

Visual Processing by a Unified Schatten- p Norm and ℓ_q Norm Regularized Principal Component Pursuit

Jing Wang, Meng Wang, Xuegang Hu

Department of Computer Science
Hefei University of Technology
Hefei, Anhui 230009, China

Shuicheng Yan

Department of Electrical and Engineering
National University of Singapore
117583, Singapore

Abstract

In this paper, we propose a non-convex formulation to recover the authentic structure from the corrupted real data. Typically, the specific structure is assumed to be low rank, which holds for a wide range of data, such as images and videos. Meanwhile, the corruption is assumed to be sparse. In the literature, such a problem is known as Robust Principal Component Analysis (RPCA), which usually recovers the low rank structure by approximating the rank function with a nuclear norm and penalizing the error by an ℓ_1 -norm. Although RPCA is a convex formulation and can be solved effectively, the introduced norms are not tight approximations, which may cause the solution to deviate from the authentic one. Therefore, we consider here a non-convex relaxation, consisting of a Schatten- p norm and an ℓ_q -norm that promote low rank and sparsity respectively. We derive a proximal iteratively reweighted algorithm (PIRA) to solve the problem. Our algorithm is based on an alternating direction method of multipliers, where in each iteration we linearize the underlying objective function that allows us to have a closed form solution. We demonstrate that solutions produced by the linearized approximation always converge and have a tighter approximation than the convex counterpart. Experimental results on benchmarks show encouraging results of our approach.

Keywords: Image processing, Denoising, Robust principal component analysis, Schatten- p norm, ℓ_q norm

1 Introduction

The popularity of webcams and mobile phone cameras has generated a large amount of visual data. However, visual data are easily corrupted by artifacts arising from imaging devices or natural factors such as illumination. The human vision system could recognize the corruption with accumulated information and knowledge. However, it will result in irrelevant or noisy information in the computer vision community. Thus, a bunch of methods have been proposed to obtain authentic data for visual processing tasks, such as image processing. Image denoising aims at reducing the noise from the original image [Lee80] [LO00] [ZYZH13] [CHH13] [SYLL08]. Specifically, some approaches focus on statistical image modeling for the purpose of optimal signal representation and transmission, such as the Gaussian scale Mixture (GSM) model, the variance-adaptive model or Bayesian estimation [SCD02] [SYLL08]. Portilla et al. presented a denoising method based on a local Gaussian scale mixture model in an overcomplete oriented pyramid representation [PSWS03]. The approaches mentioned above are based on the initial features of the visual data. Generally, better features will enhance the performance of representation. For instance, Shao et al. generated domain-adaptive global feature descriptors to obtain better performance in image classification [SLL14]. Zhu et al. utilized weakly labeled data from other domains as the feature space for the visual categorization problem [ZS14]. Based on a comprehensive feature space, some effective and promising denoising approaches are proposed by exploiting sparse and redundant representations over a trained dictionary [ER06]. Elad et al. proposed the K-SVD algorithm [AEB06]. It was the first time that sparse modeling

of image patches has been successfully applied in image denoising. Yan et al. exploited the sparsity within representation in the wavelet domain to handle high-level noises [YSL13]. One reason for the success of the algorithm is the statistical properties of noise. It is natural to assume the noise is sparse. Besides, the visual data such as images are probably of low rank structure [HZY⁺12]. For example, for a facial image taken under certain illumination conditions, the low-rank component captures the face, and the sparse component captures the light on the face [WYG⁺09]. Thus, the idea of turning the problem into a low rank matrix and a sparse matrix recovery problem has drawn considerable attention. In the following, we first describe the problem.

1.1 The Problem Description

Suppose X is an observed data matrix in $R^{m \times n}$, where m is used to denote the ambient dimension of a sample and n is the number of samples. The problem can be formulated as:

$$\min_{L,S} \text{rank}(L) + \lambda \|S\|_0, \text{ s.t. } X = L + S, \quad (1.1)$$

where $L \in R^{m \times n}$ has a low rank structure that is assumed to be the authentic structure of the observed data, and $S \in R^{m \times n}$ is assumed to be sparse representation of the noise. $\text{Rank}(L)$ is the rank of the matrix L , $\|S\|_0$ is the ℓ_0 -norm which counts the number of non-zero entries in S , and λ is a parameter balancing the two components. The goal of the above optimization problem (1.1) is called Robust Principal Component Analysis (RPCA), aiming to recover the low-rank component L and sparse component S , with the constraint of $X = L + S$.

1.2 The Reformulation and Solutions

It is challenging to solve problem (1.1), because $\text{rank}(L)$ and $\|S\|_0$ are both discontinuous and non-convex. In fact, it is NP-hard. A common strategy [CLMW11] is to relax the rank function to the convex nuclear norm $\|L\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(L)$, where σ_i denotes the i -th singular value of L , and relax the ℓ_0 -norm to the ℓ_1 -norm $\|S\|_1 = \sum_{ij} |S_{ij}|$, where $|S_{ij}|$ is the magnitude of the (i, j) -th element in S . Problem (1.1) can then be reformulated as:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1, \text{ s.t. } X = L + S. \quad (1.2)$$

Candès et al. theoretically proved that if L and S satisfy certain assumptions, they can be recovered exactly via solving a convex program called Principal Component Pursuit with $\lambda = 1/\sqrt{\max\{m, n\}}$ [CLMW11]. Unlike the formulation defined in (1.1), RPCA in (1.2) is convex, and the optimal solution is tractable. An efficient solver for (1.2) is Alternating Direction Method (ADM) [LCM09] which guarantees to obtain the optimal solution. Another well-known first-order algorithm is the Accelerated Proximal Gradient (APG), which solves an unconstrained Stable Principal Component Pursuit (SPCP) problem [ZLW⁺10] as follows:

$$\min_{L,S} \lambda_1 \|L\|_* + \lambda_2 \|S\|_1 + \frac{1}{2} \|L + S - X\|_F^2, \quad (1.3)$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are balancing parameters. APG is a fast method with a convergence rate $O(1/T^2)$, where T is the number of iterations.

1.3 Related Works

As the RPCA model is capable of recovering the low rank components from grossly corrupted data and theoretical conditions to ensure the perfect recovery have been analyzed in depth, RPCA and its extensions have been applied to many applications, including background modeling [CLMW11], image alignment [PGW⁺10] and subspace segmentation [LLY⁺13a]. Specifically, Hui et al. presented a patch-based

algorithm using low-rank matrix recovery [JLSX10]. Wang et al. studied the problem of aligning correlated images by decomposing the matrix of corrupted images as the sum of a sparse matrix of errors and a low-rank matrix of recovered aligned images [WLLL13]. Hu et al. proposed a truncated nuclear norm regularization for estimating missing values from corrupted images [HZY+12].

There are several works aimed at improving the low-rank and sparse matrix recovery. Mu et al. [MDYY11] proposed an Accelerated RPCA using random projection. Zhou and Tao [ZT11] developed a fast solver for low-rank and sparse matrix recovery with hard constraints on both L and S . To alleviate the challenges raised by coherent data, most recently, Liu et al. recovered the coherent data by Low-Rank Representation (LRR) [LLY+13b]. Aybat et al. developed a fast first-order algorithm to solve the SPCP problem [AGI11]. Fazel suggested to reformulating the rank optimization problem as a Semi-Definite Programming (SDP) problem [Faz02]. An accelerated proximal gradient optimization technique was applied to solve the nuclear norm regularized least squares [TY10] [JY09].

