A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank

Hongyang Zhang, Zhouchen Lin*, and Chao Zhang

Key Lab. of Machine Perception (MOE), School of EECS
Peking University, Beijing, China
{hy_zh,zlin}@pku.edu.cn, chzhang@cis.pku.edu.cn

Abstract. Rank minimization has attracted a lot of attention due to its robustness in data recovery. To overcome the computational difficulty, rank is often replaced with nuclear norm. For several rank minimization problems, such a replacement has been theoretically proven to be valid, i.e., the solution to nuclear norm minimization problem is also the solution to rank minimization problem. Although it is easy to *believe* that such a replacement may not always be valid, no concrete example has ever been found. We argue that such a validity checking cannot be done by numerical computation and show, by analyzing the noiseless latent low rank representation (LatLRR) model, that even for very simple rank minimization problems the validity may still break down. As a by-product, we find that the solution to the nuclear norm minimization formulation of LatLRR is *non-unique*. Hence the results of LatLRR reported in the literature may be questionable.

1 Introduction

We are now in an era of big data as well as high dimensional data. Fortunately, high dimensional data are not unstructured. Usually, they lie near low dimensional manifolds. This is the basis of linear and nonlinear dimensionality reduction [1]. As a simple yet effective approximation, linear subspaces are usually adopted to model the data distribution. Because low dimensional subspaces correspond to low rank data matrices, rank minimization problem, which models the real problem into an optimization by minimizing the rank in the objective function (cf. models (1), (3) and (4)), is now widely used in machine learning and data recovery [2–5]. Actually, rank is regarded as a sparsity measure for matrices [3]. So low rank recovery problems are studied [6–9] in parallel with the compressed sensing theories for sparse vector recovery. Typical rank minimization problems include matrix completion [2, 4], which aims at completing the entire matrix from a small sample of its entries, robust principal component analysis [3], which recovers the ground truth data from sparsely corrupted elements, and low rank representation [10, 11], which finds an affinity matrix of subspaces that has the lowest rank. All of these techniques have found wide applications,

* Corresponding author.

H. Blockeel et al. (Eds.): ECML PKDD 2013, Part II, LNAI 8189, pp. 226–241, 2013.

such as background modeling [3], image repairing [12], image alignment [12], image rectification [13], motion segmentation [10, 11], image segmentation [14], and saliency detection [15].

Since the rank of a matrix is discrete, rank minimization problems are usually hard to solve. They can even be NP hard [3]. To overcome the computational obstacle, as a common practice people usually replace rank in the objective function with nuclear norm, which is the sum of singular values and is the convex envelope of rank on the unit ball of matrix operator norm [5], to transform rank minimization problems into nuclear norm minimization problems (cf. models (2) and (5)). Such a strategy is widely adopted in most rank minimization problems [2–4, 10–15]. However, this naturally brings a *replacement validity problem* which is defined as follows.

Definition 1 (Replacement Validity Problem). *Given a rank minimization problem together with its corresponding nuclear norm formulation, the replacement validity problem investigates whether the solution to the nuclear norm minimization problem is also a solution to the rank minimization one.*

In this paper, we focus on the replacement validity problem. There is a related problem, called *exact recovery problem*, that is more widely studied by scholars. It is defined as follows.

Definition 2 (Exact Recovery Problem). *Given a nuclear norm minimization problem, the exact recovery problem investigates the sufficient conditions under which the nuclear norm minimization problem could exactly recover the real structure of the data.*

As an example of the exact recovery problem, Candès et al. proved that when the rank of optimal solution is sufficiently low and the missing data is sufficiently few or the corruption is sufficiently sparse, solving nuclear norm minimization problems of matrix completion [2] or robust PCA problems [3] can exactly recover the ground truth low rank solution with an overwhelming probability. As another example, Liu et al. [10, 16] proved that when the rank of optimal solution is sufficiently low and the percentage of corruption does not exceed a threshold, solving the nuclear norm minimization problem of low rank representation (LRR) [10, 11] can exactly recover the ground truth subspaces of the data.

We want to highlight the difference between our replacement validity problem and the exact recovery problem that scholars have considered before. The replacement validity problem is to compare the solutions between two optimization problems, while the exact recovery problem is to study whether solving a nuclear norm minimization problem can exactly recover a ground truth low rank matrix. As a result, in all the existing exact recovery problems, the scholars have to assume that the rank of the ground truth solution is sufficiently low. In contrast, the replacement validity problem does not rely on this assumption: even if the ground truth low rank solution cannot be recovered, we can still investigate whether the solution to a nuclear norm minimization problem is also the solution to the corresponding rank minimization problem.

For replacement validity problems, it is easy to *believe* that the replacement of rank with nuclear norm will break down for complex rank minimization problems. While for exact recovery problems, the existing analysis all focuses on relatively simple rank minimization problems, such as matrix completion [2], robust PCA problems [3], and LRR [10, 11], and has achieved affirmative results under some conditions. So it is also easy to *believe* that for simple rank minimization problems the replacement of rank with nuclear norm will work. This paper aims at breaking such an illusion. Here, we have to point out that replacement validity problem *cannot* be studied by numerical experiments. This is because: 1. rank is sensitive to numerical errors. Without prior knowledge, one may not correctly determine the rank of a given matrix, even if there is a clear drop in its singular values; 2. it is hard to verify whether a given solution to nuclear norm minimization problem is a *global* minimizer to a rank minimization problem, whose objective function is discrete and non-convex. So we should study replacement validity problem by purely theoretical analysis. We analyze a simple rank minimization problem, noiseless latent LRR (LatLRR) [17], to show that solutions to a nuclear norm minimization problem may not be solutions of the corresponding rank minimization problem.

