Analyzing the Weighted Nuclear Norm Minimization and Nuclear Norm Minimization based on Group Sparse Representation

Zhiyuan Zha, Xinggan Zhang, Qiong Wang, Yechao Bai and Lan Tang
Department of Electronic Science and Engineering, Nanjing University, Nanjing
zhazhiyuan.mmd@gmail.com, zhxg@nju.edu.cn, wajj131420@hotmail.com,
ychbai@nju.edu.cn, tanglan@nju.edu.cn

Abstract
The nuclear norm minimization (NNM) tends to over-shrink the rank components and treats the different rank components equally, thus limits its capability and flexibility in practical applications. Recent advances have suggested that the weighted nuclear norm minimization (WNNM) is expected to be more appropriate than NNM. However, it still lacks a plausible mathematical explanation why WNNM is more appropriate than NNM. In this paper, we analyze the WNNM and NNM from a point of the group sparse representation (GSR). Firstly, an adaptive dictionary for each group is designed. Then we show mathematically that WNNM is more appropriate than NNM. We exploit the proposed scheme to two typical low level vision tasks, including image deblurring and image compressive sensing (CS) recovery. Experimental results have demonstrated that the proposed scheme outperforms many state-of-the-art methods.

1 Introduction
Thanks to the data from many practical cases enjoying low rank property in essence, low rank matrix approximation (LRMA), which aims to recover the underlying low rank structure from its degraded/corrupted samples, has a wide range of applications in machine learning and computer vision. For instance, the Netflix customer data matrix is regarded as low rank because the customers’ choices are mostly affected by a few common factors [1]. As the matrix formed by nonlocal similar patches in a natural image is of a low rank, a flurry of matrix completion problems have been studied, such as collaborative filtering [2], reconstruction of occluded/corrupted face images [3], etc.

Generally speaking, methods dealing with LRMA can be classified into two categories: the low rank matrix factorization (LRMF) methods [4, 5] and the nuclear norm minimization (NNM) methods [2, 6, 7]. In this work we focus on the latter category. The nuclear norm of a matrix $X$ denoted by $\|X\|_*$, is the sum of its singular values, i.e., $\|X\|_* = \sum_i \sigma_i(X)$, where $\sigma_i(X)$ is the $i$-th singular value of matrix $X$. NNM aims to recover the underlying low rank matrix $X$ from its degraded observation matrix $Y$, while minimizing $\|X\|_*$. In recent years, a series of NNM-based methods have been proposed, such as robust principle component analysis (RPCA) [7], and low rank representation for subspace cluster [8].

Although NNM has been widely used for low rank matrix approximation, it often ignores the prior knowledge about the singular values of a practical data matrix among which larger ones measure the underlying principle components information. For example, the large singular values of a matrix of image similar patches characterize the major edge and texture information. This means that to recover an image from its degraded/corrupted one, we may shrink larger singular values less, while smaller ones shrink more. In a nutshell, NNM is not flexible enough since it treats singular values equally and tends to over-shrink the singular values, restricting its capability and flexibility in practical applications.

To improve the flexibility of NNM, [9, 10] proposed the weighted nuclear norm minimization (WNNM) model. The weighted nuclear norm of a data matrix $X$ is defined as $\|X\|_{w,*} = \sum_i w_i \sigma_i(X)$, where $w = [w_1, w_2, ..., w_l]$ and $w_i > 0$ is a non-negative weight assigned to $\sigma_i(X)$. Recent works have demonstrated that WNNM can obtain more accurate results than NNM in various machine learning and image processing applications [10, 11]. However, mathematically, we are still not clear why WNNM is more appropriate than NNM.

With the above question kept in mind, in this paper we analyze the WNNM and NNM from the group sparse representation (GSR) perspective. To the best of our knowledge, unfortunately, there has not been any work that analyzes why WNNM is more appropriate than NNM mathematically. The contribution of this paper is as follows. First, an adaptive dictionary for each group is designed with a low computational complexity, rather than dictionary learning from natural image dataset. Second, based on this dictionary learning scheme, we show mathematically that WNNM is more appropriate than NNM. Experimental results on two low level vision tasks, i.e., image deblurring and image compressive sensing (CS) recovery have demonstrated that the proposed scheme outperforms many state-of-the-art methods both quantitatively and qualitatively.
2 Background

2.1 Group Sparse Representation

Traditional patch-based sparse representation model usually suffers from some limits, such as dictionary learning with great computational complexity, neglecting the relationship among similar patches [31][12].

