A Comprehensive Survey for Low Rank Regularization

Zhanxuan Hu, Feiping Nie, Lai Tian, Xuelong Li, Fellow, IEEE

Abstract—Low rank regularization, in essence, involves introducing a low rank or approximately low rank assumption for matrix we aim to learn, which has achieved great success in many fields including machine learning, data mining and computer vision. Over the last decade, much progress has been made in theories and practical applications. Nevertheless, the intersection between them is very slight. In order to construct a bridge between practical applications and theoretical research, in this paper we provide a comprehensive survey for low rank regularization. We first review several traditional machine learning models using low rank regularization, and then show their (or their variants) applications in solving practical issues, such as non-rigid structure from motion and image denoising. Subsequently, we summarize the regularizers and optimization methods that achieve great success in traditional machine learning tasks but are rarely seen in solving practical issues. Finally, we provide a discussion and comparison for some representative regularizers including convex and non-convex relaxations. Extensive experimental results demonstrate that non-convex regularizers can provide a large advantage over the nuclear norm, the regularizer widely used in solving practical issues.

Index Terms—Low rank learning, non-convex relaxation, regularization, optimization, computer vision.

1 INTRODUCTION

Over the last decade, low rank regularization has attracted much attention due to its success in various fields, ranging from machine learning, computer vision, data mining to deep learning. For all of them, the fundamental assumption is that the matrix we aim to learn lie near some low-dimensional subspaces. Generally, the algorithms using low rank regularization are based on the following formulation (or its variants):

\[ \min_X \mathcal{L}(X) + \lambda \text{rank}(X) \]

\[ \text{s.t. } X \in C, \quad (1) \]

where \( \mathcal{L}(X) \) represents the loss term, rank \( X \) denotes the rank of \( X \), \( C \) represents the constraints over \( X \), and \( \lambda \) is a regularization parameter. Optimizing the problem (1) inevitably involves solving a rank minimization problem. Matrix rank minimization, a pioneer of low rank regularization, is known to be an NP-hard problem. The most widely used method is looking for a heuristic solving the original problem approximately but efficiently [1]. Correspondingly, we can obtain an approximation of problem (1) as follows:

\[ \min_X \mathcal{L}(X) + \lambda \mathcal{R}(X) \]

\[ \text{s.t. } X \in C, \quad (2) \]

where \( \mathcal{R}(X) \) represents the relaxation that we use to replace the original rank minimization problem, and the nuclear norm, i.e., trace norm is the most widely used. The nuclear norm is proposed by [1] and highlighted by [2, 3]. Recently, much progress over formulation (2) and its variants has been made. Indeed, most of them along the following three directions.

1) Use the formulation (2) or its variants to solve some particular tasks, such as machine learning tasks including Matrix Completion [3–8], Subspace Clustering [9–14], and Multi-Task Learning [15–19], and computer vision tasks including Visual Tracking [20, 21], 3D Reconstruction [22–25], Salient Object Detection [26–29], and so on.

2) Find a better approximation, i.e., a better regularizer \( \mathcal{R}(X) \) for original rank minimization problem. For instance, the TNN (Truncated Nuclear Norm) [30], WNN (Weighted Nuclear Norm) [31–33], Schatten-p norm [34–38], CNN (Capped Nuclear Norm) [39], and so on.

3) Develop efficient optimization methods for solving the problem (2) and its variants, such as the IRW (Iteratively Re-weighted method) [40, 41], ALM (Augmented Lagrangian Method) [42], ADMM (Alternating Direction Method of Multipliers) [43], APG (Accelerated Proximal Gradient) method [44], Frank-Wolfe Algorithm [45], and so on.

For each of three directions presented above, a great many of efforts have been made. But the intersection between the three groups is very slight. As presented in the rest of this paper, the nuclear norm is generally selected to serve as regularizer in solving practical issues due to its convexity, advantages in optimization and sound theoretical guarantee. Besides, ADMM method is generally used to solve the corresponding models due to its advantages in tackling complicated problems. In practice, however, the superiority of non-convex regularizers over nuclear norm has been verified.
in many traditional machine learning models, such as matrix completion [30, 34, 39], and robust PCA [37, 39]. Besides, numerous efficient optimization approaches over relaxed matrix rank minimization problems have been proposed, such as APG. Both of these two techniques are rarely seen in solving practical issues. Hence, taking a comprehensive survey for both theories and applications over low rank regularization is urgent.

Note that some investigations have been made for low rank matrix learning. For instance, a survey for low-rank matrix learning and its applications in image analysis has been provided in [50]. Furthermore, the low rank matrix learning in the visual analysis has been summarized in [51, 52]. A discussion for optimization algorithms used in RPCA or its variants has been made in [53]. Nevertheless, both of them revolve around the nuclear norm regularization or matrix factorization method, and pay little attention to the non-convex regularizers, which is the core of this paper.

To construct a hub for low rank regularization, in this paper we summarize the main process over it along three directions mentioned above. In particular, considering that low rank regularizer is one of the most significant factors influencing the performance of an algorithm, we provide a comprehensive comparison over several representative including convex and non-convex relaxations. Extensive experimental results demonstrate that non-convex regularizers can provide a large advantage over nuclear norm. Such a result may be useful for promoting the application of non-convex regularizers in solving practical issues.

The rest of this paper is organized as follows. In Section 2, we first review several traditional machine learning models using low rank regularization, and then provide a summarization for practical issues solved by them or their variants. Various regularizers for replacing the rank minimization problem are summarized in Section 3. Then, in Section 4, we discuss the optimization methods widely used in previous algorithms, and analysis their advantages and limitations. The comparison results over different regularizers are reported in Section 5. In Section 6, we conclude this paper via a short discussion for future work.

**Notations.** The notations used throughout this paper are summarized in Table I.

### Table 1: The summarization of notations

| Notation | Description |
|----------|-------------|
| \( X \in R^{m \times n} \) | Matrix with size \( m \times n \) |
| \( v_i(X) \) | i\(^{th}\) right singular vector of \( X \) |
| \( \sigma_i(X) \) | i\(^{th}\) singular value of \( X \) |
| \( u_i(X) \) | i\(^{th}\) left singular vector of \( X \) |
| \( T_r(X) \) | The trace of matrix \( X \) |
| \( x_i \in R_+^n \) | i\(^{th}\) row or column of \( X \) |
| \( \|X\|_\cdot \) | Nuclear norm of \( X \) |
| \( \|X\|_F \) | Frobenius norm of \( X \) |
| \( \|X\|_{\ell_1} \) | \( \ell_1 \) norm of \( X \) |
| \( \|X\|_{\ell_2,1} \) | \( \ell_2,1 \) norm of \( X \) |

Furthermore, these machine learning models can be roughly grouped into the following two categories according to the type of desired matrix. First, the matrix we aim to learn is a data matrix, such as Robust Principal Component Analysis (RPCA) and Robust Matrix Completion (RMC). Second, the matrix we aim to learn is a coefficient matrix, such as Subspace Clustering (SC) and Multi-Task Learning (MTL). In this section, we first describe the details of these four representative models in Sect. 2.1. And then, we show several practical issues solved by them or their variants in Sect. 2.2.

### 2.1 Representative Machine Learning Models based on Low Rank Regularization

#### 2.1.1 Robust Principal Component Analysis

In most cases, the fundamental assumption for using low rank regularization is that the data we collected lie near some low-dimensional subspaces. For instance, the genomics data in bioinformatics, users’ records (such as ratings for movies) in recommender systems, and images in computer vision [50]. In the real world, however, the data are generally corrupted by noise and outliers. Robust Principal Component Analysis (RPCA) [2] is one of the most significant tools for recovering a low-rank matrix robustly from noisy observations, and it opens the door of using low rank regularization. Mathematically, the problem of RPCA can be solved by using the following formulation:

\[
\min_X \|X - M\|_\ell + \lambda R(X),
\]

where \( M \) is the noisy observation matrix, \( \| \cdot \|_\ell \) denotes the loss function which depends on the assumption over noise distribution, and \( R(X) \) denotes the regularization term as mentioned above. In addition to the Eq. (3), a popular formulation for RPCA is as follows:

\[
\min_{X, E} \|E\|_\ell + \lambda R(X)
\]

\[
s.t. \quad M = X + E,
\]

where \( E \) is the noise matrix. Using such a model, we can learn the low rank matrix \( X \) and the noise matrix \( E \) simultaneously.

