Non-convex Lp Nuclear Norm based ADMM Framework for Compressed Sensing

Chen Zhao†, Jian Zhang†, Siwei Ma†‡and Wen Gao†‡

†Institute of Digital Media & Cooperative Medianet Innovation Center
Peking University, Beijing, 100871, P. R. China
‡Peking University Shenzhen Graduate School, Shenzhen, 518055, P. R. China
{zhaochen,jian.zhang,swma,wgao}@pku.edu.cn

Abstract

Compressed Sensing (CS) has drawn quite an amount of attention as a joint sampling and compression methodology. Recent studies further show that image prior models play an important role in image CS recovery. By exploiting the non-local self-similarity of natural images and clustering similar patches, low-rank prior model is adopted in this paper. Different from traditional nuclear norm, we extend the l_p $(0 < p < 1)$ penalty function on singular values of a matrix to characterize low-rank prior model, and propose a new non-nonvex l_p nuclear norm prior model for image CS recovery, which is able to more accurately enforce image structural sparsity and self-similarity at the same time. The proposed optimization problem is efficiently solved within the alternative direction multiplier method (ADMM) framework. Experimental results demonstrate that the proposed l_p nuclear norm based ADMM framework for image CS recovery framework exhibits good convergence and achieves significant performance improvements over the current state-of-the-art methods.

1 Introduction

Compressed Sensing (CS) has drawn quite an amount of attention as an alternative to the current methodology of sampling followed by compression [1]. By exploiting the redundancy existed in a signal, CS conducts sampling and compression at the same time. From many fewer acquired measurements than suggested by the Nyquist sampling theory, CS theory demonstrates that, a signal can be reconstructed with high probability when it exhibits sparsity in some domain.

Specifically, a signal $\mathbf{x}$ of size N is said to be sparse in domain or basis Ψ, if its transform coefficients $\boldsymbol{\alpha} = \Psi^T\mathbf{x}$ are mostly zeros, or nearly sparse if the dominant portion of coefficients are either zeros or very close to zeros. Then, given M linear measurements, the CS recovery of $\mathbf{x}$ from $\mathbf{b}$ is formulated as the following constrained optimization problem:

$$\min_{\mathbf{x}} \left\|\Psi^T\mathbf{x}\right\|_q \quad s.t. \;\; \mathbf{b} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{A}$ represents the random projections (RS). q is usually set to 1 or 0, characterizing the sparsity of the vector $\Phi^T\mathbf{x}$. $\|\cdot\|_1$ is l_1 norm, adding all the absolute values of the entries in a vector, while $\|\cdot\|_0$ is l_0 norm, counting the nonzero entries of a vector. According to [2], CS is capable of recovering K-sparse signal $\mathbf{x}$ (with an overwhelming probability) from $\mathbf{b}$ of size M, provided that the number of random

1068-0314/16 $31.00 © 2016 IEEE
DOI 10.1109/DCC.2016.104

samples meets $M \geq cK\log(N/K)$. A compressive imaging camera prototype using RS has been presented in [1].

The most attractive strength of CS-based compression is that the encoder is made signal-independent and computationally inexpensive at the cost of high decoder complexity, that is, simple encoder and complex decoder. This asymmetric design is severely desirable in some image processing applications when the data acquisition devices must be simple (e.g. inexpensive resource-deprived sensors), or when oversampling can harm the object being captured (e.g. X-ray imaging) [3].