The contributions of this paper include:

1. We use a simple rank minimization problem, noiseless LatLRR, to prove that solutions to a nuclear norm minimization problem may not be solutions of the corresponding rank minimization problem, even for very simple rank minimization problems.
2. As a by-product, we find that LatLRR is not a good mathematical model because the solution to its nuclear norm minimization formulation is *non-unique*. So the results of LatLRR reported in the literature, e.g., [10, 17], may be questionable.

2 Latent Low Rank Representation

In this section, we first explain the notations that will be used in this paper and then introduce latent low rank representation which we will analyze its closed form solutions.

2.1 Summary of Main Notations

A large amount of matrix related symbols will be used in this paper. Capital letters are used to represent matrices. Especially, I denotes the identity matrix and 0 is the all-zero matrix. The entry at the ith row and the jth column of a matrix is denoted by $[\cdot]_{ij}$. Nuclear norm, the sum of all the singular values of a matrix, is denoted by $||\cdot||_*$. Operator norm, the maximum singular value, is denoted by $||\cdot||_2$. Trace(A) represents the sum of the diagonal entries of A and $A^\dagger$ is the Moore-Penrose pseudo-inverse of A. For simplicity, we use the same letter to present the subspace spanned by the columns of a matrix. The dimension of a space V is presented by $\dim(V)$. The orthogonal complement of V is denoted by $V_\perp$. Range(A) indicates the linear space spanned by all the columns of matrix A, while Null(A) represents the null space of A. They are closely related: $(\text{Range}(A))_\perp = \text{Null}(A^T)$. Finally, we always use $U_X \Sigma_X V_X^T$ to represent the *skinny* SVD of the data matrix X. Namely, the numbers of columns in U_X and V_X are both rank(X) and Σ_X consists of all the non-zero singular values of X, making Σ_X invertible.

2.2 Low Rank Subspace Clustering Models

Low rankness based subspace clustering stems from low rank representation (LRR) [10, 11]. An interested reader may refer to an excellent review on subspace clustering approaches provided by Vidal [18]. The mathematical model of the original LRR is

$$\min_Z \text{rank}(Z), \quad \text{s.t.} \quad X = XZ, \tag{1}$$

where X is the data matrix we observe. LRR extends sparse subspace clustering [19] by generalizing the sparsity from 1D to 2D. When there is noise or corruption, a noise term can be added to the model [10, 11]. Since this paper considers closed form solutions for noiseless models, to save space we omit the noisy model. The corresponding nuclear norm minimization formulation of (1) is

$$\min_Z ||Z||_*, \quad \text{s.t.} \quad X = XZ, \tag{2}$$

which we call the heuristic LRR. LRR has been very successful in clustering data into subspaces robustly [20]. It is proven that when the underlying subspaces are independent, the optimal representation matrix is block diagonal, each block corresponding to a subspace [10, 11].

LRR works well only when the samples are sufficient. This condition may not be fulfilled in practice, particularly when the dimension of samples is large. To resolve this issue, Liu et al. [17] proposed latent low rank representation (LatLRR). Another model to overcome this drawback of LRR is fixed rank representation [21]. LatLRR assumes that the observed samples can be expressed as the linear combinations of themselves together with the unobserved data:

$$\min_Z \text{rank}(Z), \quad \text{s.t.} \quad X = [X, X_H]Z, \tag{3}$$

where X_H is the unobserved samples for supplementing the shortage of the observed ones. Since X_H is unobserved and problem (3) cannot be solved directly, by some deduction and mathematical approximation, LatLRR [17] is modeled as follows:

$$\min_{Z,L} \text{rank}(Z) + \text{rank}(L), \quad \text{s.t.} \quad X = XZ + LX. \tag{4}$$

Both the optimal Z and L can be utilized for learning tasks: Z can be used for subspace clustering, while L is for feature extraction, thus providing us with the possibility for integrating two tasks into a unified framework. We call (4) the original LatLRR. Similarly, it has a nuclear norm minimization formulation

$$\min_{Z,L} ||Z||_* + ||L||_*, \quad \text{s.t.} \quad X = XZ + LX, \tag{5}$$

which we call the heuristic LatLRR. LatLRR has been reported to have better performance than LRR [10, 17].

In this paper, we focus on studying the solutions to problems (1), (2), (4) and (5), in order to investigate the replacement validity problem.

3 Analysis on LatLRR

This section provides surprising results: both the original and heuristic LatLRR have closed form solutions! We are able to write down *all* their solutions, as presented in the following theorems.

Theorem 1. *The complete solutions to the original LatLRR problem* (4) *are as follows*

$$Z^* = V_X \tilde{W} V_X^T + S_1 \tilde{W} V_X^T \text{ and } L^* = U_X \Sigma_X (I - \tilde{W}) \Sigma_X^{-1} U_X^T + U_X \Sigma_X (I - \tilde{W}) S_2, \tag{6}$$

where $\tilde{W}$ is any idempotent matrix and S_1 and S_2 are any matrices satisfying: 1. $V_X^T S_1 = 0$ and $S_2 U_X = 0$; and 2. $\text{rank}(S_1) \le \text{rank}(\tilde{W})$ and $\text{rank}(S_2) \le \text{rank}(I - \tilde{W})$.

Theorem 2. *The complete solutions to the heuristic LatLRR problem* (5) *are as follows*

$$Z^* = V_X \widehat{W} V_X^T \text{ and } L^* = U_X (I - \widehat{W}) U_X^T, \tag{7}$$

where $\widehat{W}$ is any block diagonal matrix satisfying: 1. its blocks are compatible with Σ_X, i.e., if $[\Sigma_X]_{ii} \neq [\Sigma_X]_{jj}$ then $[\widehat{W}]_{ij} = 0$; and 2. both $\widehat{W}$ and $I - \widehat{W}$ are positive semi-definite.