Instead of using patch as the basic unit of sparse representation, group sparse representation (GSR) presents a powerful mechanism, which combines local sparsity and nonlocal self-similarity of images simultaneously in a unified framework [31][12]. To be concrete, an image $X$ with size $N$ is divided into $n$ overlapped patches of size $\sqrt{m} \times \sqrt{m}$, and each patch is denoted by the vector $x_i \in \mathbb{R}^m$, $i = 1, 2, ..., n$. Then for each patch $x_i$, its $k$ similar patches are selected from a $R \times R$ sized search window to form a set $S_i$. After this, all the patches in $S_i$ are stacked into a data matrix of size $m \times k$, denoted by $X_i$, which includes every patch in $X_i$ as its columns, i.e., $X_i = \{x_{i,1}, x_{i,2}, ..., x_{i,k}\}$. The matrix $X_i$ consisting of all the patches with similar structures is called as a group, where $x_{i,k}$ denotes the $k$-th similar patch of the $i$-th group. Similar to patch-based sparse representation [13], given a dictionary $D_i$, each group $X_i$ can be sparsely represented as $\alpha_i = D_i^* X_i$ and solved by the following $\ell_0$-norm minimization problem,

$$\alpha_i = \arg \min_{\alpha_i} \frac{1}{2} \|X_i - D_i \alpha_i\|_F^2 + \lambda_ii \|\alpha_i\|_0 \quad (1)$$

where $\|\cdot\|_F$ denotes the Frobenious norm and $\lambda_i$ is the regularization parameter. $\|\cdot\|_0$ is $\ell_0$ norm, counting the nonzero entries of $\alpha_i$.

However, since the $\ell_0$ norm minimization is discontinuous optimization and NP-hard, solving Eq. (1) is a difficult combinatorial optimization problem. For this reason, it has been suggested that the non-convex $\ell_0$ norm minimization can be replaced by its convex $\ell_1$ norm counterpart,

$$\alpha_i = \arg \min_{\alpha_i} \frac{1}{2} \|X_i - D_i \alpha_i\|_F^2 + \lambda_i\|\alpha_i\|_1 \quad (2)$$

However, a fact that is often neglected is, $\ell_1$ norm minimization is hard to achieve the desired sparsity solution in some practical problems. This raises the question of whether we can improve the sparsity of $\ell_1$ norm minimization.

2.2 Reweighted $\ell_1$ Norm Minimization

According to the above analysis, it is expected that $\ell_1$ norm which is alternative to $\ell_0$ norm might also discover the correct solution and continue to solve the convex optimization problem. Based on this scheme, [14] proposed the weighted $\ell_1$ norm to replace the $\ell_1$ norm.

$$\alpha_i = \arg \min_{\alpha_i} \frac{1}{2} \|X_i - D_i \alpha_i\|_F^2 + \lambda_i\|w_i \alpha_i\|_1 \quad (3)$$

where $w_i = [w_{1,1}, w_{2,1}, ..., w_{m,k}] \in \mathbb{R}^m \times k$ is the diagonal matrix with $w_{1,1}, w_{2,1}, ..., w_{m,k}$ on the diagonal and zeros elsewhere. Meanwhile, the proposition is given as follows,

**Proposition 1** ([14]).

$$\|w_i \alpha_i\|_1 \geq \|\alpha_i\|_1 \quad (4)$$

where $v_1 \geq v_2$ denotes that the entry $v_1$ has much more sparsity encouraging than the entry $v_2$.

3 Why WNNM is More Appropriate than NNM

As we know, the weighted nuclear norm minimization (WNNM) can achieve more accurate results than nuclear norm minimization (NNM) in various computer vision and image processing studies. However, so far, we are still not clear why WNNM is more appropriate than NNM because it lacks a feasible mathematical derivation. In this section, we analyze the WNNM and NNM from the group sparse representation (GSR) perspective. First, we briefly introduce NNM and WNNM. Second, an adaptive dictionary for each group is designed with a low computational complexity, rather than dictionary learning from natural images. Third, based on this dictionary learning scheme, we show mathematically that WNNM is more appropriate than NNM.