#### 2.1.2 Robust Matrix Completion

Robust Matrix Completion (RMC) is one of the most important variants of RPCA, which considers a general case that some entries of input data matrix \( M \) are unknown, and the known entries are corrupted by noise. The goal of RMC is utilizing the known information to estimate the values of missing entries. The basic assumption used by RMC is that the complete matrix we aim to recover is low rank or approximately low rank. Correspondingly, the problem of RMC can be solved by using the following formulation:

\[
\min_X \|P_M(X - M)\|_\ell + \lambda R(X)
\]

where \( P_M \) represents a projecting operator, \( M \) represents a set recording the indices of known entries. The entries of matrix \( P_M(X) \) are consistent with \( X \) on \( M \) and are 0 on residuals. Similarly, by introducing a new variable \( E \), the Eq. (5) can be recast as:

\[
\min_{X, E} \|P_M(E)\|_\ell + \lambda R(X)
\]

\[
s.t. \quad M = X + E.
\]
### 2.1.3 Multi-Task Learning

Both RPCA and RMC are based on the assumption that the data matrix we aim to learn is low rank or approximate low rank. That is, the data we collected is relevant. The relatedness among different samples further inspires researchers to explore the relatedness among different tasks. Given \( K \) relevant tasks \( \{ T_i \}_{i=1}^{K} \) (regression or classification) accompanied by feature matrix \( \{ X_i \}_{i=1}^{K} \in \mathbb{R}^{n_i \times d} \) and target vectors \( \{ y_i \}_{i=1}^{K} \in \mathbb{R}^{n_i \times 1} \), we can learn them simultaneously to improve the generalization performance of each task. Such a problem refers to Multi-Task Learning (MTL). Suppose \( W \in \mathbb{R}^{d \times K} \) is a weight matrix we aim to learn, where \( w_i \) is the weight vector for task \( T_i \). The relatedness among \( K \) tasks imply that the structure of \( W \) is low rank or approximate low rank [54]. Hence, a general model based on low rank regularization for MTL problems is:

\[
\min_{W} \sum_{i=1}^{K} \mathcal{L}_i(X_i, w_i; y_i) + \lambda \mathcal{R}(W),
\]

where \( \mathcal{L}_i(\cdot) \) denotes the loss function used in \( i \)th task. A particular case of MTL is multivariate regression [53], where all tasks share a common feature matrix. Under such a case, the problem \( (7) \) can be reformulated as:

\[
\min_{W} \| XW - Y \|_F + \lambda \mathcal{R}(W).
\]

#### 2.1.4 Subspace Clustering

Given a set of data points approximately drawn from a union of multiple subspaces, the goal of Subspace Clustering (SC) is partitioning the data points into their respective subspaces. To this end, [10] proposed a Low-Rank Representation (LRR) model, which seeks a low rank matrix \( Z \) consisting of the candidates of data points in a given dictionary \( D \). To obtain the matrix \( Z \), we need to solve a model as follows:

\[
\min_{Z} \| DZ - X \|_F + \lambda \mathcal{R}(Z).
\]

Here, the low rank matrix \( Z \) can be seen as a rough similarity matrix, and the final partitioning result can be obtained by conducting spectral clustering with a refined similarity matrix \( S = \frac{|Z| + |Z^T|}{2} \).

#### 2.1.5 Summarization

For each model mentioned above, various variants have been developed. For instance, RPCA [32, 33, 56], MC [3, 8, 57, 58], MTL [15, 19, 54, 59, 60], and SC [9, 14]. Nevertheless, the main differences between them are loss function and regularization term. A short discussion for loss function widely used in machine learning can be found in [61]. The details with respect to the regularization term will be discussed in next section.

Furthermore, it is worth noting that in addition to the models mentioned above, the low rank regularization has achieved success on other fields including Component Analysis [62, 63], Compressive Sensing [64], Multi-View Learning [65, 69], Self-taught [69], Transfer Learning [70, 71], Spectral Clustering [72], Metric Learning [73], and so on. Indeed, most of them including LRR and MTL are derived from RPCA and MC. For instance, the capped norm used in Multi-Stage Multi-Task Learning [15] is proposed by [39] for dealing with the RPCA problem. Besides, in [74, 75], the weight matrix was considered as a sum of a low rank matrix \( L \) and sparse matrix \( S \) (or group sparse matrix \( C \)), which is also derived from RPCA. In [72], the authors deal with the problem of clustering via a partially observed unweighted graph, where the constructed model is a variant of MC. Recently, much progress have been made for RPCA and MC in terms of low rank relaxation and optimization method, which may be useful for further improving the performance of existing algorithms. The details will be discussed in Section 3 and Section 4.

### 2.2 Practical Issues Solved by Using Low Rank Regularization

We provide a summarization for practical issues solved by using low rank regularization in Table 2. Next, we describe the details of some representatives among them.

#### 2.2.1 Background Subtraction

Video background subtraction is one of the most well known topics using low rank regularization. Suppose \( X \) is a data matrix, and its each column is a vectorized image. As illustrated in [2], the background component can be recovered by recovering a low rank matrix \( L \), and the foreground component, i.e., the moving object can be detected by learning a sparse matrix \( E \). In [2], such a problem was solved by solving a PCP model as follows:

\[
\min_{L, E} \| E \|_1 + \lambda \| L \|_*,
\]

\[
s.t. \quad X = L + E.
\]

where the \( \| E \|_1 \) is the the convex surrogate of \( \| E \|_0 \). Obviously, it can be seen as a RPCA model [3]. Subsequently, in order to achieve better splitting results, numerous improvements for Eq. (10) have been made. For instance, to cope with the nonrigid motion and dynamic background, a DECOLOR method (DEtecting Contiguous Outliers in the LOw-rank Representation) is proposed in [127]. In [32], the nuclear norm is replaced by non-convex regularizers. Recently, in order to reduce the effect caused by various illumination changes, Shakeri et al. [92] construct a low-rank and invariant sparse decomposition model. Serving as one of the indices evaluating the performance of RPCA, background subtraction has received a great many of attentions, and a comprehensive survey for it can be found in [128].

#### 2.2.2 Image Denoising

Image denoising, in essence, involves estimating the latent clean image from an noisy image, which is a fundamental problem in low level vision. The success of several state-of-the-art image denoising algorithms, including BMSD [129], LSSC [130], and NCSR [131], is based on the exploitation of image nonlocal self-similarity, which refers to the assumption that for each local patch in a natural image, one can find some similar patches to it. Intuitively, by vectoring all similar patches as column vectors and stacking them as a matrix, one can obtain a matrix with low rank structure [132].

1. A background subtraction library: https://github.com/andrewssobral/lrslibrary
Using low rank assumption, Gu et al. \cite{31} propose a low rank approximation model to tackle the problem of image denoising. The proposed model based on weighted nuclear norm is as follows:

\[
\min_X \| X - M \|_F^2 + \lambda \| X \|_{w,*} 
\]

(11)

where \( M \) refers to the noisy matrix stacked by all vectorized similar patches, \( \| X \|_{w,*} = \sum_i \| w_i \sigma_i \| \) is a non-convex relaxation for rank minimization. The Eq. (11) is also a variant of Eq. (3). Similar models with different regularizers have been proposed in \cite{64, 83, 133}, and a different model with similar regularizer has been developed in \cite{134}. Recently, the researchers are focusing on the images with particular noise structure. For instance, the images with line pattern noise \cite{52}, and the color image with different noise statistics in \( R, G, \) and \( B \) channels \cite{81}.

### 2.2.3 Recommender System

Recommender system is one of the most effective applications of matrix completion. For a recommender system, the ratings submitted by users on items are generally limited, that is some entries of the rating matrix are unknown. In order to predicate the preferences of users, one would like to recover this rating matrix. Indeed, the complete rating matrix generally has an approximately low rank structure, because the individual’s preferences are generally effected by a few factors \cite{57}. Hence, the problem of predicting ratings is equivalent to matrix completion. Similar to background subtraction, serving as one of the indices evaluating the performance of a MC algorithms, recommender system has received a great many of attention, and a comprehensive survey for it can be found in \cite{128}.

In addition to recommender system, the matrix completion model has been widely used to deal with the image restoration problems, where some pixels of a image with low rank or approximate low rank structure are missing \cite{30}.

#### 2.2.4 Image Alignment

Image Alignment (IA) involves transforming various images into a common coordinate system. Stacking all transformed images as a matrix, the matrix may be a sum of a approximately low rank matrix \( L \), corresponding to aligned images, and sparse matrix \( E \), corresponding to noise or the differences among images. In particular, the low rank component and noise component can be split via a RPCA model. In order to learn the transformation and aligned images simultaneously, Peng et al. \cite{101} introduce a transformation \( \tau \) into original RPCA model and construct a novel model for IA problems:

\[
\min_{L, E, \tau} \| L \|_1 + \lambda \| E \|_* \text{ s.t. } X \circ \tau = L + E. 
\]

(12)

where \( \tau \) denotes the transformation. Then, the method was improved by \cite{99, 110, 135}. In practice, however, both of
them select the nuclear norm to serve as regularizer and the ADMM method to optimize the corresponding models.

Besides, the Eq. (13) has been used to generate low-rank textures with transform-invariant property [116].

### 2.2.5 Multi-Image Matching

Image matching refers to finding feature correspondences between two different images, which plays a significant role in various applications such as image registration, SFM (structure from motion). The recent work mainly focuses on using the cycle consistency as a constraint to match multiple images simultaneously. In [27], the authors show that the globally consistent matches \( X \) can be considered as a product of two low rank matrices, i.e., \( X = AA^T \). Given affinity scores \( \{S_{ij}\}_{1 \leq i, j \leq n} \), one can learn the consistent matches \( X \) by solving the following problem:

\[
\begin{align*}
\min_X & \quad \|X\|_* - \lambda(S - \alpha I, X) \\
\text{s.t.} & \quad X \in \mathcal{C}
\end{align*}
\]  
(13)

where \( \mathcal{C} \) represents the constraints over \( X \). Furthermore, utilizing the following equation [136]:

\[
\|X\|_* = \min_{A,B:AB^T = X} \frac{1}{2} (\|A\|_F^2 + \|B\|_F^2)
\]  
(14)

one can recast the Eq. (15) as:

\[
\begin{align*}
\min_{A,B} & \quad \frac{1}{2} (\|A\|_F^2 + \|B\|_F^2) - \lambda(S - \alpha I, X) \\
\text{s.t.} & \quad AB^T \in \mathcal{C}.
\end{align*}
\]  
(15)

Comparing the problem (13), problem (15) can be solved efficiently. Recently, the similar ideas were used to solve the graph matching (137) as well as multi-object matching problems (138).