By Theorems 1 and 2, we can conclude that if the $\widehat{W}$ in Theorem 2 is not idempotent, then the corresponding (Z^*, L^*) is not the solution to the original LatLRR, due to the following proposition:

Proposition 1. *If the $\widehat{W}$ in Theorem 2 is not idempotent, then $Z^* = V_X \widehat{W} V_X^T$ cannot be written as $Z^* = V_X \tilde{W} V_X^T + S_1 \tilde{W} V_X^T$, where $\tilde{W}$ and S_1 satisfy the conditions stated in Theorem 1.*

The above results show that for noiseless LatLRR, nuclear norm is not a valid replacement of rank. As a by-product, since the solution to the heuristic LatLRR is *non-unique*, the results of LatLRR reported in [11, 17] may be questionable.

We provide detailed proofs of the above theorems and proposition in the following section.

4 Proofs

4.1 Proof of Theorem 1

We first provide the complete closed form solutions to the original LRR in a more general form

$$\min_{Z} \operatorname{rank}(Z), \quad \text{s.t.} \quad A = XZ, \tag{8}$$

where $A \in \operatorname{Range}(X)$ so that the constraint is feasible. We call (8) the generalized original LRR. Then we have the following proposition.

Proposition 2. *Suppose $U_A \Sigma_A V_A^T$ is the skinny SVD of A. Then the minimum objective function value of the generalized original LRR problem* (8) *is rank*(A) *and the complete solutions to* (8) *are as follows*

$$Z^* = X^\dagger A + S V_A^T, \tag{9}$$

where S is any matrix such that $V_X^T S = 0$.

Proof. Suppose Z^* is an optimal solution to problem (8). First, we have

$$\operatorname{rank}(A) = \operatorname{rank}(XZ^*) \le \operatorname{rank}(Z^*). \tag{10}$$

On the other hand, because $A = XZ$ is feasible, there exists Z_1 such that $A = XZ_1$. Then $Z_0 = X^\dagger A$ is feasible: $XZ_0 = XX^\dagger A = XX^\dagger XZ_1 = XZ_1 = A$, where we have utilized a property of Moore-Penrose pseudo-inverse $XX^\dagger X = X$. So we obtain

$$\operatorname{rank}(Z^*) \le \operatorname{rank}(Z_0) \le \operatorname{rank}(A). \tag{11}$$

Combining (10) with (11), we conclude that rank(A) is the minimum objective function value of problem (8).

Next, let $Z^* = PQ^T$ be the full rank decomposition of the optimal Z^*, where both P and Q have rank(A) columns. From $U_A \Sigma_A V_A^T = XPQ^T$, we have $V_A^T = (\Sigma_A^{-1} U_A^T X P) Q^T$. Since both V_A and Q are full column rank and $Y = \Sigma_A^{-1} U_A^T X P$ is square, Y must be invertible. So V_A and Q represent the same subspace. Because P and Q are unique up to an invertible matrix, we may simply choose $Q = V_A$. Thus $U_A \Sigma_A V_A^T = XPQ^T$ reduces to $U_A \Sigma_A = U_X \Sigma_X V_X^T P$, i.e., $V_X^T P = \Sigma_X^{-1} U_X^T U_A \Sigma_A$, and we conclude that the complete choices of P are given by $P = V_X \Sigma_X^{-1} U_X^T U_A \Sigma_A + S$, where S is any matrix such that $V_X^T S = 0$. Multiplying P with $Q^T = V_A^T$, we obtain that the entire solutions to problem (8) can be written as $Z^* = X^\dagger A + S V_A^T$, where S is any matrix satisfying $V_X^T S = 0$. □

Remark 1. Friedland and Torokhti [22] studied a similar model as (8), which is

$$\min_{Z} ||X - AZ||_F, \quad \text{s.t.} \quad \operatorname{rank}(Z) \le k. \tag{12}$$

However, (8) is different from (12) in two aspects. First, (8) requires the data matrix X to be strictly expressed as linear combinations of the columns in A. Second, (8) does not impose an upper bound for the rank of Z. Rather, (8) solves for the Z with the lowest rank. As a result, (8) has infinitely many solutions, as shown by Proposition 2, while (12) has a unique solution when k fulfills some conditions. So the results in [22] do not apply to (8).

Similar to Proposition 2, we can have the complete closed form solution to the following problem

$$\min_{Z} \operatorname{rank}(L), \quad \text{s.t.} \quad A = LX, \tag{13}$$

which will be used in the proof of Theorem 1.

Proposition 3. *Suppose $U_A \Sigma_A V_A^T$ is the skinny SVD of A. Then the minimum objective function value of problem* (13) *is rank*(A) *and the complete solutions to problem* (13) *are as follows*

$$L^* = AX^\dagger + U_A S, \tag{14}$$

where S is any matrix such that $SU_X = 0$.

Next, we provide the following propositions.

Proposition 4. *rank*(X) *is the minimum objective function value of the original LatLRR problem* (4).

Proof. Suppose (Z^*, L^*) is an optimal solution to problem (4). By Proposition 2 and fixing Z^*, we have $\operatorname{rank}(L^*) = \operatorname{rank}(X - XZ^*)$. Thus

$$\operatorname{rank}(Z^*) + \operatorname{rank}(L^*) \geq \operatorname{rank}(XZ^*) + \operatorname{rank}(X - XZ^*) \geq \operatorname{rank}(X). \tag{15}$$

On the other hand, if Z^* and L^* are adopted as $X^\dagger X$ and 0, respectively, the lower bound is achieved and the constraint is fulfilled as well. So we conclude that rank(X) is the minimum objective function value of the original LatLRR problem (4). □

Proposition 5. *Suppose (Z^*, L^*) is one of the solutions to problem* (4)*. Then there must exist another solution $(\widetilde{Z}, \widetilde{L})$, such that $XZ^* = X\widetilde{Z}$ and $\widetilde{Z} = V_X \tilde{W} V_X^T$ for some matrix $\tilde{W}$.*