3.1 Nuclear Norm Minimization

According to [15], nuclear norm is the tightest convex relaxation of the original rank minimization problem. Given a data matrix $Y \in \mathbb{R}^{m \times k}$, the goal of NNM is to find a matrix $X \in \mathbb{R}^{m \times k}$ of rank $r$ which satisfies the following objective function,

$$\hat{X} = \arg \min_{\hat{X}} \frac{1}{2} \|Y - \hat{X}\|_F^2 + \lambda \|\hat{X}\|_* \quad (5)$$

where $\lambda$ is a positive constant. [16] showed that the low rank matrix can be perfectly recovered from the degraded/corrupted data matrix with high probability by solving an NNM problem. [2] proved that NNM problem can be easily solved by imposing a soft-thresholding operation, that is, the solution of Eq. (5) which can be solved by

$$\hat{X} = US_A(\Sigma)V^T \quad (6)$$

where $Y = USV^T$ is the SVD of $Y$ and $S_A(\Sigma)$ is the soft-thresholding operator function on diagonal matrix $\Sigma$ with parameter $\lambda$. For each diagonal element $\Sigma_{ii}$ in $\Sigma$, there is

$$S_A(\Sigma)_{ii} = \text{soft}(\Sigma_{ii}, \lambda) = \max(\Sigma_{ii} - \lambda, 0).$$

More concretely, they proved the following theorem.

**Theorem 1** ([2]). For each $\lambda \geq 0$ and $Y$, the singular value shrinkage operator Eq. (5) obeys Eq. (6).

3.2 Weighted Nuclear Norm Minimization

As an alternative, [9][32] proposed the weighted nuclear norm minimization (WNNM) model. That is, the weighted nuclear norm $\|X\|_{w,*}$ is used to regularize $X$ and Eq. (5) can be rewritten as

$$\hat{X} = \arg \min_{\hat{X}} \frac{1}{2} \|Y - \hat{X}\|_F^2 + \lambda \|\hat{X}\|_{w,*} \quad (7)$$

More specifically, they proved the following theorem.

**Theorem 2** ([9]). If the singular values $\sigma_1 \geq \cdots \geq \sigma_n$ and the weights satisfy $0 \leq w_1 \leq \cdots \leq w_n$, $n_0 = \min(m, k)$, the WNNM problem in Eq. (7) has a globally optimal solution

$$\hat{X} = US_{w}(\Sigma)V^T \quad (8)$$

where $Y = USV^T$ is the SVD of $Y$ and $S_{w}(\Sigma)$ is the generalized soft-thresholding operator with the weighted vector $w$, i.e., $S_{w}(\Sigma)_{ii} = \text{soft}(\Sigma_{ii}, w_i) = \max(\Sigma_{ii} - w_i, 0)$. 

To prove that WNNM is more appropriate than NNM, we have the following theorem.

**Theorem 3.** For $0 \leq w_1 \leq \ldots w_n, n_0 = \min(m, k)$, and $Y$, the singular value shrinkage operator Eq. (3) satisfies Eq. (7).

**Proof.** For fixed weights $W, h_0(X) = \frac{1}{2}||Y - X||_F^2 + ||X||_{w,s}$, $\hat{X}$ minimizes $h_0$ if and only if it satisfies the following optimal condition,

$$0 \in \hat{X} - Y + \partial||\hat{X}||_{w,s}$$

where $\partial||X||_{w,s}$ is the set of subgradients of the weighted nuclear norm. Let matrix $X \in \mathbb{R}^{m \times k}$ be an arbitrary matrix and its SVD be $U \Sigma V^T$. It is known from [17] that the subgradient of $||X||_{w,s}$ can be derived as,

$$\partial||X||_{w,s} = \{U W, V^T + Z : Z \in \mathbb{R}^{m \times k}, U^T Z = 0\}$$

where $r$ is the rank of $X$ and $W_r$ is the diagonal matrix composed of the first $r$ rows and $r$ columns of the diagonal matrix diag ($W$).

In order to show that $\hat{X}$ obeys Eq. (10), the SVD of $Y$ can be rewritten as,

$$Y = U_0 \Sigma_0 V_0^T + U_1 \Sigma_1 V_1^T$$

where $U_0, V_0$ (resp. $U_1, V_1$) are the singular vectors associated with singular values greater than $w_j$ (resp. smaller than or equal to $w_j$). With these notations, we have

$$\hat{X} = U_0 (\Sigma_0 - W_r) V_0^T$$

Therefore,

$$Y - \hat{X} = U_0 W_r V_0^T + Z$$

where $Z = U_1 \Sigma_1 V_1^T$, by definition, $U_0^T Z = 0, Z V_0 = 0$ and since the diagonal elements of $\Sigma_1$ are smaller than $w_j + r$, it is easy to verify that $\sigma_j \leq w_{j+r}, j = 1, 2, \ldots, n_0 - r$. Thus, $Y - \hat{X} \in \partial||\hat{X}||_{w,s}$, which concludes the proof.

### 3.3 Adaptive Dictionary Learning

In this subsection, an adaptive dictionary learning method is designed, that is, for each group $X_i$, its adaptive dictionary can be learned from its observation $Y_i \in \mathbb{R}^{m \times k}$.