Since natural signals such as images are typically non-stationary, there exists no universal domain in which all parts of the signals are sparse. Traditional CS recovery methods explored a set of fixed domains (e.g. DCT, wavelet, and gradient domain), therefore are signal-independent or not adaptive, resulting in poor rate-distortion performance. To rectify the problem, many works incorporated additional prior knowledge about transform coefficients (statistical dependencies, structure, etc.) into the CS recovery framework, such as Gaussian scale mixtures (GSM) models [4]. Additionally, in [5], a projection-driven CS recovery coupled with block-based random image sampling was developed, which aims to encourage sparsity in the domain of directional transforms. Chen et al. [6] exploited multi-hypothesis predictions to generate a residual in the domain of the CS random projections, where this residual being typically more compressible than the original signal leads to improved reconstruction quality. The adaptively learned sparsifying basis was also utilized for image CS recovery [7]. Furthermore, many latest works concentrate on utilization of both local and nonlocal statistics for high quality image restoration. Zhao et. al [8] explored the structured Laplacian sparsity of DCT coefficients for high-quality CS reconstruction. Zhang et al. [9] [10] proposed a framework for CS recovery via collaborative sparsity, which enforces local 2-D sparsity and nonlocal 3-D sparsity simultaneously in an adaptive hybrid space-transform domain, thus greatly confining the CS solution space. Lately, Zhang et al. [11] further proposed a structural group sparse representation (SGSR) model, which efficiently characterizes the intrinsic sparsity and self-similarity of natural images in an adaptive group domain, achieving state-of-the-art results.

Recent studies show that, by exploiting the non-local self-similarity and clustering similar blocks, low-rank prior models achieve promising results in image restoration tasks [12]. However, in general, the rank minimization is an NP-hard problem. To obtain an approximated solution, the nuclear norm (sum of the singular values) is usually adopted as a convex surrogate of the rank. By the nuclear norm, the rank minimization problem can be efficiently solved by the classic technique of singular value thresholding (SVT) [13]. Despite exhibiting good theoretical guarantee, the nuclear norm is actually not accurate to approximate the rank.

Inspired by the success of l_p $(0 < p < 1)$ sparse optimization, in this paper, we extend the surrogate function of l_0 norm, i.e. l_p $(0 < p < 1)$ penalty function on singular values of a matrix to substitute for the nuclear norm for characterizing low-rank prior model, which is expected to be more accurate than traditional nuclear norm. Moreover, to make the optimization tractable, the alternative direction multiplier method (ADMM) framework [14] is adopted, enabling each sub-problem solved efficiently. Experimental results demonstrate that the novel CS recovery framework

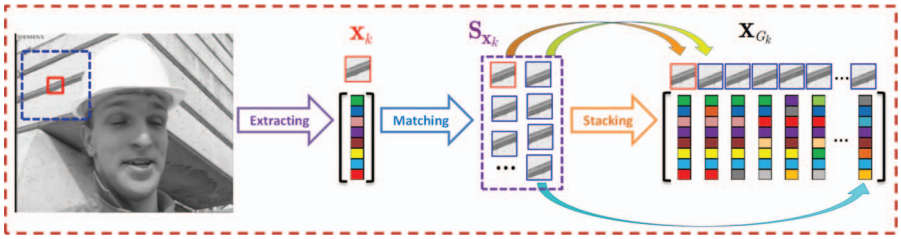

Figure 1: Illustrations for image low-rank prior. Extract each exemplar patch vector $\mathbf{x}_k$ from image $\mathbf{x}$. For each $\mathbf{x}_k$, denote $\mathbf{S}_{\mathbf{x}_k}$ the set composed of its best matched patches. Stack all the patches in $\mathbf{S}_{\mathbf{x}_k}$ to construct the data matrix, denoted by $\mathbf{X}_{G_k}$. $\mathbf{X}_{G_k}$ has the low-rank property.

exhibits good convergence and achieves significant performance improvements over the current state-of-the-art methods.

The remainder of the paper is organized as follows. Section 2 elaborates the proposed non-convex l_p $(0 < p < 1)$ nuclear norm prior model for CS. Section 3 describes the implementation details under the ADMM optimization framework. Experimental results are reported in Section 4. In Section 5, we conclude this paper.

2 Non-convex L_p Nuclear Norm Prior Model for CS

In this section, we will elaborate the proposed non-convex l_p nuclear norm prior model, which is used to characterize the low-rank property of the data matrix composed of similar patches.