### 2.2.6 Non-rigid Structure From Motion

Using low rank regularization to cope with the Non-rigid Structure From Motion (NSFM) problems is introduced by [25, 24]. Assuming that the nonrigid 3D shapes lie in a single low dimensional subspace, Dai et al. use the following model to estimate the 3D coordinates:

\[
\begin{align*}
\min_{X,E} & \quad \|X\|_* + \lambda\|E\|_\ell \\
\text{s.t.} & \quad W = RX^\# + E.
\end{align*}
\]  
(16)

where \( W \) consists of the 2D projected coordinates of all data points. In addition, the definitions of \( R \) and \( X^\# \) can be found in [25, 24]. Obviously, such a formulation is derived from RPCA. Subsequently, in order to cope with the complex nonrigid motion that 3D shapes lie in a union of multiple subspaces rather than a single subspace, Zhu et al. [24] develop an improvement of Eq. (16) as follows:

\[
\begin{align*}
\min_{X,E} & \quad \|Z\|_* + \gamma\|X\|_* + \lambda\|E\|_\ell \\
\text{s.t.} & \quad X = XZ \\
& \quad W = RX^\# + E.
\end{align*}
\]  
(17)

The goal of above formulation is simultaneously learning a 3D structure \( X \) and a affinity matrix \( Z \), and the latter can be used to cluster the 3D data points just like subspace clustering. Recently, a more complicated case that the 2D point tracks contain multiple deforming objects is considered by [22, 139, 140], and a scalable dense NSFM is considered by [141]. Note that most of them are based on subspace clustering, and the fundamental assumption is that all point trajectories are drawn from a union of spatial and temporal subspaces.

### 2.2.7 Summarization

In addition to the issues mentioned above, low rank regularization has been also applied into other fields, including Visual Tracking [20, 21, 142–144], salient object detection [25–29], face analysis [62, 76–79], Zero Shot Learning [143], Prototype Selection [146], Deep Learning [117–121] and so on. Acturally, most of the constructed models are the variants of RPCA, MC, or SC. Furthermore, observing the summarization reported in Table 2, one can find that nuclear norm serving as regularizer and ADMM serving as optimization method are in general. The reasons mainly include:

- For regularizer, the nuclear norm is convex, and close-form solution exists in the following Nuclear Norm Proximal (NNP) problem:

\[
\hat{X} = \text{Prox}_{\lambda\|\cdot\|_*}(M) = \arg\min_X \frac{1}{2} \|X - M\|_F^2 + \lambda\|X\|_*.
\]  
(18)

- For optimization method, ADMM can tackle complicated problems with various constraints.

Both nuclear norm and ADMM perform well in previous studies, but the progress made in RPCA and MC brings up the question: could we further improve the performance of existing algorithms by selecting new regularizers and optimization methods. Next, we will turn our attention to the regularizers replacing rank minimization.

### 3 Low Rank Relaxations

For matrix \( X \), suppose \( \sigma(X) = (\sigma_1(X), \ldots, \sigma_k(X)) \) is an ordered vector that consists of the singular values of \( X \). The matrix rank is equivalent to \( \|\sigma(X)\|_{\ell_0} \), i.e.,

\[
\operatorname{rank}(X) = \sum_{i=1}^{k} (\sigma_i(X))^0,
\]  
(19)

where \( \sigma_1(X) \geq \sigma_2(X), \ldots, \geq \sigma_k(X) \geq 0 \). Similar to \( \ell_0 \)-norm minimization, rank minimization is also a NP-hard problem. An alternative is selecting a relaxation to replace it. Here, we denote the relaxed regularizer by:

\[
\mathcal{R}(X) = \sum_{i=1}^{k} f(\sigma_i(X)),
\]  
(20)

where \( f(x) \) represents a relaxation function. According to the property of \( f(x) \), we roughly divide some representative regularizers used in previous studies into two groups: convex relaxations and non-convex relaxations. A summarization for both of them can be found in Table 3.
3.1 Convex Relaxations

Nuclear norm, i.e., trace norm is the most widely used regularizer for rank minimization, which corresponds to the function \( f(x) = \|x\|_* \) and is denoted as:

\[
\|X\|_* = \sum_{i=1}^{k} \sigma_i ,
\]

where \( k = \min(m,n) \). The connection between it and rank function is introduced by [157] where the authors show that nuclear norm is the convex envelope of rank function when \( \sigma_1(X) \leq 1 \), where \( \sigma_1(X) \) denotes the largest singular values of matrix \( X \). As mentioned above, a dominant advantage of nuclear norm is its convexity. In addition, the close-form solution of problem (18) can be obtained directly via a singular value thresholding operator \( \rho(x, \lambda) \) [158].

\[
\hat{X} = \text{Prox}_\lambda^\|\cdot\|_*, (M) = U \rho(S, \lambda) V^T
\]

where \( USV^T = M \) is the SVD of \( M \), and

\[
\rho(S, \lambda)_{ii} = \max(S_{ii} - \lambda, 0) .
\]

The Eq. (23) shows that nuclear norm treats all singular values equally and shrink them with the same threshold \( \lambda \). This, however, will introduce a bias to the matrix with small singular values. Recently, a particular convex regularizer was developed in [157], which aims to find a convex approximation for function \( \text{rank}(X) + \lambda \|X - X_0\|_F^2 \) rather than the rank function. So, the regularizer will be ignored in rest of this paper. In addition, a Elastic-Net Regularization of Singular Values (ERSV) has been proposed in [147], which corresponds to function \( f(x) = x + \lambda x^2/2 \). The ERSV can be seen as a generalization of \( \|x\|_1 + \mu \|x\|_2^2 \) which outperforms \( \|x\|_1 \) in sparse learning.

3.2 Non-Convex Relaxations

Although nuclear norm has achieved success in low rank matrix learning, it suffers a well-documented shortcoming that all singular values are simultaneously minimized. In practice, however, larger singular values generally quantify the main information we want to preserve. An alternative for nuclear norm is using non-convex relaxations.

The advantages of non-convex relaxations over nuclear norm are first shown in [34, 35] for dealing with the matrix completion problems. In particular, both of them generalize the nuclear norm to Schatten-p norm [3]

\[
\|X\|_p = \left( \sum_{i=1}^{k} (\sigma_i(X)^p) \right)^{\frac{1}{p}} .
\]

where \( 0 < p < 1 \), and when \( p = 1 \) Schatten p-Norm is equivalent to nuclear norm. Considering that larger singular values should not be punished, Hu [30] proposes an Truncated Nuclear Norm (TNN) regularizer to cope with the matrix completion problem. The TNN is denoted as:

\[
\|X\|_r = \sum_{i=1}^{k} \sigma_i(X) .
\]

A similar regularizer, namely Partial Sum Nuclear Norm (PSNN), has been developed in [148]. Indeed, both TNN and PSNN can be considered as special cases of capped nuclear norm used in [39]. Here, we provide a generalization of capped nuclear norm, namely capped Schatten-p norm, as follows:

\[
\|X\|_{pC} = \sum_{i=1}^{k} \min(\sigma_i(X)^p, c) .
\]

3. Actually, Schatten-p norm refers to the Schatten-p quasi norm when \( 0 < p < 1 \).
If we set $p = 1$ and $c = \sigma_i(M)$, $\|X\|_{p,c}$ is equivalent to $\|X\|_p$. To alleviate rather than abandon the punishment on larger singular values, Gu et al. [31] propose a weighted nuclear norm:

$$\|X\|_{w,s} = \sum_{i=1}^{k} w_i \sigma_i(X). \quad (27)$$

where $w = [w_1, w_2, \ldots, w_k]$ and $w_i \geq 0$ is a weight value assigned to $\sigma_i(X)$. Replacing the nuclear norm of NNP [18] by weighted nuclear norm, one can obtain a weighted nuclear norm proximal (WNNP) operator:

$$\hat{X} = Prox_\lambda^w(M) = arg \min_X \frac{1}{2}\|X - M\|_F^2 + \lambda \|X\|_{w,s}. \quad (28)$$

It has been shown that when the entries of weight vector $w$ are non-descending, the close-form solution of problem [23] can be easily obtained by a weighted soft-thresholding operator defined as:

$$\rho_w(S, \lambda)_{ii} = \max(S_{ii} - w_i \lambda, 0). \quad (29)$$

Here, $\rho_w$ is a generalization of $\rho$. Furthermore, we have:

$$\hat{X} = Prox_\lambda^w(M) = U \rho_w(S, \lambda)V^T. \quad (30)$$

It is obvious that the punishment bias between larger values and small values can be alleviated by assigning small weights to former and larger weights to latter. Note that TNN, PSNN, and capped nuclear norm can be considered as the special cases of WNN with weight vector:

$$w = [0, 0, \ldots, 0, 1, \ldots, 1], \quad k-r.$$  

In addition to the regularizers mentioned above, numerous non-convex regularizers derived from sparse learning have been proposed, such as nuclear norm [149], Log Nuclear Norm (LNN) [150], ETP [152], Logarithm [151], Geman [153], Laplace [154], MCP [155], and so on. The details of using them to tackle the matrix completion problems can be found in [7, 47, 160, 161]. A significant result reported in [161] is Generalized Singular Value Thresholding (GSVT) operator $Prox_\lambda^w(M)$ defined as:

$$\hat{X} = Prox_\lambda^w(M) = arg \min_X \frac{1}{2}\|X - M\|_F^2 + \lambda \sum_{i=1}^{k} f(\sigma_i(X)), \quad (31)$$

where $f(x)$ can be anyone continuous function satisfying the Assumption [1]  

**Assumption 1.** Function $f(x)$ is concave, nondecreasing, and differentiable on $[0, +\infty)$. Besides, $f(0) = 0$ and $\nabla f$ is convex.