More specifically, we apply the singular value decomposition (SVD) to $Y_i$,

$$Y_i = U_i \Sigma_i V_i^T = \sum_{i,j=1}^{n_0} \sigma_{i,j} u_{i,j} v_{i,j}^T$$

where $\mu_i = [\sigma_{i,1}, \sigma_{i,2}, \ldots, \sigma_{i,n_0}], n_0 = \min(m, k), \Sigma_i = \text{diag(}$ $\mu_i)$ is a diagonal matrix whose non-zero elements are represented by $\mu_i$, and $u_{i,j}, v_{i,j}$ are the columns of $U_i$ and $V_i$, respectively.

Moreover, we define each dictionary atom $d_{i,j}$ of the adaptive dictionary $D_i$ for each group $Y_i$ as follows:

$$d_{i,j} = u_{i,j} v_{i,j}^T, \quad j = 1, 2, \ldots, n_0$$

Finally, learning an adaptive dictionary from each group $Y_i$ can be constructed by $D_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,n_0}]$. The proposed dictionary learning method is efficient due to the fact that it only requires one SVD operator for each group.

### 3.4 WNNM is More Appropriate than NNM

To prove that WNNM is more appropriate than NNM, we need the following lemmas:

**Lemma 1 (18).** The minimization problem

$$x = \arg \min_{x} \frac{1}{2} ||x - a||_2^2 + \tau ||x||_1$$

has a closed form, which can be expressed as

$$\hat{x} = \text{soft}(a, \tau) = \text{sgn}(a, \tau) \max(\text{abs}(a) - \tau, 0)$$

**Lemma 2.** For the following optimization problem

$$\min_{x_i \geq 0} \frac{1}{2} \sum_{i=1}^{n} (x_i - a_i)^2 + w_i x_i$$

If $a_1 \geq \ldots \geq a_n \geq 0$ and the weights satisfy $0 \leq w_1 \leq \ldots \leq w_n$, then the global optimum of Eq. (18) is $\hat{x}_i = \text{soft}(a_i, w_i) = \max(a_i - w_i, 0)$.

**Proof.** Without considering the constraint, the optimization problem Eq. (18) can be rewritten as the following unconstrained form,

$$\min_{x_i \geq 0} \frac{1}{2} (x_i - a_i)^2 + w_i x_i$$

Thus, it can be easily derived that the global optimization solution of Eq. (19) is $\hat{x}_i = \max(a_i - w_i, 0)$.

Now, the following $\ell_1$ norm-based group sparse representation problem can be expressed as,

$$\alpha_i = \arg \min_{\alpha_i} \frac{1}{2} ||Y_i - D_i \alpha_i||_F^2 + \lambda_i ||\alpha_i||_1$$

where $\lambda_i$ is the regularization parameter. According to the above design of adaptive dictionary $D_i$, we have the following theorem.

**Theorem 4.**

$$||Y_i - X_i||_F^2 = ||\mu_i - \alpha_i||_2^2$$

where $Y_i = D_i \mu_i$ and $X_i = D_i \alpha_i$.

**Proof.** Since the adaptive dictionary $D_i$ is constructed by Eq. (15), and the unitary property of $U_i$ and $V_i$, we have

$$||Y_i - X_i||_F^2 = ||D_i(\mu_i - \alpha_i)||_F^2 = ||U_i \text{diag}(\mu_i - \alpha_i) V_i^T||_F^2$$

$$= \text{trace}(U_i \text{diag}(\mu_i - \alpha_i) V_i^T \text{diag}(\mu_i - \alpha_i) V_i^T)$$

$$= \text{trace}(U_i \text{diag}(\mu_i - \alpha_i) \text{diag}(\mu_i - \alpha_i) U_i^T)$$

$$= \text{trace}(\text{diag}(\mu_i - \alpha_i) U_i U_i^T \text{diag}(\mu_i - \alpha_i))$$

$$= \text{trace}(\text{diag}(\mu_i - \alpha_i) \text{diag}(\mu_i - \alpha_i))$$

$$= ||\mu_i - \alpha_i||_2^2$$

**Proposition 2.** WNNM is more appropriate than NNM.
Proof. On the basis of Theorem 4, we have
\[ \alpha_i = \arg \min_{\alpha_i} \frac{1}{2} || Y_i - D_i \alpha_i ||_F^2 + \lambda || \alpha_i ||_1 \]
\[ = \arg \min_{\alpha_i} \frac{1}{2} || \mu_i - \alpha_i ||_2^2 + \lambda || \alpha_i ||_1 \] (23)
Thus, based on Lemma 1, we have
\[ \alpha_i = \text{soft}(\mu_i, \lambda_i) = \max(|\mu_i| - \lambda_i, 0) \] (24)

