IEICE TRANS. INF. & SYST., VOL.E99–D, NO.5 MAY 2016

PAPER

Non-Convex Low-Rank Approximation for Image Denoising and Deblurring

Yang LEI[†a)], **Zhanjie SONG**[†], ***Nonmembers*, *and*** **Qiwei SONG**[†], ***Student Member***

SUMMARY Recovery of low-rank matrices has seen significant activity in many areas of science and engineering, motivated by theoretical results for exact reconstruction guarantees and interesting practical applications. Recently, numerous methods incorporated the nuclear norm to pursue the convexity of the optimization. However, this greatly restricts its capability and flexibility in dealing with many practical problems, where the singular values have clear physical meanings. This paper studies a generalized non-convex low-rank approximation, where the singular values are in l_p−heuristic. Then specific results are derived for image restoration, including denoising and deblurring. Extensive experimental results on natural images demonstrate the improvement of the proposed method over the recent image restoration methods.

key words: *low-rank approximation, nuclear norm, image restoration, non-convex optimization*

1. Introduction

Low-rank approximation, as a dimensionality reduction technique for data analysis, which aims to recover the underlying low-rank matrix from its degraded observation, is taking on increasing importance in myriad applications such as machine learning, control, and computer vision. In various low-level vision tasks such as image restoration, the matrix formed by nonlocal similar patches collected from natural images is of low-rank, i.e., the low-rank prior exists to characterize the nonlocal self-similarity for a wide range of natural images [1]–[3]. Although a flurry of studies on low-rank matrix approximation have been reported due to the rapid development of convex and non-convex optimization techniques, comparatively little attention has been paid to low-level vision. In this paper, we intend to concentrate on a generalized non-convex low-rank approximation approach to the image restoration problem.

The literature on low-rank matrix approximation with respect to the Frobenius norm falls into two categories: the low-rank matrix factorization (LRMF) methods [3]–[6] and the nuclear norm minimization (NNM) methods [7]. In the presence of noise-corrupted observation, LRMF aims to find a matrix, which can be factorized into the product of two low-rank matrices. The subject of matrix factorization has been extensively studied, ranging from the classical singular value decomposition (SVD) to the many l_1-norm robust LRMF algorithms. Unfortunately, LRMF is generally time consuming due to the non-convexity and very sensitive to initialization. By employing the nuclear norm as a convex surrogate for the non-convex LRMF problem, NNM has been attracting great research interest in recent years [7]. Recent progress [8] proves that NNM can recover most low-rank matrices by singular value thresholding (SVT) method under some general constraints. While many researchers have investigated NNM, it still has some problems in practice. Compared to the basic rank function in which all nonzero singular values have equal contributions, the standard nuclear norm treats each singular value by adding them together. This ignores the prior knowledge we often have on the matrix's singular values. For instance, in the application of computer vision, the column vectors in the matrix often lie in a two-dimensional subspace, the larger singular values are generally associated with the major projection orientations, and thus they should be shrunk to a lesser extent to preserve the major data components. Hu et al., [9], proposed a truncated nuclear norm regularization (TNNR) method, which is given by the nuclear norm subtracted by the sum of the largest few singular values. However, TNNR is not flexible enough since it makes the decision of whether to regularize a singular value or not. To improve the flexibility of nuclear norm, Gu et al., [3] proposed the weighted nuclear norm method (WNNM) and study its minimization. However, this kind of relaxation suppresses the low-rank components and shrinks the reconstructed data.

Note that when X is restricted to be a diagonal matrix, the nuclear norm minimization problem reduces to the compressed sensing problem. Wang et al., [10] proved that in sparse compressed sensing l_p-norm minimization is more like l_0-norm minimization than l_1-norm. Actually, both theoretical analysis and numerical experiments have shown that the solution of l_p-norm sparse coding ($0 \leq p < 1$) is close to that of the l_1-minimization and it is sparser. In image restoration, it has been shown that the image gradients of the natural images can be better modeled with hyper-Laplacian distribution with $0.5 \leq p \leq 0.8$ [27]. This work, following previous research, studies in detail a generalized non-convex heuristic defined by $\mathcal{J}(X; p) := \sum_i |\sigma_i(X)|^p$, $(0 \leq p \leq 1)$. Based on a generalized singular value thresholding (GSVT) operator, the proposed non-covnex low-rank (NCLR) algorithm is as efficient as that of the NNM problem, and greatly improves the flexibility of WNNM.

The contribution of this paper is threefold. First, we analyze in detail the non-covnex low-rank optimization problem and provide the solutions by generalized singular value

Manuscript received August 6, 2015.
Manuscript revised December 24, 2015.
Manuscript publicized February 4, 2016.
[†]The authors are with Tianjin University, China.
a) E-mail: lei000yang@sina.com (Corresponding author)
DOI: 10.1587/transinf.2015EDP7307

thresholding operator. Second, we adopt the proposed NCLR algorithm to image restoration to demonstrate its great potentials in low level vision applications. Third the experimental results showed that NCLR outperforms state-of-the-art image restoration algorithms not only in PSNR index but also in local structure preservation, leading to visually more pleasant outputs.

2. Related Work

As reviewed to the Sect. 1, various algorithms have been proposed for the low-rank approximation. Based on the problems in nuclear norm minimization, we provide a brief survey and discussion on NNM [7] and WNNM [3]. The nuclear norm of a matrix X, denoted by $\|X\|_*$, is defined as the sum of its singular values, i,e,. $\|X\|_* = \sum_i |\sigma_i(X)|_1$, where $\sigma_i(X)$ means the i-th singular value of X. Given matrix Y, NNM aims to approximate Y by X, while minimizing the nuclear norm of X,

$$\min_X \lambda \sum_i |\sigma_i(X)| + \frac{1}{2}\|X - Y\|_F^2, \tag{1}$$

where λ is a positive constant. Theoretical studies showed that the nuclear norm is the tightest convex lower bound of the rank function of the matrices [11]. The relationship between nuclear norm and the rank of matrices is similar to the relationship between the l_1-norm and the l_0-norm of vectors. Cai et al., [7] proved that the NNM based low-rank matrix approximation problem with F-norm data fidelity can be easily solved by a soft-thresholding operator on the singular values of observation matrix i.e., $\hat{X} = U\mathcal{S}_\lambda(\Sigma)V^T$, where $Y = U\Sigma V^T$ is the SVD of Y and $\mathcal{S}_\lambda(\Sigma)$ is the soft-thresholding function on diagonal matrix Σ with parameter λ. The above singular value thresholding (SVT) method has been widely adopted to solve many NNM based problems, such as matrix completion, robust principle component analysis (RPCA), low-rank textures, and low rank representation (LRR) for subspace clustering.

Gu et al., [3] proposed the WNNM method to improve the flexibility of nuclear norm. The weighted nuclear norm of a matrix X is defined as $\|X\|_{w,*} = \sum_i |w_i\sigma_i(X)|$, where $w = [w_i, \ldots, w_n]$ is a non-negative weight vector assigned to $\sigma_i(X)$.

$$\min_X \lambda \sum_i w_i|\sigma_i(X)| + \frac{1}{2}\|X - Y\|_F^2, \tag{2}$$

The weighted nuclear norm minimization (WNNM) is not convex in general case, and it is more difficult to solve than NNM. So far little work has been reported on the WNNM problem. Gu et al analyzed the WNNM optimization problem in detail and provide the solutions under different weight conditions (see [3]) Unfortunately, WNNM suppresses the low-rank components and shrinks the reconstructed data, and was only used in image denoising.

In order to improve the capability and flexibility in dealing with image restoration problems, we consider a generalized non-convex low-rank (NCLR) approximation in this paper. The remainder of this paper is organized as follows. Section 3 investigates the NCLR algorithm with non-convex low-rank regularization. Section 4 presents the modeling of image restoration, and then provides the proposed NCLR algorithm for solving the image restoration problem. Some numerical results on both simulated and real data are given in Sect. 5. Section 6 summarizes our work.