However, existing algorithms may lead to solutions that deviate from the original problem. Most previous works use the convex nuclear norm as a surrogate of the rank function and the ℓ_1 -norm as a surrogate of the ℓ_0 -norm, and then instead solve the new problem. But the nuclear norm is the sum of the singular values, while the rank function is the number of the non-zero singular values in which each singular value contributes equally. There are similar differences between the ℓ_0 -norm and ℓ_1 -norm when performing a theoretical analysis [RFP10]. Hence, the solution to the relaxed problem may be far from the original one. Some researchers instead consider non-convex surrogate functions.

The smoothed Schatten- p norm is a popular non-convex surrogate of the rank function defined as [MF12][NWC+12]:

$$\begin{aligned} \ell_p(X) &= \text{Tr}(X^T X + \epsilon I)^{p/2} \\ &= \sum_{i=1}^n (\sigma_i^2(X) + \epsilon)^{p/2} \end{aligned} \quad (1.4)$$

where I is the identity matrix with the same size as X , and $\ell_p(X)$ is differentiable for $p > 0$ and nonconvex for $p < 1$. Mohan and Fazel used the Schatten- p norm to replace the rank function and considered the problem [MF12]:

$$\begin{aligned} \min \ell_p(X) \\ \text{s.t. } A(X) = b, \end{aligned} \quad (1.5)$$

where $A : R^{m \times n} \rightarrow R^p$ is a linear map, and $b \in R^p$ denotes the measurements. They also proposed the Iterative Reweighted Least Squares (IRLS) algorithm for rank minimization. Under certain conditions, IRLS-1 converges to the global minimum of the smoothed nuclear norm and IRLS- p converges to a stationary point of the corresponding non-convex yet smooth approximation to the rank function. Nie et al. [NHD12] proposed the extended Schatten- p norm as an efficient surrogate of the rank function defined as:

$$\begin{aligned} \ell_p &= \left(\sum_{i=1}^{\min(m,n)} \sigma_i^p \right)^{1/p} \\ &= \left(\text{Tr}(X^T X)^{p/2} \right)^{1/p}. \end{aligned} \quad (1.6)$$

They derived an efficient algorithm to solve the above problem.

For the ℓ_0 -norm, many non-convex surrogate functions have been proposed, e.g., ℓ_q -norm with $0 < q < 1$ [FL09], and Smoothly Clipped Absolute Deviation (SCAD) [FL01]. Nie et al. [NWC+12] used the Alternate Direction Method (ADM) to solve a similar problem for the non-convex matrix completion

problem. Candés et al. [CWB08] proposed an algorithm to solve the reweighted ℓ_1 minimization problem, which could better recover the ℓ_0 -norm. The condition of sparse vector recovery has been given in [FL09].

The major drawback to the above approaches is that previous iteratively reweighted algorithms can only approximate either the low-rank component or the sparse one with a non-convex surrogate [CZ][LXY13]. One important reason for this is that it is difficult to solve a problem whose objective function contains two or more nonsmooth terms. Thus, in this paper, we simultaneously approximate the low rank and sparse functions with non-convex surrogates.

1.4 Introducing Our Approach

In this paper, we propose a new formulation with the Schatten- p norm and ℓ_q -norm regularized Principal Component Pursuit (p, q -PCP) ($0 < p, q < 1$) for recovering the low-rank and sparse matrices. We also provide an algorithm to solve such a non-convex problem with two non-smooth components. Empirically, our proposed Proximal Iteratively Reweighted Algorithm (PIRA) can solve p, q -PCP effectively without loss of efficiency. In each iteration, PIRA provides closed-form solutions which make the algorithm efficient. To the best of our knowledge, this is the first time that the $\ell_{p,q}$ norm has been used to approximate the RPCA problem. We are also the first to provide corresponding solutions. Experimental results demonstrate that the solutions can tightly approximate the RPCA problem and the objective function can converge in several iterations. The main contributions of this study are summarized as follows.

- We propose a joint Schatten- p norm and ℓ_q -norm regularized Principal Component Pursuit (p, q -PCP) model for low-rank and sparse matrix recovery.
- A new Proximal Iteratively Reweighted Algorithm (PIRA) is presented to solve the p, q -PCP problem. We demonstrate the effectiveness and efficiency of our algorithm.
- We empirically show that our solutions can approximate the original problem and the objective function will converge with a few iterations.
- Extensive experiments on synthetic data and real world data show that our proposed algorithm outperforms state-of-the-art algorithms.

1.5 Overview of the Paper

The rest of the paper is organized as follows. In Section 2, we give detailed information about our proposed non-convex p, q -PCP model and an iteratively reweighted algorithm (PIRA) to solve p, q -PCP. We provide a detailed analysis of the optimization algorithm in Section 3. Experimental results are presented in Section 4. We conclude this paper in Section 5.

2 Non-convex Principal Component Pursuit

In this section, we first present the non-convex principal component pursuit model. We then propose a new iteratively reweighted algorithm to solve the non-convex principal component pursuit problem.

2.1 The p, q -PCP Model

The motivation for approximating the rank function with ℓ_p -norm is to obtain better empirical performance, in terms of recovering low-rank matrices, than the nuclear norm [MF12] when $0 \leq p < 1$. Mohan and Fazel theoretically prove that the ℓ_p -norm is similar to the nuclear norm minimization when $p = 1$ [MF12].

The ℓ_q -norm is used with $0 \leq q < 1$ as a surrogate of the ℓ_0 -norm because it generalizes and improves the ℓ_1 -norm [FL09]. The ℓ_q -norm degenerates into the ℓ_0 -norm when $q \rightarrow 0$. A similar property holds for the Schatten- p norm as a surrogate of the rank function. It is of interest to consider the non-convex principal component pursuit by using the Schatten- p norm and ℓ_q -norm jointly:

$$\begin{aligned} \min_{L,S} \quad & \lambda_1 \sum_{i=1}^m \sigma_i^p(L) + \lambda_2 \sum_{i=1}^m \sum_{j=1}^n |S_{ij}|^q, \\ \text{s.t.} \quad & X = L + S, \end{aligned} \quad (2.1)$$

where $X \in \mathbb{R}^{m \times n}$ (we assume $m \leq n$ in this paper) is the observed data matrix, and $\sigma_i(L)$, $i = 1, \dots, m$, denotes the singular values of L . $p, q \in [0, 1]$, $\lambda_1, \lambda_2 > 0$. More generally, we further consider the stable model as follows:

$$\min_{L,S} \lambda_1 \sum_{i=1}^m \sigma_i^p(L) + \lambda_2 \sum_{i=1}^m \sum_{j=1}^n |S_{ij}|^q + \frac{1}{2} \|X - L - S\|_F^2. \quad (2.2)$$

When $p = q = 1$, the above p, q -PCP model degenerates into the convex PCP as in (1.2) or (1.3).

