SNIG Property of Matrix Low-rank Factorization Model

Hong Wang · Xin Liu · Xiaojun Chen · Ya-xiang Yuan

Abstract Recently, the matrix factorization model attracts increasing attentions in handling large-scale rank minimization problems, which is essentially a nonconvex minimization problem. Specifically, it is a quardratic least squares problem and consequently a quartic polynomial optimization problem. In this paper, we introduce a concept of the SNIG ("<u>S</u>econd-order <u>N</u>ecessary optimality <u>I</u>mplies <u>G</u>lobal optimality") condition which stands for the property that any second-order stationary point of the matrix factorization model must be a global minimizer. Some scenarios under which the SNIG condition holds are presented. Furthermore, we illustrate by an example when the SNIG condition may fail.

Keywords low rank factorization, nonconvex optimization, second-order optimality condition, global minimizer

Mathematics Subject Classification (2010) 15A18, 15A83, 65K05, 90C26

1 Introduction

1.1 Problem description

Consider the following matrix factorization problem

$$\underset{Y \in \mathbb{R}^{n \times k}, Z \in \mathbb{R}^{m \times k}}{\text{minimize}} \quad f(Y, Z) := \frac{1}{2} \|\mathcal{A}(YZ^\top) - b\|_2^2 = \frac{1}{2} \sum_{i=1}^{p} \left(\langle A_i, YZ^\top \rangle - b_i \right)^2, \quad (1)$$

H. Wang
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China, E-mail: hong.wang@connect.polyu.hk

X. Liu
State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China, E-mail: liuxin@lsec.cc.ac.cn.

X. Chen
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China, E-mail: xiaojun.chen@polyu.edu.hk

Ya-xiang Yuan
State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China, E-mail: yyx@lsec.cc.ac.cn

where $b \in \mathbb{R}^p$ is a column vector, $\mathcal{A} \in \mathcal{B}(\mathbb{R}^{n \times m}, \mathbb{R}^p)$ is a bounded linear operator mapping $n \times m$ matrices onto p-dimensional Euclidean space. Namely,

$$\mathcal{A}(X) = (\langle A_1, X \rangle, ..., \langle A_p, X \rangle)^\top ,$$

where $A_i \in \mathbb{R}^{n \times m} (i = 1, ..., p)$ are the p column matrices of $\mathcal{A}$ and $\langle W_1, W_2 \rangle := \text{tr}(W_1^\top W_2)$ designates the inner product of two matrices W_1 and W_2 with the same size. Denoting the adjoint operator of $\mathcal{A}$ by $\mathcal{A}^\top : \mathbb{R}^p \mapsto \mathbb{R}^{n \times m}$, it is not difficult to verify that

$$\mathcal{A}^\top(y) = \sum_{i=1}^{p} y_i A_i.$$

1.2 Existing works

Model (1) appears to be a practical and efficient way for solving low-rank optimization problem. It is also arisen from many areas of scientific and engineering applications including matrix completion, principle component analysis (PCA) and others [1,2,8]. LMaFit [27], for instance, using a series of matrix factorization models with different k (the approximation of the optimal rank) to describe the matrix completion problem, turns out to be an efficient and robust alternative to the convex relaxation model [3,7,10,16] based on nuclear norm relaxation [4,5,6,12,17,24]. Matrix factorization is also used to tackle semidefinite programs (SDP) problems. For instance, [18,19] introduced an equivalent factorization model for SDP through the Cholesky decomposition. Mishra in [17] used a factorization to make the trace norm differentiable in the search space and the duality gap numerically computable, which is a similar approach to SVD.

However, the factorization model (1) is nonconvex. More specifically, it is a quartic polynomial optimization problem. It may contain exponential number of local minimizers or saddle points. Hence, solving problem (1) to the global optimality is usually unachievable.

Recently, Candès and Li in [9] proposed a so-called Wirtinger Flow (WF) method to solve the phase retrieval problem, which is, like (1), essentially a quardratic least squares problem and quartic polynomial problem. The WF algorithm consists of two phases, one is a careful initialization stage realized by a spectral method, and the other is the local minimization stage invoking a gradient descent algorithm with a restricted stepsize. The authors proved that if the random sampling vectors obey certain distribution and there is no noise in the observation, the sequence generated by the gradient descent scheme will converge linearly to a global solution with high probability. Sun and Luo in [20] applied a similar idea to analyze the matrix completion problems described by factorization formulation, in which an initialization step is followed by a general first-order algorithm framework. Under the standard assumptions on incoherence condition [4] and the random observations similar to [9], the authors of [20] showed their framework can converge to a global solution linearly.

1.3 Our contributions

Even if the linear operator $\mathcal{A}$ of problem (1) does not involve any random property, it is observed that some local optimal solvers can often find a global solution of (1) by starting from a randomly chosen initial point. In this paper, we theoretically investigate the relationship between the global optimality of problem (1) and its second-order optimality under certain scenarios, which can partly explain the above mentioned phenomenon.

Note that if there exists a nonzero vector $c \in \mathbb{R}^p$ such that $\mathcal{A}^\top(c) = \mathbf{0}$, the linear operator $\mathcal{A}$ is row linearly dependent which implies the redundancy of the observations $\mathcal{A}(YZ^\top) = b$. To simply invoke the second-order optimality condition in further analysis, it is reasonable for us to exclude the row rank deficient cases. So in this paper, we are only interested in those operators with linearly independent rows, that is, we only consider those operators with full row rank. For simplicity of notations, we denote such operators as

$$\Pi^{(n,m,p)} := \mathcal{B}(\mathbb{R}^{n \times m}, \mathbb{R}^p) \cap \{\mathcal{A} \mid \mathcal{A}^\top(c) \neq 0, \ \forall c \neq 0, c \in \mathbb{R}^p\}.$$

Clearly, problem (1) is bounded from below.

Assumption 1 *Problem* (1) *has zero residual solution.*

It is clear that the global optimality of (1) becomes checkable: $f(Y, Z) = 0$ once Assumption 1 holds. For the convenience of our analysis, throughout this paper, we assume Assumption 1 holds.

Let $K := \{1, 2, \cdots, \min(n, m)\}$. Note that once a triplet $(\mathcal{A}, b, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ is given, a specific instance of problem (1) is immediately determined. For convenience, once we say $(\mathcal{A}, b, k)$ satisfies Assumption 1, it refers to the fact that problem (1) satisfies Assumption 1. Moreover, if $(\mathcal{A}, b, k)$ satisfies Assumption 1, there must exist at least one rank-k matrix $W \in \mathbb{R}^{n \times m}$ such that $\mathcal{A}(W) = b$. Denote

$$r^* = \operatorname*{argmin}_r \{r = \operatorname{rank}(W) \mid \mathcal{A}(W) = b\},$$

we have $r^* \leq k \leq \min(n, m)$.

Although solving the nonlinear least squares problem (1) to the global optimality is NP-hard in general, but obtaining a second-order stationary point can be achieved in polynomial time and there is no gap between them in quite some scenarios.

Definition 1 Given a triplet $(\mathcal{A}, b, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ satisfying Assumption 1. Let $\mathcal{C}$ be a cone in $\mathbb{R}^{n \times k} \times \mathbb{R}^{m \times k}$. Then, if for any $(Y, Z) \in \mathcal{C}$ satisfying the second-order necessary optimality condition of (1), (Y, Z) is a global optimizer of (1), we call the SNIG (<u>S</u>econd-order <u>N</u>ecessary optimality <u>I</u>mplies <u>G</u>lobal optimality) condition holds at the triplet $(\mathcal{A}, b, k)$ over $\mathcal{C}$. Particularly, if $\mathcal{C} = \mathbb{R}^{n \times k} \times \mathbb{R}^{m \times k}$, we say the SNIG condition holds at the triplet $(\mathcal{A}, b, k)$.