3. The Basic Non-Convex Low-Rank Minimization

In this section, we study a novel and general regularization defined by $\mathcal{J}(X;p) := \sum_i |\sigma_i(X)|^p$ $(0 \le p \le 1)$, the NCLR optimization can be expressed as follows:

$$\min_X \frac{1}{2}\|Y - X\|_F^2 + \lambda\mathcal{J}(X;p). \tag{3}$$

Before proceeding with the solution of the NCLR we establish a key result (see Theorem 1 below) that will be crucial for the NCLR.

Theorem 1: For any $Y \in \mathbb{R}^{m\times n}$, $m \ge n$, and let $Y = U\Sigma V^T$ be the singular value decomposition of Y, where $\Sigma = \begin{pmatrix} diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \\ 0 \end{pmatrix}$. The solution of the non-convex problem in (3) can be expressed as $\hat{X} = UDV^T$, where $D = \begin{pmatrix} diag(d_1, d_2, \ldots, d_n) \\ 0 \end{pmatrix}$ is a diagonal non-negative matrix and $(d_1, d_2, \ldots, d_n)$ is the solution of the following non-convex optimization problem:

$$\min_{d_1,d_2,\ldots,d_n} \sum_{i=1}^n \frac{1}{2}(d_i - \sigma_i)^2 + \lambda|d_i|^p. \tag{4}$$

(The proof of Theorem 1 see Appendix.)

In [12] the author proposed a generalized soft-thresholding (GST) function for solving the non-convex l_p−minimization problem $\min_x \frac{1}{2}(y-x)^2 + \lambda|x|^p$ with $0 \le p \le 1$,

$$\mathcal{D}_p^G(y;\lambda) = \begin{cases} 0 & \text{if } |y| \le \tau_p^G(\lambda) \\ \operatorname{sgn}(y)\mathcal{S}_p^G(|y|;\lambda) & \text{if } |y| > \tau_p^G(\lambda) \end{cases} \tag{5}$$

where $\tau_p^G(\lambda) = (2\lambda(1-p))^{1/(2-p)} + \lambda p(2\lambda(1-p))^{(p-1)/(2-p)}$, and $\mathcal{S}_p^G(|y|;\lambda)$ is obtained by solving $\mathcal{S}_p^G(|y|;\lambda) - y + \lambda p\left(\mathcal{S}_p^G(|y|;\lambda)\right)^{p-1} = 0$. Following the research of [12], the optimization solution of (3) can be expressed by

$$\hat{X} = U\mathcal{G}_\lambda^p(\Sigma)V^T, \tag{6}$$

where $Y = U\Sigma V^T$ is SVD of Y, and $\mathcal{G}_\lambda^p(\Sigma)$ is the generalized singular value thresholding (GSVT) operator $\mathcal{G}_\lambda^p(\Sigma)_{ii} = \mathcal{D}_p^G(\Sigma_{ii};\lambda)$. In fact the convergence of the GST operator is analyzed in [12]. It is also applied in this non-convex low-rank scenario. In [12], a generalized iterated shrinkage algorithm (GISA) was proposed for l_p−norm minimization. GISA was used to solve the vector optimization with a fixed redundant matrix for sparse coding based image restoration.

However, in this paper, the redundant matrix for sparse representation is unfixed. We propose a novel low-rank optimization framework in Eq. (3) to simultaneously learn the adaptive redundant matrix and the sparse coefficients. On the one hand, the novel framework extend NNM to a non-convex relaxation which depends on the sum of singular values in l_p−based heuristic $\mathcal{J}(X; p)$. The extended GST operator in Eq. (6) is derived to solve the new objective and followed by the rigorous theoretical proof in Appendix. On the other hand, by blocking the vectorized nonlocal similar image patches into an approximate low-rank matrix, in Sect. 4.1, NCLR is used to simultaneously learn the patch-wise sparse coefficients and the adaptive sub-dictionary.

4. NCLR for Image Restoration

As important applications, in this section we adopt the proposed NCLR algorithm to the classic image restoration problem: an ideal image x is measured in the presence of a linear system with an additive zero-mean white and homogeneous Gaussian noise

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{7}$$

where $\mathbf{y}$ is the observed measurement, $\mathbf{H}$ is a degradation matrix, $\mathbf{x}$ is the unknown original image vector and $\mathbf{n}$ is the additive noise vector with standard deviation σ_n. With different settings of degradation matrix, Eq. (7) can represent different IR problems: when $\mathbf{H}$ is an identity matrix, the problem is image denoising; and the problem becomes deblurring when $\mathbf{H}$ is specified by a blurring kernel[13]–[15]. Image restoration is not only an important pre-processing step for many vision applications, but also an ideal test bed for evaluating statistical image modeling methods. The seminal work of nonlocal means [16] triggers the wide study of nonlocal self-similarity (NSS) based methods for image restoration. NSS refers to the fact that there are many repeated local patterns across a natural image, and those nonlocal similar patches to a given patch can help much the reconstruction of it. Recently, efforts have been made to improve the restored image quality by exploiting the NSS, such as the nonlocal means (NLM) methods [17], Block-matching 3D filtering (BM3D) [18], [19], centralized sparse representation (CSR) [20] and nonlocal centralized sparse representation (NCSR) [21]. Intuitively, by stacking the nonlocal similar patch vector into a matrix, this matrix should be a low-rank matrix and has sparse singular values. This assumption is validated by Wang et al., [17] where they called it the nonlocal spectral prior. In [1], Dong et al., combined NNM and $L_{2,1}$−norm group sparsity for image restoration, and demonstrated very competitive results. In [2], a simultaneous sparse coding (SSC) scheme was proposed to code similar patches simultaneously and achieved impressive restoration results. In this section, we introduce the use of NCLR to solve the SSC model.

4.1 Simultaneous Sparse Coding by NCLR

Instead of sparsely coding each patch individually, the previous NNM model simultaneously codes a set of patches. Mathematically, the model can be expressed as follows:

$$(\mathbf{U}_i, \mathbf{A}_i) = \arg\min_{\mathbf{A}_i} \left\{ \|\mathbf{X}_i - \mathbf{U}_i\mathbf{A}_i\|_F^2 + \tau_i J(\mathbf{A}_i) \right\} \tag{8}$$

where $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \ldots, \mathbf{x}_{i,m}] \in \mathbf{R}^{n\times m}$, $\mathbf{A}_i = [\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,m}] \in \mathbf{R}^{n\times m}$ denote dataset and sparse coefficients matrix, respectively [16], $\mathbf{U}_i \in \mathbf{R}^{n\times m}$ denotes the dictionary, τ_i is a regularization parameter and $J(A_i)$ denotes the regularization term. The dataset $\mathbf{X}_i$ is constructed by grouping the K-nearest-neighbor (K-NN) nonlocal similar patches to $\mathbf{x}_{i,1}$ (i.e., $m = K + 1$). In this paper, $J(\mathbf{A}_i)$ is defined as $J(\mathbf{A}_i) = \|\mathbf{A}_i\|_{p,q} = \sum_{l=1}^{n} \left\|\mathbf{A}_i^l\right\|_q^p$ with $0 \le p \le 1$ and $q = 2$, where $\mathbf{A}_i^l$ denotes the vector of the l-th row of matrix $\mathbf{A}_i$. Note that, if $p = 1$, model in Eq. (8) will be recast to the conventional NNM model in Eq. (2). In Sect. 5.1 we show that $0.6 \le p \le 0.8$ is better than that with $p = 1$ in terms of suppressing noise.