The basic assumption is that the well-known non-local self-similarity, which depicts the repetitiveness of higher level patterns (e.g. textures and structures) globally positioned in natural images, implies that many similar patches can be searched for any exemplar patch [15] [16] To be concrete, as illustrated in Fig. 1, first, divide the image $\mathbf{x} \in \mathbb{R}^N$ with size N into K overlapped patches of size $\sqrt{B_s} \times \sqrt{B_s}$, and each patch is denoted by the vector $\mathbf{x}_k \in \mathbb{R}^{B_s}$, i.e. $k = 1, 2, \ldots, K$. Then, for each exemplar patch $\mathbf{x}_k$, denoted by small red square in Fig. 1, within the $W_s \times W_s$ training window (big blue square), search its c best matched patches, which comprise the group $\mathbf{S}_{\mathbf{x}_k}$. Here, Euclidean distance is selected as the similarity criterion between different patches. Next, all the patches in each group $\mathbf{S}_{\mathbf{x}_k}$ are stacked into a data matrix of size $B_s \times c$, denoted by $\mathbf{X}_{G_k}$, which includes every patch in $\mathbf{S}_{\mathbf{x}_k}$ as its columns, i.e. $\mathbf{X}_{G_k}$=$[\mathbf{x}_{G_k \otimes 1}, \mathbf{x}_{G_k \otimes 2}, \ldots, \mathbf{x}_{G_k \otimes c}]$. Since all the patches in each data matrix have similar structures, the constructed data matrix $\mathbf{X}_{G_k}$ has a low-rank property. By incorporating the low-rank prior into the image CS recovery framework Eq. (1), we easily have the following optimization problem:

$$\hat{\mathbf{x}} = arg\,\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{k=1}^{K} rank(\mathbf{X}_{G_k}). \tag{2}$$

In general, the rank minimization is an NP-hard problem. To obtain an approxi-

mated solution, the nuclear norm (sum of the singular values) is usually adopted as a convex surrogate of the rank. By the nuclear norm, the rank minimization problem can be efficiently solved by the classic technique of singular value thresholding (SVT) [13]. Despite exhibiting good theoretical guarantee, the nuclear norm is actually not accurate to approximate the rank. Inspired by the success of l_p $(0 < p < 1)$ sparse optimization, we believe that the non-convex optimization toward rank minimization could give rise to better recovery results.

In this paper, to approximate matrix rank more accurately, we extend the non-convex l_p $(0 < p < 1)$ penalty function on singular values of the data matrix to substitute the convex nuclear norm. Concretely, the rank function can be approximately solved by the following function:

$$F(\mathbf{X}_{G_k}) = \sum_{i=1}^{c} |\sigma_i(\mathbf{X}_{G_k})|^p, \tag{3}$$

where $0 < p < 1$ and $\sigma_i(\mathbf{X})$ denotes the i-th singular value of a matrix $\mathbf{X} \in \mathbb{R}^{B_s \times c}$ (assuming $B_s \leq c$ in this work).

Accordingly, considering all the matrices $\mathbf{X}_{G_k}$, the proposed non-convex low-rank prior model for CS recovery is formulated as

$$\hat{\mathbf{x}} = arg\,\underset{\mathbf{x}}{min}\, \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{k=1}^{K} F(\mathbf{X}_{G_k}), \tag{4}$$

where $\mathbf{X}_{G_k} = \mathbf{R}_{G_k}\mathbf{x} = [\mathbf{x}_{G_k \otimes 1}, \mathbf{x}_{G_k \otimes 2}, \ldots, \mathbf{x}_{G_k \otimes c}]$ and $\mathbf{R}_{G_k}$ is the matrix operator that extracts the matrix $\mathbf{X}_{G_k}$ from $\mathbf{x}$. Obviously, the proposed non-convex low-rank model is able to exploit both the group sparsity of similar patches and the non-convexity of rank minimization simultaneously, which is expected to achieve better recovery than previous methods.

3 L_p Nuclear Norm based ADMM Framework for CS

In this paper, we adopt the alternative direction multiplier method (ADMM) framework [14] to solve the proposed CS recovery optimization problem with the non-convex l_p nuclear norm, which is verified to be quite effective, making each sub-problem solved efficiently.