Proof. According to the constraint of problem (4), we have $XZ = (I - L)X$, i.e., $(XZ)^T \in \operatorname{Range}(X^T)$. Since $V_X V_X^T$ is the projection matrix onto $\operatorname{Range}(X^T)$, we have

$$XZ^* V_X V_X^T = XZ^*. \tag{16}$$

On the other hand, given the optimal Z^*, L^* is the optimal solution to

$$\min_{L} \operatorname{rank}(L) \quad \text{s.t.} \quad X(I - Z^*) = LX. \tag{17}$$

So by Proposition 2 we get

$$\operatorname{rank}(L^*) = \operatorname{rank}(X(I - Z^*)X^\dagger). \tag{18}$$

As a result,

$$\begin{aligned} \operatorname{rank}(X) &= \operatorname{rank}(Z^*) + \operatorname{rank}(L^*) \\ &= \operatorname{rank}(Z^*) + \operatorname{rank}(X(I - Z^*)X^\dagger) \\ &= \operatorname{rank}(Z^*) + \operatorname{rank}(X(I - V_X V_X^T Z^* V_X V_X^T)X^\dagger) \\ &\geq \operatorname{rank}(V_X V_X^T Z^* V_X V_X^T) + \operatorname{rank}(X(I - V_X V_X^T Z^* V_X V_X^T)X^\dagger) \\ &\geq \operatorname{rank}(X), \end{aligned} \tag{19}$$

where the last inequality holds since $(V_X V_X^T Z^* V_X V_X^T, X(I - V_X V_X^T Z^* V_X V_X^T) X^\dagger)$ is a feasible solution to problem (4) and $\text{rank}(X)$ is the minimum objective according to Proposition 4. (19) shows that $(V_X V_X^T Z^* V_X V_X^T, X(I - V_X V_X^T Z^* V_X V_X^T) X^\dagger)$ is an optimal solution. So we may take $\widetilde{Z} = V_X V_X^T Z^* V_X V_X^T$ and write it as $\widetilde{Z} = V_X \tilde{W} V_X^T$, where $\tilde{W} = V_X^T Z^* V_X$.

Finally, combining with equation (16), we conclude that

$$X\widetilde{Z} = U_X \Sigma_X V_X^T V_X V_X^T Z^* V_X V_X^T = X Z^* V_X V_X^T = X Z^*. \tag{20}$$

□

Proposition 5 provides us with a great insight into the structure of problem (4): we may break (4) into two subproblems

$$\min_Z \text{rank}(Z), \quad \text{s.t.} \quad X V_X \tilde{W} V_X^T = XZ, \tag{21}$$

and

$$\min_L \text{rank}(L), \quad \text{s.t.} \quad X - X V_X \tilde{W} V_X^T = LX, \tag{22}$$

and then apply Propositions 2 and 3 to find the complete solutions to problem (4).

For investigating the properties of $\tilde{W}$ in (21) and (22), the following lemma is critical.

Lemma 1. *For $A, B \in \mathbb{R}^{n \times n}$, if $AB = BA$, then the following inequality holds*

$$\mathit{rank}(A + B) \leq \mathit{rank}(A) + \mathit{rank}(B) - \mathit{rank}(AB). \tag{23}$$

Proof. On the basis of $AB = BA$, it is easy to check that

$$\text{Null}(A) + \text{Null}(B) \subset \text{Null}(AB), \tag{24}$$

and

$$\text{Null}(A) \cap \text{Null}(B) \subset \text{Null}(A + B). \tag{25}$$

On the other hand, according to the well-known dimension formula

$$\dim(\text{Null}(A)) + \dim(\text{Null}(B)) = \dim(\text{Null}(A) + \text{Null}(B)) + \dim(\text{Null}(A) \cap \text{Null}(B)), \tag{26}$$

by combining (26) with (24) and (25), we get

$$\dim(\text{Null}(A)) + \dim(\text{Null}(B)) \leq \dim(\text{Null}(AB)) + \dim(\text{Null}(A + B)). \tag{27}$$

Then by the relationship $\text{rank}(S) = n - \dim(\text{Null}(S))$ for any $S \in \mathbb{R}^{n \times n}$, we arrive at the inequality (23). □

Based on the above lemma, the following proposition presents the sufficient and necessary condition on $\tilde{W}$.

Proposition 6. *Let L^* be any optimal solution to subproblem* (22)*, then* $(V_X\tilde{W}V_X^T, L^*)$ *is optimal to problem* (4) *if and only if the square matrix* $\tilde{W}$ *is idempotent.*

Proof. Obviously, $(V_X\tilde{W}V_X^T, L^*)$ is feasible based on the constraint in problem (22). By considering the optimality of L^* for (22) and replacing Z^* with $V_X\tilde{W}V_X^T$ in equation (18), we have

$$\mathrm{rank}(L^*) = \mathrm{rank}(X(I - V_X\tilde{W}V_X^T)X^\dagger). \tag{28}$$

First, we prove the sufficiency. According to the property of idempotent matrices, we have

$$\mathrm{rank}(\tilde{W}) = \mathrm{trace}(\tilde{W}) \text{ and } \mathrm{rank}(I - \tilde{W}) = \mathrm{trace}(I - \tilde{W}). \tag{29}$$

By substituting $(V_X\tilde{W}V_X^T, L^*)$ into the objective function, the following equalities hold

$$\begin{aligned}\mathrm{rank}(V_X\tilde{W}V_X^T) + \mathrm{rank}(L^*) &= \mathrm{rank}(\tilde{W}) + \mathrm{rank}(X(I - V_X\tilde{W}V_X^T)X^\dagger)\\ &= \mathrm{rank}(\tilde{W}) + \mathrm{rank}(U_X\Sigma_X(I - \tilde{W})\Sigma_X^{-1}U_X^T)\\ &= \mathrm{rank}(\tilde{W}) + \mathrm{rank}(I - \tilde{W})\\ &= \mathrm{trace}(\tilde{W}) + \mathrm{trace}(I - \tilde{W})\\ &= \mathrm{rank}(X).\end{aligned} \tag{30}$$

So $(V_X\tilde{W}V_X^T, L^*)$ is optimal since it achieves the minimum objective function value of problem (4).