Different from [9] and [20], we do not impose any random properties on the operator "$\mathcal{A}$". On the other hand, we assume that it is of some special structures, for instance, the structure mentioned in [11]. Our main contributions are to provide some scenarios of $(\mathcal{A}, b, k)$ under which the SNIG condition holds:

1. the SNIG condition always holds over the cone consisting of all the rank deficient points in $\mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$;
2. the SNIG condition holds when the number of observation entries is equal to the total number of unknowns;
3. the SNIG condition holds if the operator $\mathcal{A}$ enjoys the special structure mentioned in [11], more specifically, $\mathcal{A}$ maps a matrix to a part of it and the missing part is a block of the matrix.

The above situations can be viewed as sufficient conditions of the SNIG. Moreover, we also demonstrate by a simple instance that the SNIG does not always hold.

1.4 Organization

The rest of this paper is organized as follows. In Section 2, we first give the optimality conditions of problem (1) and then present the proofs of the above mentioned first two sufficient conditions of the SNIG. We assert the third sufficient condition in Section 3. In Section 4, an example is constructed to illustrate that the SNIG condition may be violated. Finally, we conclude this paper in Section 5.

2 Preliminaries and two special scenarios

In this section, we first present the optimality conditions for the factorization model (1). Then we deliver two sufficient conditions for the SNIG.

2.1 Optimality conditions

The gradient of $f(Y, Z)$ can be expressed as

$$\nabla f(Y, Z) = \begin{bmatrix} \nabla_Y f(Y, Z) \\ \nabla_Z f(Y, Z) \end{bmatrix},$$

where

$$\nabla_Y f(Y, Z) = \mathcal{A}^\top(\mathcal{A}(YZ^\top) - b)Z;$$
$$\nabla_Z f(Y, Z) = (\mathcal{A}^\top(\mathcal{A}(YZ^\top) - b))^\top Y.$$

The Hessian of $f(Y, Z)$ can be expressed as

$$\nabla^2 f(Y, Z) = \begin{bmatrix} \nabla^2_{YY} f(Y, Z) & \nabla^2_{YZ} f(Y, Z) \\ \nabla^2_{ZY} f(Y, Z) & \nabla^2_{ZZ} f(Y, Z) \end{bmatrix}, \tag{2}$$

where

$$\nabla^2_{YY} f(Y, Z)[S_Y] = \mathcal{A}^\top(\mathcal{A}(S_Y Z^\top))Z; \tag{3a}$$

$$\nabla^2_{YZ} f(Y, Z)[S_Z] = \mathcal{A}^\top(\mathcal{A}(Y S_Z^\top))Z + \mathcal{A}^\top((\mathcal{A}(YZ^\top) - b))S_Z; \tag{3b}$$

$$\nabla^2_{ZY} f(Y, Z)[S_Y] = (\mathcal{A}^\top(\mathcal{A}(S_Y Z^\top)))^\top Y + (\mathcal{A}^\top(\mathcal{A}(YZ^\top) - b)))^\top S_Y; \tag{3c}$$

$$\nabla^2_{ZZ} f(Y, Z)[S_Z] = (\mathcal{A}^\top(\mathcal{A}(Y S_Z^\top)))^\top Y, \tag{3d}$$

for all $S_Y \in \mathbb{R}^{n \times k}$ and $S_Z \in \mathbb{R}^{m \times k}$.

Since (1) is a twice continuously differentiable unconstrained optimization problem, we can directly give its first-order and second-order necessary optimality condition as follows, respectively.

Definition 2 A pair of matrices $(Y^*, Z^*) \in \mathbb{R}^{n \times k} \times \mathbb{R}^{m \times k}$ is called a stationary point of (1) if $\nabla f(Y^*, Z^*) = \mathbf{0}$.

Proposition 1 *Let (Y^*, Z^*) be a local minimizer of (1), then it must be a stationary point and $\nabla^2 f(Y^*, Z^*)$ is positive semi-definite. Namely,*

$$||\mathcal{A}(Y^* S_Z^\top + S_Y (Z^*)^\top)||_2^2 + 2tr(S_Y^\top \mathcal{A}^\top(\mathcal{A}(Y^*(Z^*)^\top - UV^\top))S_Z) \geq 0, \tag{4}$$

for all $S_Y \in \mathbb{R}^{n \times k}$, $S_Z \in \mathbb{R}^{m \times k}$ where (U, V) is a fixed point satisfying $\mathcal{A}(UV^\top) = b$.

Proof The first part directly follows from the first-order optimality condition. Next, we give the proof of the second part. The Hessian $\nabla^2 f(Y^*, Z^*)$ is positive semi-definite due to the second-order optimality necessary condition. We derive from the positive semi-definiteness of $\nabla^2 f(Y^*, Z^*)$ and the relation (2) that

$$
\langle S_Y, \nabla_{YY}^2 f(Y^*, Z^*)[S_Y] \rangle + \langle S_Y, \nabla_{YZ}^2 f(Y^*, Z^*)[S_Z] \rangle
$$
$$
+ \langle S_Z, \nabla_{ZY}^2 f(Y^*, Z^*)[S_Y] \rangle + \langle S_Z, \nabla_{ZZ}^2 f(Y^*, Z^*)[S_Z] \rangle \geq 0 \tag{5}
$$

holds for all $S_Y \in \mathbb{R}^{n \times k}$ and $S_Z \in \mathbb{R}^{m \times k}$. Substituting relations (3) into (5), we can obtain (4) which completes the proof. $\qquad \square$

Definition 3 A pair of matrices $(Y^*, Z^*) \in \mathbb{R}^{n \times k} \times \mathbb{R}^{m \times k}$ is called a second-order stationary point of (1), if it is a stationary point and the second-order necessary optimality condition (4) holds for all $S_Y \in \mathbb{R}^{n \times k}$, $S_Z \in \mathbb{R}^{m \times k}$.

Apparently, a local minimizer must be a second-order stationary point, but not necessarily vice versa.

2.2 Rank deficient second-order stationary point

The SNIG condition holds at a rather general collection of triplets $(\mathcal{A}, b, k)$ over a particular cone in $\mathbb{R}^{n \times k} \times \mathbb{R}^{m \times k}$.

Theorem 2 *Given a triplet $(\mathcal{A}, b, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ satisfying Assumption 1. Then the SNIG condition holds at the triplet $(\mathcal{A}, b, k)$ over the cone*

$$
\bar{\mathcal{C}} := \{(X, Y) \in \mathbb{R}^{n \times k} \times \mathbb{R}^{m \times k} \mid \text{at least one of } X \text{ and } Y \text{ is rank deficient}\}.
$$

Proof Let (Y, Z) be any second-order stationary point of (1) satisfying that at least one of Y and Z is rank deficient.

Denote (U, V) as a global optimizer of (1). It follows from Assumption 1 and $k \geq r^*$ that $f(U, V) = 0$. Namely, $\mathcal{A}(UV^\top) = b$ holds. Therefore we can rewrite the objective function of (1) as

$$
f(Y, Z) = \frac{1}{2} \|\mathcal{A}(YZ^\top) - \mathcal{A}(UV^\top)\|_2^2. \tag{6}
$$

Without loss of generality, we assume that Y is rank deficient, *i.e.*, there exists a nonzero vector $\tilde{y} \in \mathbb{R}^k$ satisfying $Y\tilde{y} = 0$. Without loss of generality, we assume $\tilde{y}_l \neq 0$ for some $l \in \{1, 2, \ldots, k\}$. We use reduction to absurdity. Suppose there exists $(s, t) \in \{(i, j) \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ satisfying $\eta = (\mathcal{A}^\top(\mathcal{A}(YZ^\top - UV^\top)))_{st} \neq 0$.