Solving the problem [31] is equivalent to solving the following problem with $b = \sigma_i(M)$, and $i = 1, 2, \ldots, k$.

$$Prox_\lambda^w(b) = arg \min_{x \geq 0} g_b(x) = \lambda f(x) + \frac{1}{2}(x - b)^2. \quad (32)$$

A general solver for finding the optimal solution of the problem [32] has been provided in [161]. To demonstrate the difference between different regularizers including nuclear norm, we report the shrinkage results returned by them in Figure [1] where the difference between $b$ and $Prox_\lambda^w(b)$ represents the shrink. Figure [1] shows that when $b$ takes a small value the shrinkage effect of different regularizers are similar. Nevertheless, when $b$ takes a large value the difference between non-convex regularizers and nuclear norm are significant. In particular, the shrink of non-convex relaxations on larger singular values are very small, which is contrast with the nuclear norm taking serious shrinks on larger singular values. Hence, using non-convex regularizers can preserve the main information of $M$. In addition, one can find that non-convex relaxations prefer to generate 0 singular values, i.e., the low rank solution when regularization parameter $\lambda$ takes the same value.

Matrix factorization is another method for low-rank regularization, which represents the expected low rank matrix $X$ with rank $r$ as $X = UV^T$, where $U \in R^{m \times r}$ and $V \in R^{n \times r}$. Moreover, the equation [33] has been adopted by [162] for dealing with matrix completion problem and [22] for dealing with Multi-Image Matching problem.

$$\|X\|_r = \min_{A,B} \frac{1}{2} \|A\|_F + \|B\|_F \quad (33)$$

where $A \in R^{m \times d}$, $B \in R^{n \times d}$ and $d \geq r$. Recently, Shang et al. [36] develop two variants of Eq. [33], namely Double Nuclear norm penalty ($\|X\|_{D-N}$) and Frobenius/nuclear hybrid norm penalty ($\|X\|_{F-N}$), as follows:

$$\|X\|_{D-N} = \|X\|_{S_\frac{r}{2}}$$

$$\|X\|_{F-N} = \|X\|_{S_\frac{r}{2}}$$

$$\min_{A,B:AB^T = X} \frac{1}{2} \|A\|_F + \|B\|_F \quad (34)$$

where $\|X\|_{S_\frac{r}{2}}$ denotes the Schatten-$p$ quasi-norm of $X$. In addition to RPCA and MC, both of them have been also used to cope with the subspace clustering problem [163]. One of the advantages of using matrix factorization for low rank regularization is low computation and storage consumption.

The superiorities of non-convex regularizers over nuclear norm have been shown in a great many of studies w.r.t RPCA and MC, but the following two reasons prevent it from solving practical issues.

- **R.1:** The resultant optimization problem is non-convex and much more challenging to be solved.
- **R.2:** Facing various non-convex regularizers the researchers may be confused about selecting a reasonable one for their algorithms.

For R.1, we in next section review several representative optimization methods used in low rank approximation, and provide a short discussion for them. For R.2, we in Section [5] provide a comprehensive comparison between several representative regularizers.

### 4 Optimization

Solving a problem with low rank regularization has drawn significant attention, and a great many of specialized
optimization approaches have been proposed. Nevertheless, most of them are developed for solving the RPCA or MC problems. In this section we review three representatives, and discuss their superiorities and deficiencies. Some traditional methods such as FrankWolfe Algorithm [49] and SDP are omitted due to the limitations for solving practical issues. Without specific description, our discussion is based on the following problem:

$$\min_X F(X) \equiv L(X) + \lambda \sum_{i=1}^k f(\sigma_i(X)).$$

(35)

where \(L(X)\) represents the loss function which depends on the specific issue.\(^5\)

4.1 Proximal Gradient Algorithm

Proximal Gradient algorithm (PG) is one of the earliest first-order approaches for solving the problem \(^{35}\) with nuclear norm regularizer \(^{164}\). Suppose \(f(x) = x\). Instead of directly solving the original problem \(F(X)\), PG method aims to iteratively minimize a quadratic approximation of it \(^{51}\). Concretely, it generates a solution sequence \(X_t\) as follows:

$$X_{t+1} = \arg \min_X L(X_t) + (X - X_t)^T \nabla L(X_t)$$

$$+ \frac{L}{2} \| X - X_t \|_F^2 + \lambda \sum_{i=1}^k f(\sigma_i(X))$$

(36)

$$= \text{Prox}^\lambda_X(Y_t - \frac{1}{L} \nabla L(Y_t)).$$

where \(L\) is the Lipschitz constant, and

$$\text{Prox}^\lambda_X(Z) = \| X - Z \|_F^2 + \lambda \sum_{i=1}^k f(\sigma_i(X)).$$

(37)

As presented in Eq. (22), the problem (37) has a close form solution when \(f(x) = x\). The convergence rate of PG is \(O(\frac{1}{t})\).

And it can be accelerated to \(O(\frac{1}{\sqrt{t}})\) by incorporating Nesterov’s technique \(^{165}\) and generating a solution sequence \(X_t\) as:

$$X_{t+1} = \arg \min_X L(Y_t) + (X - Y_t)^T \nabla L(Y_t)$$

$$+ \frac{L}{2} \| X - Y_t \|_F^2 + \sum_{i=1}^k f(\sigma_i(X))$$

(38)

$$= \text{Prox}^\lambda_X(Y_t - \frac{1}{L} \nabla L(Y_t)).$$

where \(Y_t\) is an extrapolation of \(X_t\) and \(X_{t-1}\), i.e.,

$$Y_t = X_t + \frac{\alpha_{t-1} - 1}{\alpha_t} (X_t - X_{t-1})$$

(39)

and \(\alpha_0 = \alpha_1, \alpha_t = \frac{1}{2} (\sqrt{4\alpha_t^2 + 1} + 1)\). As Accelerated Proximal Gradient (APG) cannot guarantee that \(F(X_{t+1}) \leq F(X_t)\), Amir et al. develop a monotone APG in \(^{166}\).

Both PG and APG require that the relaxation function \(f(x)\) is convex. Recently, the case that \(f(x)\) or \(L(X)\) is non-convex was considered by \(^{45-48}\). In \(^{45}\), the authors develop two APG-type algorithms, named monotone APG (mAPG) and non-mono- tone APG (nmAPG), respectively, which replace the descent condition used in \(^{166}\) by a sufficient descent condition. Subsequently, an Inexact Proximal Gradient algorithm (IPG) was developed in \(^{48}\) to reduce the computation cost caused by two proximal mappings, and a fast proximal algorithm was developed in \(^{37}\) to reduce the computation cost caused by conducting SVT over a large scale matrix. Generally, the GSVD should be used for proximal operator \(^{37}\) when \(f(x)\) is a non-convex function.

Note that most of the existing PG-type algorithms are based on the assumption that the problem is unconstrained and has only one variable to optimize.\(^6\) Hence, although it has sound theoretical guarantee in terms of convergence as well as convergence speed, it is rarely used in solving the practical issues where various constraints must be considered and multiple variables must be optimized simultaneously.

4.2 Iteratively Re-Weighted Algorithm

The Iteratively Re-Weighted algorithm (IRW), primitively designed for sparse learning problems \(^{40-42}\), is derived from the Majorization-Minimization (MM) approaches \(^{41}\). The essence of MM is iteratively solving a convex optimization problem that is amenable to existing first-order methods. Iterative Reweighted Least Squares algorithms (IRLS) proposed in \(^{34}\) is a seminal work using IRW to handle the rank minimization problem, and a similar algorithm was developed in \(^{35}\). Both of them aim to solve the matrix completion problems with Schatten-p norm \((0 < p \leq 1)\) regularizer. Then, a variant of IRW was developed in \(^{167}\), which is designed for optimizing a matrix completion model with celebrated Huber loss function and nuclear norm regularizer. In \(^{38}\), IRW is generalized to deal with the subspace clustering problem where both the regularizer and

\(5\) Without any specific description we suppose that \(L(X)\) is convex and \(L\)-Lipschitz smooth.

\(6\) In \(^{47}\), the authors extend it to cope with the problem involved two separable parameter blocks.
with single variable (without constraints). So, its feasibility
Algorithm 1 should be changed when $L$ provided by IRW is a stationary point of original problem.

4.3 Alternating Direction Method of Multipliers

Before describing Alternating Direction Method of Multipliers (ADMM), we want to introduce the Augmented Lagrangian method (ALM) which is developed for solving RPCA problem and has been generalized to various issues. Introducing a lagrange multiplier $Y$ to the equality constraint, one can obtain the augmented lagrangian function of problem (10) as follows:

$$L(X, S; Y) = \|E\|_1 + \lambda \|X\|_*$
$$< Y, M - X - E > + \frac{\mu}{2} \|M - X - E\|^2_F,$$

where $\mu > 0$ represents a penalty parameter. A typical iterating procedure of ALM is:

$$(X^{t+1}, E^{t+1}) = \arg \min_{X,E} L(X, E; Y^t)$$
$$Y^{t+1} = Y^t + \mu(M - X - E).$$

In 43, two versions of ALM including Exact ALM (EALM) and Inexact ALM (IALM) are proposed. The difference between them is that EALM solves the first step of Eq. (43) exactly while IALM solves the first step of Eq. (43) inexactly. In practice, IALM is more efficient than EALM method because the time consuming of solving the subproblem exactly usually is expensive.