It is expected that smaller values of p and q can help p, q -PCP approximate the RPCA in (1.1). It is worth mentioning that many non-convex penalty functions can be applied for the non-convex principal component pursuit model. We use the Schatten- p norm and the ℓ_q -norm in this study, because compared with other non-convex penalty functions, they are matrix norms which bear more similar special properties such as the nuclear norm and ℓ_1 -norm. The low-rank matrix and the sparse matrix recovery conditions based on the Null Space Property (NSP) have been presented in previous works [FL09][MF12]. In fact, they are extended from the nuclear norm and the convex ℓ_1 -norm. It is easy to tune the parameters of p and q within $(0, 1)$. Many previous works empirically showed that the ℓ_q -norm improved the recovery performance by comparison with the convex ℓ_1 -norm [CWB08]. A similar improvement was recently found in the Schatten- p norm by comparison with the convex nuclear norm [MF12]. It is expected that jointly combining them in a model will surpass the recovery performance of the convex PCP.

2.2 Proximal Iteratively Reweighted Algorithm

In this section, we demonstrate how to solve problem (2.2) using the Schatten- p norm and ℓ_q -norm regularizers. In fact, the ℓ_q -minimization is non-smooth, non-Lipschitz continuous, and NP-hard [CGWY11]. We use the strategy of shifting σ_i^p to $(\sigma_i + \varepsilon)^p$, and $|S_{ij}|^q$ to $(|S_{ij}| + \varepsilon)^q$, with $0 < \varepsilon \ll 1$, and solve the relaxed problem as follows:

$$\begin{aligned} \min_{L,S} \quad & \lambda_1 \sum_{i=1}^m (\sigma_i(L) + \varepsilon)^p + \lambda_2 \sum_{i=1}^m \sum_{j=1}^n (|S_{ij}| + \varepsilon)^q \\ & + \frac{1}{2} \|L + S - X\|_F^2. \end{aligned} \quad (2.3)$$

ε ensures that the zero singular values and nonzero entries in the sparse component have corresponding weightings. The above problem is non-smooth and has two variables. We present a Proximal Iteratively Reweighted Algorithm (PIRA) to solve it.

Intuitively, we need to update L and S alternately. For fixed $S = S^k$ in the k -th iteration, problem (2.3) can be described as:

$$L^{k+1} = \arg \min_L \lambda_1 \sum_{i=1}^m (\sigma_i(L) + \varepsilon)^p + \frac{1}{2} \|L + S^k - X\|_F^2. \quad (2.4)$$

To solve the above problem, we linearize the objective function of (2.3) using the Taylor expansion w.r.t. L at $L = L^k$ and add a proximal term. L^{k+1} is then updated by minimizing the relaxed function:

$$\begin{aligned}
L^{k+1} &= \arg \min_L \lambda_1 \sum_{i=1}^m (\sigma_i(L^k) + \varepsilon)^p + w_i^k (\sigma_i(L) - \sigma_i(L^k)) \\
&\quad + \|L^k + S^k - X\|_F^2 + \left\langle L^k + S^k - X, L - L^k \right\rangle \\
&\quad + \frac{\mu_1}{2} \|L - L^k\|_F^2 \\
&= \arg \min_L \frac{\lambda_1}{\mu_1} \sum_{i=1}^m w_i^k \sigma_i(L) \\
&\quad + \frac{1}{2} \left\| L - \left(L^k - \frac{1}{\mu_1} (L^k + S^k - X) \right) \right\|_F^2,
\end{aligned} \tag{2.5}$$

where

$$w_i^k = \frac{p}{(\sigma_i(L^k) + \varepsilon)^{1-p}}, \quad i = 1, \dots, m, \tag{2.6}$$

are the weights corresponding to L^k . They are actually the gradients of $(\sigma_i(L) + \varepsilon)^p$ w.r.t $\sigma_i(L)$ at $L = L^k$, $i = 1, \dots, m$. The backtracking rule can be used to estimate in each iteration [BT09]. Note that problem (2.5) is non-convex. Fortunately, it has a closed form solution as shown in [CDC13].

Lemma 1. *Given $Y \in \mathbb{R}^{m \times n}$, $0 \leq w_1 \leq w_2 \leq \dots \leq w_s$ ($s = \min(m, n)$), and $\lambda > 0$, the optimal solution to the following problem:*

$$\mathcal{D}_{\lambda w}(Y) = \arg \min_Y \lambda \sum_{i=1}^s w_i \sigma_i(X) + \frac{1}{2} \|X - Y\|_F^2, \tag{2.7}$$

is given by the weighted singular value threshold:

$$\mathcal{D}_{\lambda w}(Y) = U \mathcal{S}_{\lambda w}(\Sigma) V^T, \tag{2.8}$$

where $Y = U \Sigma V^T$ is the SVD of Y , and $\mathcal{S}_{\lambda w}(\mathbf{m}\Sigma) = \text{Diag}\{(\Sigma_{ii} - \lambda w_i)_+\}$.

Using Lemma 1, L^{k+1} can be updated by:

$$L^{k+1} = D_{\lambda_1 w^k / \mu_1} \left(L^k - \frac{1}{\mu_1} (L^k + S^k - X) \right). \tag{2.9}$$

Note that the main computation for each iteration is one SVD. The iteration is expected to obtain a better estimation of the rank successfully. Even with large rank initially, some small singular values will have large weights and will themselves become zero in the following iterations. Thus, the rank of L decreases with each iteration. For fixed $L = L^k$ in the k -th iteration, the solution of problem (2.3) can be derived as:

$$S^{k+1} = \arg \min_S \lambda_2 \sum_{i=1}^m \sum_{j=1}^n (|S_{ij}| + \varepsilon)^q + \frac{1}{2} \|L^k + S - X\|_F^2. \tag{2.10}$$

To solve the above problem, we linearize the objective function of (2.3) using the Taylor expansion w.r.t S

Algorithm 1 Solving Problem (2.3) using PIRA

Input: $X \in \mathbb{R}^{m \times n}$, $0 < p, q < 1$, $\mu_1 > 1$ and $\mu_2 > \frac{1}{2}$.