To be concrete, by exploiting ADMM and introducing a variable $\mathbf{z}$ and a constraint $\mathbf{x} = \mathbf{z}$, the minimization of Eq. (4) is transformed into three iterative steps:

$$\mathbf{x}^{(t+1)} = arg\,\underset{\mathbf{x}}{min}\, \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\beta}{2}\left\|\mathbf{x} - \mathbf{z}^{(t)} - \mathbf{b}^{(t)}\right\|_2^2; \tag{5}$$

$$\mathbf{z}^{(t+1)} = arg\,\underset{\mathbf{z}}{min}\, \frac{\beta}{2}\left\|\mathbf{x}^{(t+1)} - \mathbf{z} - \mathbf{b}^{(t)}\right\|_2^2 + \lambda \sum_{k=1}^{K} F(\mathbf{Z}_{G_k}); \tag{6}$$

$$\mathbf{b}^{(t+1)} = \mathbf{b}^{(t)} - \left(\mathbf{x}^{(t+1)} - \mathbf{z}^{(t+1)}\right). \tag{7}$$

Thus, by ADMM, we obtain two sub-problems, namely $\mathbf{x}$ and $\mathbf{z}$ sub-problems. In the following, we will provide the implementation details to obtain the efficient solutions to each separated sub-problem.

3.1 $\mathbf{x}$ *Sub-problem*

Given $\mathbf{z}^{(t)}$, the $\mathbf{x}$ sub-problem denoted by Eq. (5) is essentially a minimization problem of strictly convex quadratic function. Since $\mathbf{A}$ is a random projection matrix without special structure, it is too costly to compute the matrix inverse to solve Eq. (5). Instead, in this paper, the gradient descent method is utilized by iteratively applying

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta^{(t)}\mathbf{d}^{(t)}, \tag{8}$$

where $\mathbf{d}^{(t)}$ is the gradient direction of the objective function and $\eta^{(t)}$ represents the optimal step. Therefore, solving the $\mathbf{x}$ sub-problem for image CS recovery only requires computing the following equation iteratively

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta^{(t)}\left(\mathbf{A}^T\mathbf{A}\mathbf{x}^{(t)} - \mathbf{A}^T\mathbf{y} + \beta\left(\mathbf{x}^{(t)} - \mathbf{z}^{(t)} - \mathbf{b}^{(t)}\right)\right), \tag{9}$$

where $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{y}$ can be calculated before, making above computation more efficient.

3.2 $\mathbf{z}$ *Sub-problem*

Given $\mathbf{x}^{(t+1)}$, the $\mathbf{z}$ sub-problem is formulated as

$$\mathbf{z}^{(t+1)} = arg\min_{\mathbf{z}} \frac{1}{2}\left\|\mathbf{z} - \mathbf{w}^{(t)}\right\|_2^2 + \frac{\lambda}{\beta}\sum_{k=1}^{K} F(\mathbf{Z}_{G_k}), \tag{10}$$

where $\mathbf{w}^{(t)} = \mathbf{x}^{(t+1)} - \mathbf{b}^{(t)}$. To enable solving Eq. (10) tractable, furthermore, like [11] we also make a reasonable and general assumption that each element of $\mathbf{e}^{(t)} = \mathbf{x} - \mathbf{w}^{(t)}$ follows an independent zero-mean distribution with the variance, leading to the following Theorem 1 [11].

Theorem 1 *Let* $\mathbf{x}, \mathbf{w} \in \mathbb{R}^N$, $\mathbf{X}_{G_k}, \mathbf{W}_{G_k} \in \mathbb{R}^{B_s \times c}$, *and denote the error vector by* $\mathbf{e} = \mathbf{x} - \mathbf{w}$ *and each element of* $\mathbf{e}$ *by* $\mathbf{e}(j), j = 1, \ldots, N$. *Assume that* $\mathbf{e}(j)$ *is independent and comes from a distribution with zero mean and variance* σ^2. *Then, for any* $\epsilon > 0$, *we have the following property to describe the relationship between* $\|\mathbf{x} - \mathbf{w}\|_2^2$ *and* $\sum_{k=1}^{K}\|\mathbf{X}_{G_k} - \mathbf{W}_{G_k}\|_F^2$, *that is,*