Comparing to ALM, ADMM is more powerful. As presented in Table 2, ADMM occupies a dominant position in solving practical issues. The reason is that it can deal with various complicated problems with multiple non-separable variables accompanied by multiple equality constraints as follows:

$$\min_{\{X_i\}^K_1} \{F(X_1, X_2, \ldots, X_K) \equiv \mathcal{L}(X_1, X_2, \ldots, X_K)$$
$$+ \sum_{i=1}^K \lambda_i R_i(X_i)$$
$$s.t. A_i(X_i) = B_i, \quad i = 1, 2, \ldots, K. \quad (44)$$

where $R_i(X_i)$ and $A_i(X_i) = B_i$ represents the regularizer and constraint over variable $X_i$, respectively. For PG-type algorithm, the obstacle of solving the problem (14) is that proximal operator is generally unsuitable to the subproblem with equality constraints and multiple non-separable variables. For IRNN or IRW, the obstacle is that the optimal solution of relaxing subproblem with equality constraints and the multiple non-separable variables cannot be solved directly. Nevertheless, in ADMM method, the problem (44) can be recast as:

$$\min_{\{X_i\}^K_1, \{C_i\}^K_1} \{F(\{X_i\}^K_1, \{C_i\}^K_1) \equiv \mathcal{L}(X_1, X_2, \ldots, X_K)$$
$$+ \sum_{i=1}^K \lambda_i R_i(C_i)$$
$$s.t. A_i(X_i) = B_i, \quad C_i = X_i, \quad i = 1, 2, \ldots, K. \quad (45)$$

Further, the augmented lagrangian function of problem (45)
Algorithm 2 ADMM for solving problem (46)

Require: The known information
Ensure: The matrices \(\{X\}^K_{i=1}\), \(\{Y\}^K_{i=1}\), \(\{\Lambda_i\}^K_{i=1}\)

Initialization: Initialize \(\{C\}^K_{i=1}\), \(\{Y\}^K_{i=1}\), \(\{\Lambda_i\}^K_{i=1}\).

while not convergent do
  1. Update \(X_i\) while fix other variables, where \(i = 1, 2, \ldots, K\);
  2. Update \(C_i\) while fix other variables, where \(i = 1, 2, \ldots, K\);
  3. Update \(Y_i\) by \(Y_i^{t+1} = Y_i^t + \mu_i^t(A_i(X_i^{t+1}) - B_i)\), where \(i = 1, 2, \ldots, K\); 
  4. Update \(\Lambda_i\) by \(\Lambda_i^{t+1} = \Lambda_i^t + \nu_i^t(C_i^{t+1} - X_i^{t+1})\), where \(i = 1, 2, \ldots, K\);
  5. Update parameters \(\mu_i\) by \(\mu_i^{t+1} = \rho \mu_i^t\), where \(i = 1, 2, \ldots, K\);
  6. Update parameters \(\nu_i\) by \(\nu_i^{t+1} = \rho \nu_i^t\), where \(i = 1, 2, \ldots, K\);
  7. Convergence checking.

end while

The details of using ADMM to solve the problem (46) have been outlined in Algorithm 2, where \(\{\mu_i\}^K_{i=1}\) and \(\{\nu_i\}^K_{i=1}\) are penalty parameters, and \(\rho\) is the step of updating. The difficulty occurs in updating the variables \(\{C_i\}^K_{i=1}\). Removing the irrelevant terms at \((t+1)th\) iteration, we update the variables \(C_i\) by solving the following problem:

\[
C_i = \arg\min_{C_i} R(C_i) + \frac{\nu_i}{2} \|C_i - X_i^t\|_F^2 + \frac{\lambda_i^t}{\nu_i^t} \|\Lambda_i - X_i^t\|_F^2. \tag{47}
\]

It is obvious that the optimal solution of the problem (47) can be obtained directly by using proximal operator (SVT, WSVT or GSVT).

It is worth noting that most theoretical analysis over ADMM are based on the assumption that original problem is convex \([165, 169]\). In \([12, 148]\), ADMM is used to deal with the non-convex problem where weighted nuclear norm and partial sum nuclear norm serve as regularizer, respectively. The theoretical analysis over ADMM for non-convex problems is absent yet. In practice, however, it works well empirically. A practical variant of ADMM, namely Relaxed ADMM, was discussed in \([170, 171]\). Besides, a comprehensive survey and an useful tool for ADMM can be found in \([44]\), where an unified optimization framework and theoretical analysis over ADMM for convex problems are provided \([44]\).

4.4 Summarization

We have described three representative methods for solving the problems involving low rank regularization. We can find that both of them require that the close-form solution of subproblem can be obtained directly. PG can deal with the problems with single variable or multiple separable variables, and has convergence guarantee. IRW is generally suitable to the problems with single variable. For multiple variables problems, IRW can ensure the objective function value of original problem being monotonously decreasing when we use the alternating minimization method to update the variables in subproblem. Comparing with PG and IRW, ADMM is more suitable to complicated problem. Experimentally, ADMM performs well, however, the theoretical guarantee of it for nonconvex problems is blank.

To compute the weighted matrix of \(X \in R^{m \times n}\), IRW generally need to compute the full SVD of \(X\) or the eigenvalue decomposition of \(XX^T\). \((X^T X)\) \([59]\). So, the time consuming of IRW on large scale data is very high. In the intermediate step, no matter which regularizer, ADMM and PG have to solve a subproblem defined as Eq. (31). It is obvious that \(i\)th singular value of \(X\) is zero when \(\sigma_i(M)\) is smaller than a specific threshold. That is, only serval leading singular values and singular vectors of \(M\) are needed in SVM and GSVM \([47]\). Such that the computational complexity of SVM or GSVM can be reduced from \(O(mn^2)\) to \(O(mnr)\) by using PROPACK \([172]\), where \(r\) represents the number of leading singular values. In addition, an useful tool for further reducing the time consuming of algorithm is using approximate SVM or GSVM \([172, 173]\). Both of them based on the following proposition \([47]\).

Proposition 1. Assume that \(Q \in R^{m \times \hat{r}}\), where \(\min(m, n) > \hat{r} > r\), is an orthogonal matrix and span \((U_r) \subseteq \text{span}(Q)\). Then, for problem (31) we have \(\text{Prox}_\nu^Q(M) = Q \text{Prox}_\nu^Q(Q^T M)\).

In Proposition 1, \(U, S, V^T = M\) is the rank-\(r\) approximation of \(M\). The matrix \(Q\) is generally obtained by using the power method \([174]\). Such a technique is very useful for solving large scale data.

5 Experiments

A huge number of models based on low rank regularization have been developed for solving various problems in the past decade. In practice, however, loss function
and regularization term are the main differences between them. In particular, the selection of loss function depends on the problem we aim to solve. Taking a comprehensive comparison for all problems beyond the scope of our ability. The main goal of this paper is promoting the application of non-convex regularizers in solving practical issues. Hence, in this section, we would like to take a comprehensive investigation for several non-convex regularizers listed in Table 3. Our investigation is based on image denoising, a low level computer version problem solved by the following model:

$$\min_X \|X - M\|_p^2 + \lambda \sum_{i=1}^{k} f(\sigma_i(X)).$$ \hspace{1cm} (48)

The reasons include:

1) As discussed above, the problem \( (48) \) is a basic model that we cannot avoid when dealing with majority complicated models;

2) The model \( (48) \) has only one parameter needed to be tuned. Such that the results can be analysed conveniently.

3) The structure of original matrix we aim to learn is approximately low rank. Recovering a approximately low rank rather than exactly low rank matrix is more reasonable for solving practical issues.

In addition, note that the relaxation functions that we select, including: Weighted Nuclear Norm (WNN), Log Nuclear Norm (LNN), and Truncated Nuclear Norm (TNN) are equivalent to the Schatten-p norm with \( p = 1, 0.5, 0.01 \), respectively. Particularly, Nuclear Norm (NN) is equivalent to the Schatten-p norm with \( p = 1 \). For each regularizer we can obtain the solution directly via SVT or GSVT. Here, we abandon the Capped Nuclear Norm, ETP, Laplace, Geman and so on, for all of them have additional parameters needed to be tuned.

5.1 Experimental Setting

We select 8 widely used images with size \( 256 \times 256 \) to evaluate the competing regularizers. The thumbnails are shown in Fig 2. Similarly, we corrupt the original image by Gaussian noise with distribution \( N(0, \sigma^2) \).

We use the codes provided by Gu et al. \cite{32} \footnote{https://sites.google.com/site/shuhanggu/home} for each patch, in \cite{32} the authors run \( K \) iterations of this approximation process to enhance the quality of denoising. Nevertheless, in this paper to avoid introducing additional parameters used in iterations, we fix \( K = 1 \) and other parameters as authors suggested. More implementation details over experiments can be found in \cite{32}.

For all regularizers except for WNN we set \( \lambda = \tau \eta \sqrt{\pi \sigma^2} \), where \( \eta \) is the number of similar parts, \( \sigma \) represents the noise level, \( \tau \) controls the large scale range of varying \( \lambda \), and \( \eta \) controls the small scale range of varying \( \lambda \). For WNN, we set \( \lambda_w = C + \lambda \), where \( C = \sqrt{2} \) as authors suggested.

