P)^T \text{vec} U & = \text{X vec}(P^T U) = \text{vec}(\hat{X} U^T P + P^T U \hat{X}^T) \\
& = \text{vec}(P^T(XX^T + U^T X)P) = P^T X \text{vec} U,
\end{align*}
\]

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

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

**Proof.** Let \((\delta, \hat{H})\) be an arbitrary feasible solution to the optimization problem defining \( \text{LMI}(\hat{X}, \hat{Z}) \) with \( \delta \leq 1 \).

It is desirable to show that \((\delta, H)\) with

\[
H = PHP^T + (I_{n^2} - PP^T)
\]
is a feasible solution to the optimization problem defining LMI($X, Z$), which directly proves the lemma. To this end, notice that
\[
H - (1 - \delta)I_{n^2} = P(\hat{H} - (1 - \delta)I_d)P^T + \delta(I_{n^2} - PP^T),
\]
which is positive semidefinite because
\[
I_{n^2} - PP^T = (PP^T + P_\perp P_\perp^T) \otimes (PP^T + P_\perp P_\perp^T) - (PP^T) \otimes (PP^T) = (PP^T) \otimes (P_\perp P_\perp^T) + (P_\perp P_\perp^T) \otimes (PP^T) + (P_\perp P_\perp^T) \otimes (P_\perp P_\perp^T) \succeq 0.
\]
Similarly,
\[
H - (1 + \delta)I_{n^2} \succeq 0,
\]
and therefore the last constraint in (2) is satisfied and $H$ is always positive semidefinite. Next, since
\[
X^THe = X^THPe = X^TPHe = (I_r \otimes P)\hat{X}^T\hat{H}e,
\]
we have
\[
\|X^THe\|^2 = (\hat{X}^T\hat{H}e)^T(I_r \otimes P^T)(I_r \otimes P)(\hat{X}^T\hat{H}e) = \|\hat{X}^T\hat{H}e\|^2,
\]
and thus the first constraint in (2) is satisfied. Finally, by letting $W \in \mathbb{R}^{d \times d}$ be the vector satisfying vec $W = \hat{H}e$, one can write
\[
\text{vec}(PW^TP^T) = P \text{vec} W = PHe.
\]
Hence,
\[
2I_r \otimes \mathcal{M}(\mathcal{H}e) = 2I_r \otimes \mathcal{M}(\mathcal{H}Pe) = 2I_r \otimes \mathcal{M}(\mathcal{P}He) = I_r \otimes (P(W + W^T)P^T)
= 2I_r \otimes (P \mathcal{M}(\hat{H}e)P^T) = 2(I_r \otimes P)(I_r \otimes \mathcal{M}(\hat{H}e)) (I_r \otimes P)^T.
\]
In addition,
\[
X^THX(I_r \otimes P) = X^THPX = X^TPHX = (I_r \otimes P)\hat{X}^T\hat{H}X.
\]
Therefore, by defining
\[
S := 2I_r \otimes \mathcal{M}(\mathcal{H}e) + X^THX + bI_{nr},
\]
we have
\[
(I_r \otimes P)^T S(I_r \otimes P) = 2I_r \otimes \mathcal{M}(\hat{H}e) + \hat{X}^T\hat{H}X + bI_{dr} \succeq 0,
\]
\[
(I_r \otimes P_\perp)^T S(I_r \otimes P_\perp) = (I_r \otimes P_\perp)^T X^THX(I_r \otimes P_\perp) + bI_{(n-d)r} \succeq 0,
\]
\[
(I_r \otimes P_\perp)^T S(I_r \otimes P) = 0.
\]
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 $S$ is positive semidefinite, and thus the second constraint in (2) is satisfied.

\textbf{Lemma 3.} The inequality $\text{LMI}(X, Z) \succeq \text{LMI}(\hat{X}, \hat{Z})$ holds.

\textbf{Proof.} The dual problem of the optimization problem defining LMI($\hat{X}, \hat{Z}$) can be expressed as
\[
\max_{\bar{U}_1, \bar{U}_2, V, \hat{\lambda}, \hat{\gamma}} \quad \text{tr}(\bar{U}_1 - \bar{U}_2) - \text{tr}(\hat{G}) - a^2\hat{\lambda} - b\text{tr}(\hat{V})
\text{ s.t. } \quad \text{tr}(\bar{U}_1 + \bar{U}_2) = 1,
\sum_{j=1}^{r} (\hat{X}^T - \text{vec} \hat{V}_{j,j})\hat{e}^T + \sum_{j=1}^{r} (\hat{X}^T - \text{vec} \hat{V}_{j,j})^T - \hat{X}^T \hat{V} \hat{X}^T = \hat{U}_1 - \hat{U}_2,
\begin{bmatrix}
\hat{G} \\
-\hat{y}^T \\
\hat{\lambda}
\end{bmatrix} \succeq 0,
\hat{U}_1 \succeq 0, \quad \hat{U}_2 \succeq 0, \quad \hat{V} = \begin{bmatrix}
\hat{V}_{1,1} & \cdots & \hat{V}_{r,1} \\
\vdots & \ddots & \vdots \\
\hat{V}_{r,1}^T & \cdots & \hat{V}_{r,r}
\end{bmatrix} \succeq 0.
\]
On the Absence of Spurious Local Minima in Nonlinear Low-Rank Matrix Recovery Problems

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

\[ M = r((\text{vec } I_d) \hat{e}^T + \hat{e}(\text{vec } I_d)^T) + XX^T, \]

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 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 LMI(\( X, Z \)) given by:

\[
\max_{U_1,U_2,V,G,\lambda,y} \quad \text{tr}(U_1 - U_2) - \text{tr}(G) - a^2 \lambda - b \text{tr}(V)
\]

s.t. \( \text{tr}(U_1 + U_2) = 1, \)

\[
\sum_{j=1}^{r} (Xy - \text{vec } V_{j,j})e^T + \sum_{j=1}^{r} e(Xy - \text{vec } V_{j,j})^T - XVX^T = P(U_1 - U_2)P^T,
\]

\[
\begin{bmatrix}
G & -y \\
-y^T & \lambda
\end{bmatrix} \succeq 0,
\]

\( U_1 \succeq 0, \quad U_2 \succeq 0, \quad V = \begin{bmatrix} V_{1,1} & \cdots & V_{r,1} \\ \vdots & \ddots & \vdots \\ V_{r,1} & \cdots & V_{r,r} \end{bmatrix} \succeq 0. \)

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

\[ U_1 = \hat{U}_1, \quad U_2 = \hat{U}_2, \quad V = (I_r \otimes P)(I_r \otimes P)^T,
\]

\[
\begin{bmatrix}
G & -y \\
-y^T & \lambda
\end{bmatrix} = \begin{bmatrix} I_r \otimes P & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix}
\hat{G} & -\hat{y} \\
-\hat{y}^T & \hat{\lambda}
\end{bmatrix} \begin{bmatrix} I_r \otimes P & 0 \\ 0 & 1 \end{bmatrix}
\]

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