Second, we prove the necessity. Suppose $(V_X\tilde{W}V_X^T, L^*)$ is optimal to problem (4). Substituting it into the objective follows

$$\begin{aligned}\mathrm{rank}(X) &= \mathrm{rank}(V_X\tilde{W}V_X^T) + \mathrm{rank}(X(I - V_X\tilde{W}V_X)X^\dagger)\\ &= \mathrm{rank}(\tilde{W}) + \mathrm{rank}(I - \tilde{W})\\ &\geq \mathrm{rank}(X).\end{aligned} \tag{31}$$

Hence $\mathrm{rank}(\tilde{W}) + \mathrm{rank}(I - \tilde{W}) = \mathrm{rank}(X)$. On the other hand, as $\tilde{W}$ and $I - \tilde{W}$ are commutative, by Lemma 1 we have $\mathrm{rank}(X) \leq \mathrm{rank}(\tilde{W}) + \mathrm{rank}(I - \tilde{W}) - \mathrm{rank}(\tilde{W} - \tilde{W}^2)$. So $\mathrm{rank}(\tilde{W} - \tilde{W}^2) = 0$ and thus $\tilde{W} = \tilde{W}^2$. □

We are now ready to prove Theorem 1.

Proof. Solving problems (21) and (22) by using Propositions 2 and 3, where $\tilde{W}$ is idempotent as Proposition 6 shows, we directly get

$$Z^* = V_X\tilde{W}V_X^T + \widetilde{S}_1V_A^T \text{ and } L^* = U_X\Sigma_X(I - \tilde{W})\Sigma_X^{-1}U_X^T + U_B\widetilde{S}_2, \tag{32}$$

where $U_A\Sigma_AV_A^T$ and $U_B\Sigma_BV_B^T$ are the skinny SVDs of $U_X\Sigma_X\tilde{W}V_X^T$ and $U_X\Sigma_X(I - \tilde{W})V_X^T$, respectively, and $\widetilde{S}_1$ and $\widetilde{S}_2$ are matrices such that $V_X^T\widetilde{S}_1 = 0$

and $\widetilde{S}_2 U_X = 0$. Since we have Range$((\tilde{W}V_X^T)^T) =$ Range(V_A) and Range$(U_X \Sigma_X (I - \tilde{W})) =$ Range(U_B), there exist full column rank matrices M_1 and M_2 satisfying $V_A = (\tilde{W}V_X^T)^T M_1$ and $U_B = U_X \Sigma_X (I - \tilde{W}) M_2$, respectively. The sizes of M_1 and M_2 are rank$(X) \times$ rank$(\tilde{W})$ and rank$(X) \times$ rank$(I - \tilde{W})$, respectively. We can easily see that a matrix S_1 can be decomposed into $S_1 = \widetilde{S}_1 M_1^T$, such that $V_X^T \widetilde{S}_1 = 0$ and M_1 is full column rank, if and only if $V_X^T S_1 = 0$ and rank$(S_1) \leq$ rank$(\tilde{W})$. Similarly, a matrix S_2 can be decomposed into $S_2 = M_2 \widetilde{S}_2$, such that $\widetilde{S}_2 U_X = 0$ and M_2 is full column rank, if and only if $S_2 U_X = 0$ and rank$(S_2) \leq$ rank$(I - \tilde{W})$. By substituting $V_A = (\tilde{W}V_X^T)^T M_1$, $U_B = U_X \Sigma_X (I - \tilde{W}) M_2$, $S_1 = \widetilde{S}_1 M_1^T$, and $S_2 = M_2 \widetilde{S}_2$ into (32), we obtain the conclusion of Theorem 1. □

4.2 Proof of Theorem 2

We first quote two results from [10].

Lemma 2. *Assume $X \neq 0$ and $A = XZ$ have feasible solution(s), i.e., $A \in$ Range(X). Then*

$$Z^* = X^\dagger A \tag{33}$$

is the unique minimizer to the generalized heuristic LRR problem:

$$\min_Z ||Z||_*, \quad s.t. \quad A = XZ. \tag{34}$$

Lemma 3. *For any four matrices B, C, D and F of compatible dimensions, we have the inequalities*

$$\left\| \begin{bmatrix} B & C \\ D & F \end{bmatrix} \right\|_* \geq ||B||_* + ||F||_* \text{ and } \left\| \begin{bmatrix} B & C \\ D & F \end{bmatrix} \right\|_* \geq ||B||_*, \tag{35}$$

where the second equality holds if and only if $C = 0$, $D = 0$, and $F = 0$.

Then we prove the following lemma.

Lemma 4. *For any square matrix $Y \in \mathbb{R}^{n \times n}$, we have $||Y||_* \geq$ trace(Y), where the equality holds if and only if Y is positive semi-definite.*

Proof. We prove by mathematical induction. When $n = 1$, the conclusion is clearly true. When $n = 2$, we may simply write down the singular values of Y to prove.

Now suppose for any square matrix $\widetilde{Y}$, whose size does not exceed $n - 1$, the inequality holds. Then for any matrix $Y \in \mathbb{R}^{n \times n}$, using Lemma 3, we get

$$\begin{aligned} ||Y||_* &= \left\| \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \right\|_* \\ &\geq ||Y_{11}||_* + ||Y_{22}||_* \\ &\geq \text{trace}(Y_{11}) + \text{trace}(Y_{22}) \\ &= \text{trace}(Y), \end{aligned} \tag{36}$$

where the second inequality holds due to the inductive assumption on the matrices Y_{11} and Y_{22}. So we always have $||Y||_* \geq \text{trace}(Y)$.