$$P\left(\left|\|\mathbf{x} - \mathbf{w}\|_2^2 - \frac{N}{L}\sum_{k=1}^{K}\|\mathbf{X}_{G_k} - \mathbf{W}_{G_k}\|_F^2 < \epsilon\right|\right) = 1, \tag{11}$$

where $P(\cdot)$ *represents the probability and* $L = B_s \times c \times K$.

With **Theorem 1**, incorporating Eq. (11) into Eq. (6) leads to

$$\mathbf{z}^{(t+1)} = arg\min_{\mathbf{z}} \frac{1}{2}\sum_{k=1}^{K}\left(\left\|\mathbf{Z}_{G_k} - \mathbf{W}_{G_k}^{(t))}\right\|_F^2 + \tau F(\mathbf{Z}_{G_k})\right), \tag{12}$$

where $\tau = \frac{\lambda N}{\beta L}$. It is obvious to see that Eq. (12) can be efficiently minimized by solving K sub-problems for all the $\mathbf{Z}_{G_k}$, each of which is formulated below:

$$\mathbf{Z}_{G_k}^{(t+1)} = arg\min_{\mathbf{Z}_{G_k}} \frac{1}{2}\|\mathbf{Z}_{G_k} - \mathbf{W}_{G_k}^{(t)}\|_F^2 + \sum_{i=1}^{c} g_\tau(\sigma_i(\mathbf{Z}_{G_k})), \tag{13}$$

where the penalty function $g_\tau(\cdot)$ is defined as $g_\tau(x) = \tau|x|^p, (0 < p < 1)$.

Note that $g_\tau(x)$ is concave and monotonically increasing on $[0\ \infty)$, with its gradients being decreasing. Based on this property, recently, Lu et. al proposed an iteratively reweighted nuclear norm (IRNN) algorithm to solve the general non-convex non-smooth low-rank minimization with convergence guarantee [17]. Here, borrowing the wisdom of IRNN, we will show how to solve the problem (13).

For simplicity of notation, denote $h(\mathbf{Z}_{G_k}) = \frac{1}{2}\|\mathbf{Z}_{G_k} - \mathbf{W}_{G_k}^{(t)}\|_F^2$, $\sigma_i = \sigma_i(\mathbf{Z}_{G_k})$, and $\sigma_i^{(l)} = \sigma_i(\mathbf{Z}_{G_k}^{(l)})$. Owing to that $g_\tau(x)$ is concave and differentiable on $[0\ \infty)$, we obtain

$$g_\tau(\sigma_i) \leq g_\tau(\sigma_i^{(l)}) + w_i^{(l)}(\sigma_i - \sigma_i^{(l)}). \tag{14}$$

Here $w_i^{(l)} = g'_\tau(\sigma_i^{(l)})$, and $g'_\tau(x) = \tau p x^{p-1}, x \in [0\ \infty)$. Then, by the anti-monotone gradient property of $g_\tau(x)$, since $\sigma_1^{(l)} \geqslant \sigma_2^{(l)} \geq ... \geq \sigma_c^{(l)} \geq 0$, we have $0 \leq w_1^{(l)} \leq w_2^{(l)} \leq ... \leq w_c^{(l)}$. According to (14), we minimize its right hand instead of $g_\tau(\sigma_i)$, leading to the following relaxed problem

$$\mathbf{Z}_{G_k}^{(l+1)} = arg\min_{\mathbf{Z}_{G_k}} g_\tau(\sigma_i^{(l)}) + w_i^{(l)}(\sigma_i - \sigma_i^{(l)}) + h(\mathbf{Z}_{G_k}). \tag{15}$$