Suppose $\sigma_{i,l}$ denotes the standard deviation of the sparse coefficients $\mathbf{A}_i^l$ in the l-th row, then

$$J(\mathbf{A}_i) = \sum_{l=1}^{n} \omega_{i,l} \left(\sum_{k=1}^{m} \alpha_{i,k}(l)^2 \right)^{p/2} = \sum_{l=1}^{n} m^{p/2} \sigma_{i,l}^p. \tag{9}$$

Substituting Eq. (9) into Eq. (8), we obtain the following

$$(\mathbf{U}_i, \mathbf{A}_i) = \arg\min_{\mathbf{A}_i} \|\mathbf{X}_i - \mathbf{U}_i\mathbf{A}_i\|_F^2 + \tau_i m^{p/2} \sum_{l=1}^{n} \sigma_{i,l}^p. \tag{10}$$

Now we consider the decomposition of the matrix $\mathbf{A}_i$ first appeared in [11]: given a diagonal matrix $\Lambda_i = diag\{\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,r}\}$ and a matrix $\mathbf{V}_i \in \mathbf{R}^{K\times r}$ defined as $\mathbf{v}_{i,l} = \frac{1}{\lambda_{i,l}}(\mathbf{A}_i^l)^T$, $r = \max\{n, m\}$, then $\mathbf{A}_i = \Lambda_i \mathbf{V}_i^T$. Due to the unitary property of $\mathbf{v}_{i,l}$, we have

$$m\sigma_{i,l}^2 = \|\mathbf{A}_i^l\|_2^2 = \|\lambda_{i,l}\mathbf{v}_{i,l}^T\|_2^2 = \lambda_{i,l}^2. \tag{11}$$

Following Eq. (10) and Eq. (11),

$$(\mathbf{U}_i, \Lambda_i, \mathbf{V}_i) = \arg\min_{\mathbf{U}_i,\Lambda_i,\mathbf{V}_i} \left\|\mathbf{X}_i - \mathbf{U}_i\Lambda_i{\mathbf{V}_i}^T\right\|_F^2 + \tau_i \sum_{l=1}^{r} \lambda_{i,l}^p \tag{12}$$

where $\lambda_{i,l}$ and $\mathbf{v}_{i,l}$ denote the l-th singular value and l-th column of matrix $\mathbf{V}_i$, respectively. Equation (12) is a non-convex low-rank approximation problem which can be solved by the proposed GSVT operator (see Eq. (6)). The sparsity of Λ_i enforce the coding coefficients of exemplar patch to be spatially sparse.

4.2 Adaptive Regularization Parameter

An important issue of the NNM is the selection of the regularization parameters τ_i. According to previous subsection, the sparsity of $\mathbf{A}_i = \Lambda_i \mathbf{V}_i$ is equivalent to the non-convex regularization of the sparse coefficients, i.e.,

$$\alpha_i = \arg\min_{\alpha_i} \sum_{j=1}^{m} \left(\left\| \mathbf{x}_{i,j} - \mathbf{U}_i \alpha_{i,j} \right\|_2^2 + \tau_i \|\alpha_{i,j}\|_p^p \right). \tag{13}$$

In Eq. (13), the parameter τ_i should be adaptively determined for better restoration performance. This subsection provides a Bayesian interpretation to set an adaptive regularization parameter τ_i. In the literature of wavelet denoising, the Maximum a Posterior (MAP) estimator for the regularization parameter has been derived [22]. In this paper, we extend the derivation to the proposed model.

For the convenience of expression, we define $\theta_i := \alpha_i$, where α_i are the concatenation of $\alpha_{i,j}$. The MAP estimation of θ_i can be formulated as:

$$\theta_{\mathbf{X}_i} = \arg\max_{\theta_i} \{\log \mathrm{P}(\mathbf{X}_i|\theta_i) + \log \mathrm{P}(\theta_i)\}. \tag{14}$$

The likelihood term is characterized by the Jointly Gaussian distribution:

$$\begin{aligned} &\log \mathrm{P}(\mathbf{X}_i|\theta_i) = \log \mathrm{P}(\mathbf{X}_i|\alpha_i) \\ &= \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left(-\frac{1}{2\sigma_w{}^2} \left\| \mathbf{x}_{i,j} - \mathbf{U}_i \alpha_{i,j} \right\|_2^2 \right), \end{aligned} \tag{15}$$

where σ_w denotes the standard variance of the additive Gaussian noise, and θ_i are assumed to be independent. As [8] and [10] discussed, the sparse coding noise θ_i can be well characterized by the Laplacian distribution:

$$\mathrm{P}(\theta_i) = \prod_{j=1}^{m} \prod_{l=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}\sigma_i(l)} \exp\left(-\frac{|\theta_i(l)|}{\sigma_i(l)} \right) \right\}, \tag{16}$$

where $\theta_{i,j}(l)$ is the l-th elements of $\theta_{i,j}$, and $\sigma_i(l)$ is the standard deviation of $\theta_{i,j}(l)$.

Following Eq. (14), Eq. (15) and Eq. (16), we obtain:

$$\begin{aligned} &\theta_{\mathbf{X}_i} \\ &= \arg\min_{\theta_i} \sum_{j=1}^{m} (\left\| \mathbf{x}_{i,j} - \mathbf{U}_i \alpha_{i,j} \right\|_2^2 + 2\sqrt{2}\sigma_w^2 \sum_{l=1}^{n} \frac{|\theta_{i,j}(l)|^p}{\sigma_i(l)|\theta_{i,j}(l)|^{p-1}}). \end{aligned} \tag{17}$$

Then, the sparse codes α can be estimated by:

$$\begin{aligned} &\alpha_{\mathbf{X}_i} \\ &= \arg\min_{\alpha_i} \sum_{j=1}^{m} \left(\left\| \mathbf{x}_{i,j} - \mathbf{U}_i \alpha_{i,j} \right\|_2^2 + \sum_{l=1}^{n} \frac{2\sqrt{2}\sigma_w^2}{\sigma_i^p(l)} \left| \alpha_{i,j}(l) \right|^p \right). \end{aligned} \tag{18}$$

Comparing Eq. (18) and Eq. (13), we have

$$\tau_{i,l} = 2\sqrt{2}\sigma_w^2 / \sigma_i^p(l), \tag{19}$$

where $\sigma_i(l)$ denotes standard deviation of the locally estimated variance at the position l.

4.3 Modeling of Image Restoration

According to Eq. (14), the objective function of the restoration method can be formulated as:

$$\begin{aligned} &(\hat{\mathbf{x}}, \hat{\mathbf{U}}_i, \hat{\alpha}_i) \\ &= \arg\min_{\mathbf{x}, \mathbf{U}_i, \alpha_i} (\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \\ &+ \beta \sum_i \sum_{j=1}^{m} \left(\left\| R_{i,j}\mathbf{x} - \mathbf{U}_i \alpha_{i,j} \right\|_2^2 + \tau_i \left\| \alpha_{i,j} \right\|_p^p \right)), \end{aligned} \tag{20}$$

where $R_{i,j}$ is the matrix extracting patch $\mathbf{x}_{i,j}$ from $\mathbf{x}$ at location (i, j), β is an adaptive regularization parameter. Equation (20) can be solved by using a variable-splitting scheme that separates the image $\mathbf{x}$ and sparse coefficients $\{\alpha_i\}$.

Given $\hat{\mathbf{x}}^{(k)}$, the estimates of $\alpha_{i,j}$ and $\mathbf{U}_i$ by solving Eq. (13), then we are alternatively concerned with the following optimization:

$$\hat{\mathbf{x}}^{(k+1)} = \arg\min_{\mathbf{x}} (\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \beta \sum_i \sum_{j=1}^{m} \left\| R_{i,j}\mathbf{x} - \hat{\mathbf{U}}_i \hat{\alpha}_{i,j} \right\|_2^2) \tag{21}$$

which is a quadratic optimization problem that admits a closed form solution:

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{H}^T\mathbf{y} + \beta \sum_i \sum_{j=1}^{m} \left(R_{i,j}\mathbf{x} - \mathbf{U}_i \alpha_{i,j} \right)}{\mathbf{H}^T\mathbf{H} + \beta \sum_i \sum_{j=1}^{m} R_{i,j}{}^T R_{i,j}} \tag{22}$$

The above alternative optimization process can be iterated until convergence.

4.4 Summary of the Algorithm

As we mentioned in Sect. 4, in our NCLR algorithm the non-convex low-rank approximation method is used to code each patch. First, we cluster the patches of observed image $\mathbf{y}$ into M clusters. For a given patch $\mathbf{y}_j$, we check which cluster it falls into by calculating its distances to means of the clusters, i.e., find the similar patch group $\mathbf{Y}_j$. Total L similar patches are selected for each chosen exemplar. Second, the non-convex low-rank approximation method is used to simultaneously learn the adaptive sub-dictionary and sparse coefficients. Last, by Eq. (22) we aggregate the clusters to restore the clean image. Finally, the proposed NCLR algorithm is summarized in **Algorithm 1**.

