# Compressive Sensing via Nonlocal Low-rank Regularization

Weisheng Dong, Guangming Shi, *Senior member, IEEE*, Xin Li, Yi Ma, *Fellow, IEEE*, and Feng Huang

***Abstract*— Sparsity has been widely exploited for exact reconstruction of a signal from a small number of random measurements. Recent advances have suggested that structured or group sparsity often leads to more powerful signal reconstruction techniques in various compressed sensing (CS) studies. In this paper, we propose a nonlocal low-rank regularization (NLR) approach toward exploiting structured sparsity and explore its application into CS of both photographic and MRI images. We also propose the use of a nonconvex $\log\det(\boldsymbol{X})$ as a smooth surrogate function for the rank instead of the convex nuclear norm; and justify the benefit of such a strategy using extensive experiments. To further improve the computational efficiency of the proposed algorithm, we have developed a fast implementation using the alternative direction multiplier method (ADMM) technique. Experimental results have shown that the proposed NLR-CS algorithm can significantly outperform existing state-of-the-art CS techniques for image recovery.**

***Index Terms*— compresses sensing, low-rank approximation, structured sparsity, nonconvex optimization, alternative direction multiplier method.**

## I. Introduction

The theory of compressive sensing (CS) [1], [2] has attracted considerable research interests from signal/image processing communities. By achieving perfect reconstruction of a sparse signal from a small number of random measurements, generated with either random Gaussian matrixes or partial Fourier measurement matrixes, the theory has the potential of significantly improving the energy efficiency of sensors in real-world applications. For instance, several CS-based imaging systems have been built in recent years, e.g., the single-pixel camera [3], compressive spectral imaging system [4], and high-speed video camera [5]. Among those rapidly-growing applications of the CS theory, the compressive sensing Magnetic Resonance Imaging (CS-MRI) [6] is arguably among the high-impact ones because of its promise in significantly reducing the acquisition time of MRI scanning. Since long acquisition time remains one of the primary obstacles in the clinical practice of MRI, any technology related to faster scanning could be valuable. Moreover, the combination of CS-MRI with conventional fast MRI methods (e.g. SMASH [35], SENSE [33], etc.) for further speed-up has drawn increasingly more attention from the MRI field.

W. Dong and G. Shi are with School of Electronic Engineering, Xidian University, Xi'an, 710071, China (e-mail: wsdong@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

X. Li is with the Lane Department of CSEE, West Virginia University, Morgantown, WV 26506-6109 USA (e-mail: xin.li@ieee.org).

Y. Ma is with the School of Information Science and Technology, ShanghaiTech University, China. He is also affiliated with the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, USA (e-mail: mayi@shanghaitech.edu.cn).

Feng Huang is with Philips Research China, Shanghai, China (e-mail: f.huang@philips.com).

Since exploiting a prior knowledge of the original signals (e.g., sparsity) is critical to the success of CS theory, numerous studies have been performed to build more realistic models for real-world signals. Conventional CS recovery exploits the $l_1$-norm based sparsity of a signal and the resulting convex optimization problems can be efficiently solved by the class of surrogate-function based methods [7], [8], [9]. More recently, the concept of sparsity has evolved into various sophisticated forms including model-based or Bayesian [18], nonlocal sparsity [10], [21], [11] and structured/group sparsity [19], [20], where exploiting higher-order dependency among sparse coefficients has shown beneficial to CS recovery. On the other hand, several experimental studies have shown that nonconvex optimization based approach toward CS [22], [23] often leads to more accurate reconstruction than their convex counterpart though at the cost of higher computational complexity. Therefore, it is desirable to pursue computationally more efficient solutions that can exploit the benefits of both high-order dependency among sparse coefficients and nonconvex optimization.