Removing the constant terms, the problem (15) equivalently becomes

$$\mathbf{Z}_{G_k}^{(l+1)} = arg\min_{\mathbf{Z}_{G_k}} w_i^{(l)}(\sigma_i(\mathbf{Z}_{G_k})) + h(\mathbf{Z}_{G_k}). \tag{16}$$

Furthermore, instead of updating $\mathbf{Z}_{G_k}^{(l+1)}$ by solving (16) directly, we linearize $h(\mathbf{Z}_{G_k})$ at $\mathbf{Z}_{G_k}^{(l)}$ and add a proximal term:

$$h(\mathbf{Z}_{G_k}) \approx h(\mathbf{Z}_{G_k}^{(l)}) + < \nabla h(\mathbf{Z}_{G_k}^{(l)}), \mathbf{Z}_{G_k} - \mathbf{Z}_{G_k}^{(l)} > + \frac{\mu}{2}\|\mathbf{Z}_{G_k} - \mathbf{Z}_{G_k}^{(l)}\|_F^2, \tag{17}$$

where $\mu > L(h)$, and $L(h) > 0$ is the Lipschitz constant of ∇h. Hence, $\mathbf{Z}_{G_k}^{(l+1)}$ is

$$\mathbf{Z}_{G_k}^{(l+1)} = arg\min_{\mathbf{Z}_{G_k}} w_i^{(l)}(\sigma_i(\mathbf{Z}_{G_k})) + \frac{\mu}{2}\|\mathbf{Z}_{G_k} - (\mathbf{Z}_{G_k}^{(l)} - \frac{1}{\mu}\nabla h(\mathbf{Z}_{G_k}^{(l)}))\|_F^2. \tag{18}$$

Although (18) is still non-convex, it actually has a closed-form solution based on Theorem 1 in [18]. Therefore, by iteratively computing $w_i^{(l)}$ and updating $\mathbf{Z}_{G_k}^{(l+1)}$, we achieve the efficient solution to the non-convex low-rank minimization (13). The convergence analysis is provided by Theorem 2 in [17]. This process is applied for all $\mathbf{Z}_{G_k}^{(l+1)}$, which are then returned to their original positions and averaged at each pixel to obtain the solution for Eq. (10).

In light of all derivations above, the complete description of proposed l_p nuclear norm based ADMM framework for image CS recovery is given in **Algorithm 1**.

Algorithm 1 L_p Nuclear Norm based ADMM Framework for CS

Input: The observed measurement $\mathbf{b}$, the measurement matrix $\mathbf{A}$
Initialization: Set initial estimate $\mathbf{x}^{(0)}$
Repeat
  Update $\mathbf{x}^{(t+1)}$ by computing Eq. (9);
  For Each $\mathbf{Z}_{G_k}$
    Update $\mathbf{Z}_{G_k}^{(t+1)}$ by solving problem (18);
  End For
  Update $\mathbf{z}^{(t+1)}$ by averaging all $\mathbf{Z}_{G_k}^{(t+1)}$;
  Update $\mathbf{b}^{(t+1)}$ by computing Eq. (7);
  $t \leftarrow t + 1$;
Until maximum iteration number is reached
Output: Final recovered image $\hat{\mathbf{x}}$.

Figure 2: Test images: *House, Barbara, Leaves, Monarch, Parrots, Vessels.*

4 Experimental Results

In this section, experimental results are presented to evaluate the performance of the proposed l_p nuclear norm based ADMM framework for image CS recovery. Six test images are shown in Fig. 2. In our experiments, the CS measurements are obtained by applying a Gaussian random projection matrix to the original image signal at block level, i.e., block-based CS with block size of 32×32 [5]. The proposed algorithm is compared with four representative CS recovery methods in literature, i.e., DWT [5], MH [6], CoS [9], and SGSR [11], which deal with image signals in the dual-tree wavelet domain, random projection residual domain, hybrid space-transform domain, and adaptive structural group domain, respectively. It is worth emphasizing that SGSR is known as the state-of-the-art algorithm for image CS recovery [1].