5. Experimental Results

In this section, we report our experimental results on image denoising and deblurring with NCLR algorithm described in the previous section. These experimental results are used to support the effectiveness of the proposed NCLR image restoration algorithm. All the experiments are conducted and timed on the same workstation with an Intel Xeon E5-2620 2.00GHz CPU that has 4 cores and 32.0GB memory, running Windows 7 and Matlab R2013b.

5.1 Image Denoising

We compare the proposed NCLR based denoising method with several recently developed denoising methods, including the block-matching 3D filtering (BM3D) method [18], [19], expected patch log likehood (EPLL) method [23], the learned simultaneous sparse coding (LSSC) method [24], the nonlocally centralized sparse representation (NCSR) method [21], the spatially adaptive iterative singular-value thresholding (SAIST) method [1] and the weighted nuclear norm minimization (WNNM) method [3]. A set of 14 natural images which commonly used in the literature of image denoising are tested for the comparison. The patch size is 6×6, 7×7, 8×8 and 9×9 for different noise level, respectively. Here we set $\beta = 0.1$ and total $L = 30$ similar patches which are selected for each chosen exemplar.

Table 1 lists the PSNR results of NCLR algorithm on several test images by using $p = 0.5, 0.6, \ldots, 1$. Table 1 shows that $0.7 \le p \le 0.9$ is better than that with $p = 1$ in terms of suppressing noise, which indicates that non-convex low-rank image denoising can much improve the denoising performance.

By setting $p = 0.8$, the PSNR results of the test methods are reported in Table 2–5. The highest PSNR result for each image and on each noise level is highlighted in bold. We have the following observations. The proposed NCLR achieves the highest PSNR in almost every case. It achieves 1.34dB-2.23dB improvement over the NNM method in average and outperforms the WNNM method by 0.1dB-0.2dB in average (up to 0.36dB on image *Straw* with noise level $\sigma_n = 50$) consistently on all the four noise levels. We also list the average CPU time of each algorithm in Table 6. From Table 6, we can observe that the proposed NCLR is faster than LSSC, NCSR, and SAIST. Although NCLR is slower than WNNM in some noise level ($\sigma = 10$, $\sigma = 30$), it achieves better performance than WNNM. In Algorithm 1, let m denotes the width of patch $\mathbf{y}_j$, n denotes the hight of patch $\mathbf{y}_j$, L denotes the number of similar patches, M is the number of clusters and K denotes the iterations, then the computation cost of NCLR is $O((L^2mn + m^3n^3)MK)$. In fact, the main computational cost of NCLR, WNNM, NNM, SAIST and LSSC in each iteration is the computation of the singular value decomposition. Since non-convex relaxation provides better approximation to the original low-rank assumption, the proposed NCLR algorithm is converged with a few iterations. This reduces the computational cost of NCLR. We also illustrate our comparisons graphically by plotting in Fig. 1 the PSNR value versus the number of iterations taken by these algorithms for denoising of image *House* with noise corruption $\sigma = 50$. From Fig. 1 we can see that NCLR outperforms WNNM, NNM, SAIST and LSSC.

In Fig. 2 and Fig. 3, we show the visual quality of the

Algorithm 1

Input: Noisy image
1: Initialize $\hat{\mathbf{x}}^{(0)} = \mathbf{y}$, $\hat{\mathbf{y}}^{(0)} = \mathbf{y}$, the maximum number of iteration K;
2: **for** $k = 1 : K$ **do**
3: Iterative regularization $\mathbf{y}^{(k+1)} = \hat{\mathbf{x}}^{(k)} + \delta^T \mathbf{H}^T(\mathbf{y} - \mathbf{H}\hat{\mathbf{y}}^{(k)})$, where relaxation parameter δ is the pre-determined constant
4: **for** each patch $\mathbf{y}_j^{(k+1)} \in \mathbf{y}^{(k+1)}$ **do**
5: Find similar patch group $\mathbf{Y}_j$
6: Singular value decomposition $[\hat{\mathbf{U}}_j^{(k+1)}, \hat{\Lambda}_j^{(k+1)}, \hat{\mathbf{V}}_j^{(k+1)}] = \mathrm{SVD}(\mathbf{Y}_j)$
7: Get the estimation: $\hat{\mathbf{X}}_j^{(k+1)} = \hat{\mathbf{U}}_j^{(k+1)} \mathcal{G}_\lambda^p(S)(\hat{\mathbf{V}}_j^{(k+1)})^T$
8: **end for**
9: Reconstructed image by aggregate $\hat{\mathbf{X}}_j^{(k+1)}$ to form the clean image $\hat{\mathbf{x}}^{(k+1)}$ by Eq. (22)
10: **end for**
Output: Clean image $\hat{\mathbf{x}}^{(K)}$

Table 1 NCLR based image denoising PSNR(dB) results by different p.

	$p=0.5$	$p=0.6$	$p=0.7$	$p=0.8$	$p=0.9$	$p=1$
	$\sigma_n = 10$					
House	36.95	37.02	37.05	**37.06**	37.03	36.97
Leaves	35.06	35.18	**35.23**	35.21	35.09	34.92
	$\sigma_n = 30$					
Monarch	28.89	28.94	29.02	29.06	**29.08**	28.97
Barbara	30.17	30.21	30.24	**30.26**	30.19	30.14
	$\sigma_n = 50$					
Peppers	26.97	27.03	**27.09**	**27.09**	**27.09**	27.04
F.print	24.73	24.78	24.80	**24.81**	24.75	24.71
	$\sigma_n = 100$					
Starfish	22.32	22.34	22.37	**22.38**	22.37	22.36
Straw	19.87	19.91	19.96	**20.03**	19.97	19.87

Table 2 PSNR(dB) results by different image denoising methods ($\sigma_n = 10$).

	NNM	BM3D	EPLL	LSSC	NCSR	SAIST	WNNM	NCLR
C.Man	32.87	34.18	34.02	34.24	34.18	34.30	34.44	**34.74**
House	35.97	36.71	35.75	36.95	36.80	36.66	36.95	**37.06**
Peppers	33.77	34.68	34.54	34.80	34.68	34.82	34.95	**35.08**
Leaves	33.55	34.04	33.29	34.52	34.53	34.92	35.20	**35.21**
Starfish	32.62	33.30	33.29	33.74	33.65	33.72	33.99	**34.13**
Monarch	33.54	34.12	34.27	34.44	34.51	34.76	35.03	**35.16**
Lena	35.19	35.93	35.58	35.83	35.85	35.90	36.03	**36.15**
Barbara	34.40	34.98	33.61	34.98	35.00	35.24	**35.51**	35.51
Boat	33.05	33.92	33.66	34.01	33.91	33.91	34.09	**34.14**
Hill	32.89	33.62	33.48	33.66	33.69	33.65	33.79	**33.82**
F.print	31.38	32.46	32.12	32.57	32.68	32.69	32.82	**32.85**
Man	32.99	33.98	33.97	34.10	34.05	34.12	34.23	**34.26**
Couple	32.97	34.04	33.85	34.01	34.00	33.96	**34.14**	34.14
Straw	29.84	30.89	30.74	31.25	31.35	31.49	31.62	**31.78**
AVE	32.216	34.061	33.726	34.221	34.206	34.296	34.485	**34.652**

Table 3 PSNR(dB) results by different image denoising methods (σ_n = 30).