Then, we set $S_Y \in \mathbb{R}^{n \times k}$ as follows

$$
(S_Y)_{i_1 i_2} = \begin{cases} \eta, & \text{if } i_1 = s \text{ and } i_2 = l, \\ 0, & \text{otherwise.} \end{cases} \tag{7}
$$

Let $S_Z = \tilde{z}\tilde{y}^\top$, and set $\tilde{z} \in \mathbb{R}^m$ as follows

$$
\tilde{z}_j = \begin{cases} -\xi \tilde{y}_l, & \text{if } j = t, \\ 0, & \text{otherwise,} \end{cases} \tag{8}
$$

where $\xi > \|\mathcal{A}(r_s z_l^\top)\|_2^2 / 2\tilde{y}_l^2$. Here $r_s \in \mathbb{R}^n$ is the s-th column of identity matrix I_n and z_l denotes the l-th column of Z.

Recall Proposition 1, we have

$$\|\mathcal{A}(Y S_Z^\top + S_Y Z^\top)\|_2^2 + 2\mathrm{tr}(S_Y^\top \mathcal{A}^\top (\mathcal{A}(YZ^\top - UV^\top))S_Z) \geq 0, \tag{9}$$

for all $S_Y \in \mathbb{R}^{n \times k}$, $S_Z \in \mathbb{R}^{m \times k}$ and $\|S_Y\|_\mathrm{F}^2 + \|S_Z\|_\mathrm{F}^2 \neq 0$.

Plugging (7) and (8) into (9), we obtain

$$\begin{aligned}
&\|\mathcal{A}(Y S_Z^\top + S_Y Z^\top)\|_2^2 + 2\mathrm{tr}(S_Y^\top \mathcal{A}^\top (\mathcal{A}(YZ^\top - UV^\top))S_Z) \\
&= \|\mathcal{A}(Y \tilde{y} \tilde{z}^\top + S_Y Z^\top)\|_2^2 + 2\mathrm{tr}(S_Y^\top \mathcal{A}^\top (\mathcal{A}(YZ^\top - UV^\top))\tilde{z}\tilde{y}^\top) \\
&= \|\mathcal{A}(S_Y Z^\top)\|_2^2 + 2\mathrm{tr}(\tilde{y}^\top S_Y^\top \mathcal{A}^\top (\mathcal{A}(YZ^\top - UV^\top))\tilde{z}) \\
&= \eta^2 \|\mathcal{A}(r_s z_l^\top)\|_2^2 - 2\xi \eta^2 \tilde{y}_l^2 < 0.
\end{aligned}$$

Hence, the second-order necessary optimality condition is violated, which is contrary to the fact that (Y, Z) is a second-order stationary point. Therefore $\mathcal{A}^\top (\mathcal{A}(YZ^\top - UV^\top)) = 0$, which implies $\mathcal{A}(YZ^\top - UV^\top) = 0$ due to the full rankness of $\mathcal{A}$. Namely, $f(Y, Z) = 0$. We complete the proof.   □

2.3 The scenario when $\mathcal{A}$ is special bijection

Consider $p = nm$ and a special class of triplets $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$ where

$$A_{i+(j-1)n}^{(C,D)} = C_i D_j^\top \quad \text{and} \quad b_{i+(j-1)n}^{(B)} = B_{i,j}, \quad i = 1, ..., n, \, j = 1, ..., m, \tag{10}$$

with C_i and D_j the i-th column of C and j-th column of D, respectively. And $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{n \times n}$ and $D \in \mathbb{R}^{m \times m}$ are three given matrices. Then it is not difficult to verify that

$$\|\mathcal{A}^{(C,D)}(X) - b^{(B)}\|_2^2 = \|C^\top X D - B\|_\mathrm{F}^2. \tag{11}$$

Then we have the following result.

Theorem 3 *Suppose that the triplet $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$ defined by (10) satisfies Assumption 1. If both $C \in \mathbb{R}^{n \times n}$ and $D \in \mathbb{R}^{m \times m}$ are nonsingular, then the SNIG condition holds at the triplet $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$.*

Proof It follows from the triplet $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$ satisfying Assumption 1 that $k \geq \mathrm{rank}(B)$ and there exist $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{m \times k}$ such that $b^{(B)} = \mathcal{A}^{(C,D)}(C^\top (UV^\top)D)$. Namely, (U, V) is a global optimizer of (1) corresponding to the triplet $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$. By (11), the objective function can be reformulated as

$$f(Y, Z) = \frac{1}{2} \|C^\top (YZ^\top)D - B\|_\mathrm{F}^2.$$

Assume that (Y, Z) is a second-order stationary point of problem (1) with respect to $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$. The first-order optimality condition can be written as

$$\begin{cases} Y^\top C C^\top (YZ^\top - UV^\top)DD^\top = \mathbf{0}, \\ C C^\top (YZ^\top - UV^\top)DD^\top Z = \mathbf{0}. \end{cases} \tag{12}$$

Denote $\tilde{Y} = CC^\top Y$, $\tilde{Z}^\top = Z^\top DD^\top$, $\tilde{U} = CC^\top U$, $\tilde{V}^\top = V^\top DD^\top$, (12) can be rearranged as

$$\begin{cases} (\tilde{Y}\tilde{Z}^\top - \tilde{U}\tilde{V}^\top)Z = \mathbf{0}, \\ Y^T(\tilde{Y}\tilde{Z}^\top - \tilde{U}\tilde{V}^\top) = \mathbf{0}. \end{cases}$$

First we consider the case that both $Y^\top\tilde{Y}$ and $\tilde{Z}^\top Z$ are nonsingular, we have

$$\begin{cases} \tilde{Y} = \tilde{U}\tilde{V}^\top Z(\tilde{Z}^\top Z)^{-1}, \\ \tilde{Z} = ((Y^\top\tilde{Y})^{-1}Y^\top\tilde{U}\tilde{V}^\top)^\top, \end{cases} \tag{13}$$

and then

$$Y^\top(\tilde{Y}\tilde{Z}^\top - \tilde{U}\tilde{V}^\top)Z = \mathbf{0} \quad \Rightarrow \quad Y^\top\tilde{Y}\tilde{Z}^\top Z = Y^\top\tilde{U}\tilde{V}^\top Z,$$

which indicates that $Y^\top\tilde{U}$ and $\tilde{V}^\top Z$ are both nonsingular as well. Moreover, we can obtain

$$(\tilde{U}^\top Z)^{-1}(Y^\top\tilde{V})^{-1} = (\tilde{Z}^\top Z)^{-1}(Y^\top\tilde{Y})^{-1}. \tag{14}$$

Together with (13) and (14), we have

$$\begin{aligned} (\mathcal{A}^{(C,D)})^\top(\mathcal{A}^{(C,D)}(YZ^\top - UV^\top)) &= \tilde{Y}\tilde{Z}^\top - \tilde{U}\tilde{V}^\top \\ &= \tilde{U}\tilde{V}^\top Z(\tilde{Z}^\top Z)^{-1}(Y^\top\tilde{Y})^{-1}Y^\top\tilde{U}\tilde{V}^\top - \tilde{U}\tilde{V}^\top \\ &= \tilde{U}(\tilde{V}^\top Z(\tilde{Z}^\top Z)^{-1}(Y^\top\tilde{Y})^{-1}Y^\top\tilde{U} - I_k)\tilde{V}^\top \\ &= \mathbf{0}, \end{aligned}$$

which indicates $\mathcal{A}^{(C,D)}(YZ^\top - UV^\top) = \mathbf{0}$ due to the full rankness of $\mathcal{A}^{(C,D)}$.