It is easy to check that any positive semi-definite matrix Y, it satisfies $||Y||_* = \text{trace}(Y)$. On the other hand, just following the above proof by choosing Y_{22} as 2×2 submatrices, we can easily get that $||Y||_* > \text{trace}(Y)$ strictly holds if $Y \in \mathbb{R}^{n\times n}$ is asymmetric. So if $||Y||_* = \text{trace}(Y)$, then Y must be symmetric. Then the singular values of Y are simply the absolute values of its eigenvalues. As trace(Y) equals the sum of all eigenvalues of Y, $||Y||_* = \text{trace}(Y)$ holds only if all the eigenvalues of Y are non-negative. □

Using Lemma 2, we may consider the following unconstrained problem

$$\min_Z f(Z) \triangleq ||Z||_* + ||X(I-Z)X^\dagger||_*, \tag{37}$$

which is transformed from (5) be eliminating L therein. Then we have the following result.

Proposition 7. *Unconstrained optimization problem* (37) *has a minimum objective function value rank*(X)*.*

Proof. Recall that the sub-differential of the nuclear norm of a matrix Z is [23]

$$\partial_Z ||Z||_* = \{U_Z V_Z^T + R | U_Z^T R = 0, RV_Z = 0, ||R||_2 \leq 1\}, \tag{38}$$

where $U_Z \Sigma_Z V_Z^T$ is the skinny SVD of the matrix Z. We prove that $Z^* = 1/2X^\dagger X$ is an optimal solution to (37). It is sufficient to show that

$$\begin{aligned} 0 \in \partial_Z f(Z^*) &= \partial_Z ||Z^*||_* + \partial_Z ||X(I-Z^*)X^\dagger||_* \\ &= \partial_Z ||Z^*||_* - X^T \partial_{X(I-Z)X^\dagger} ||X(I-Z^*)X^\dagger||_* (X^\dagger)^T. \end{aligned} \tag{39}$$

Notice that $X(I-Z^*)X^\dagger = U_X(1/2I)U_X^T$ is the skinny SVD of $X(I-Z^*)X^\dagger$ and $Z^* = V_X(1/2I)V_X^T$ is the skinny SVD of Z^*. So $\partial_Z f(Z^*)$ contains

$$V_X V_X^T - X^T(U_X U_X^T)(X^\dagger)^T = V_X V_X^T - V_X \Sigma_X U_X^T U_X U_X^T U_X \Sigma_X^{-1} V_X^T = 0. \tag{40}$$

Substituting $Z^* = 1/2X^\dagger X$ into (37), we get the minimum objective function value rank(X). □

Next, we have the form of the optimal solutions to (37) as follows.

Proposition 8. *The optimal solutions to the unconstrained optimization problem* (37) *can be written as* $Z^* = V_X \widehat{W} V_X^T$*.*

Proof. Let $(V_X)_\perp$ be the orthogonal complement of V_X. According to Proposition 7, rank(X) is the minimum objective function value of (37). Thus we get

$$\begin{aligned}
\text{rank}(X) &= ||Z^*||_* + ||X(I - Z^*)X^\dagger||_* \\
&= \left\| \begin{bmatrix} V_X^T \\ (V_X)_\perp^T \end{bmatrix} Z^* \left[V_X, (V_X)_\perp\right] \right\|_* + ||X(I - Z^*)X^\dagger||_* \\
&= \left\| \begin{bmatrix} V_X^T Z^* V_X & V_X^T Z^* (V_X)_\perp \\ (V_X)_\perp^T Z^* V_X & (V_X)_\perp^T Z^* (V_X)_\perp \end{bmatrix} \right\|_* + ||X(I - Z^*)X^\dagger||_* \\
&\geq ||V_X^T Z^* V_X||_* + ||U_X \Sigma_X V_X^T (I - Z^*) V_X \Sigma_X^{-1} U_X^T||_* \\
&= ||V_X V_X^T Z^* V_X V_X^T||_* + ||U_X \Sigma_X V_X^T (I - V_X V_X^T Z^* V_X V_X^T) V_X \Sigma_X^{-1} U_X^T||_* \\
&= ||V_X V_X^T Z^* V_X V_X^T||_* + ||X(I - V_X V_X^T Z^* V_X V_X^T)X^\dagger||_* \\
&\geq \text{rank}(X),
\end{aligned} \tag{41}$$

where the second inequality holds by viewing $Z = V_X V_X^T Z^* V_X V_X^T$ as a feasible solution to (37). Then all the inequalities in (41) must be equalities. By Lemma 3 we have

$$V_X^T Z^* (V_X)_\perp = (V_X)_\perp^T Z^* V_X = (V_X)_\perp^T Z^* (V_X)_\perp = 0. \tag{42}$$

That is to say

$$\begin{bmatrix} V_X^T \\ (V_X)_\perp^T \end{bmatrix} Z^* \left[V_X, (V_X)_\perp\right] = \begin{bmatrix} \widehat{W} & 0 \\ 0 & 0 \end{bmatrix}, \tag{43}$$

where $\widehat{W} = V_X^T Z^* V_X$. Hence the equality

$$Z^* = \left[V_X, (V_X)_\perp\right] \begin{bmatrix} \widehat{W} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_X^T \\ (V_X)_\perp^T \end{bmatrix} = V_X \widehat{W} V_X^T \tag{44}$$

holds. □

Based on all the above lemmas and propositions, the following proposition gives the whole closed form solutions to the unconstrained optimization problem (37). So the solution to problem (37) is non-unique.