	NNM	BM3D	EPLL	LSSC	NCSR	SAIST	WNNM	NCLR
C.Man	27.43	28.64	28.36	28.63	28.59	28.36	28.80	**28.90**
House	30.99	32.09	31.23	32.41	32.07	32.30	32.52	**32.60**
Peppers	28.11	29.28	29.16	29.25	29.10	29.24	29.49	**29.58**
Leaves	26.81	27.81	27.18	27.65	28.14	28.29	28.60	**28.71**
Starfish	26.62	27.65	27.52	27.70	27.78	27.92	28.08	**28.16**
Monarch	27.44	28.36	28.35	28.20	28.46	28.65	28.92	**29.06**
Lena	30.15	31.26	30.79	31.18	31.06	31.27	**31.43**	31.43
Barbara	28.59	29.81	27.57	29.60	29.62	30.14	**30.31**	30.26
Boat	27.82	29.12	28.89	29.06	28.94	28.98	29.24	**29.26**
Hill	28.11	29.16	28.90	29.09	28.97	29.06	**29.25**	29.24
F.print	25.84	26.83	26.19	26.68	26.92	26.95	26.99	**27.07**
Man	27.87	28.86	28.83	28.87	28.78	28.81	29.00	**29.01**
Couple	27.36	28.87	28.62	28.77	28.57	28.72	28.98	**29.03**
Straw	23.52	24.84	24.64	24.99	25.00	25.23	25.27	**25.55**
AVE	27.619	28.756	28.374	28.720	28.714	28.851	29.063	**29.134**

Table 4 PSNR(dB) results by different image denoising methods (σ_n = 50).

	NNM	BM3D	EPLL	LSSC	NCSR	SAIST	WNNM	NCLR
C.Man	24.88	26.12	26.02	26.35	26.14	26.15	26.42	**26.45**
House	27.84	29.69	28.76	29.99	29.62	30.17	30.32	**30.46**
Peppers	25.29	26.68	26.63	26.79	26.82	26.73	26.91	**27.08**
Leaves	23.36	24.68	24.38	24.81	25.04	25.25	25.47	**25.50**
Starfish	23.83	25.04	25.04	25.12	25.07	25.29	25.44	**25.50**
Monarch	24.46	25.82	25.78	25.88	25.73	26.10	26.32	**26.40**
Lena	27.74	29.05	28.42	28.95	28.90	29.01	29.24	**29.31**
Barbara	25.75	27.23	24.82	27.03	26.99	27.51	27.79	**27.84**
Boat	25.39	26.78	26.65	26.77	26.66	26.63	26.97	**26.99**
Hill	25.94	27.19	26.96	27.14	26.99	27.04	**27.34**	27.31
F.print	23.37	24.53	23.59	24.26	24.48	24.52	24.67	**24.81**
Man	25.66	26.81	26.72	26.72	26.67	26.68	26.94	**27.01**
Couple	24.84	26.46	26.24	26.35	26.19	26.30	26.65	**26.67**
Straw	20.99	22.29	21.93	22.51	22.30	22.65	22.74	**23.10**
AVE	24.953	26.312	25.853	26.334	26.257	26.431	26.659	**26.745**

Table 5 PSNR(dB) results by different image denoising methods (σ_n = 100).

	NNM	BM3D	EPLL	LSSC	NCSR	SAIST	WNNM	NCLR
C.Man	21.49	23.07	22.86	23.15	22.93	23.09	23.36	**23.53**
House	23.65	25.87	25.19	25.71	25.56	26.53	26.68	**26.91**
Peppers	21.24	23.39	23.08	23.20	22.84	23.32	23.46	**23.54**
Leaves	18.73	20.91	20.25	20.58	20.86	21.40	21.57	**21.75**
Starfish	20.58	22.10	21.92	21.77	21.91	22.10	22.22	**22.38**
Monarch	20.22	22.52	22.23	22.24	22.11	22.61	22.95	**23.04**
Lena	24.41	25.95	25.30	25.96	25.71	25.93	26.20	**26.25**
Barbara	22.14	23.62	22.14	23.54	23.20	24.07	24.37	**24.40**
Boat	22.48	23.97	23.71	23.87	23.68	23.80	**24.10**	24.10
Hill	23.32	24.58	24.43	24.47	24.36	24.29	**24.75**	24.70
F.print	20.01	21.61	19.85	21.30	21.39	21.62	21.81	**21.90**
Man	22.88	24.22	24.07	23.98	24.02	24.01	24.36	**24.41**
Couple	22.07	23.51	23.32	23.27	23.15	23.21	23.55	**23.58**
Straw	18.33	19.43	18.84	19.43	19.10	19.42	19.67	**19.96**
AVE	21.539	23.196	22.656	23.033	22.916	23.243	23.504	**23.604**

Table 6 CPU time(min) results by different image denoising methods.

	σ_n = 10							
	NNM	BM3D	EPLL	LSSC	NCSR	SAIST	WNNM	NCLR
AVE	4.65	0.03	1.19	18.64	7.15	10.28	4.93	5.49
	σ_n = 30							
AVE	5.80	0.04	1.19	21.84	8.53	12.48	6.35	6.04
	σ_n = 50							
AVE	4.61	0.06	1.24	10.70	16.08	7.20	4.76	2.91
	σ_n = 100							
AVE	7.24	0.06	1.22	13.78	13.31	9.91	7.36	3.08

denoised images by the competing algorithms on two typical images with moderate noise corruption $\sigma = 30$ and strong noise corruption $\sigma = 100$. Figure 2 shows an example in low level noise corruption, we can see that all the competing algorithms can achieve good denoising output in low level noise corruption. Figure 3 demonstrates that the proposed NCLR is very effective in reconstructing image details from the strong noisy observation. As can be seen in the highlighted window, NCLR generates much less artifacts and

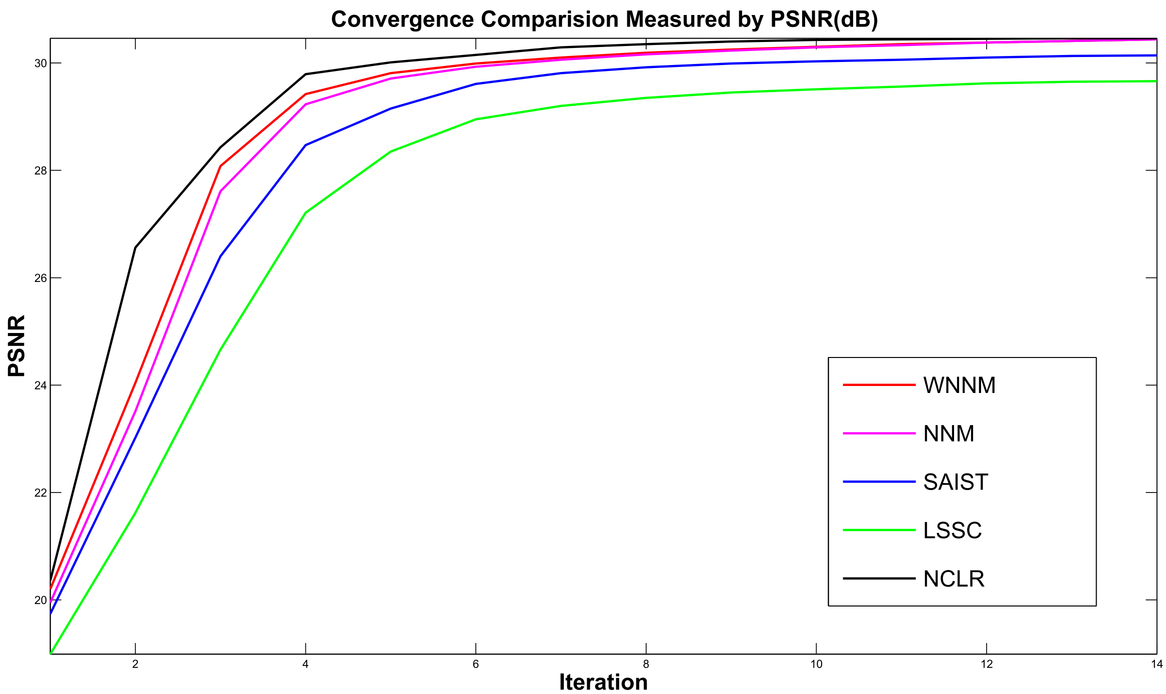

Fig. 1 Convergence result on image *House* by different methods (noise level $\sigma_n = 50$).