Finally, we notice that $Y^\top\tilde{Y} = (CY)^\top(CY)$ and $\tilde{Z}^\top Z = (DZ)^\top(DZ)$. It then follows from the nonsingularity of C and D that the nonsingularity of $Y^\top\tilde{Y}$ or $\tilde{Z}^\top Z$ implies the rank deficiency of Y or Z. Then we recall Theorem 2, and complete the proof. $\qquad\square$

3 The scenario when $\mathcal{A}$ takes a special form

Consider the triplet $(\mathcal{A}^\Omega, b^\Omega, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ with $(\mathcal{A}^\Omega, b^\Omega)$ defined by

$$\mathcal{A}^\Omega(X) = (\langle A_1^\Omega, X\rangle, \ldots, \langle A_p^\Omega, X\rangle)^\top, \quad b^\Omega = (b_1^\Omega, \ldots, b_p^\Omega)^\top, \tag{15}$$

where $\Omega \subset \{(i,j) | 1 \leq i \leq n, 1 \leq j \leq m\}$, $p = |\Omega|$ is the cardinality of Ω, and $A_t^\Omega = E_{i_t j_t}$ $((i_t, j_t) \in \Omega, 1 \leq t \leq p)$. Here $E_{ij} \in \mathbb{R}^{n \times m}$ is a matrix with (i,j)-th entry equal to 1 and 0 otherwise. We call Ω the observation-index set.

Any triplet $(\mathcal{A}^\Omega, b^\Omega, k)$ determines a concrete instance of (1) as the following

$$\underset{Y \in \mathbb{R}^{n \times k}, Z \in \mathbb{R}^{m \times k}}{\text{minimize}} \quad f(Y, Z) := \frac{1}{2}||\mathcal{A}^\Omega(YZ^\top) - b^\Omega||_2^2. \tag{16}$$

If the triplet $(\mathcal{A}^\Omega, b^\Omega, k)$ satisfies Assumption 1, as mentioned before, there exists at least one matrix M such that

$$\mathcal{A}^\Omega(M) = b^\Omega,$$

or equivalently,

$$M_{i_t j_t} = b_t^\Omega, \quad (i_t, j_t) \in \Omega, \ 1 \leq t \leq p.$$

We call such matrix M an observation matrix with respect to the triplet $(\mathcal{A}^\Omega, b^\Omega, k)$.

3.1 Main results and proof outline

Suppose the observation-index set Ω in (15) has the following form

$$\Omega = I_1 \cup I_2 \cup I_3, \tag{17}$$

where

$$\begin{aligned}
I_1 &= \{(i_s, j_t) | 1 \le s \le \tilde{n}, 1 \le t \le \tilde{m}\}, \\
I_2 &= \{(i_s, j_t) | 1 \le s \le \tilde{n}, \tilde{m} + 1 \le t \le m\}, \\
I_3 &= \{(i_s, j_t) | \tilde{n} + 1 \le s \le n, 1 \le t \le \tilde{m}\},
\end{aligned} \tag{18}$$

for given $\tilde{n} \in \{1, \cdots, n\}$, $\tilde{m} \in \{1, \cdots, m\}$. Clearly, Ω defined by (17) consists of $\tilde{n}$ rows and $\tilde{m}$ columns of the observation matrix M.

Our main results can be stated as follows.

Theorem 4 *Let $(\mathcal{A}^\Omega, b^\Omega)$ be defined as (15) and Ω be in the form of (17). Suppose the triplet $(\mathcal{A}^\Omega, b^\Omega, k)$ satisfies Assumption 1 and M is its corresponding observation matrix with rank r. If the rank of the submatrices M_{I_i} indexed by I_i satisfies $\mathrm{rank}(M_{I_i}) = \mathrm{rank}(M) = r$ where I_i $(i = 1, 2, 3)$ is given by (18), then the SNIG condition holds at the triplet $(\mathcal{A}^\Omega, b^\Omega, k)$.*

Remark 1 We notice that the scenario discussed in Theorem 4 is much more general than the one discussed in Theorem 3. The number of observations (the number of known entries of the observation matrix M) here is $p = \tilde{m}n + \tilde{n}(m - \tilde{m}) < mn$. The smallest choice of p is $k(m + n)$ when $\tilde{m} = \tilde{n} = k$, which is much less than the lowest requirement, i.e. $(m + n)\log(m + n)$, on the number of observations to guarantee the exact recovery.

The proof of Theorem 4 will be divided into two parts:

i) Firstly, we prove that the SNIG condition holds at the triplet $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ with the observation set $\tilde{\Omega}$ being a special case of (17);

ii) Secondly, we show that for any Ω defined as (15), there exists $\tilde{\Omega}$ being of the structure mentioned above and the same cardinality with Ω, and satisfying that problems (1) determined by $(\mathcal{A}^\Omega, b^\Omega, k)$ and $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ share the same optimality properties, i.e. function values, optimality conditions.

3.2 The situation that $\tilde{\Omega}$ is of special structure

In particular, suppose the special triplet $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ satisfies Assumption 1 where $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}})$ is defined by (15) and the observation-index set is of the form

$$\tilde{\Omega} = \tilde{I}_1 \cup \tilde{I}_2 \cup \tilde{I}_3 \tag{19}$$

where

$$\begin{aligned}
\tilde{I}_1 &= \{(i, j) | 1 \le i \le \tilde{n}, 1 \le j \le \tilde{m}\}, \\
\tilde{I}_2 &= \{(i, j) | 1 \le i \le \tilde{n}, \tilde{m} + 1 \le j \le m\}, \\
\tilde{I}_3 &= \{(i, j) | \tilde{n} + 1 \le i \le n, 1 \le j \le \tilde{m}\},
\end{aligned} \tag{20}$$

which contains all the indices of the first $\tilde{n}$ rows and the first $\tilde{m}$ columns of the corresponding observation matrix, say $\tilde{M}$.

Hence, problem (16) can further be reformulated as,

$$\underset{\tilde{Y}\in\mathbb{R}^{n\times k},\tilde{Z}\in\mathbb{R}^{m\times k}}{\text{minimize}} \quad \tilde{f}(\tilde{Y},\tilde{Z}) := \frac{1}{2}\|\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}\|_2^2. \tag{21}$$

Lemma 1 *Let $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}})$ be defined by (15) and $\tilde{\Omega}$ be in the form of (19). Suppose the triplet $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ satisfies Assumption 1 and $\tilde{M}$ is the corresponding observation matrix with rank $\tilde{r}$. If the rank of the submatrices $\tilde{M}_{\tilde{I}_i}$ indexed by $\tilde{I}_i$ satisfies $\text{rank}(\tilde{M}_{\tilde{I}_i}) = \text{rank}(\tilde{M}) = \tilde{r}$ where $\tilde{I}_i \, (i = 1, 2, 3)$ is given by (20), then the SNIG condition holds at the triplet $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$.*

Proof Suppose $(\tilde{U}, \tilde{V}) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ is the global minimizer of (21), namely, $\tilde{M} = \tilde{U}\tilde{V}^\top$ and $b^{\tilde{\Omega}} = \mathcal{A}^{\tilde{\Omega}}(\tilde{U}\tilde{V}^\top)$. Let $(\tilde{Y}, \tilde{Z}) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ be a second-order stationary point of problem (21). Next, we are going to prove the lemma by the following two folds.
I: Either $\tilde{Y}$ or $\tilde{Z}$ is rank deficient. In this case, the SNIG condition holding at $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ directly follows from Theorem 2 and hence the proof is completed.
II: Both of $\tilde{Y}$ and $\tilde{Z}$ are of full column rank. According to the structure of $\tilde{\Omega}$ in the form of (19), we rewrite the matrices $\tilde{U}, \tilde{V}, \tilde{Y}$ and $\tilde{Z}$ as follows

$$\tilde{U} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, \quad \tilde{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad \tilde{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix},$$

where $U_1, Y_1 \in \mathbb{R}^{\tilde{n}\times k}$, $U_2, Y_2 \in \mathbb{R}^{(n-\tilde{n})\times k}$, $V_1, Z_1 \in \mathbb{R}^{\tilde{m}\times k}$ and $V_2, Z_2 \in \mathbb{R}^{(m-\tilde{m})\times k}$. Then it follows from straightforward calculations that

$$(\mathcal{A}^{\tilde{\Omega}})^\top(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}) = (\mathcal{A}^{\tilde{\Omega}})^\top\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top - \tilde{U}\tilde{V}^\top) = \begin{bmatrix} Y_1 Z_1^\top - U_1 V_1^\top & Y_1 Z_2^\top - U_1 V_2^\top \\ Y_2 Z_1^\top - U_2 V_1^\top & \mathbf{0} \end{bmatrix}.$$