Comparison 1. In this test, we fix \( \sigma = 50 \) and vary \( \tau \) and \( \eta \) in the sets \( S_\tau = \{0.1, 1, 10\} \) and \( S_\eta = \{1, 2, \ldots, 9\} \), respectively. We select only Lena and Peppers two images to test. The PSNR results under different parameters for all competing regularizers are shown in Figure 3 and Figure 4 respectively, where \( R_0 \) refers to the PSNR of input noisy image. \{\( R_1, R_2, \ldots, R_6 \)\} refer to the regularizers WNN, LNN, TNN, \( S_p \) with \( p = 0.1 \), \( S_p \) with \( p = 0.5 \), and nuclear norm, respectively. According to the definitions of SVT and GSVT, we can find that all regularizers, and especially nuclear norm, will shrink all singular values to zero when \( \lambda \) takes a large value. (Under this case, we set \( PSNR = 1 \) for visualization). The information delivered by Figure 3 and Figure 4 can be summarized as fellows:

- Non-convex regularizers outperform nuclear norm in terms of the best result;
- We should select a smaller regularization parameter \( \lambda \) for nuclear norm, for it prefers to shrink all singular values to zero when \( \lambda \) takes a large value (see the cases that \( \tau > 1 \) and \( \eta > 0.5 \)). In practice, however, the shrink for larger singular values may be insufficient when \( \lambda \) takes small value (see the case that \( \tau = 0.01 \) and \( 1 \leq \eta \leq 4 \)). Such a contradiction limits the performance of nuclear norm;
- TNN performs well even when \( \lambda \) takes a large value, for it prevents the \( r \) largest singular values from being shrunk. Nevertheless, two limitations of it cannot be ignored. First, the value of \( r \) must be estimated. Second, the \( r \) largest singular values generally carry noise information we want to remove, and preserving them may degenerate the performance of algorithm;
- We should select a larger regularization parameter \( \lambda \) for non-convex regularizers. When \( \lambda \) takes a small value, the shrink on all singular values is vary slight, so, the PSNRs of them are very close to original images. Comparing nuclear norm, non-convex regularizers can reduce the shrink on larger singular values and enhance the shrink on smaller singular values simultaneously.

In this test, we ignore the numerical difference (PSNR values) between different non-convex regularizers, because such a difference can be compressed by carefully selecting regularization parameter for each regularizer.

Comparison 2. In this test, for each regularizer we fix \( \lambda \) as the value achieving the best performance in above test. More specifically, we fix \( \tau \eta = 5 \) for WNN, \( \tau \eta = 2 \) for LNN, \( \tau \eta = 0.03 \) for TNN and nuclear norm, \( \tau \eta = 8 \) for \( S_p \) with \( p = 0.1 \), \( \tau \eta = 0.5 \) for \( S_p \) with \( p = 0.5 \). Besides, we select the WNN with multiple Iterations (WNN-I) just like \cite{32} for baseline. Particularly, all parameters of the algorithm are selected as author suggested. The noise level is controlled by varying the parameter \( \sigma \) in the set \( S_\sigma = \{10, 30, 50, 70, 100\} \). The PSNR results for alternative regularizers are reported in Table 4. The information delivered by Table 4 can be summarized as follows:

- WNN-I achieves the best result in all cases due to conducting the reconstruction process iteratively. But, it provides only a small advantage over others that conduct only one iteration, especially the best one highlighted in bold.
- The performance of NN is the worst in most cases. Particularly, when \( \sigma \) takes a small value, the difference between NN and TNN is very small, while other non-
Although the nuclear norm with solid theoretical guarantee non-convex relaxations. In order to promote the application significantly. And, such a deviation can be alleviated by using returned by it may deviate from the original problem to the regularizers, especially the non-convex relaxations. Differing from previous investigations over low rank matrix learning, we pay more attention and optimization methods. Differing from previous investigations, regularizers, low rank regularization including applications, regularizers, and non-convex relaxations in solving practical issues, we give a detailed summarization for non-convex regularizers used in previous studies, and discuss the properties of them via a comparison.

It is worth mentioning that the theoretical research over non-convex regularizers is very limited. In practice, however, its advantages over nuclear norm has been shown in numerous experiments including the results reported in this paper. We believe that the difference between different non-convex regularizers is very slight in terms of performance. The most significant thing we need to consider is choosing a reasonable regularization parameter, and a larger value is generally required by non-convex regularizers.

An inevitable thing of using low rank regularization (without considering matrix factorization) is conducting SVD, which is time consuming for large scale data. So, introducing the techniques over approximate SVD is an important direction in the future.47

Recently, more efforts have been made in tensor learning 175–184. But most of them are based on the convex relaxations. The problem of tensor completion based on non-convex regularization is considered in 185. Generalizing tensor learning based on non-convex regularizers to more

---

**6 Conclusion and Discussion**

In this paper, we provided a comprehensive survey for low rank regularization including applications, regularizers, and optimization methods. Differing from previous investigations over low rank matrix learning, we pay more attention to the regularizers, especially the non-convex relaxations. Although the nuclear norm with solid theoretical guarantee has been widely used to solve various problems, the solution returned by it may deviate from the original problem significantly. And, such a deviation can be alleviated by using non-convex relaxations. In order to promote the application of non-convex relaxations in solving practical issues, we give a detailed summarization for non-convex regularizers used in previous studies, and discuss the properties of them via a comparison.

It is worth mentioning that the theoretical research over non-convex regularizers is very limited. In practice, however, its advantages over nuclear norm has been shown in numerous experiments including the results reported in this paper. We believe that the difference between different non-convex regularizers is very slight in terms of performance. The most significant thing we need to consider is choosing a reasonable regularization parameter, and a larger value is generally required by non-convex regularizers.

An inevitable thing of using low rank regularization (without considering matrix factorization) is conducting SVD, which is time consuming for large scale data. So, introducing the techniques over approximate SVD is an important direction in the future.47

Recently, more efforts have been made in tensor learning 175–184. But most of them are based on the convex relaxations. The problem of tensor completion based on non-convex regularization is considered in 185. Generalizing tensor learning based on non-convex regularizers to more
TABLE 4: Experimental results (PSNR,dB) of alternative regularizers on the 8 test images. Here, WNN-I refers to the algorithm proposed in [22], which iteratively conduct the reconstruction process on all patches. The best results (without considering WNN-I) are highlighted in bold.
of subspaces,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pp. 1542–1549.

[25] Y. Dai, H. Li, and M. He, “A simple prior-free method for non-rigid structure-from-motion factorization,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012, pp. 2018–2025.

[26] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, “Salient object detection via structured matrix decomposition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 818–832, 2017.

[27] X. Shen and Y. Wu, “A unified approach to salient object detection via low rank matrix recovery,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012, 2012, pp. 853–860.

[28] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, “Segmentation driven low-rank matrix recovery for saliency detection,” in British Machine Vision Conference, BMVC 2013, Bristol, UK, September 9–13, 2013, 2013.

[29] Z. Gao, L.-F. Cheong, and Y.-X. Wang, “Block-sparse rPCA for salient motion detection,” IEEE transactions on pattern analysis and machine intelligence, vol. 36, no. 10, pp. 1975–1987, 2014.

[30] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, “Fast and accurate matrix completion via truncated nuclear norm regularization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 9, pp. 2117–2130, 2013.

[31] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pp. 2862–2869.

[32] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, “Weighted nuclear norm minimization and its applications to low level vision,” International Journal of Computer Vision, vol. 121, no. 2, pp. 183–208, 2017.

[33] X. Zhong, L. Xu, Y. Li, Z. Liu, and E. Chen, “A nonconvex relaxation approach for rank minimization problems,” in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., pp. 1980–1987.

[34] K. Mohan and M. Fazel, “Iterative reweighted algorithms for matrix rank minimization.”

[35] F. Nie, H. Huang, and C. H. Q. Ding, “Low-rank matrix recovery via efficient schatten p-norm minimization,” in Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada., 2012.

[36] F. Shang, J. Cheng, Y. Liu, Z.-Q. Luo, and Z. Lin, “Bilinear factor matrix norm minimization for robust pca: Algorithms and applications,” IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 1, pp. 1–1, 2017.

[37] F. Shang, Y. Liu, and J. Cheng, “Scalable algorithms for tractable schatten quasi-norm minimization,” in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA., 2016, pp. 2016–2022.

[38] C. Lu, Z. Lin, and S. Yan, “Smoothed low rank and
sparse matrix recovery by iteratively reweighted least squares minimization,” IEEE Trans. Image Processing, vol. 24, no. 2, pp. 646–654, 2015.

[39] Q. Sun, S. Xiang, and J. Ye, “Robust principal component analysis via capped norms,” in The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013, pp. 311–319.

[40] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted \(\ell_1\) minimization,” Journal of Fourier analysis and applications, vol. 14, no. 5-6, pp. 877–905, 2008.

[41] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, “On iteratively reweighted algorithms for nonsmooth non-convex optimization in computer vision,” SIAM Journal on Imaging Sciences, vol. 8, no. 1, pp. 331–372, 2015.

[42] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Guntürk, “Iteratively reweighted least squares minimization for sparse recovery,” Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 63, no. 1, pp. 1–38, 2010.

[43] Z. Lin, M. Chen, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” arXiv preprint arXiv:1009.5055, 2010.

[44] C. Lu, J. Feng, S. Yan, and Z. Lin, “A unified alternating direction method of multipliers by majorization minimization,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 3, pp. 527–541, 2018.

[45] H. Li and Z. Lin, “Accelerated proximal gradient methods for nonconvex programming,” in Advances in neural information processing systems, 2015, pp. 379–387.