It is also necessary to stress that the choice for all the parameters is general, and can be generalized to other natural images, which has been verified in our experiments. In this paper, we set $p = 0.2$ and exploit the results of MH as initialization of the proposed algorithm for image CS recovery. The PSNR comparisons for all the test images in the cases of 25% to 45% measurements are provided in Table 1. The proposed algorithm provides quite promising results, achieving the highest PSNR among the five comparative algorithms over all the cases, which can improve roughly 8.2 dB, 5.8 dB, 5.4 dB, and 2.9 dB on average, compared with DWT, MH, CoS, and SGSR respectively.

The visual results of the recovered image *Monarch* by various algorithms are presented in Fig. 3. Obviously, DWT generates the worst perceptual results. The CS

[1]We would like to thank the authors of [5], [6], [9] and [11] for kindly providing their software or codes. Our source code is available on the website http://124.207.250.90/staff/zhangjian/.

Table 1: PSNR comparisons with different CS recovery methods (dB)

subrate	*Algorithm*	*House*	*Barbara*	*Leaves*	*Monarch*	*Parrots*	*Vessels*	*Average*
0.25	DWT	34.29	25.43	23.60	27.88	29.51	26.49	27.87
	MH	34.94	31.97	26.38	28.25	30.07	28.69	30.05
	RCoS	35.07	27.45	29.25	30.36	29.50	28.55	30.03
	SGSR	36.53	34.84	31.07	30.48	31.13	32.75	32.05
	Proposed	**38.22**	**35.76**	**34.17**	**34.12**	**32.31**	**36.31**	**35.15**
0.35	DWT	35.99	27.22	25.26	29.96	32.12	28.21	29.79
	MH	36.41	34.31	28.95	30.39	32.12	31.69	32.31
	RCoS	36.16	30.75	32.49	32.69	31.18	33.15	32.74
	SGSR	38.21	36.82	34.44	33.47	33.78	36.26	35.50
	Proposed	**40.25**	**38.21**	**37.97**	**37.34**	**35.28**	**40.70**	**38.29**
0.45	DWT	37.36	29.17	27.02	31.72	34.21	30.58	31.68
	MH	37.80	36.01	30.96	31.98	34.67	34.41	34.31
	RCoS	38.05	33.61	35.25	34.95	33.23	36.05	35.19
	SGSR	39.87	38.55	37.05	35.85	36.72	38.67	37.79
	Proposed	**41.89**	**40.06**	**41.01**	**39.89**	**37.65**	**43.80**	**40.72**

recovered images by MH and CoS possess much better visual quality than DWT, but still suffer from some undesirable artifacts, such as ringing effects and lost details. SGSR produces better results than MH and CoS, almost eliminating the ringing effects. However, the proposed algorithm preserves sharper edges and finer details, showing much clearer and better visual results than the other competing methods. which fully demonstrates the effectiveness of the l_p nuclear norm. In addition, we provide empirical evidence to illustrate the nice convergence of the proposed CS recovery scheme. Fig. 4 plots the evolutions of PSNR versus iteration numbers for image *Monarch* with various subrates of measurements. One can clearly see that with the growth of iteration number, all the PSNR curves increase monotonically and ultimately become flat and stable, exhibiting good convergence.

The complexity of the proposed framework is discussed below. Assume that the number of image pixels is N, that the average time to compute similar patches for each exemplar patch is T_s . The SVD of each $\mathbf{X}_{G_k}$ with size of $B_s \times c$ is $\mathcal{O}\left(B_s c^2\right)$. Hence, the total complexity is $\mathcal{O}\left(N\left(B_s c^2 + T_s\right)\right)$. For a 256×256 image, the proposed algorithm requires about 9$\sim$10 minutes for CS recovery on an Intel(R) Core(TM) 3.40GHz PC under Matlab R2015a environment, while CoS and SGSR require about 6$\sim$7 minutes.