Hence, the first-order optimality condition of (21) can be expressed as follows,

$$\begin{cases} ((\mathcal{A}^{\tilde{\Omega}})^\top(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}))\tilde{Z} = \begin{bmatrix} (Y_1 Z_1^\top - U_1 V_1^\top)Z_1 + (Y_1 Z_2^\top - U_1 V_2^\top)Z_2 \\ (Y_2 Z_1^\top - U_2 V_1^\top)Z_1 \end{bmatrix} = \mathbf{0}, \\ ((\mathcal{A}^{\tilde{\Omega}})^\top(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}))^\top\tilde{Y} = \begin{bmatrix} (Z_1 Y_1^\top - V_1 U_1^\top)Y_1 + (Z_1 Y_2^\top - V_1 U_2^\top)Y_2 \\ (Z_2 Y_1^\top - V_2 U_1^\top)Y_1 \end{bmatrix} = \mathbf{0} \end{cases}$$

that is

$$\begin{cases} (Y_1 Z_1^\top - U_1 V_1^\top)Z_1 + (Y_1 Z_2^\top - U_1 V_2^\top)Z_2 = \mathbf{0}, & \text{(22a)} \\ (Y_2 Z_1^\top - U_2 V_1^\top)Z_1 = \mathbf{0}, & \text{(22b)} \\ (Z_1 Y_1^\top - V_1 U_1^\top)Y_1 + (Z_1 Y_2^\top - V_1 U_2^\top)Y_2 = \mathbf{0}, & \text{(22c)} \\ (Z_2 Y_1^\top - V_2 U_1^\top)Y_1 = \mathbf{0}. & \text{(22d)} \end{cases}$$

By rearranging (22a) and (22c), we have

$$\begin{cases} Y_1(\tilde{Z}^\top\tilde{Z}) = U_1(\tilde{V}^\top\tilde{Z}), & \text{(23a)} \\ Z_1(\tilde{Y}^\top\tilde{Y}) = V_1(\tilde{U}^\top\tilde{Y}). & \text{(23b)} \end{cases}$$

Combining (23), (22b) and (22d) together, we obtain

$$
\begin{cases}
Y_1 = U_1(\tilde{V}^\top \tilde{Z})(\tilde{Z}^\top \tilde{Z})^{-1}, & \text{(24a)} \\
Z_1 = V_1(\tilde{U}^\top \tilde{Y})(\tilde{Y}^\top \tilde{Y})^{-1}, & \text{(24b)} \\
Y_2(Z_1^\top Z_1) = U_2(V_1^\top Z_1), & \text{(24c)} \\
Z_2(Y_1^\top Y_1) = V_2(U_1^\top Y_1). & \text{(24d)}
\end{cases}
$$

As we know, the second-order necessary optimality condition of problem (21) can be formulated as

$$
\|\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}S_{\tilde{Z}}^\top + S_{\tilde{Y}}\tilde{Z}^\top)\|_2^2 + 2\mathrm{tr}(S_{\tilde{Y}}^\top((\mathcal{A}^{\tilde{\Omega}})^\top(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top - \tilde{U}\tilde{V}^\top)))S_{\tilde{Z}}) \geq 0, \qquad (25)
$$

for all $S_{\tilde{Y}} \in \mathbb{R}^{n\times k}$, $S_{\tilde{Z}} \in \mathbb{R}^{m\times k}$.

We further prove our argument through discussing the following four cases of different structures of (Y_1, Z_1),

  i. Y_1 is rank deficient and Z_1 is of full column rank;
 ii. Y_1 is of full column rank and Z_1 is rank deficient;
iii. both Y_1 and Z_1 are rank deficient;
 iv. both Y_1 and Z_1 are of full column rank.

i. Y_1 is rank deficient and Z_1 is of full column rank. Due to the rank deficiency of Y_1, there exists a nonzero vector $\tilde{y} = (y_1, \ldots, y_k)^\top \in \mathbb{R}^k$ satisfying $Y_1\tilde{y} = 0$. Without loss of generality, we assume that $y_l \neq 0$ for some $l \in \{1, \cdots, k\}$. Then we can conclude that $Y_1 Z_2^\top = U_1 V_2^\top$. Otherwise, suppose that there exists $(s, t) \in \{1, \cdots, \tilde{n}\} \times \{\tilde{m}+1, \cdots, m\}$ satisfying $\eta = (Y_1 Z_2^\top - U_1 V_2^\top)_{st} \neq 0$.

Then we set

$$
S_{\tilde{Y}} = \begin{bmatrix} S_{Y_1} \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad S_{\tilde{Z}} = \begin{bmatrix} \mathbf{0} \\ S_{Z_2} \end{bmatrix} \qquad (26)
$$

where

$$
(S_{Y_1})_{i_1 i_2} = \begin{cases} \eta, & \text{if } i_1 = s \text{ and } i_2 = l, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad S_{Z_2} = \tilde{z}\tilde{y}^\top
$$

with $\tilde{z} \in \mathbb{R}^{m-\tilde{m}}$ given by

$$
\tilde{z}_j = \begin{cases} -\xi y_l, & \text{if } j = t, \\ 0, & \text{otherwise,} \end{cases}
$$

and $\xi > \|\mathcal{A}^\Omega(r_s z_l^\top)\|_2^2/2y_l^2$. Here $r_s \in \mathbb{R}^{\tilde{n}}$ is the s-th column of identity matrix $I_{\tilde{n}}$ and z_l denotes the l-th column of Z.

Substituting (26) into (25), we obtain

$$
\begin{aligned}
&\|\mathcal{A}^\Omega(YS_{Z_2}^\top + S_{Y_1}Z^\top)\|_2^2 + 2\mathrm{tr}(S_{Y_1}^\top(Y_1 Z_2^\top - U_1 V_2^\top)S_{Z_2}) \\
=&\|\mathcal{A}^\Omega(S_{Y_1}Z^\top)\|_2^2 + 2\mathrm{tr}(\tilde{y}^\top S_{Y_1}^\top(Y_1 Z_2^\top - U_1 V_2^\top)\tilde{z}) \\
=&\eta^2\|\mathcal{A}^\Omega(r_s z_l^\top)\|_2^2 - 2\xi\eta^2\tilde{y}_l^2 < 0,
\end{aligned}
$$

which implies that the second-order necessary condition is violated at $(\tilde{Y}, \tilde{Z})$. Therefore, it holds that $Y_1 Z_2^\top = U_1 V_2^\top$. Together with (22a), (22b), we obtain $\tilde{Y}(Z_1^\top Z_1) = \tilde{U}(V_1^\top Z_1)$.

From the assumption $\text{rank}(\tilde{M}_{\tilde{I}_1}) = \tilde{r} = k$, we know $\text{rank}(U_1 V_1^\top) = \tilde{r} = k$, which implies U_1 and V_1 are full column rank. Thus, $V_1^\top Z_1$ is nonsingular. Consequently,

$$Y_1 = U_1(V_1^\top Z_1)(Z_1^\top Z_1)^{-1}$$

is of full column rank, which contradicts to the assumption that Y_1 is rank deficient.

ii. Y_1 is of full column rank and Z_1 is rank deficient. We can show the same contradiction as case i. in the same manner and hence omit the detailed proof.

iii. Both Y_1 and Z_1 are rank deficient. By the same argument in case i, it follows from the second-order optimality (25) that

$$\begin{cases} Y_1 Z_2^\top = U_1 V_2^\top, \\ Y_2 Z_1^\top = U_2 V_1^\top, \end{cases} \tag{27}$$

due to the rank deficiency of Y_1 and Z_1.