[46] S. Ghadimi and G. Lan, “Accelerated gradient methods for nonconvex nonlinear and stochastic programming,” Mathematical Programming, vol. 156, no. 1-2, pp. 59–99, 2016.

[47] Q. Yao, J. T. Kwok, T. Wang, and T. Liu, “Large-scale low-rank matrix learning with nonconvex regularizers,” CoRR, vol. abs/1708.00146, 2017.

[48] Q. Yao, J. T. Kwok, F. Gao, W. Chen, and T-Y. Liu, “Efficient inexact proximal gradient algorithm for nonconvex problems,” arXiv preprint arXiv:1612.09069, 2016.

[49] X. Zhang, D. Schuurmans, and Y.-I. Yu, “Accelerated training for matrix-norm regularization: A boosting approach,” in Advances in Neural Information Processing Systems, 2012, pp. 2906–2914.

[50] X. Zhou, C. Yang, H. Zhao, and W. Yu, “Low-rank modeling and its applications in image analysis,” ACM Comput. Surv., vol. 47, no. 2, pp. 36:1–36:33, 2014.

[51] Z. Lin and H. Zhang, Low-rank Models in Visual Analysis: Theories, Algorithms, and Applications. Academic Press, 2017.

[52] Z. Lin, “A review on low-rank models in data analysis,” Big Data & Information Analytics, vol. 1, no. 2/3, pp. 139–161, 2017.

[53] S. Ma and N. S. Aybat, “Efficient optimization algorithms for robust principal component analysis and its variants,” arXiv preprint arXiv:1806.03430, 2018.

[54] T. K. Pong, P. Tseng, S. Ji, and J. Ye, “Trace norm regularization: Reformulations, algorithms, and multi-task learning,” SIAM Journal on Optimization, vol. 20, no. 6, pp. 3465–3489, 2010.

[55] H. Liu, L. Wang, and T. Zhao, “Calibrated multivariate regression with application to neural semantic basis discovery,” Journal of Machine Learning Research, vol. 16, pp. 1579–1606, 2015.

[56] T. H. Oh, Y. Tai, J. C. Bazin, H. Kim, and I. S. Kweon, “Partial sum minimization of singular values in robust PCA: algorithm and applications,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 4, pp. 744–758, 2016.

[57] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” Commun. ACM, vol. 55, no. 6, pp. 111–119, 2012.

[58] Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, “Matrix completion with column manipulation: Near-optimal sample-robustness-rank tradeoffs,” IEEE Transactions on Information Theory, vol. 62, no. 1, pp. 503–526, 2016.

[59] F. Nie, Z. Hu, and X. Li, “Calibrated multi-task learning,” in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018, pp. 2012–2021.

[60] Y. Kong, M. Shao, K. Li, and Y. Fu, “Probabilistic low-rank multitask learning,” IEEE transactions on neural networks and learning systems, vol. 29, no. 3, pp. 670–680, 2018.

[61] F. Nie, Z. Hu, and X. Li, “An investigation for loss functions widely used in machine learning,” Communications in Information and Systems, vol. 18, no. 1, pp. 37–52, 2018.

[62] C. Sagonas, Y. Panagakis, A. Leidinger, and S. Zafeiriou, “Robust joint and individual variance explained,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21–26, 2017, pp. 5739–5748.

[63] Y. Panagakis, M. A. Nicolaou, S. Zafeiriou, and M. Pantic, “Robust correlated and individual component analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 8, pp. 1665–1678, 2016.

[64] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, “Compressive sensing via nonlocal low-rank regularization,” IEEE Trans. Image Processing, vol. 23, no. 8, pp. 3618–3632, 2014.

[65] H. Gao, F. Nie, X. Li, and H. Huang, “Multi-view subspace clustering,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 4238–4246.

[66] M. Liu, Y. Luo, D. Tao, C. Xu, and Y. Wen, “Low-rank multi-view learning in matrix completion for multi-label image classification,” in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., 2015, pp. 2778–2784.

[67] R. Xia, Y. Pan, L. Du, and J. Yin, “Robust multi-view spectral clustering via low-rank and sparse decomposition,” in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 –31, 2014, Québec City, Québec, Canada., 2014, pp. 2149–2155.

[68] Y. Wang, W. Zhang, L. Wu, X. Lin, M. Fang, and S. Pan, “Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering,” in Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016, 2016, pp. 2153–2159.
[69] S. Li, K. Li, and Y. Fu, “Self-taught low-rank coding for visual learning,” IEEE transactions on neural networks and learning systems, vol. 29, no. 3, pp. 645–656, 2018.

[70] Z. Ding, M. Shao, and Y. Fu, “Latent low-rank transfer subspace learning for missing modality recognition,” in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 –31, 2014, Québec City, Québec, Canada., 2014, pp. 1192–1198.

[71] M. Shao, D. Kit, and Y. Fu, “Generalized transfer subspace learning through low-rank constraint,” International Journal of Computer Vision, vol. 109, no. 1-2, pp. 74–93, 2014.

[72] Y. Chen, A. Jalali, S. Sanghavi, and H. Xu, “Clustering partially observed graphs via convex optimization,” The Journal of Machine Learning Research, vol. 15, no. 1, pp. 2213–2238, 2014.

[73] Z. Huo, F. Nie, and H. Huang, “Robust and effective metric learning using capped trace norm: Metric learning via capped trace norm,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, 2016, pp. 1605–1614.

[74] J. Chen, J. Liu, and J. Ye, “Learning incoherent sparse and low-rank patterns from multiple tasks,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 5, no. 4, p. 22, 2012.

[75] J. Chen, J. Zhou, and J. Ye, “Integrating low-rank and group-sparse structures for robust multi-task learning,” in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011, pp. 42–50.

[76] N. Xue, J. Deng, S. Cheng, Y. Panagakis, and S. Zafeiriou, “Side information for face completion: a robust pca approach,” arXiv preprint arXiv:1801.07580, 2018.

[77] N. Xue, Y. Panagakis, and S. Zafeiriou, “Side information in robust principal component analysis: Algorithms and applications,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 4327–4335.

[78] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, “Robust statistical frontalization of human and animal faces,” International Journal of Computer Vision, vol. 122, no. 2, pp. 270–291, 2017.

[79] P. Snape, Y. Panagakis, and S. Zafeiriou, “Automatic construction of robust spherical harmonic subspaces,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 91–100.

[80] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, “Multi-task learning with low rank attribute embedding for multi-camera person re-identification,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 5, pp. 1167–1181, 2018.

[81] J. Xu, L. Zhang, D. Zhang, and X. Feng, “Multi-channel weighted nuclear norm minimization for real color image denoising,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 1105–1113.

[82] Y. Chang, L. Yan, and S. Zhong, “Transformed low-rank model for line pattern noise removal,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 1735–1743.

[83] T. Huang, W. Dong, X. Xie, G. Shi, and X. Bai, “Mixed noise removal via laplacian scale mixture modeling and nonlocal low-rank approximation,” IEEE Trans. Image Processing, vol. 26, no. 7, pp. 3171–3186, 2017.

[84] R. Wang and E. Trucco, “Single-patch low-rank prior for non-pointwise impulse noise removal,” in IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013, pp. 1073–1080.

[85] P. K. G. and R. R. Sahay, “Accurate structure recovery via weighted nuclear norm: A low rank approach to shape-from-focus,” in 2017 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2017, Venice, Italy, October 22-29, 2017, pp. 563–574.

[86] W. Ren, J. Tian, Z. Han, A. Chan, and Y. Tang, “Video desnowing and deraining based on matrix decomposition,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 2838–2847.

[87] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, “Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 1, pp. 156–171, 2017.

[88] J. Lezama, Q. Qiu, and G. Sapiro, “Not afraid of the dark: NIR-VIS face recognition via cross-spectral hallucination and low-rank embedding,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 6807–6816.

[89] Y. Li, J. Liu, Z. Li, Y. Zhang, H. Lu, S. Ma et al., “Learning low-rank representations with classwise block-diagonal structure for robust face recognition,” in AAAI, 2014, pp. 2810–2816.

[90] T. H. Oh, J. Lee, Y. Tai, and I. Kweon, “Robust high dynamic range imaging by rank minimization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 6, pp. 1219–1232, 2015.

[91] H. Zhao, Z. Ding, and Y. Fu, “Pose-dependent low-rank embedding for head pose estimation,” in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA., 2016, pp. 1422–1428.

[92] M. Shakeri and H. Zhang, “Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 5133–5141.

[93] B. Han and J. Sim, “Reflection removal using low-rank matrix completion,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 3872–3880.

[94] Z. Ding, M. Shao, and Y. Fu, “Low-rank embedded ensemble semantic dictionary for zero-shot learning,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 6005–6013.

[95] L. Zhu, C. Fu, M. S. Brown, and P. Heng, “A non-local...
low-rank framework for ultrasound speckle reduction,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, 2017, pp. 493–501.

[96] M. Li, J. Liu, Z. Xiong, X. Sun, and Z. Guo, “Marlow: A joint multiplanar autoregressive and low-rank approach for image completion,” in Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VII, 2016, pp. 819–834.

[97] X. Zhou, M. Zhu, and K. Daniilidis, “Multi-image matching via fast alternating minimization,” in 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, 2015, pp. 4032–4040.

[98] C. Li, L. Lin, W. Zuo, S. Yan, and J. Tang, “Sold: Sub-optimal low-rank decomposition for efficient video segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5519–5527.