Initialize: $k = 0$, $w^k \in \mathbb{R}^m$, $M^k \in \mathbb{R}^{m \times n}$, $L^k \in \mathbb{R}^{m \times n}$, $S^k \in \mathbb{R}^{m \times n}$, and $\varepsilon > 0$.

while not converge **do**

1. Update L^{k+1} by

$$L^{k+1} = D_{\lambda_1 w^k / \mu_1} \left(L^k - \frac{1}{\mu_1} (L^k + S^k - X) \right).$$

2. Update S^{k+1} by

$$S^{k+1} = S_{\lambda_2 M / \mu_2} \left(S^k - \frac{1}{\mu_2} (L^k + S^k - X) \right).$$

3. Update the weight vector w_i^{k+1} , $i = 1, \dots, m$, by

$$w_i^{k+1} = \frac{p}{(\sigma_i(L^{k+1}) + \varepsilon)^{1-p}}.$$

4. Update the weight matrix M_{ij}^{k+1} , $i = 1, \dots, m$, $j = 1, \dots, n$, by (2.12).

$$M_{ij}^{k+1} = \frac{q}{(|S_{ij}^{k+1}| + \varepsilon)^{1-q}}.$$

end while

Output: L^* , S^* .

at $S = S^k$ and add a proximal term. S^{k+1} is then updated by minimizing the relaxed function:

$$\begin{aligned} S^{k+1} &= \arg \min_S \lambda_2 \sum_{i=1}^m \sum_{j=1}^n (|S_{ij}^k| + \varepsilon)^q + M_{ij}^k (|S_{ij}| - |S_{ij}^k|) \\ &\quad + \|L^k + S^k - X\|_F^2 + \left\langle L^k + S^k - X, S - S^k \right\rangle \\ &\quad + \frac{\mu_2}{2} \|S - S^k\|_F^2 \\ &= \arg \min_S \frac{\lambda_2}{\mu_2} \sum_{i=1}^m \sum_{j=1}^n M_{ij}^k |S_{ij}| \\ &\quad + \frac{1}{2} \left\| S - \left(S^k - \frac{1}{\mu_2} (L^k + S^k - X) \right) \right\|_F^2, \end{aligned} \tag{2.11}$$

where

$$M_{ij}^k = \frac{q}{(|S_{ij}^k| + \varepsilon)^{1-q}}, \quad i = 1, \dots, m, j = 1, \dots, n, \tag{2.12}$$

are the weights corresponding to S^k . They are actually the gradients of $(|S_{ij}| + \varepsilon)^q$ w.r.t. $|S_{ij}|$ at $S = S^k$, $i = 1, \dots, m$, $j = 1, \dots, n$. Note that this problem requires $O(mn)$ flops. The value of μ_2 will influence the convergence of the iterations [HYZ08], but we empirically set it to be 2.1 which shall hold. Note that problem (2.5) is separable. Each S_{ij} can be solved separately using the following closed form solution [HYZ08]:

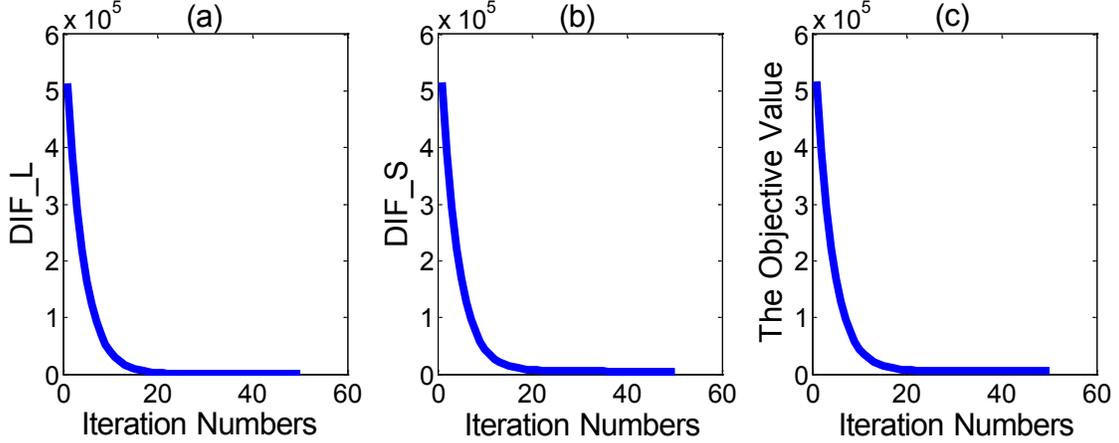


Figure 1: The convergences of the subproblems (2.4) and (2.10) are shown in (a) and (b) respectively. The convergence of the global objective function (2.2) is shown in (c).

Lemma 2. Given $y, w \in \mathbb{R}$, $w \geq 0$ and $\lambda > 0$, the optimal solution to the following problem

$$\mathcal{S}_{\lambda w}(y) = \arg \min_x \lambda w |x| + \frac{1}{2}(x - y)^2, \quad (2.13)$$

is given by

$$\mathcal{S}_{\lambda w}(y) = \begin{cases} y - \lambda w, & \text{if } y > \lambda w, \\ y + \lambda w, & \text{if } y < -\lambda w, \\ 0, & \text{otherwise.} \end{cases} \quad (2.14)$$

Using Lemma 2, S^{k+1} can be updated by

$$S^{k+1} = \mathcal{S}_{\lambda_2 M / \mu_2} \left(S^k - \frac{1}{\mu_2} (L^k + S^k - X) \right). \quad (2.15)$$

Alternately updating L by (2.9), S by (2.15) and their weights w by (2.6) and M by (2.12) leads to the proposed Proximal Iteratively Reweighted Algorithm (PIRA), as shown in Algorithm 1. Note that in each iteration, PIRA solves a weighted nuclear norm minimization problem and a weighted ℓ_1 -norm minimization problem. Both have closed form solutions, and their computational costs are the same as for convex optimization.

The detailed procedure of our algorithm is shown in Algorithm 1. We first use a common strategy which relaxes it to the form (2.3) by introducing ε , and then fix the L and S separately to obtain the optimal solution (step 2-3). Updated weights w and M are used for the next iteration based on the current solutions (step 4-7). We will provide the further analysis of our algorithm in the following section.

3 Algorithmic Analysis

In this section, we give a detailed analysis of our algorithm. We first illustrate that the obtained solution L^* is the stationary point for problem (2.5), and the solution S^* is optimal for problem (2.11). We then show that the obtained solutions can approximate the optimal solutions of the original problems (2.4) and (2.10). Finally, the numerical results show that PIRA decreases the objective function (2.3) monotonically.

The experimental data are $X = L + S + \xi$. $L = UV^T$ is a rank-10 matrix where U and V are 200×5 generated by the Matlab function `randn`. Each element of S is set to zero with probability 0.8 and non-zero

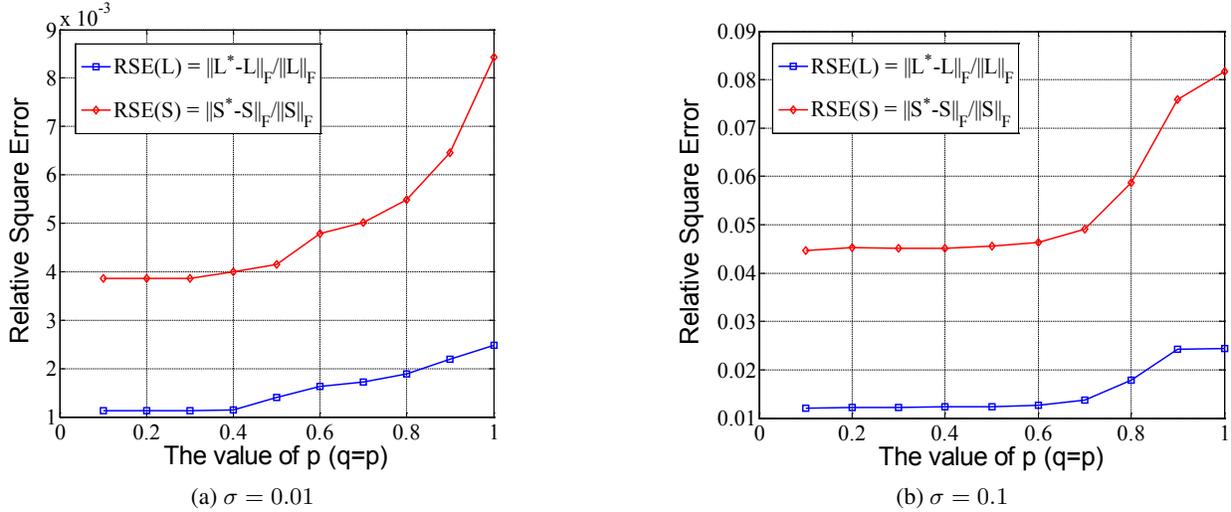


Figure 2: Comparison of low-rank and sparse matrix recovery with varying noise levels $\sigma = \{0.01, 0.1\}$, $n = 200$, $r = 10$.