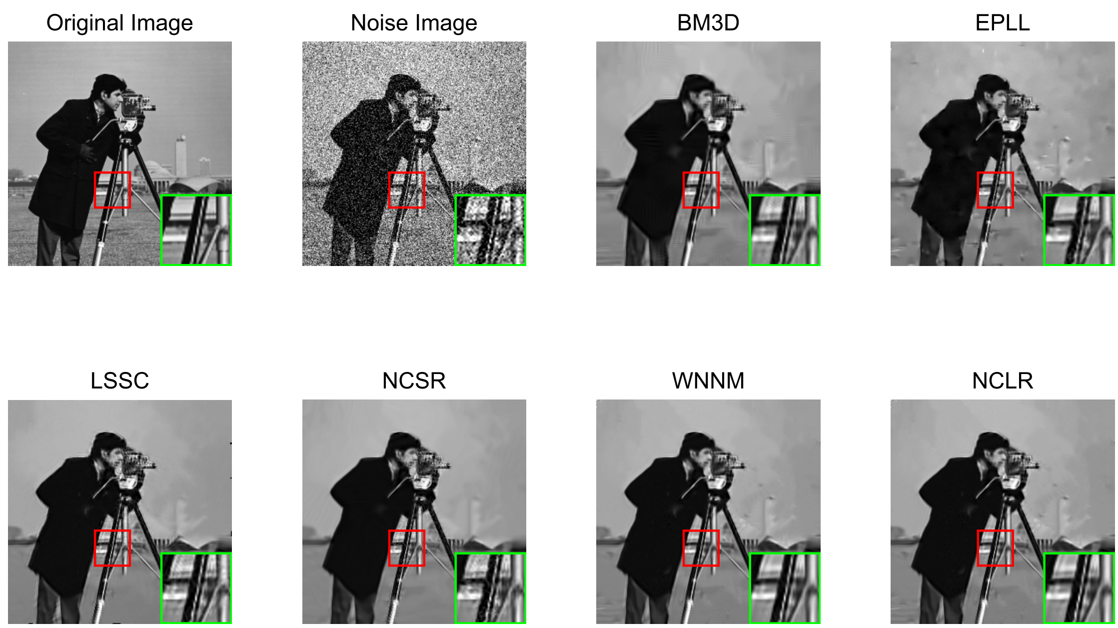

Fig. 2 Denoising results on image *Cameraman* by different methods (noise level $\sigma_n = 30$). From left to right and top to bottom: original image, noisy image, and denoised image by BM3D [19] (PSNR = 28.64dB), EPLL [23] (PSNR = 28.36dB), LSSC [24] (PSNR = 28.63dB), NCSR [21] (PSNR = 28.59dB), WNNM [3] (PSNR = 28.80dB), and NCLR (PSNR = 28.90dB).

preserves much better the image edge structures than other competing methods. In summary, NCLR shows strong denoising capability, producing visually much more pleasant denoising outputs while having higher PSNR indices.

5.2 Image Deblurring

In the experiments of image deblurring, two types of blur kernels include a 9×9 uniform blur and a Gaussian kernel of standard deviation $\sigma_n = 1.6$ were used for simula-

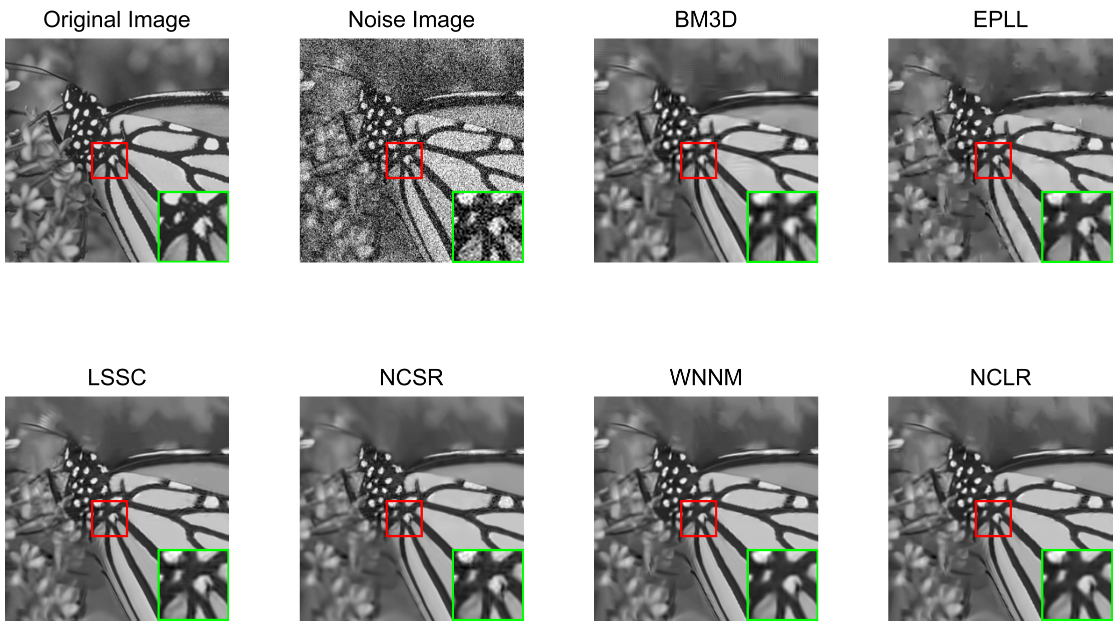

Fig. 3 Denoising results on image *Monarch* by different methods (noise level $\sigma_n = 30$). From left to right and top to bottom: original image, noisy image, and denoised image by BM3D [19] (PSNR = 22.52dB), EPLL [23] (PSNR = 22.23dB), LSSC [24] (PSNR = 22.24dB), NCSR [21] (PSNR = 22.11dB), WNNM [3] (PSNR = 22.95dB), and NCLR (PSNR = 23.04dB).

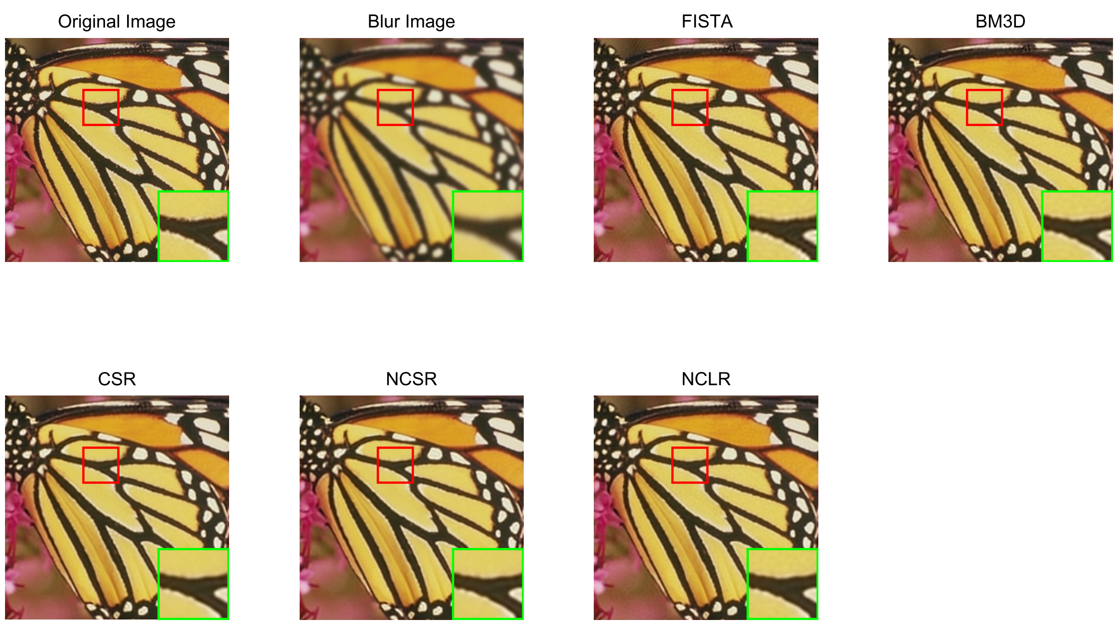

Fig. 4 Deblurring performance comparison on the *Butterfly* image. From left to right and top to bottom: original image, noisy and blurred image (9×9 uniform blur, $\sigma_n = \sqrt{2}$), and deblurred image by FISTA [25] (PSNR = 28.37dB), BM3D [26] (PSNR = 29.21dB), CSR [20] (PSNR = 29.75dB), NCSR [21] (PSNR = 29.68dB), and NCLR (PSNR = 30.75dB).

tions. Additive Gaussian white noises with standard deviation $\sigma_n = \sqrt{2}$ was added to the blurred images. The basic parameters of NCLR are as follows: the patch size is 6×6, total $L = 40$ similar patches are selected for each chosen exemplar and $p = 0.8$. We compare the proposed NCLR deblurring method with four recently image deblurring methods, including the constrained TV deblurring (denoted by FISTA) method [25], the IDD-BM3D deblurring method [26], the centralized sparse representation deblurring (CSR) method [20] and the nonlocally centralized sparse representation deblurring (NCSR) method [21].