Obviously, according to Eqs. (14) and (15), we have
\[ \hat{X}_i = D_i \alpha_i = \sum_{j=1}^{n_0} \text{soft}(\mu_{i,j}, \lambda_i) d_{i,j} \]
\[ = \sum_{j=1}^{n_0} \text{soft}(\mu_{i,j}, \lambda_i) u_{i,j} v_{i,j}^T \]
\[ = U_i D_{\lambda_i} (\Sigma_i) V_i^T \] (25)
where \( \alpha_{i,j} \) represents the \( j \)-th element of the \( i \)-th group sparse coefficient \( \alpha_i \), and \( \Sigma_i \) is the singular value matrix of the \( i \)-th group \( Y_i \).

Thus, based on Theorem 1, we prove that the \( \ell_1 \)-norm based GSR problem (Eq. (20)) is equivalent to the NNM problem (Eq. (5)).

Similarly, for the weighted \( \ell_1 \) norm-based group sparse representation problem, we have
\[ \alpha_i = \arg \min_{\alpha_i} \frac{1}{2} || Y_i - D_i \alpha_i ||_F^2 + \lambda || \alpha_i ||_1 \]
\[ = \arg \min_{\alpha_i} \frac{1}{2} || \mu_i - \alpha_i ||_2^2 + \lambda || \alpha_i ||_1 \] (26)
Thus, based on Lemma 2, we have
\[ \alpha_i = \text{soft}(\mu_i, w_i) = \max(\mu_i - w_i, 0) \] (27)

Based on Eqs. (14) and (15), we have
\[ \hat{X}_i = D_i \alpha_i = \sum_{j=1}^{n_0} \text{soft}(\mu_{i,j}, w_i) d_{i,j} \]
\[ = \sum_{j=1}^{n_0} \text{soft}(\mu_{i,j}, w_i) u_{i,j} v_{i,j}^T \]
\[ = U_i D_{\lambda_i} (\Sigma_i) V_i^T \] (28)

Therefore, based on Theorem 3, we prove that the weighted \( \ell_1 \) norm-based GSR problem (Eq. (24)) is equivalent to the WNNM problem (Eq. (7)).

Proposition 1 showed that the weighted \( \ell_1 \) norm has much more sparsity encouraging than the \( \ell_1 \) norm [14] and based on group sparse representation, we have proven that WNNM, NNM are equivalent to the weighted \( \ell_1 \) norm minimization and \( \ell_1 \) norm minimization, respectively. Thus, we prove that WNNM is more appropriate than NNM.

4 Group-based Sparse Representation with the Weighted \( \ell_1 \) Norm Minimization for Low-Level Vision Tasks

In this section, we demonstrate the proposed scheme in the application of low-level vision tasks, including image deblurring and image compressive sensing (CS) recovery. The goal of low-level vision tasks is to reconstruct a high quality image \( X \) from its degraded observation \( Y \), which is a typical ill-posed linear inverse problem and it can be mathematically expressed as
\[ Y =HX + V \] (29)
where \( X, Y \) are lexicographically ordered vector representation of the original image and the degraded image, respectively. \( H \) is the degraded operator and \( V \) is usually assumed to be additive white Gaussian noise.

In the scenario of low level vision, what we observed is the degraded image \( Y \) via Eq. (29), and thus we utilize GSR model to reconstruct \( X \) from \( Y \) by the following weighted \( \ell_1 \) norm minimization problem,
\[ \alpha = \arg \min_{\alpha} \frac{1}{2} || Y - HD\alpha ||_2^2 + \lambda || w\alpha ||_1 \] (30)
where \( \lambda \) is a regularization parameter.

The objective function of Eq. (30) can be solved by the iterative shrinkage (IS) algorithm [19]. However, the major drawback of the IS algorithm is its low convergence speed. In this paper, we exploit the algorithm framework of the alternating direction method of multipliers (ADMM) [20] to solve Eq. (30). To be concrete, we introduce an auxiliary variable \( U \) and Eq. (30) can be rewritten as
\[ \alpha = \arg \min_{\alpha} \frac{1}{2} || Y -HU||_2^2 + \lambda ||w\alpha||_1, \text{ s.t. } U = D\alpha \] (31)