It follows from the relationship (27) and $\text{rank}(\tilde{M}_{\tilde{I}_i}) = \text{rank}(\tilde{M}) = \tilde{r}$ $(i = 1, 2, 3)$ that $\tilde{r}(= \text{rank}(\tilde{U}\tilde{V}^\top) = \text{rank}(\tilde{M})) < k$. Hence there exist $\bar{U} \in \mathbb{R}^{n \times \tilde{r}}$ and $\bar{V} \in \mathbb{R}^{m \times \tilde{r}}$ such that $\bar{U}\bar{V}^\top = \tilde{U}\tilde{V}^\top = \tilde{M}$. Denote $\bar{U}$ and $\bar{V}$ as

$$\bar{U} = \begin{bmatrix} \bar{U}_1 \\ \bar{U}_2 \end{bmatrix} \quad \text{and} \quad \bar{V} = \begin{bmatrix} \bar{V}_1 \\ \bar{V}_2 \end{bmatrix}.$$

Similar to (24) and (27), we obtain

$$\begin{cases} Y_1 = \bar{U}_1(\bar{V}^\top \tilde{Z})(\tilde{Z}^\top \tilde{Z})^{-1}, & \text{(28a)} \\ Z_1 = \bar{V}_1(\bar{U}^\top \tilde{Y})(\tilde{Y}^\top \tilde{Y})^{-1}, & \text{(28b)} \\ Y_2(Z_1^\top Z_1) = \bar{U}_2(\bar{V}_1^\top Z_1), & \text{(28c)} \\ Z_2(Y_1^\top Y_1) = \bar{V}_2(\bar{U}_1^\top Y_1). & \text{(28d)} \end{cases}$$

and

$$\begin{cases} Y_1 Z_2^\top = \bar{U}_1 \bar{V}_2^\top, \\ Y_2 Z_1^\top = \bar{U}_2 \bar{V}_1^\top. \end{cases} \tag{29}$$

It follows from (28b), the second equation of (29) and the full column rankness of $\bar{U}_2$, $\bar{V}_1$ that

$$\text{Span}\{Z_1\} = \text{Span}\{\bar{V}_1\}. \tag{30}$$

Define $\bar{H} := Y_1 Z_1^\top - \bar{U}_1 \bar{V}_1^\top \in \mathbb{R}^{\tilde{n} \times \tilde{m}}$. Due to (30), it holds that $\bar{H}y = 0$, for any $y \perp \text{Span}\{Z_1\}$. On the other hand, for any $y \in \text{Span}(Z_1)$, there exists $\lambda \in \mathbb{R}^{\tilde{m}}$ satisfying $y = Z_1\lambda$. Using (22a) and the first equation of (29), we have $\bar{H}y = 0$. Consequently, we can conclude that $\bar{H}y = 0$, for all $y \in \mathbb{R}^{\tilde{m}}$, which implies $\bar{H} = 0$. Together with (29), we have $(\mathcal{A}^{\tilde{\Omega}})^\top(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}) = 0$. Due to the full rankness of $A^{\tilde{\Omega}}$, we have $\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}} = \mathbf{0}$.

iv. Both Y_1 and Z_1 are of full column rank. Define

$$\hat{H} = Y_2 Z_1^\top - U_2 V_1^\top \in \mathbb{R}^{(n-\tilde{n}) \times \tilde{m}}.$$

The full column rankness of Z_1 and equation (24b) imply

$$\text{Span}\{Z_1\} = \text{Span}\{V_1\}. \tag{31}$$

Therefore, we can prove $\hat{H}y = 0$ holds for any $y \perp \mathrm{Span}Z_1$. By the optimality condition (22b), we have $\hat{H}y = 0$ holds for any $y \in \mathrm{Span}Z_1$. Thus, for any $y \in \mathbb{R}^{\tilde{m}}$, it holds that $\hat{H}y = 0$, which further implies

$$\hat{H} = 0. \tag{32}$$

Similarly,

$$Y_1 Z_2^\top - U_1 V_2 = 0. \tag{33}$$

Together with (22a), we have $(Y_1 Z_1^\top - U_1 V_1^\top)Z_1 = 0$. On the other hand, (31) implies $(Y_1 Z_1^\top - U_1 V_1^\top)z = 0$ holds for any $z \perp \mathrm{Span}\{Z_1\}$. Therefore,

$$Y_1 Z_1^\top - U_1 V_1^\top = 0. \tag{34}$$

Collecting (32), (33) and (34), we obtain $(\mathcal{A}^{\tilde{\Omega}})^\top (\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}) = 0$. Due to the full rankness of $\mathcal{A}^{\tilde{\Omega}}$, we have $\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}} = \mathbf{0}$.

To sum up, we conclude that if $(\tilde{Y}, \tilde{Z})$ is a second-order stationary point of problem (21), then $\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}} = \mathbf{0}$, i.e., $(\tilde{Y}, \tilde{Z})$ is a global minimizer of problem (21), which implies that the SNIG condition holds at the triplet $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$. We complete the proof. $\quad\square$

Remark 2 Let $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}})$ be defined by (15) with $b_i^{\tilde{\Omega}} = 1$ $(i = 1, \ldots, p)$ and $\tilde{\Omega}$ of the form

$$\tilde{\Omega} = \{(i,j) | 1 \le i \le 2, 1 \le j \le 2\} \cup \{(i,j) | 1 \le i \le 2, 3 \le j \le 4\} \cup \{(i,j) | 3 \le i \le 4, 1 \le j \le 2\},$$

which implies $n = m = 4$ and $p = |\Omega| = 12$. Set

$$\tilde{U} = \tilde{V} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}^\top,$$

and denote $\tilde{M} = \tilde{U}\tilde{V}^\top$ as the obervation matrix, we can easily verify that $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ satisfies Assumption 1 and the SNIG condition holds at $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, 2)$.

Set

$$\tilde{Y} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 \end{bmatrix}^\top, \quad \tilde{Z} = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 0 & 0 \end{bmatrix}^\top.$$

Clearly, $(\tilde{Y}, \tilde{Z})$ is also a global minimizer of (21). However, $\tilde{Y}\tilde{Z}^\top \ne \tilde{M}$, which means the exact recovery does not hold at $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$.

3.3 Permutation between (16) and (21)

Let Ω be any observation-index set defined by (17), and $(\mathcal{A}^{\Omega}, b^{\Omega})$ be defined by (15). Suppose M is the corresponding observation matrix of triplet $(\mathcal{A}^{\Omega}, b^{\Omega}, k)$, which implies that Assumption 1 is satisfied. Denote $S = (S_{ij})$ by

$$S_{ij} = \begin{cases} 1, & (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

It is not difficult to verify that there exist two permutation matrices, say P and Q, such that

$$\tilde{S} = PSQ^\top, \tag{35}$$

and the observation-index set $\tilde{\Omega}$ corresponding to $\tilde{S}$, i.e.

$$\tilde{S}_{ij} = \begin{cases} 1, & (i,j) \in \tilde{\Omega}, \\ 0, & \text{otherwise}, \end{cases}$$

satisfies (19). Let $\mathcal{A}^{\tilde{\Omega}}$ be defined by (15), and $b^{\tilde{\Omega}}$ be defined by

$$b_t^{\tilde{\Omega}} = \tilde{M}_{i_t j_t}, \qquad (i_t, j_t) \in \tilde{\Omega}, t = 1, 2, ..., p := |\Omega|,$$

where $\tilde{M} = PMQ^{\top}$.