The PSNR results on a set of 7 photographic images are reported in Table 7. From Table 7, we can see that the proposed NCLR algorithm significantly outperforms the FISTA

Table 7 PSNR(dB) results by different image deblurring methods

	9×9 Uniform Blur, $\sigma_n = \sqrt{2}$										
Images	Butterfly	Boats	C. Man	House	Parrot	Lena	Barbara	Starfish	Peppers	Leaves	Average
FISTA	28.37	29.04	26.82	31.99	29.11	28.33	25.75	25.75	28.43	26.49	28.008
BM3D	29.21	31.20	28.56	34.44	31.06	29.70	27.98	29.48	29.62	29.38	30.063
CSR	29.75	31.10	28.55	30.30	32.09	29.95	27.93	30.31	29.64	29.97	29.959
NCSR	29.68	31.08	28.62	34.31	31.95	**29.96**	28.10	30.28	29.66	29.98	30.362
NCLR	**30.75**	**31.13**	**28.67**	**34.51**	**32.31**	29.88	**28.70**	**30.81**	**29.81**	**31.01**	**30.758**
	Gaussian Blur, $\sigma_n = \sqrt{2}$										
FISTA	30.36	29.36	26.81	31.50	31.23	29.47	25.03	29.56	29.42	29.36	29.210
BM3D	30.73	**31.68**	28.17	34.08	32.89	31.45	27.19	31.66	29.99	31.40	30.924
CSR	30.75	31.40	28.24	32.31	33.44	31.23	27.81	32.25	**30.17**	31.44	30.904
NCSR	30.84	31.49	28.43	33.63	33.39	31.26	27.91	32.27	30.16	31.57	31.095
NCLR	**31.56**	31.64	**28.73**	**33.83**	**33.60**	**31.43**	**28.71**	**32.74**	30.17	**32.17**	**31.458**

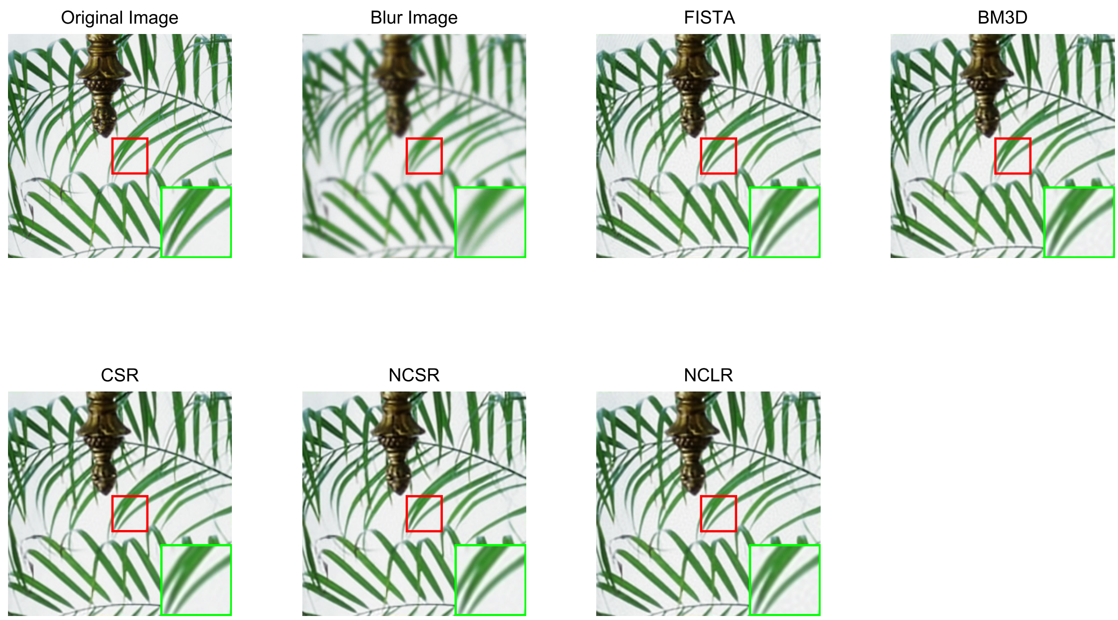

Fig. 5 Deblurring performance comparison on the *Leaves* image. From left to right: original image, noisy and blurred image (PSF = $[1 \quad 4 \quad 6 \quad 4 \quad 1]^T[1 \quad 4 \quad 6 \quad 4 \quad 1]/256$, $\sigma_n = 1.6$), deblurred image by FISTA [25] (PSNR = 29.36dB), BM3D [26] (PSNR = 31.40dB), CSR [20] (PSNR = 31.44dB), NCSR [21] (PSNR = 31.57dB), and NCLR (PSNR = 32.17dB).

Table 8 CPU time(min) by different image deblurring methods

	FISTA	BM3D	CSR	NCSR	NCLR
9×9 Uniform Blur	6.75	0.08	8.00	5.95	2.83
Gaussian Blur	6.67	0.07	5.58	8.51	2.72

and BM3D method, and also performs better than NCSR and CSR on most test images. In average, the NCLR gain over NCSR can be up to 0.396dB. The visual comparison of the deblurring methods are shown in Fig. 4 and Fig. 5, from which we can see that the NCLR method produces much cleaner and sharper image edges and textures than other methods. The average CPU time of each algorithm is listed in Table 8, and we can see that the proposed NCLR is faster than FISTA, CSR, and NCSR.

6. Conclusion

This paper has studied the non-convex low-rank (NCLR) approximation as a significant generalization of the nuclear norm minimization problem. The l_p non-convex relaxation, which improves the sparsity of a singular matrix value, can be solved by generalized singular value thresholding (GSVT) operator. Then the proposed NCLR algorithm was applied to image restoration. The Bayesian interpretation approach was provided to estimate an adaptive regularization parameter. Then the whole image is reconstructed by the proposed NCLR algorithm. Experimental results on image denoising and deblurring demonstrated that the NCLR can achieve highly competitive performance to the recently denoising and deblurring methods.

References

[1] W. Dong, G. Shi, and X. Li, Nonlocal image restoration with bilateral variance estimation: a low-rank approach, IEEE Trans. Image Process., vol.22, no.2, pp.700–711, 2013.

[2] W. Dong, G. Shi, and X.Li, Image Deblurring with Low-rank Ap-

proximation Structured Sparse Representation, in Signal and Information Processing Association Annual Summit and Conference, Ed., 1-5, 2012.

[3] S. Gu, L. Zhang, W. Zuo, and X. Feng, Weighted Nuclear Norm Minimization with Application to Image Denoising, in IEEE Conference on Computer Vision and Pattern Recongnition, Ed., 2862-2869, 2014.

[4] A.M. Buchanan and A.W. Fitzgibbon, Damped Newton Algorithms for Matrix Factorization with Missing Data, in IEEE Conference on Computer Vision and Pattern Recongnition, Ed., 316-322, 2005.

[5] Q. Ke and T. Kanade, Robust l_1 Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming, in IEEE Conference on Computer Vision and Pattern Recongnition, Ed., 739-746, 2005.