By defining \( f(U) = \frac{1}{2} || Y -HU||_2^2 \), and \( g(\alpha) = \lambda ||w\alpha||_1 \), we have
\[ U^{\ell+1} = \arg \min_{U} \frac{1}{2} || Y -HU||_2^2 + \frac{\rho}{2} ||U - D\alpha^\ell - C||_2^2 \] (32)
\[ \alpha^{\ell+1} = \arg \min_{\alpha} \lambda ||w\alpha||_1 + \frac{\rho}{2} ||U^{\ell+1} - D\alpha - C||_2^2 \] (33)
and
\[ C^{\ell+1} = C^\ell - (U^{\ell+1} - D\alpha^{\ell+1}) \] (34)

It can be seen that the minimization for Eq. (31) can be separated into two minimize sub-problems, i.e., the \( U \) and \( \alpha \) sub-problem. We will show that there is an efficient solution to each sub-problem. To avoid confusion, the subscribe \( \ell \) may be omitted for conciseness.

4.1 Subproblem of Minimizing \( U \) for fixed \( \alpha \)

Given \( \alpha \), the \( U \) sub-problem denoted by Eq. (32) becomes
\[ \min_{U} Q_1(U) = \min_{U} \frac{1}{2} || Y -HU||_2^2 + \frac{\rho}{2} ||U -D\alpha - C||_2^2 \] (35)
Clearly, Eq. (35) has a closed-form solution and its solution can be expressed as
\[ \hat{U} = (H^TH + \rho I)^{-1}(H^TY + \rho(D\alpha + C)) \] (36)
where \( I \) represents the identity matrix.

4.2 Subproblem of Minimizing \( \alpha \) for fixed \( U \)

Given \( U \), similarly, according to Eq. (33), the \( \alpha \) sub-problem can be written as
\[ \min_{\alpha} Q_2(\alpha) = \min_{\alpha} \frac{1}{2} ||D\alpha - L||_2^2 + \frac{\lambda}{\rho} ||w\alpha||_1 \] (37)
where \( L = Z - C \).
However, due to the complex definition of $w\alpha$, it is difficult to solve Eq. (37) directly. Let $X = D\alpha$, Eq. (37) can be rewritten as

$$
\min_{\alpha} Q_2(\alpha) = \min_{\alpha} \frac{1}{2} ||X - L||_2^2 + \frac{\lambda}{\rho} ||w\alpha||_1 \tag{38}
$$

To make Eq. (38) tractable, a general assumption is made, with which even a closed-form of Eq. (38) can be achieved. More specifically, $L$ can be regarded as some type of noisy observation of $X$, and thus $R = X - L$ follows an independent zero mean distribution with variances $\sigma^2$. The following theorem can be proven with this assumption.

**Theorem 5 ([21])**. Defining $X, L \in \mathbb{R}^N, X_i, L_i$, and $e(j)$ is each element of error vector $e$, where $e = X - L, j = 1, ..., N$. Assume that $e(j)$ follows an independent zero mean distribution with variance $\sigma^2$; and thus for any $\varepsilon > 0$, we can express the relationship between $\frac{1}{N}||X - L||_2^2$ and $\frac{1}{K} \sum_{i=1}^n ||X_i - L_i||_2^2$ by the following property,

$$
\lim_{N \rightarrow \infty \atop K \rightarrow \infty} P\left(\frac{1}{N}||X - L||_2^2 - \frac{1}{K} \sum_{i=1}^n ||X_i - L_i||_2^2 < \varepsilon\right) = 1
$$

where $P(\bullet)$ represents the probability and $K = m \times k \times n$.

Thus, based on Theorem 3, we have the following equation with a very large probability (restricted 1) at each iteration,

$$
\frac{1}{N}||X - L||_2^2 = \frac{1}{K} \sum_{i=1}^n ||X_i - L_i||_2^2
$$

And Eq. (38) can be rewritten as

$$
\min_{\alpha} \frac{1}{2} ||X - L||_2^2 + \frac{\lambda}{\rho} ||w\alpha||_1 = \min_{\alpha} \left( \sum_{i=1}^n \left( \frac{1}{2} ||X_i - L_i||_2^2 + \tau_i ||w_i\alpha_i||_1 \right) \right) \tag{41}
$$

where $\tau_i = \lambda_i K / \rho N$ and $D_i$ is a dictionary. It can be seen that Eq. (21) can be regarded as a group sparse coding problem by solving $n$ sub-problems for all the group $X_i$. Based on Theorem 4, Eq. (41) can be rewritten as

$$
\hat{\alpha}_i = \min_{\alpha_i} \sum_{i=1}^n \left( \frac{1}{2} ||\gamma_i - \alpha_i||_2^2 + \tau_i ||w_i\alpha_i||_1 \right)
$$

where $L_i = D_i\gamma_i$ and $X_i = D_i\alpha_i$. Therefore, based on Lemma 2, a closed-form solution of Eq. (42) can be computed as

$$
\hat{\alpha}_i = \text{soft}(\gamma_i, \tau_i w_i) = \max(\text{abs}(\gamma_i) - \tau_i w_i, 0) \tag{43}
$$