Problem (1) determined by $(\mathcal{A}^{\Omega}, b^{\Omega}, k)$ and $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ can be reformulated as

$$\underset{Y \in \mathbb{R}^{n \times k}, Z \in \mathbb{R}^{m \times k}}{\text{minimize}} \quad f(Y, Z) = \frac{1}{2}\|S \circ (YZ^{\top} - M)\|_{\text{F}}^2, \tag{36}$$

and

$$\underset{\tilde{Y} \in \mathbb{R}^{n \times k}, \tilde{Z} \in \mathbb{R}^{m \times k}}{\text{minimize}} \quad \tilde{f}(\tilde{Y}, \tilde{Z}) = \frac{1}{2}\|\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\top} - \tilde{M})\|_{\text{F}}^2, \tag{37}$$

respectively, where $\circ$ designates the Hadamard product of two matrices with the same size.

We first reveal the relationship between the rank of the submatrices of M and $\tilde{M}$.

Lemma 2 *Let M be the observation matrix with respect to $(\mathcal{A}^{\Omega}, b^{\Omega}, k)$ and $\tilde{M} = PMQ^{\top}$ with P and Q defined by (35). Suppose that I_i and $\tilde{I}_i$ are defined by (18) and (20) for $i = 1, 2, 3$, then the submatrix M_{I_i} and $\tilde{M}_{\tilde{I}_i}$ have the same rank, that is,*

$$\text{rank}(M_{I_i}) = \text{rank}(\tilde{M}_{\tilde{I}_i}),\ i = 1, 2, 3.$$

Proof It is well known that multiplying a permutation matrix from the left and right side of a matrix is only reorder the rows and columns of that matrix. From $\tilde{M} = PMQ^{\top}$, we know that actually $\tilde{M}_{\tilde{I}_i} = M_{I_i}$, $\forall i = 1, 2, 3$. Thus, it holds that $\text{rank}(\tilde{M}_{\tilde{I}_i}) = \text{rank}(M_{I_i})$ ($\forall i = 1, 2, 3$). $\qquad\square$

To obtain the equivalence of (36) and (37), we need the following relationship.

Lemma 3 *Let P and Q be two permutation matrices and $S \in \mathbb{R}^{n \times m}$ be a $0 - 1$ matrix. Then we have*

$$S \circ W = P^{\top}(PSQ \circ PWQ)Q^{\top},\ \forall W \in \mathbb{R}^{n \times m}. \tag{38}$$

The proof of Lemma 3 directly follows from the definition of Hadamard product and the basic properties of permutation matrix, and hence be omitted here.

Now, we arrive at our main theorem.

Theorem 5 *Problem (36) and problem (37) share the following optimality properties:*

1. *$f(Y, Z) = \tilde{f}(\tilde{Y}, \tilde{Z})$, for any $Y \in \mathbb{R}^{n \times k}$, $Z \in \mathbb{R}^{m \times k}$, $\tilde{Y} = PY$ and $\tilde{Z} = QZ$;*
2. *if (Y^*, Z^*) is a second-order stationary point of problem (36), then (PY^*, QZ^*) is a second-order stationary point of problem (37), and vice versa.*

Proof 1. By the equality (38) and the orthogonal invariance of Frobenius norm, it is not difficult to verify that

$$\tilde{f}(\tilde{Y}, \tilde{Z}) = \frac{1}{2}\|\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\top} - \tilde{M})\|_{\text{F}}^2 = \frac{1}{2}\|S \circ (YZ^{\top} - M)\|_{\text{F}}^2 = f(Y, Z).$$

2. Again by the equality (38) and some basic arguments, we can conclude that (Y^*, Z^*) satisfying the first-order and second-order optimality conditions of problem (36) is equivalent to (PY^*, QZ^*) satisfying the first-order and second-order optimality conditions of problem (37), respectively. We omit the tedious details and complete the proof. $\qquad\square$

3.4 Proof of Theorem 4

Proof From Lemma 2, $\mathrm{rank}(\tilde{M}_{\tilde{I}_i}) = \mathrm{rank}(M_{I_i})$ $(i = 1, 2, 3)$. Then for any triplet $(\mathcal{A}^\Omega, b^\Omega, k)$, we can find a corresponding triplet $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ according to the procedure introduced in the previous subsection, and hence $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ satisfies the assumptions of Theorem 4 and Theorem 1. Suppose that $(\tilde{Y}, \tilde{Z})$ is a second-order stationary point of problem (37), it follows from Theorem 5 that $(Y, Z) = (P^\top \tilde{Y}, Q^\top \tilde{Z})$ is a second-order stationary point of problem (36). By (38) and Theorem 1, we have

$$(\mathcal{A}^\Omega)^\top(\mathcal{A}^\Omega(YZ^\top) - b^\Omega) = S \circ (YZ^\top - M) = P^\top(\tilde{S} \circ (\tilde{Y}\tilde{Z}^\top - \tilde{M}))Q$$
$$= P^\top((\mathcal{A}^{\tilde{\Omega}})^\top(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^\top) - b^{\tilde{\Omega}}))Q = \mathbf{0},$$

which implies $\mathcal{A}^\Omega(YZ^\top) - b^\Omega = \mathbf{0}$ due to the full rankness of $\mathcal{A}$, namely, $f(Y, Z) = 0$. Therefore, the SNIG condition holds the triplet $(\mathcal{A}^\Omega, b^\Omega, k)$. The proof is completed. □

4 A special example violating the SNIG condition

In this section, we illustrate that the SNIG condition may be violated through a concrete instance.

We describe the instance as follows. Let $\mathcal{A}^\Omega$ and b^Ω be defined by (15), Ω be an observation-index set and $n = m = 6$. Suppose that $(\mathcal{A}^\Omega, b^\Omega, 1)$ satisfies Assumption 1 and M is its observation matrix. As aforementioned, problem determined by $(\mathcal{A}^\Omega, b^\Omega, 1)$ can be reformulated as follows,

$$\underset{y \in \mathbb{R}^{6 \times 1}, z \in \mathbb{R}^{6 \times 1}}{\mathrm{minimize}} \quad f(y, z) = \frac{1}{2}\|S \circ (yz^\top - M)\|_\mathrm{F}^2, \tag{39}$$

where $S \in \mathbb{R}^{6 \times 6}$ is the $0 - 1$ matrix corresponding to the observation-index set Ω.

We set

$$S = \begin{bmatrix} E & I \\ I & E \end{bmatrix} \in \mathbb{R}^{6 \times 6}, \quad \text{and} \quad M := \bar{x}\bar{y}^\top \in \mathbb{R}^{6 \times 6},$$

where $\bar{x} = \bar{y} = (\mathbf{e}^\top, \mathbf{e}^\top)^\top$, $E = \mathbf{e}\mathbf{e}^\top$ and $\mathbf{e} = (1, 1, 1)^\top$.