[6] A. Eriksson and A.V.D. Hengel, Efficient Computation of Robust Low-Rank Matrix Approximations in the Presence of Missing Data Using the l_1 Norm, in IEEE International Conference on Computer Vision and Pattern Recongnition, Ed., 771-778, 2010.

[7] J.-F. Cai, E.J. Candès, and Z. Shen, A Singular Value Thresholding Algorithm for Matrix Completion, SIAM Journal on Opitimization, vol.20, no.4, pp.1956–1982, 2010.

[8] E.J. Candès and B. Recht, Exact Matrix Completion via Convex Optimization, Foundations of Computational Mathematics, vol.9, no.6, pp.717–772, 2009.

[9] 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.

[10] Y. Wang, J. Wang, and Z. Xu, Restricted p−Isometry Properties of Nonconvex Block-Sparse Compressive Sensing, Signal Procssing, 104, 188-196, 2014.

[11] B. Recht, M. Fazel, and P.A. Parrilo, Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization, SIAM, vol.52, no.3, pp.471–501, 2010.

[12] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, A Generalized Iterated Shrinkage Algorithm for Non-convex Sparse Coding, in IEEE International Conference on Computer Vision, Ed., 217-224, 2013.

[13] E. Candès, and T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. Information Theory, vol.52, no.12, pp.5406–5425, 2006.

[14] D. Donoho, Compressed sensing, IEEE Trans. Information Theory, vol.52, no.4, pp.1289–1306, 2006.

[15] E. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Information Theory, vol.52, no.2, pp.489–509, 2006.

[16] M. Zibulevsky and M. Elad, l_1-l_2 optimization in signal and image processing, IEEE Signal Process. Mag., vol.27, no.3, pp.76–88, 2010.

[17] A. Buades, B. Coll, and J.M. Morel, A non-local algorithm for image denoising, in IEEE Computer Society Conference on Comput. Vis. Pattern Recognit., Ed., 60-65, 2005.

[18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, IEEE Trans. Image Process., vol.16, no.8, pp.2080–2095, 2007.

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, Image restoration by sparse 3D transform-domain collaborative filtering, SPIE Electronic Imageing, 6812, 681207-1-681207-12, 2008.

[20] W.Dong, L. Zhang, and G. Shi, Centralized sparse represenation for image resotoration, in IEEE International Conference on Computer Vision, Ed., 1259-1266, 2011.

[21] W. Dong, L. Zhang, G. Shi, and X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Trans. Image Processing, vol.22, no.4, pp.1620–1630, 2013.

[22] L. Sendur, and I.W. Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency, IEEE Trans. Signal Processing, vol.20, no.11, pp.2744–2756, 2002.

[23] D. Zoran and Y. Weiss, From Learning Models of Natural Image Patches to Whole Image Restoration, in IEEE International Conference on Computer Vision, Ed., 479-486, 2011.

[24] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, Non-local sparse models for image restoration, in IEEE International Conference on Computer Vision, Ed., 2272-2279, 2009.

[25] J. Portilla, Image restoration through l0 analysis-based sparse optimization in tight frames, in IEEE International Conference on Image Processing, Ed., 3909-3912, 2009.

[26] A. Danielyan, V. Katkovnik, and K. Egiazarian, BM3D frames and variational image deblurring, IEEE Trans. Image Process., vol.21, no.4, pp.1715–1728, 2012.

[27] A. Levin, R. Fergus, F. Durand, and W. Freeman, Image and Depth from a Conventional Camera with a Coded Aperture, ACM Trans. Graphics, 26 (3), 2007.

Appendix: Proof of Theorem 1

In order to proof Theorem 1, we need to introduce Lemma 1 as follows.

Lemma 1: For any $A \in \mathbb{R}^{m\times n}$ and a diagonal non-negative matrix $W \in \mathbb{R}^{m\times n}$, let $A = X\Phi Y^T$ be the singular value decomposition of A, we have $\sum_{i=1}^{n} \sigma_i(A)\sigma_i(W) = \max_{U^T U=I, V^T V=I} \mathrm{trace}(WV^T AU)$, where I is the identity matrix, $\sigma_i(A)$ and $\sigma_i(W)$ are the i-th singular values of matrices A and W, respectively. When $U = X$ and $V = Y$, $\mathrm{trace}(WV^T AU)$ reaches its maximum value.

Proof. The proof is immediate from properties of trace.

We are now in the position to proof the Theorem 1.

Proof. For any $X \in \mathbb{R}^{m\times n}$, its singular value decomposition can be expressed as $X = \bar{U}D\bar{V}^T$, where $\bar{U}$ and $\bar{V}$ are unitary matrices, and $D = \begin{pmatrix} diag(d_1, d_2, \ldots, d_n) \\ 0 \end{pmatrix}$ with $d_1 \geq d_2 \geq \ldots \geq d_n \geq 0$. Then we have

$$\min_X \frac{1}{2}\|Y - X\|_F^2 + \lambda \sum_i |\sigma_i(X)|^p \Leftrightarrow$$
$$\min_{\bar{U},\bar{V},D} \frac{1}{2}\|Y - \bar{U}D\bar{V}^T\|_F^2 + \lambda \sum_i |\sigma_i(\bar{U}D\bar{V}^T)|^p \Leftrightarrow$$
$$\min_{\bar{U}\bar{V}D} \frac{1}{2}\|Y\|_F^2 - \mathrm{trace}(Y\bar{V}^T D\bar{U}) + \frac{1}{2}\|D\|_F^2 + \lambda \sum_i |\sigma_i(\bar{U}D\bar{V}^T)|^p \Leftrightarrow$$
$$\min_D \left(\frac{1}{2}\|D\|_F^2 + \lambda \sum_i |\sigma_i(\bar{U}D\bar{V}^T)|^p - \max_{\substack{\bar{U}^T\bar{U}=I \\ \bar{V}^T\bar{V}=I}} \mathrm{trace}(Y\bar{V}^T D\bar{U}) \right),$$

where $d_1 \geq d_2 \geq \ldots \geq d_n \geq 0$. According to Lemma 1, we have

$$\max_{\bar{U}^T\bar{U}=I, \bar{V}^T\bar{V}=I} \mathrm{trace}(Y\bar{V}^T D\bar{U}) = \max_{U^T U=I, V^T V=I} \mathrm{trace}(U\Sigma V^T \bar{V}^T D\bar{U})$$

$$= \max_{(U\bar{U})^T(U\bar{U})=I, (V\bar{V})^T(V\bar{V})=I} \mathrm{trace}(\Sigma(V\bar{V})^T D(U\bar{U})^T) = \sum_{i=1}^{n} d_i\sigma_i,$$

and the optimization is obtained at $\bar{U} = U$ and $\bar{V} = V$. We then have

$$\min_X \frac{1}{2}\|Y - X\|_F^2 + \lambda \sum\nolimits_i |\sigma_i(X)|^p \Leftrightarrow$$
$$\min_D \left(\frac{1}{2}\|D\|_F^2 + \lambda \sum\nolimits_i |\sigma_i(\bar{U}D\bar{V}^T)|^p - \sum_{i=1}^{n} d_i\sigma_i\right) \Leftrightarrow$$
$$\min_D \left(\sum_{i=1}^{n} \frac{1}{2}{d_i}^2 - d_i\sigma_i + \lambda\, |d_i|^p\right) \Leftrightarrow$$
$$\min_D \left(\sum_{i=1}^{n} \frac{1}{2}(d_i - \sigma_i)^2 + \lambda|d_i|^p\right).$$

Form the above derivation, we can see that the optimal solution of the l_p nuclear norm problem in Eq. (3) is $\hat{X} = UDV^T$, where D is the optimum of the constrained optimization problem in Eq. (3). The proof is then completed.

Yang Lei was born in 1987. He is a Ph.D candidate in School of Electronic Information Engineering, Tianjin University.

Zhanjie Song was born in 1965. He is is a Professor with the School of Science, Tianjin University, and a vice-director with the Institute of TV and Image Information, Tianjin University.

Qiwei Song was born in 1987. He received the Ph.D degree form the School of Electronic Information Engineering, Tianjin University. He is a student member of IEICE