entries are sampled in the interval $[-5, 5]$ with probability 0.2. ξ is the Gaussian noise with i.i.d. $N(0, 0.01^2)$. We set p to be 0.5. As shown in the synthetic data experiment later, $p = 0.5$ is representative.

For the low rank part recovery, the original problem is in the form of (2.4). With linearization, the problem (2.4) can be derived as (2.5), which falls into the general nonconvex low-rank minimization form,

$$\min_{X \in R^{m \times n}} F(X) = \sum_{i=1}^m g_{\lambda}(\sigma_i(X)) + f(X). \quad (3.1)$$

We observe that the penalty function g_{λ} in our problem (2.5) is Schatten- p norm, which is continuous, concave and monotonically increasing on $[0, \infty)$. The loss function f in Eqn. (2.5) is smooth and continuously differentiable with a Lipschitz continuous gradient, that is

$$\|\nabla f(X) - \nabla f(Y)\|_F \leq L(f)\|X - Y\|_F, \quad (3.2)$$

where $X = L, Y = L^k - \frac{1}{\mu_1}(L^k + S^k - X)$ in the k -th iteration, $L(f) > 0$ is the Lipschitz constant of ∇f , and ∇ denotes the gradient of f [CSZ14]. The solution obtained by PIRA then has the attractive properties defined in the following lemmas.

Lemma 3. *The sequence L^k generated by Algorithm 1 satisfies the following properties: (1) $F(L^k)$ decreases monotonically; (2) the sequence L^k is bounded; (3) $\sum_{k=1}^{\infty} \|L^k - L^{k+1}\|_F^2 \leq \frac{2F(L^1)}{\mu_1 - L_f}$.*

Lemma 4. *Let L^k generated by Algorithm 1 be bounded as shown in Lemma 3. Any accumulation point L^* of L^k is then a stationary point.*

Thus, the objective function value (2.5) monotonically decreases, and any limit point of L^k is a stationary point. Then, we empirically demonstrate the convergence of the subproblems 2.4 2.10 and the global problem 2.2. To this end, let f denote the subproblem to recover L , g denotes the subproblem to recover S , and W is the global problem. In the k -th iteration, the objective function of L problem is $f(L^{k+1}; S^k)$, the objective function of S is $g(S^{k+1}; L^k)$, and the global objective function is $W(L^k, S^k)$. The Figure 1 reports the convergence of the objective function values in respond to iteration numbers. Specifically, in the k -th iteration, the y-value of Figure 1 (a) is $\|f(L^{k+1}; S^k) - f(L^k; S^{k-1})\|_F$. It is shown that the algorithm converges with limited rounds of iterations.

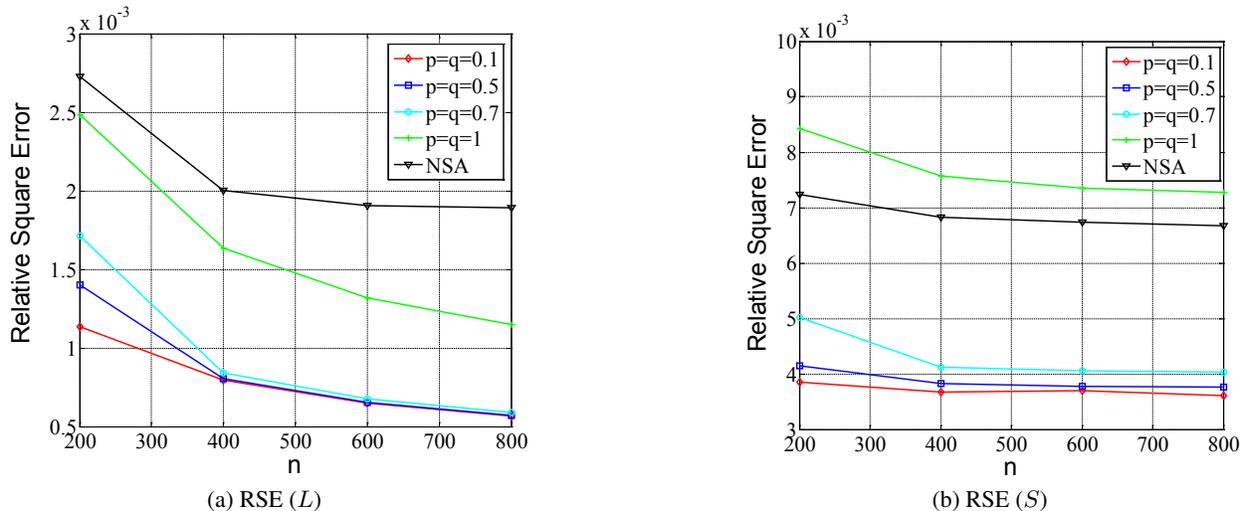


Figure 3: Comparison of low-rank and sparse matrix recovery with varying matrix sizes $n = \{200, 400, 600, 800\}$, $\sigma = 0.01$, $r = 10$. In the case of $p = q = 1$, our method p, q -PCP is the same as SPCP.

4 Experiments

In this section, we conduct experiments on both synthetic and real visual data to validate the effectiveness of our proposed method p, q -PCP. In the experiment on synthetic data, we mainly discuss the influence of various p, q values and noise levels. For the real-world data sets, we test our method in multiple tasks, such as image denoising and light/shadow removal.

The comparative algorithms include the classical convex solution SPCP [ZLW⁺10], Non-Smooth Augmented Lagrangian algorithm (NSA) [AGI11] and Truncated Nuclear Norm Regularization (TNNR) [HZY⁺12]. We use the solver based on ADMM to solve a subproblem of TNNR in the release codes (denoted as TNNR-ADMM). As TNNR-ADMM could only recover the authentic structure, we compare our method with it in the application of image denoising. For the parameters in our algorithm, μ_1 is set the same as μ_2 (2.1) (empirical value). ε is initialized to be $1e - 3$ and decreases to ε/p ($p=1.1$) after each iteration. λ_1 and λ_2 are tuned using 3-fold cross validation. Similarly, we tune the parameters for the comparative algorithms.

All the experiments are conducted with Matlab on a PC with Intel Core2 Q9550 2.8G CPU and 8G RAM.