For each weight $w_i$, large values of sparse coefficient $\alpha_i$ usually transmit major edge and texture information. This implies that to reconstruct $X_i$ from its degraded one, we should shrink large values less, while shrinking smaller ones more, and thus we have $w_i = \tau_i / |\gamma_i| + \epsilon$, where $\epsilon$ is a small constant. Inspired by [23], the regularization parameter $\lambda_i$ of each group $L_i$ is set as: $\lambda_i = 2\sqrt{2\sigma^2/\delta_i + \epsilon}$, where $\delta_i$ denotes the estimated variance of $\gamma_i$, and $\epsilon$ is a small constant. After solving the two sub-problems, we summarize the overall algorithm for Eq. (30) in Table 1.

| Table 1: ADMM method for Eq. (30) |
|-----------------------------------|
| **Input:** the observed image $Y$ and the measurement matrix $H$. |
| **Initialization:** $C, U, \alpha, R, m, k, \rho, \sigma, \varepsilon, \varepsilon, \ell$; |
| **Repeat** |
| For Each group $L_i$; |
| - Construct dictionary $D_i$ by computing Eq. (15); |
| - Update $\lambda_i^{\ell+1}$ by computing $\lambda_i = 2\sqrt{2\sigma^2/\delta_i + \varepsilon}$; |
| - Update $\tau_i^{\ell+1}$ computing by $\tau_i = \lambda_i K / \rho N$; |
| - Update $w_i^{\ell+1}$ computing by $w_i = \tau_i / |\gamma_i| + \varepsilon$; |
| - Update $\alpha_i^{\ell+1}$ computing by Eq. (43); |
| **End For** |
| Update $D^{\ell+1}$ by concatenating all $D_i$ ; |
| Update $\alpha^{\ell+1}$ by concatenating all $\alpha_i$ ; |
| Update $\epsilon^{\ell+1}$ by computing Eq. (34) ; |
| $\ell \leftarrow \ell + 1$; |
| **Until** |
| maximum iteration number is reached. |
| **Output:** |
| The final restored image $\hat{X} = D\hat{\alpha}$. |

$\quad$

![Figure 1: All test images.](image-url)
deblurring methods is shown in Fig. 2. It can be found out that the NCSR, FPD, MSEP LL and GSR-NNM still generated some undesirable artifacts, while IDD-BM3D resulted in over-smooth phenomena. By contrast, the GSR-WNNM was able to preserve the sharpness of edges and suppress undesirable artifacts more effectively than the other methods.

In image CS recovery, we generated the CS measurements at the block level by utilizing a Gaussian random projection matrix to test images, i.e., the block-based CS recovery with block size of $32 \times 32$. The size of training window $R \times R$ is set to be $20 \times 20$ and searching matched patches $k = 60$, $\rho = 0.03$ for $0.1N$ measurements, and $\rho = 0.1$ for other measurements. $\varepsilon$ and $\epsilon = 0.35$. The size of each patch $\sqrt{m} \times \sqrt{m}$ is set to be $6 \times 6$. To verify the performance of the GSR-WNNM, we compared it with some competitive CS recovery methods including the BCS [26], BM3D-CS [27], ADS-CS [28], SGSR [29], MRK [30] and GSR-NNM methods.

The average PSNR results of gray images are shown in Table 3. It can be seen that GSR-WNNM outperforms many state-of-the-art methods.

### 6 Conclusion

This paper analyzed the weighted nuclear norm minimization (WNNM) and nuclear norm minimization (NNM) based on group sparse representation (GSR). We designed an adaptive dictionary for each group with a low computational complexity, rather than dictionary learning from natural images. Based on this dictionary learning scheme, we proved that WNNM is more appropriate than NNM. Experimental results on two low level vision tasks, i.e., image deblurring and image CS recovery have demonstrated that the proposed scheme outperforms many state-of-the-art methods.

### References

[1] Srebro N, Salakhutdinov R R. Collaborative filtering in a non-uniform world: Learning with the weighted trace norm[C]//Advances in Neural Information Processing Systems. 2010: 2056-2064.

[2] Cai J F, Candes E J, Shen Z. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.
[3] Zheng Y, Liu G, Sugimoto S, et al. Practical low-rank matrix approximation under robust \(\ell_1\)-norm[C]/Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 1410-1417.