5 Conclusion

In this paper, the l_p $(0 < p < 1)$ penalty function is extended on singular values of a matrix to characterize low-rank prior model, and a novel non-nonvex l_p nuclear norm based ADMM framework for image CS recovery is proposed. Extensive experiments manifest that the developed general CS recovery framework is able to increase image recovery quality by a large margin compared with the current existing state-of-the-art methods, further offering a successful instance to corroborate the CS theory applied for real signals (i.e., natural images). Ongoing work includes the extensions on a variety of other image restoration applications by taking advantage of the proposed l_p

(a) Original (b) DWT (27.88 dB) (c) MH (28.25 dB)

(d) CoS (30.36 dB) (e) SGSR (30.48 dB) (f) Proposed (34.12 dB)

Figure 3: Visual comparison of CS recovered results for *Monarch* (*subrate* = 0.25).

nuclear norm prior and the utilization of GPU hardware to address the parallelization.

6 Acknowledgement

This work was supported in part by the National Basic Research Program of China (2015CB351800), National Natural Science Foundation of China (No. 61572047 and No. 61421062), and Shenzhen Peacock Plan, which are gratefully acknowledged.

7 References

[1] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. E. Kelly, R. G. Baraniuk *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, p. 83, 2008.

[2] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[4] Y. Kim, M. S. Nadar, and A. Bilgin, "Compressed sensing using a gaussian scale mixtures model in wavelet domain," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 3365–3368.

[5] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 3021–3024.

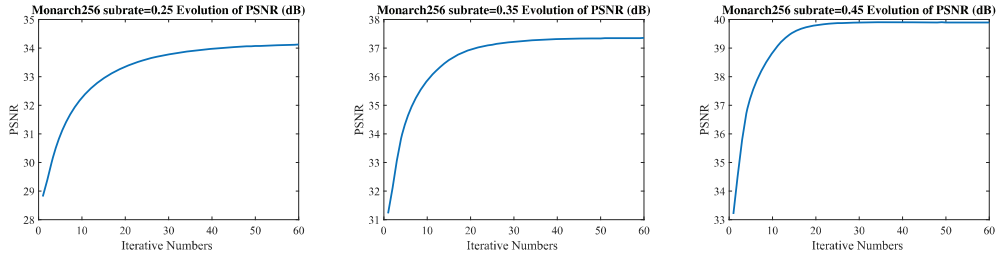

Figure 4: Progression of the PSNR (dB) results achieved by the proposed algorithm for images *Monarch* with respect to iteration number in different CS subrates.

[6] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on.* IEEE, 2011, pp. 1193–1198.

[7] J. Zhang, C. Zhao, D. Zhao, and W. Gao, "Image compressive sensing recovery using adaptively learned sparsifying basis via l0 minimization," *Signal Processing*, vol. 103, pp. 114–126, 2014.

[8] C. Zhao, S. Ma, and W. Gao, "Image compressive-sensing recovery using structured laplacian sparsity in dct domain and multi-hypothesis prediction," in *International Conference on Multimedia and Expo (ICME).* IEEE, 2014, pp. 1–6.

[9] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, and W. Gao, "Compressed sensing recovery via collaborative sparsity," in *Data Compression Conference (DCC), 2012.* IEEE, 2012, pp. 287–296.

[10] ——, "Image compressive sensing recovery via collaborative sparsity," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 380–391, 2012.

[11] J. Zhang, D. Zhao, F. Jiang, and W. Gao, "Structural group sparse representation for image compressive sensing recovery," in *Data Compression Conference (DCC), 2013.* IEEE, 2013, pp. 331–340.

[12] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Computer Vision and Pattern Recognition (CVPR)*, 2014.

[13] J. F. Cai, E. J. Cands, and Z. Shen, "A singular value thresholding algorithm for matrix completion." *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[14] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, 2010.

[15] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, 2014.

[16] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, 2014.

[17] C. Lu, J. Tang, S. Yan, and Z. Lin, "Generalized nonconvex nonsmooth low-rank minimization," in *Computer Vision and Pattern Recognition (CVPR)*, 2014.

[18] K. Chen, H. Dong, and K.-S. Chan, "Reduced rank regression via adaptive nuclear norm penalization," *Biometrika*, pp. 901–920, 2013.