4.1 Synthetic Data

In this experiment, we verify the effectiveness and robustness of our algorithm by comparing with NSA and SPCP. For each setting of parameters, we report the average result over 10 trials. We generate a rank- r matrix as $L = UV^T$, where U and V are $n \times r$ generated following Gaussian distribution. The zero elements of S are sampled with probability $1 - \rho_s$ and the non-zero entries are sampled in the interval $[-5, 5]$ with probability ρ_s ($\rho_s = 0.2$). We further add Gaussian noise ξ with i.i.d. $N(0, \sigma^2)$. $X = L + S + \xi$ is then the observed recovered matrix.

We first examine our proposed p, q -PCP algorithm with different p and q . The data are generated with different noise levels, i.e., $\sigma = 0.01$ and $\sigma = 0.1$. For simplicity, we set $p = q$. When $p = q = 1$, our p, q -PCP model is actually the SPCP model. We measure the recovery performance based on the Relative

Table 1: Comparison of low-rank and sparse matrix recovery with varying underlying ranks of data.

Rank(L)	Algorithm	rank(L^*)	$\ E\ _0$	$\frac{\ L^* - L\ _F}{\ L\ _F} (\times 10^{-3})$	$\frac{\ S^* - S\ _F}{\ S\ _F} (\times 10^{-3})$	$\sum I(E_{ij} = E_{ij}^*)$
$r = 5$	NSA	118	23041	3.32	6.50	0.9986
	SPCP	24	9635	2.33	7.43	0.8262
	p, q -PCP	5	8023	1.12	3.75	0.9911
$r = 15$	NSA	118	23209	4.67	14.36	0.9978
	SPCP	34	10348	3.20	10.87	0.7704
	p, q -PCP	15	7969	1.14	4.06	0.9919
$r = 20$	NSA	120	23452	14.66	50.42	0.9941
	SPCP	50	12569	6.83	23.78	0.6297
	p, q -PCP	20	7921	1.25	4.87	0.9870
$r = 25$	NSA	122	23825	36.82	142.39	0.9775
	SPCP	83	18483	21.95	84.89	0.4249
	p, q -PCP	26	8077	3.25	14.27	0.9740

Squared Error (RSE) of the low rank part L and sparse part S as

$$\text{RSE}(L) = \frac{\|L - L^*\|_F}{\|L\|_F}, \quad (4.1)$$

$$\text{RSE}(S) = \frac{\|S - S^*\|_F}{\|S\|_F}, \quad (4.2)$$

where L^* and S^* are the recovered matrices. The experimental results are shown in Figure 2. It can be seen that p, q -PCP achieved better recovery performance with smaller values of p and q . The performance of p, q -PCP with $p = q = 0.5$ is compatible with that for $p = q = 0.1$.

To verify the effectiveness and robustness of our algorithm, we further design two experiments for comparison with other methods. One is to vary the underlying rank r of the observed data, and the other to vary the dimension n of the matrix. The experimental settings are as follows:

- We fix $\sigma = 0.01$ and $n = 200$, and vary r in the set $\{5, 15, 20, 25\}$;
- We fix $r = 10$ and $\sigma = 0.01$, and vary n in the set $\{200, 400, 600, 800\}$.

Following the two directions n and r , the experimental results for the different algorithms are shown in Table 1 and Figure 3. In all cases, our algorithm outperforms NSA and SPCP in terms of the rank of L and the sparsity of E . Specifically, Figure 3 records the relative square error of low-rank matrices and sparse matrices. It is obvious that our algorithm with three representative p, q values all outperforms NSA and SPCP. For our own algorithm p, q -PCP, the differences between $p, q = 0.1$ and $p, q = 0.5$ are limited. Thus, in the following experiments, we adopt $p, q = 0.5$ to test our algorithm. Table 1 shows comprehensive results of the recovery errors related to L and E (column 4-5), and the accuracy of the captured sparse location (column 6). L^* and S^* are the solutions obtained using different algorithms, and L, S are the groundtruth matrices. $\|E^*\|_0$ represents the number of non-zero entries of E^* . $\sum I(E_{ij} = E_{ij}^*)$ records the percentage of correctly located entries. From the results shown in Table 1, we see that our algorithm and NSA both perform much better than SPCP in capturing the sparse locations in matrix E . For recovering the low-rank and sparse matrix, our algorithm obtains the best approximation. In particular, in the case of $r = 25$, compared with NSA (36.82×10^{-3}) and SPCP (21.95×10^{-3}), we obtain 3.25×10^{-3} relative error of L . Furthermore, our algorithm achieves 14.27×10^{-3} RSE of E , which is much better than NSA (142.39×10^{-3}) and SPCP (84.89×10^{-3}).



Figure 4: Comparison of image recovery using different low rank approximation algorithms. The images in the second and last row are the amplified patches circled by the red Q4 line in the previous row.

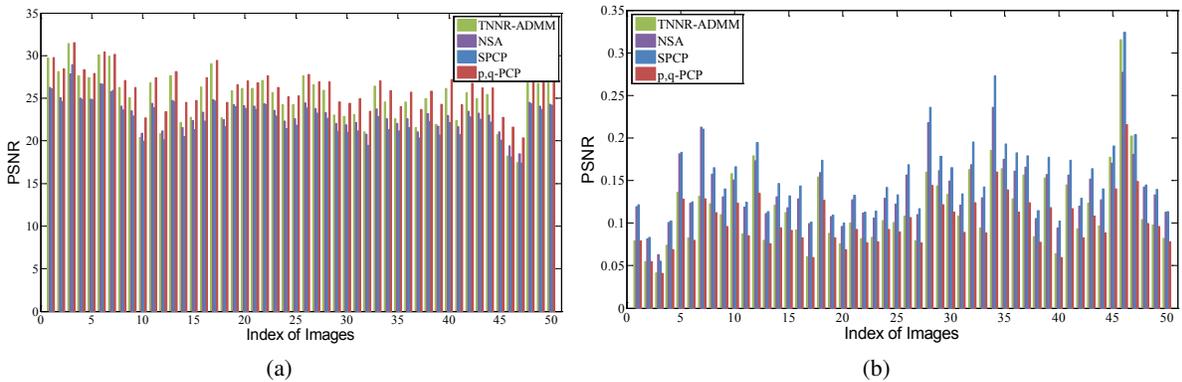


Figure 5: Comparison of the relative square error (a) and PSNR values (b) on the 50 natural images.