Theorem 6 *The SNIG condition is violated at $(\mathcal{A}^\Omega, b^\Omega, 1)$.*

Proof Consider the point (x, y) with $x = y = \frac{\sqrt{2}}{2}(\mathbf{e}^\top, -\mathbf{e}^\top)^\top$. First, we prove that (x, y) is a second-order stationary point of problem (36) for the specific triplet $(\mathcal{A}^\Omega, b^\Omega, 1)$. It follows from the definition of S that

$$S \circ (xy^\top - \bar{x}\bar{y}^\top) = \begin{bmatrix} -\frac{1}{2}E & -\frac{3}{2}I \\ -\frac{3}{2}I & -\frac{1}{2}E \end{bmatrix}. \tag{40}$$

It can be easily verified that the first order optimality conditions of (1) are satisfied at (x, y):

$$(\mathcal{A}^\Omega)^\top(\mathcal{A}^\Omega(xy^\top) - b^\Omega)y = S \circ (xy^\top - \bar{x}\bar{y}^\top)y = \mathbf{0};$$
$$z^\top(\mathcal{A}^\Omega)^\top(\mathcal{A}^\Omega(xy^\top) - b^\Omega) = x^\top S \circ (xy^\top - \bar{x}\bar{y}^\top) = \mathbf{0}.$$

Hence (x, y) is a stationary point of (1). Next, we check the second-order optimality condition. For any $u = (u_1^\top, u_2^\top)^\top$ and $v = (v_1^\top, v_2^\top)^\top$, where $u_i \in \mathbb{R}^3$ and $v_i \in \mathbb{R}^3$ $(i = 1, 2)$, we obtain

$$
\begin{aligned}
h(u, v) &= \|S \circ (xv^\top + uy^\top)\|_F^2 + 2u^\top S \circ (xy^\top - \bar{x}\bar{y}^\top)v \\
&= \frac{1}{2}\|\mathbf{e}v_1^\top + u_1\mathbf{e}^\top\|_F^2 + \frac{1}{2}\|\mathbf{e}v_2^\top + u_2\mathbf{e}^\top\|_F^2 + \frac{1}{2}\|\mathrm{Diag}(\mathbf{e}v_2^\top - u_1\mathbf{e}^\top)\|_F^2 \\
&\quad + \frac{1}{2}\|\mathrm{Diag}(u_2\mathbf{e}^\top - \mathbf{e}v_1^\top)\|_F^2 - (u_1^\top E v_1 + u_2^\top E v_2 + 3u_1^\top v_2 + 3u_2^\top v_1) \\
&= 2(\|u_1 - v_2\|_2^2 + \|u_2 - v_1\|_2^2) \geq 0,
\end{aligned}
$$

where $\mathrm{Diag}(X)$ refers to the diagonal matrix consisting of the diagonal of the square matrix X. Hence, (x, y) is a second-order stationary point.

On the other hand, it follows from (40) that

$$
f(x, y) = \|\mathcal{A}^\Omega(x, y) - b^\Omega\|_2^2 = \|S \circ (xy^\top - \bar{x}\bar{y}^\top)\|_2^2 > 0 = f(\bar{x}, \bar{y}).
$$

Namely, (x, y) is not a global minimizer. Therefore, the SNIG condition is violated at $(\mathcal{A}^\Omega, b^\Omega, 1)$. We complete the proof. $\qquad\square$

Remark 3 It is not difficult to verify that the zero residual global optimizer $(\hat{x}, \hat{y})$ of problem (39) must be of the following form

$$
\hat{x} = a(\mathbf{e}^\top, \mathbf{e}^\top)^\top, \quad \hat{y} = \frac{1}{a}(\mathbf{e}^\top, \mathbf{e}^\top)^\top, \quad \forall a \neq 0.
$$

Therefore, this example implies that the exact recovery holds at $(\mathcal{A}^\Omega, b^\Omega, 1)$.

5 Conclusion

In this paper, we propose the conception of the SNIG condition, namely, the second-order necessary optimality condition implies the global optimality. We theoretically prove the SNIG holds in some classes of matrix factorization problem (1). Such results may help us to understand the structure of the matrix factorization problem and the performance of the algorithms for solving it. Moreover, we also illustrate that the SNIG condition does not always hold. We emphasize that the conception of the SNIG condition is not related with the exact recovery. Combining Remark 2 and Remark 3, we can conclude that the set of $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ satisfying the SNIG condition and the set of $(\mathcal{A}^{\tilde{\Omega}}, b^{\tilde{\Omega}}, k)$ satisfying the exact recovery do not contain each other.

References

1. Singh, Ajit P., Gordon, Geoffrey J.: A unified view of matrix factorization models. Springer, 358-373(2008).
2. Koren Y., Bell R., Volinsky C.: Matrix factorization techniques for recommender systems. IEEE, 30-37(2009).
3. Cai J., Candès E., Shen Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim., **20**, 1956-1982(2010).
4. Candès E., Recht B.: Exact matrix completion via convex optimization. Found. Comput. Math., **9**, 717-772(2008).
5. Candès E., Tao T.: The power of convex relaxation: near-optimal matrix completion. IEEE Trans. Inform. Theory **56**, 2053-2080(2009).
6. Chandrasekaran V., Sanghavi S., Parrilo P.A., Willsky A.S.: Rank-sparsity incoherence for matrix decomposition. SIAM J. Optim., **21**, 572-596(2011).
7. Chen C., He B., Yuan X.: Matrix completion via alternating direction methods. IMA J. Numer. Anal., **32**, 227-245(2012).
8. Takács G., Pilászy I., Németh B., Tikk D.: Major components of the gravity recommendation system. ACM SIGKDD Explorations Newsletter, **9**, 80-83(2007).
9. Candès E., Li X., Soltanolkotabi M.: Phase retrieval via Wirtinger flow: Theory and algorithms. IEEE Trans. on Inform. Theory, **61**, 1985-2007(2015).
10. Goldfarb D., Ma S.: Convergence of fixed-point continuation algorithms for matrix rank minimization. Found. Comput. Math., **11**, 183-210(2011).
11. Zhu Z., So A., Ye Y.: Fast and near–optimal matrix completion via randomized basis pursuit. Proceed. of the Fifth International Congress of Chinese Mathematicians, ser. AMS/IP Studi. Adv. Math., **51**, 859-882(2009).
12. Hardt M.: Understanding alternating minimization for matrix completion. arXiv:1312.0925.
13. Jain P., Netrapalli P., Sanghavi S.: Low-rank matrix completion using alternating minimization. In Proc. 45th Symposium on Theory of Computing, 665-674(2013).
14. Liu X., Wen Z., Zhang Y.: Limited memory block Krylov subspace optimization for computing dominant singular value decompositions. SIAM J. Sci. Comput., **35**, A1641-A1668(2013).
15. Liu X., Wen Z., Zhang Y.: An efficient Gauss-Newton algorithm for symmetric low-rank product matrix approximations. SIAM J. Optim., **25**, 15711608(2015).
16. Ma S., Goldfarb D., Chen L.: Fixed point and Bregman iterative methods for matrix rank minimization. Math. Program. Ser. A, **128**, 321-353(2011).
17. Mishra B., Meyer G., Bach F., Sepulchre R.: Low-rank optimization with trace norm penalty. SIAM J. Optim., **23**, 2124–2149(2013).
18. Journe M., Bach F., Absil P.-A., Sepulchre R.: Low-rank optimization on the cone of positive semidefinite matrices. SIAM J. Optim., **20**, 2327-2351(2010)
19. Burer S., Monteiro R.D.C: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Math. Program. Ser. B, **95**, 329–357(2013)
20. Sun R., Luo Z.: Guaranteed matrix completion via non-convex factorization. arXiv:1411.8003.
21. Lu Z., Zhang Y.: Penalty decomposition methods for rank minimization. under review.
22. Recht B., Fazel M., Parrilo A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review, **52**, 471-501(2010).
23. Oymak S., Mohan K., Fazel M., Hassibi B.: A simplified approach to recovery conditions for low rank matrices. IEEE International Symposium on Information Theory(2011).
24. Oymak S., Hassibi B.: New null space results and recovery thresholds for matrix rank minimization. IEEE International Symposium on Information Theory(2011).
25. Wang Z., Lai M., Lu Z., Ye J.: Orthogonal rank-one matrix pursuit for matrix completion. under review.
26. Wen Z., Yang C., Liu X., Zhang Y.: Trace-penalty minimization for large-scale eigenspace computation. under review at J. Sci. Comput..
27. Wen Z., Yin W., Zhang Y.: Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm, Math. Program. Comput., **4**, 333-361(2012).
28. Murty K., Kabadi S.: Some NP-complete problems in quadratic and nonlinear programming. Math. Program., **39**, 117-129(1987).