Proposition 9. *The solutions to the unconstrained optimization problem* (37) *are $Z^* = V_X \widehat{W} V_X^T$, where $\widehat{W}$ satisfies: 1. it is block diagonal and its blocks are compatible with Σ_X*[1]*; 2. both $\widehat{W}$ and $I - \widehat{W}$ are positive semi-definite.*

Proof. First, we prove the sufficiency. Suppose $Z^* = V_X \widehat{W} V_X^T$ satisfies all the conditions in the theorem. Substitute it into the objective function, we have

$$\begin{aligned}
||Z^*||_* + ||X(I - Z^*)X^\dagger||_* &= ||\widehat{W}||_* + ||\Sigma_X (I - \widehat{W}) \Sigma_X^{-1}||_* \\
&= ||\widehat{W}||_* + \text{trace}(\Sigma_X (I - \widehat{W}) \Sigma_X^{-1}) \\
&= ||\widehat{W}||_* + \text{trace}(I - \widehat{W}) \\
&= ||\widehat{W}||_* + \text{rank}(X) - \text{trace}(\widehat{W}) \\
&= \text{rank}(X) \\
&= \min_Z ||Z||_* + ||X(I - Z)X^\dagger||_*,
\end{aligned} \tag{45}$$

[1] Please refer to Theorem 2 for the meaning of "compatible with Σ_X."

where based on Lemma 4 the second and the fifth equalities hold since $I - \widehat{W} = \Sigma_X(I - \widehat{W})\Sigma_X^{-1}$ as $\widehat{W}$ is block diagonal and both $I - \widehat{W}$ and $\widehat{W}$ are positive semi-definite.

Next, we give the proof of the necessity. Let Z^* represent a minimizer. According to Proposition 8, Z^* could be written as $Z^* = V_X\widehat{W}V_X^T$. We will show that $\widehat{W}$ satisfies the stated conditions. Based on Lemma 4, we have

$$\begin{aligned} \operatorname{rank}(X) &= ||Z^*||_* + ||X(I - Z^*)X^\dagger||_* \\ &= ||\widehat{W}||_* + ||\Sigma_X(I - \widehat{W})\Sigma_X^{-1}||_* \\ &\geq ||\widehat{W}||_* + \operatorname{trace}(\Sigma_X(I - \widehat{W})\Sigma_X^{-1}) \\ &= ||\widehat{W}||_* + \operatorname{trace}(I - \widehat{W}) \\ &= ||\widehat{W}||_* + \operatorname{rank}(X) - \operatorname{trace}(\widehat{W}) \\ &\geq \operatorname{rank}(X). \end{aligned} \tag{46}$$

Thus all the inequalities above must be equalities. From the last equality and Lemma 4, we directly get that $\widehat{W}$ is positive semi-definite. By the first inequality and Lemma 4, we know that $\Sigma_X(I - \widehat{W})\Sigma_X^{-1}$ is symmetric, i.e.,

$$\frac{\sigma_i}{\sigma_j}[I - \widehat{W}]_{ij} = \frac{\sigma_j}{\sigma_i}[I - \widehat{W}]_{ij}, \tag{47}$$

where σ_i represents the ith entry on the diagonal of Σ_X. Thus if $\sigma_i \neq \sigma_j$, then $[I - \widehat{W}]_{ij} = 0$, i.e., $\widehat{W}$ is block diagonal and its blocks are compatible with Σ_X. Notice that $I - \widehat{W} = \Sigma_X(I - \widehat{W})\Sigma_X^{-1}$. By Lemma 4, we get that $I - \widehat{W}$ is also positive semi-definite. Hence the proof is completed. □

Now we can prove Theorem 2.

Proof. Let $\widehat{W}$ satisfy all the conditions in the theorem. According to Proposition 8, since the row space of $Z^* = V_X\widehat{W}V_X^T$ belongs to that of X, it is obvious that $(Z^*, X(I - Z^*)X^\dagger)$ is feasible to problem (5). Now suppose that (5) has a better solution $(\widetilde{Z}, \widetilde{L})$ than (Z^*, L^*), i.e.,

$$X = X\widetilde{Z} + \widetilde{L}X, \tag{48}$$

and

$$||\widetilde{Z}||_* + ||\widetilde{L}||_* < ||Z^*||_* + ||L^*||_*. \tag{49}$$

Fixing Z in (5) and by Lemma 2, we have

$$||\widetilde{Z}||_* + ||(X - X\widetilde{Z})X^\dagger||_* \leq ||\widetilde{Z}||_* + ||\widetilde{L}||. \tag{50}$$

Thus

$$||\widetilde{Z}||_* + ||(X - X\widetilde{Z})X^\dagger||_* < ||Z^*||_* + ||X(I - Z^*)X^\dagger||_*. \tag{51}$$

So we obtain a contradiction with respect to the optimality of Z^* in Proposition 9, hence proving the theorem. □

4.3 Proof of Proposition 1

Proof. Suppose the optimal formulation $Z^* = V_X \widehat{W} V_X^T$ in Theorem 2 could be written as $Z^* = V_X \tilde{W} V_X^T + S_1 \tilde{W} V_X^T$, where $\tilde{W}$ is idempotent and S_1 satisfies $\tilde{W} V_X^T S_1 = 0$. Then we have

$$V_X \widehat{W} V_X^T = V_X \tilde{W} V_X^T + S_1 \tilde{W} V_X^T. \tag{52}$$

By multiplying both sides with V_X^T and V_X on the left and right, respectively, we get

$$\widehat{W} = \tilde{W} + V_X^T S_1 \tilde{W}. \tag{53}$$

As a result, $\widehat{W}$ is idempotent:

$$\begin{aligned} \widehat{W}^2 &= (\tilde{W} + V_X^T S_1 \tilde{W})(\tilde{W} + V_X^T S_1 \tilde{W}) \\ &= \tilde{W}^2 + V_X^T S_1 \tilde{W}^2 + \tilde{W} V_X^T S_1 \tilde{W} + V_X^T S_1 \tilde{W} V_X^T S_1 \tilde{W} \\ &= \tilde{W} + V_X^T S_1 \tilde{W} + \tilde{W} V_X^T S_1 \tilde{W} + V_X^T S_1 \tilde{W} V_X^T S_1 \tilde{W} \\ &= \tilde{W} + V_X^T S_1 \tilde{W} = \widehat{W}, \end{aligned} \tag{54}$$

which is contradictory to the assumption. □

5 Conclusions

Based on the expositions in Section 3 and the proofs in Section 4, we conclude that even for rank minimization problems as simple as noiseless LatLRR, replacing rank with nuclear norm is not valid. We have also found that LatLRR is actually problematic because the solution to its nuclear norm minimization formation is not unique. We can also have the following interesting connections between LRR and LatLRR. Namely, LatLRR is indeed an extension of LRR because its solution set strictly includes that of LRR, no matter for the rank minimization problem or the nuclear norm minimization formulation. So we can summarize their relationship as Figure 1.

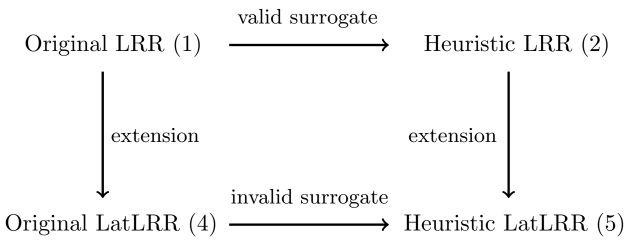

Fig. 1. The detailed relationship among the original LRR (1), the heuristic LRR (2), the original LatLRR (4), and the heuristic LatLRR (5) in the sense of their solution sets

Although the existing formulation of LatLRR is imperfect, since some scholars have demonstrated its effectiveness in subspace clustering by using a solution which is randomly chosen in some sense, in the future we will consider how to choose the best solution in the solution set in order to further improve the performance of LatLRR.

Acknowledgments. Hongyang Zhang and Chao Zhang are supported by National Key Basic Research Project of China (973 Program) 2011CB302400 and National Nature Science Foundation of China (NSFC Grant, no. 61071156). Zhouchen Lin is supported by National Nature Science Foundation of China (Grant nos. 61272341, 61231002, and 61121002).

References

1. Wang, J.: Geometric Structure of High-Dimensional Data and Dimensionality Reduction. Springer (2012)
2. Candès, E., Recht, B.: Exact matrix completion via convex optimization. Foundations of Computational Mathematics 9(6), 717–772 (2009)
3. Candès, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM 58(3), 1–37 (2009)
4. Candès, E.: Matrix completion with noise. Proceedings of the IEEE 98(6), 925–936 (2010)
5. Fazel, M.: Matrix Rank Minimization with Applications. PhD thesis, Standford University (2002), http://search.proquest.com/docview/305537461
6. Gross, D.: Recovering low-rank matrices from few coefficients in any basis. IEEE Transactions on Information Theory 57(3), 1548–1566 (2011)
7. Wright, J., Ganesh, A., Min, K., Ma, Y.: Compressive principal component pursuit. In: IEEE International Symposium on Information Theory Proceedings, pp. 1276–1280 (2012)
8. Waters, A.E., Sankaranarayanan, A.C., Baraniuk, R.G.: SpaRCS: Recovering low-rank and sparse matrices from compressive measurements. In: Advances in Neural Information Processing Systems, pp. 1089–1097 (2011)
9. Liu, Y.K.: Universal low-rank matrix recovery from pauli measurements. In: Advances in Neural Information Processing Systems, pp. 1638–1646 (2011)
10. Liu, G., Lin, Z., Yan, S., Sun, J., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(1), 171–184 (2013)
11. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: International Conference on Machine Learning, vol. 3, pp. 663–670 (2010)
12. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 34, pp. 2233–2246 (2010)
13. Zhang, Z., Ganesh, A., Liang, X., Ma, Y.: TILT: Transform invariant low-rank textures. International Journal of Computer Vision 99(1), 1–24 (2012)
14. Cheng, B., Liu, G., Huang, Z., Yan, S.: Multi-task low-rank affinities pursuit for image segmentation. In: IEEE International Conference on Computer Vision, pp. 2439–2446 (2011)

15. Lang, C., Liu, G., Yu, J., Yan, S.: Saliency detection by multi-task sparsity pursuit. IEEE Transactions on Image Processing 21(3), 1327–1338 (2012)
16. Liu, G., Xu, H., Yan, S.: Exact subspace segmentation and outlier detection by low-rank representation. In: International Conference on Artificial Intelligence and Statistics (2012)
17. Liu, G., Yan, S.: Latent low-rank representation for subspace segmentation and feature extraction. In: IEEE International Conference on Computer Vision, pp. 1615–1622 (2011)
18. Vidal, R.: Subspace clustering. IEEE Signal Processing Magazine 28(2), 52–68 (2011)
19. Elhamifar, E., Vidal, R.: Sparse subspace clustering. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2790–2797 (2009)
20. Adler, A., Elad, M., Hel-Or, Y.: Probabilistic subspace clustering via sparse representations. IEEE Signal Processing Letters 20(1), 63–66 (2013)
21. Liu, R., Lin, Z., Torre, F.D.L., Su, Z.: Fixed-rank representation for unsupervised visual learning. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 598–605 (2012)
22. Friedland, S., Torokhti, A.: Generalized rank-constrained matrix approximations. SIAM Journal on Matrix Analysis and Applications 29(2), 656–659 (2007)
23. Cai, J., Candès, E., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM Journal of Optimization 20(4), 1956–1982 (2010)