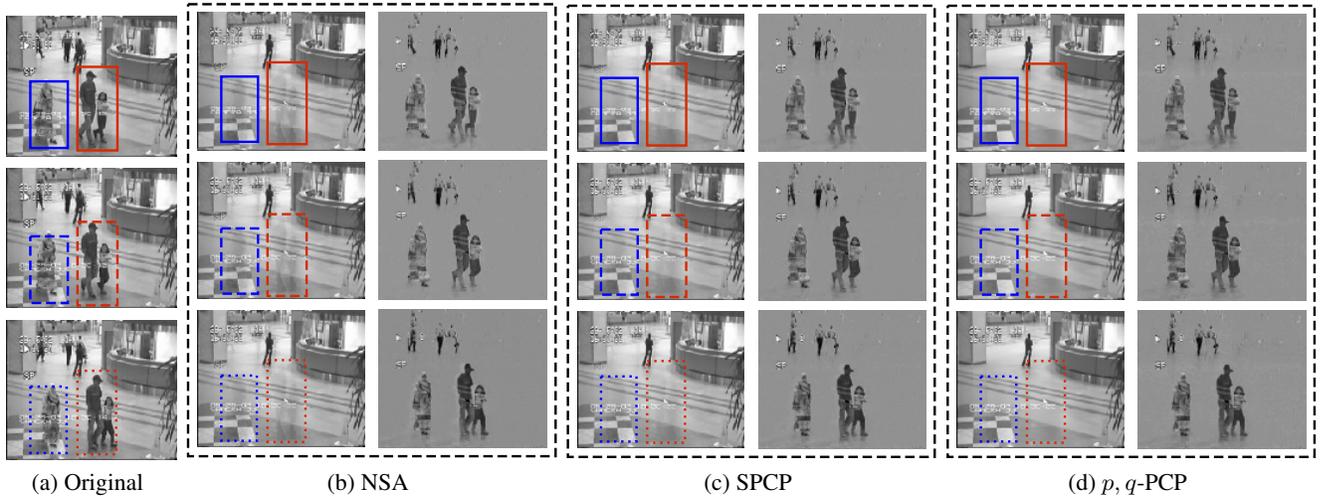


Figure 6: Background extraction using different algorithms. Because of limited space, we only plot the results of three frames (corresponding to three rows).

4.2 Image Denoising

The real images may be corrupted by different types of noise. In this experiment, we apply the low-rank approximation algorithms for image denoising. We download 50 images from the Google image search engine. The images are with three channels and of the same size 300×300 . We add Gaussian noise $N(0, 0.2^2)$ to 50% of the pixels of each image. Note that the color image consists of three channels: L_{red} , L_{green} and $L_{\text{blue}} \in R^{m \times n}$. We implement the principal component analysis algorithms for each channel respectively. The image is then recovered by combining the recovered results from the three channels. Some recovered images are shown in Figure 4. It can be seen that our method achieves the best image recovery performance. The images recovered by NSA and SPCP are blurred and some important details are missing. The recovered images of TNNR-ADMM are much clear than NSA and SPCP, but still not as good as our method.

We measure the recovery performance based on the RSE(L) defined in (4.1) and the PSNR (Peak Signal Noise Ratio) [HTG08] value. Figure 5 plots the RSE and PSNR values on 50 tested images. It can be seen that our algorithm obtains the highest PSNR values and the smallest RSE for all the images. Such results indicate that the our low-rank approximation is better than the traditional nuclear norm heuristic in this situation. Though TNNR-ADMM is non-convex method, it is still inferior to our method. It demonstrates the effectiveness of our model and the optimization method.

4.3 Other Applications

To test the generalization of our method, we design two algorithms for two more applications, separation of foreground and background for a surveillance video and removal of light/shadow from facial images.

Specifically, we first apply different approaches to background separation and detection of objects in the foreground in an airport surveillance video [LHGT04]. It contains a sequence of 201 grayscale frames of 144×176 pixels during an observation time period. The size of the observed data is $X \in R^{25344 \times 201}$.

Figure 6 shows the background extraction results using NSA, SPCP and p, q -PCP algorithms. It can be seen that all three methods are able to separate the background and foreground. To further compare the background recovery results clearly, we mark the ghosting parts of the NSA with rectangles, and mark the same parts with the same rectangles in the original frame, as well as in the recovered background of SPCP

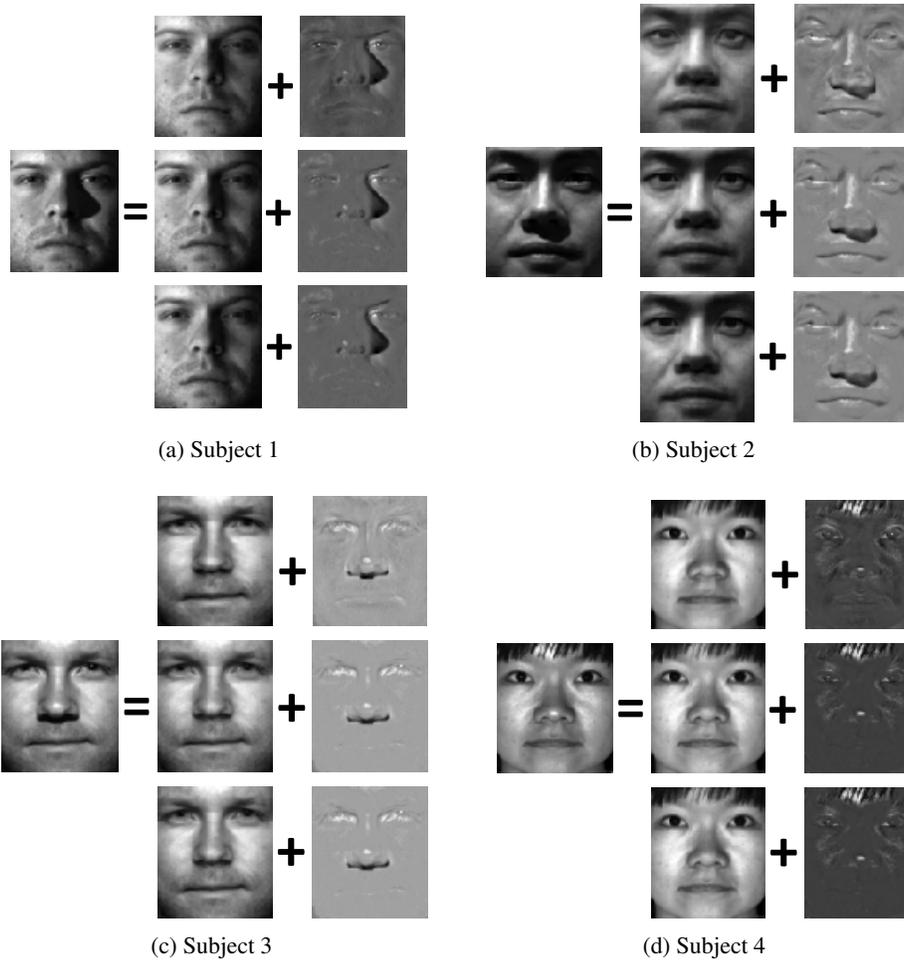


Figure 7: Shadow removal results for facial images of Subject 1 (a) and Subject 2 (b); Light removed from facial images of Subject 3 (c) and Subject 4 (d). All the images are from the YaleB database. In each subfigure, the first row is the recovered low rank and sparse part by NSA. The second and third rows are the results of SPCP and our method.

and p, q -PCP. Although the SPCP has a better recovered background than NSA, there are ghosting shadows in the recovered background. Our algorithm best recovers the background, removing almost all the shadows.