[99] X. Peng, S. Zhang, Y. Yu, and D. N. Metaxas, “Toward personalized modeling: Incremental and ensemble alignment for sequential faces in the wild,” International Journal of Computer Vision, vol. 126, no. 2-4, pp. 184–197, 2018.

[100] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, “RAPS: robust and efficient automatic construction of person-specific deformable models,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, 2014, pp. 1789–1796.

[101] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2233–2246, 2012.

[102] C. Zhao, W. Cham, and X. Wang, “Joint face alignment with a generic deformable face model,” in The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011, 2011, pp. 561–568.

[103] J. Li, X. Chen, D. Zou, B. Gao, and W. Teng, “Conformal and low-rank sparse representation for image restoration,” in 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, 2015, pp. 235–243.

[104] R. S. Cabral, F. D. la Torre, J. P. Costeira, and A. Bernardino, “Matrix completion for weakly-supervised multi-label image classification,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 1, pp. 121–135, 2015.

[105] T. Zhang, B. Ghanem, S. Liu, C. Xu, and N. Ahuja, “Low-rank sparse coding for image classification,” in IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013, 2013, pp. 281–288.

[106] Z. Zhang, F. Li, M. Zhao, L. Zhang, and S. Yan, “Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification,” IEEE Trans. Image Processing, vol. 25, no. 6, pp. 2429–2443, 2016.

[107] X. Zhu, X. Jing, F. Wu, D. Wu, L. Cheng, S. Li, and R. Hu, “Multi-kernel low-rank dictionary pair learning for multiple features based image classification,” in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA., 2017, pp. 2970–2976.

[108] Z. Jiang, P. Guo, and L. Peng, “Locality-constrained low-rank coding for image classification,” in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27–31, 2014, Québec City, Québec, Canada., 2014, pp. 2780–2786.

[109] Y. Zhang, Z. Jiang, and L. S. Davis, “Learning structured low-rank representations for image classification,” in 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013, 2013, pp. 676–683.

[110] X. Cheng, S. Sridharan, J. Saraghi, and S. Lucey, “Rank minimization across appearance and shape for AAM ensemble fitting,” in IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013, 2013, pp. 577–584.

[111] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, “Multi-task low-rank affinity pursuit for image segmentation,” in Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 2439–2446.

[112] G. Liu and S. Yan, “Latent low-rank representation for subspace segmentation and feature extraction,” in IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011, 2011, pp. 1615–1622.

[113] Q. Yao and J. T. Kwok, “Colorization by patch-based local low-rank matrix completion,” in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., 2015, pp. 1959–1965.

[114] S. Wang and Z. Zhang, “Colorization by matrix completion,” in Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada., 2012.

[115] H. Fan, Y. Luo, L. Qi, N. Wang, J. Dong, and H. Yu, “Robust photometric stereo in a scattering medium via low-rank matrix completion and recovery,” in 9th International Conference on Human System Interactions, HSI 2016, Portsmouth, United Kingdom, July 6-8, 2016, 2016, pp. 323–329.

[116] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, “TILT: transform invariant low-rank textures,” International Journal of Computer Vision, vol. 99, no. 1, pp. 1–24, 2012.

[117] Z. Ding, M. Shao, and Y. Fu, “Deep robust encoder through locality preserving low-rank dictionary,” in Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI, 2016, pp. 567–582.

[118] J. Lezama, Q. Qiu, P. Musé, and G. Sapiro, “Ole: Orthogonal low-rank embedding, a plug and play geometric loss for deep learning,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[119] X. Yu, T. Liu, X. Wang, and D. Tao, “On compressing deep models by low rank and sparse decomposition,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26,
2017, 2017, pp. 67–76.
[120] T. N. Sainath, B. Kingsbury, V. Sindhwani, E. Arisoy, and B. Ramabhadran, “Low-rank matrix factorization for deep neural network training with high-dimensional output targets,” in IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26–31, 2013, pp. 6655–6659.

[121] M. Denil, B. Shakibi, L. Dinh, M. Ranzato, and N. de Freitas, “Predicting parameters in deep learning,” in Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States., 2013, pp. 2148–2156.

[122] C. Georgakis, Y. Panagakis, and M. Pantic, “Dynamic behavior analysis via structured rank minimization,” International Journal of Computer Vision, vol. 126, no. 2-4, pp. 333–357, 2018.

[123] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, “Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, 2016, pp. 2396–2404.

[124] W. Li, Z. Xu, D. Xu, D. Dai, and L. V. Gool, “Domain generalization and adaptation using low rank exemplar svms,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 5, pp. 1114–1127, 2018.

[125] L. Shi, “Sparse additive text models with low rank background,” in Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States., 2013, pp. 172–180.

[126] Y. Pan, H. Lai, C. Liu, Y. Tang, and S. Yan, “Rank aggregation via low-rank and structured-sparse decomposition,” in Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA., 2013.

[127] X. Zhou, C. Yang, and W. Yu, “Moving object detection by detecting contiguous outliers in the low-rank representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 597–610, 2013.

[128] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E. Zahzah, “Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset,” Computer Science Review, vol. 23, pp. 1–71, 2017.

[129] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” IEEE Trans. Image Processing, vol. 16, no. 8, pp. 2080–2095, 2007.

[130] J. Mairal, F. R. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local sparse models for image restoration,” in IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009, 2009, pp. 2272–2279.

[131] W. Dong, L. Zhang, and G. Shi, “Centralized sparse representation for image restoration,” in IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011, 2011, pp. 1259–1266.

[132] S. Wang, L. Zhang, and Y. Liang, “Nonlocal spectral prior model for low-level vision,” in Computer Vision - ACCV 2012 - 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part III, 2012, pp. 231–244.

[133] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, “Weighted schatten p-norm minimization for image denoising and background subtraction,” IEEE Trans. Image Processing, vol. 25, no. 10, pp. 4842–4857, 2016.

[134] N. Yair and T. Michaeli, “Multi-scale weighted nuclear norm image restoration,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3165–3174.

[135] X. Cheng, S. Sridharan, J. Saraghi, and S. Lucey, “Anchored deformable face ensemble alignment,” in Computer Vision - ECCV 2012. Workshops and Demonstrations - Florence, Italy, October 7-13, 2012, Proceedings, Part I, 2012, pp. 133–142.

[136] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” SIAM Review, vol. 52, no. 3, pp. 471–501, 2010.

[137] J. Yan, H. Xu, H. Zha, X. Yang, H. Liu, and S. M. Chu, “A matrix decomposition perspective to multiple graph matching,” in 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, 2015, pp. 199–207.

[138] N. Hu, Q. Huang, B. Thibert, U. Alpes, and L. Guibas, “Distributable consistent multi-object matching,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2463–2471.

[139] A. Agudo and F. Moreno-Noguer, “DUST: dual union of spatio-temporal subspaces for monocular multiple object 3d reconstruction,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, 2017, pp. 1513–1521.

[140] S. Kumar, Y. Dai, and H. Li, “Spatial-temporal union of subspaces for multi-body non-rigid structure-from-motion,” CoRR, vol. abs/1705.04916, 2017.

[141] S. Kumar, A. Cherian, Y. Dai, and H. Li, “Scalable dense non-rigid structure-from-motion: A grassmannian perspective,” arXiv preprint arXiv:1803.00233, 2018.

[142] T. Zhang, S. Liu, N. Ahuja, M. Yang, and B. Ghanem, “Robust visual tracking via consistent low-rank sparse learning,” International Journal of Computer Vision, vol. 111, no. 2, pp. 171–190, 2015.

[143] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, “Low-rank sparse learning for robust visual tracking,” in Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI, 2012, pp. 470–484.

[144] Y. Sui, G. Wang, Y. Tang, and L. Zhang, “Tracking completion,” in Computer Vision - ECCV 2016 - 14th European Conference on Computer Vision, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII, 2016, pp. 194–209.

[145] Y. Liu, Q. Gao, J. Li, J. Han, and L. Shao, “Zero shot learning via low-rank embedded semantic autoencoder,” in IJCAI, 2018, pp. 2490–2496.
[177] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, “Low-rank tensor constrained multiview subspace clustering,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1582–1590.

[178] T. Yokota, B. Erem, S. Guler, S. K. Warfield, and H. Hontani, “Missing slice recovery for tensors using a low-rank model in embedded space,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8251–8259.

[179] C.-Y. Ko, K. Batselier, W. Yu, and N. Wong, “Fast and accurate tensor completion with tensor trains: A system identification approach,” arXiv preprint arXiv:1804.06128, 2018.

[180] A. Zare, A. Ozdemir, M. A. Iwen, and S. Aviyente, “Extension of pca to higher order data structures: An introduction to tensors, tensor decompositions, and tensor pca,” Proceedings of the IEEE, no. 99, 2018.

[181] T. Yokota and H. Hontani, “Simultaneous visual data completion and denoising based on tensor rank and total variation minimization and its primal-dual splitting algorithm,” in Proceedings of CVPR, 2017, pp. 3732–3740.

[182] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, “Tensor robust principal component analysis with a new tensor nuclear norm,” submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (major revision), 2018.

[183] P. Zhou, C. Lu, and J. Feng, “Tensor low-rank representation for data recovery and clustering,” submitted to International Journal of Computer Vision (IJCV) (major revision), 2018.

[184] C. Lu, J. Feng, Z. Lin, and S. Yan, “Exact low tubal rank tensor recovery from gaussian measurements,” in International Joint Conference on Artificial Intelligence (IJCAI), 2018.

[185] Q. Yao, “Scalable tensor completion with nonconvex regularization,” arXiv preprint arXiv:1807.08725, 2018.