[4] Eriksson A, Van Den Hengel A. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the \(\ell_1\) norm[C]/Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 771-778.

[5] Ke Q, Kanade T. Robust \(\ell_1\) norm factorization in the presence of outliers and missing data by alternative convex programming[C]/Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005: 1: 739-746.

[6] Ji H, Liu C, Shen Z, et al. Robust video denoising using low rank matrix completion[C]/Computer Vision and Pattern Recognition, 2010 IEEE Conference on. IEEE, 2010: 1791-1798.

[7] Cands E J, Li X, Ma Y, et al. Robust principal component analysis?[J]. Journal of the ACM (JACM), 2011, 58(3): 11.

[8] Liu G, Lin Z, Yu Y. Robust subspace segmentation by low-rank representation[C]/Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 663-670.

[9] Gu S, Zhang L, Zuo W, et al. Weighted nuclear norm minimization with application to image denoising[C]/Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2862-2869.

[10] Sadigh D, Ohlsson H, Sastry S, et al. Robust subspace system identification via weighted nuclear norm optimization[J]. IFAC Proceedings Volumes, 2014, 47(3): 9510-9515.

[11] Lu C, Tang J, Yan S, et al. Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm[J]. IEEE Transactions on Image Processing, 2016, 25(2): 829-839.

[12] Zhang J, Zhao D, Gao W. Group-based sparse representation for image restoration[J]. IEEE Transactions on Image Processing, 2014, 23(8): 3336-3351.

[13] Aharon M, Elad M, Bruckstein A. k-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on signal processing, 2006, 54(11): 4311-4322.

[14] Candes E J, Wakin M B, Boyd S P. Enhancing sparsity by reweighted \(\ell_1\) minimization[J]. Journal of Fourier analysis and applications, 2008, 14(5-6): 877-905.

[15] Fazel M. Matrix rank minimization with applications[D]. PhD thesis, Stanford University, 2002.

[16] Candes E, Recht B. Exact matrix completion via convex optimization[J]. Communications of the ACM, 2012, 55(6): 111-119.

[17] Lewis A S. The convex analysis of unitarily invariant matrix functions[J]. Journal of Convex Analysis, 1995, 2(1): 173-183.

[18] Li C, Yin W, Zhang Y. Users guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms[J]. CAAM report, 2009, 20: 46-47.

[19] Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on pure and applied mathematics, 2004, 57(11): 1413-1457.

[20] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends? in Machine Learning, 2011, 3(1): 1-122.

[21] Zha Z, Liu X, Huang X, et al. Analyzing the group sparsity based on the rank minimization methods[J]. arXiv preprint arXiv:1611.08983, 2016.

[22] Danielyan A, Katkovnik V, Egiazarian K. BM3D frames and variational image deblurring[J]. IEEE Transactions on Image Processing, 2012, 21(4): 1715-1728.

[23] Dong W, Zhang L, Shi G, et al. Nonlocally centralized sparse representation for image restoration[J]. IEEE Transactions on Image Processing, 2013, 22(4): 1620-1630.

[24] Xu Y, Yin W. A fast patch-dictionary method for whole image recovery[J]. arXiv preprint arXiv:1408.3740, 2014.

[25] Papyan V, Elad M. Multi-scale patch-based image restoration[J]. IEEE Transactions on image processing, 2016, 25(1): 249-261.

[26] Mun S, Fowler J E. Residual reconstruction for block-based compressed sensing of video[C]/Data Compression Conference (DCC), 2011. IEEE, 2011: 183-192.

[27] Egiazarian K, Foi A, Katkovnik V. Compressed sensing image reconstruction via recursive spatially adaptive filtering[C]/Image Processing. 2007. ICIP 2007. IEEE International Conference on. IEEE, 2007, 1: I-549-I-552.

[28] Dong W, Shi G, Li X, et al. Image reconstruction with locally adaptive sparsity and nonlocal robust regularization[J]. Signal Processing: Image Communication, 2012, 27(10): 1109-1122.

[29] Zhang J, Zhao D, Jiang F, et al. Structural group sparse representation for image compressive sensing recovery[C]/Data Compression Conference (DCC), 2013. IEEE, 2013: 331-340.

[30] Canh T N, Dinh K Q, Jeon B. Multi-scale/multi-resolution Kroncker compressive imaging[C]/Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, 2015: 2700-2704.

[31] Mairal J, Bach F, Ponce J, et al. Non-local sparse models for image restoration[C]/Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 2272-2279.

[32] Gu S, Xie Q, Meng D, et al. Weighted nuclear norm minimization and its applications to low level vision[J]. International Journal of Computer Vision, 2016: 1-26.