We further apply the principal component pursuit algorithms for removing light and shadow on facial images. Such processing is usually very important for face recognition. The reason that PCP can be applied for removing light/shadow on a face is that the captured facial image can be regarded as the sum of the low rank part (common face) and sparse errors (e.g., light/shadow). We test the competing algorithms on the YaleB data set, which consists of 38 subjects under different illumination. Each subject has 64 images with resolution 192×168 . Thus the size of the observed matrix is 32256×2432 . We apply NSA, SPCP and p, q -PCP on this data set, and plot four example images (1 per individual) in Figure 7. It can be seen that the shadow and light on the faces are removed, and the recovered facial images are very clear. This verifies the effectiveness of our proposed algorithm. The three methods perform well in this situation.

5 Conclusions and Future Work

This study investigates the use of the Schatten- p norm and the ℓ_q -norm regularized non-convex principal component pursuit. We further develop an iteratively reweighted algorithm PIRA to solve the non-convex problem. In each iteration, PIRA solves two sub-problems which have closed form solutions. The obtained L and R are optimal for the linearized sub-problem. We demonstrate the convergence of the objective function based on the calculated low rank and sparse parts by experiments. We present extensive experiments on both synthetic and real-world data to demonstrate the attractive properties and effectiveness of our algorithm. Interesting future work will be applying the joint Schatten- p norm and ℓ_q -norm to other low-rank and sparse problems in the area of video denoising.

References

- [AEB06] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.
- [AGI11] Necdet Serhat Aybat, Donald Goldfarb, and Garud Iyengar. Fast first-order methods for stable principal component pursuit. *arXiv preprint arXiv:1105.2126*, 2011.
- [BT09] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [CDC13] Kun Chen, Hongbo Dong, and Kung-Sik Chan. Reduced rank regression via adaptive nuclear norm penalization. *Biometrika*, 2013.
- [CGWY11] Xiaojun Chen, Dongdong Ge, Zizhuo Wang, and Yinyu Ye. Complexity of unconstrained $\ell_2 - \ell_p$ minimization. *Mathematical Programming*, pages 1–13, 2011.
- [CHH13] Hsien-Hsin Chou, Ling-Yuan Hsu, and Hwai-Tsu Hu. Turbulent-pso-based fuzzy image filter with no-reference measures for high-density impulse noise. *IEEE Transactions on Cybernetics*, 43(1):296–307, 2013.
- [CLMW11] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), 2011.
- [CSZ14] Lu Canyi, Yan Shuicheng, and Lin Zhouchen. Generalized nonconvex nonsmooth low-rank minimization. In *CVPR*, 2014.
- [CWB08] Emmanuel J Candès, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5-6):877–905, 2008.
- [CZ] Xiaojun Chen and Weijun Zhou. Convergence of the reweighted ℓ_1 minimization algorithm for $\ell_2 - \ell_p$ minimization. *to appear in Comp. Optim. Appl.*
- [ER06] Ramin Eslami and Hayder Radha. Translation-invariant contourlet transform and its application to image denoising. *IEEE Transactions on Image Processing*, 15(11):3362–3374, 2006.
- [Faz02] Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.
- [FL01] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.
- [FL09] S. Foucart and M.J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q < 1$. *Applied and Computational Harmonic Analysis*, 26(3):395–407, 2009.
- [HTG08] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.
- [HYZ08] Elaine T Hale, Wotao Yin, and Yin Zhang. Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

- [HZY⁺12] Yao Hu, Debing Zhang, Jieping Ye, Xuelong Li, and Xiaofei He. Fast and accurate matrix completion via truncated nuclear norm regularization. *Pacific Journal of Optimization*, 35(9):2117–213, 2012.
- [JLSX10] Hui Ji, Chaoqiang Liu, Zuwei Shen, and Yuhong Xu. Robust video denoising using low rank matrix completion. In *CVPR*, pages 1791–1798, 2010.
- [JY09] Shuiwang Ji and Jieping Ye. An accelerated gradient method for trace norm minimization. In *ICML*, pages 457–464, 2009.
- [LCM09] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *UIUC Technical Report UILU-ENG-09-2215*, 2009.
- [Lee80] Jong-Sen Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):165–168, 1980.
- [LHGT04] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu, and Qi Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472, 2004.
- [LLY⁺13a] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.
- [LLY⁺13b] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.
- [LO00] Xin Li and Michael T Orchard. Spatially adaptive image denoising under overcomplete expansion. In *ICIP*, volume 3, pages 300–303, 2000.
- [LXY13] Ming-Jun Lai, Yangyang Xu, and Wotao Yin. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization. *SIAM Journal on Numerical Analysis*, 51(2):927–957, 2013.
- [MDYY11] Yadong Mu, Jian Dong, Xiaotong Yuan, and Shuicheng Yan. Accelerated low-rank visual recovery by random projection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2609–2616, 2011.
- [MF12] Karthik Mohan and Maryam Fazel. Iterative reweighted algorithms for matrix rank minimization. *Journal of Machine Learning Research*, 13:3441–3473, 2012.
- [NHD12] Feiping Nie, Heng Huang, and Chris HQ Ding. Low-rank matrix recovery via efficient Schatten p -norm minimization. In *AAAI*, 2012.
- [NWC⁺12] Feiping Nie, Hua Wang, Xiao Cai, Heng Huang, and Chris Ding. Robust matrix completion via joint Schatten p -norm and ℓ_p -norm minimization. In *ICDM*, pages 566–574, 2012.
- [PGW⁺10] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *CVPR*, pages 763–770, 2010.

- [PSWS03] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003.
- [RFP10] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.
- [SCD02] J-L Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2002.
- [SLL14] Ling Shao, Li Liu, and Xuelong Li. Feature learning for image classification via multiobjective genetic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7):1359 – 1371, 2014.
- [SYLL08] Ling Shao, Ruomei Yan, Xuelong Li, and Yan Liu. From heuristic optimization to dictionary learning: a review and comprehensive comparison of image denoising algorithms. *IEEE Transactions on Cybernetics*, 44(7):1001–1013, 2008.
- [TY10] Kim-Chuan Toh and Sangwoon Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(15):615–640, 2010.
- [WLLL13] Zhangyang Wang, Houqiang Li, Qing Ling, and Weiping Li. Robust temporal-spatial decomposition and its applications in video processing. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(3):387–400, 2013.
- [WYG⁺09] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.
- [YSL13] Ruomei Yan, Ling Shao, and Yan Liu. Nonlocal hierarchical dictionary learning using wavelets for image denoising. *IEEE Transactions on Image Processing*, 22(12):4689 – 4698, 2013.
- [ZLW⁺10] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel Candes, and Yi Ma. Stable principal component pursuit. In *ISIT*, pages 1518–1522, 2010.
- [ZS14] Fan Zhu and Ling Shao. Weakly-supervised cross-domain dictionary learning for visual recognition. *International Journal of Computer Vision*, 109(1-2):42–59, 2014.
- [ZT11] Tianyi Zhou and Dacheng Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *ICML*, pages 33–40, 2011.
- [ZYZH13] Haichao Zhang, Jianchao Yang, Yanning Zhang, and Thomas S Huang. Image and video restorations via nonlocal kernel regression. *IEEE Transactions on Cybernetics*, 43(3):1035–1046, 2013.