In [12] we have shown an intrinsic relationship between simultaneous sparse coding (SSC) [20] and low-rank approximation. Such connection has inspired us to solve the challenging SSC problem by the singular-value thresholding (SVT) method [24], leading to state-of-the-art image denoising results. In this paper, we propose a unified variational framework for nonlocal low-rank regularization of CS recovery. To exploit the nonlocal sparsity of natural or medical images, we propose to regularize the CS recovery by patch grouping and low-rank approximation. Specifically, for each exemplar image patch we group a set of similar image patches to form a data matrix $\boldsymbol{X}$. Since each patch contain similar structures, the rank of this data matrix $\boldsymbol{X}$ is low implying a useful image prior. To more efficiently solve the problem of rank minimization, we propose to use the $\log\det(\boldsymbol{X})$ as a smooth surrogate function for the rank (instead of using the convex nuclear norm), which lends itself to iterative singular-value thresholding. Experimental results on both natural images and complex-valued MRI images show that our low-rank approach is capable of achieving dramatically more accurate reconstruction (PSNR gain over $> 2dB$) than other competing approaches. To the best of our knowledge, this work represents the current state-of-the-art in making the CS theory to work for the class of images containing diverse and realistic structures.

## II. Background

In the theory of CS , one seeks the perfect reconstruction of a signal $\boldsymbol{x} \in \mathbb{C}^N$ from its $M$ randomized linear measurements, i.e., $\boldsymbol{y} = \boldsymbol{\Phi x}, \boldsymbol{y} \in \mathbb{C}^M$, where $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}, M < N$ is the measurement matrix. Since $M < N$, the matrix $\boldsymbol{\Phi}$ is rank-deficient, there generally exists more than one $\boldsymbol{x} \in \mathbb{C}^n$ that can yield the same measurements $\boldsymbol{y}$. The CS theory guarantees perfect reconstruction of a sparse (or compressive) signal $\boldsymbol{x}$ if $\boldsymbol{\Phi}$ satisfies the so called *restricted isometry property* (RIP) [1], [2]. It has been known that a large class of random matrixes have the RIP with high probabilities. To recover $\boldsymbol{x}$ from the measurement $\boldsymbol{y}$, prior knowledge of $\boldsymbol{x}$ is needed. Standard CS method recovers the unknown signal by pursuing the sparsest signal $\boldsymbol{x}$ that satisfies $\boldsymbol{y} = \boldsymbol{\Phi x}$, namely

$$\boldsymbol{x} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_0, \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Phi x}, \tag{1}$$

where $\|\cdot\|_0$ is a pseudo norm counting the number of non-zero entries of $\boldsymbol{x}$. In theory a $K$-sparse signal can be perfectly recovered from as low as $M = 2K$ measurements [1].

However, since $\|\cdot\|_0$ norm minimization is a difficult combinatorial optimization problem, solving Eq.(1) directly is both NP-hard and unstable in the presence of noise. For this reason, it has been proposed to replace the nonconvex $l_0$ norm by its convex $l_1$ counterpart - i.e.,

$$\boldsymbol{x} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_1, \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Phi x}, \tag{2}$$

It has been shown in [1] that solving this $l_1$ norm optimization problem can recover a $K$-sparse signal from $M = O(Klog(N/K))$ random measurements. The optimization problem in Eq.(2) is convex and corresponds to linear programming known as basis pursuit [1], [2]. By selecting an appropriate regularization parameter $\lambda$, Eq.(2) is equivalent to the following unconstrained optimization problem:

$$\boldsymbol{x} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1, \tag{3}$$

The above $l_1$-minimization problem can be efficiently solved by various methods, such as iterative shrinkage algorithm [7], Bregman Split algorithm [9] and alternative direction multiplier method (ADMM) [13]. Recent advances have also shown that better CS recovery performance can be obtained by replacing the $l_1$ norm with a non-convex $l_p (0 < p < 1)$ norm though at the price of higher computational complexity [22].

In addition to the standard $l_p (0 \leq p \leq 1)$ sparsity, recent advances in CS theory use structured sparsity to model a richer class of signals. By modeling high-order dependency among sparse coefficients, one can greatly reduce the uncertainty about the unknown signal leading to more accurate reconstruction [18]. Structured sparsity is particularly important to the modeling of natural signals (e.g., photographic images) that often exhibit rich self-repetitive structures. Exploiting the so-called nonlocal self-similarity has led to the well-known nonlocal means methods [10], block-matching 3D denoising [21] and simultaneous sparse coding (SSC) [20]. Most recently, a clustering-based structured sparse model is proposed in [15], which unified the local sparsity and nonlocal sparsity into a variational framework. Both [20], [15] have shown state-of-the-art image restoration results. However, their effectiveness in CS applications has not been documented in the open literature to the best of our knowledge. In this paper, we will present a unified variational framework for CS recovery exploiting the nonlocal structured sparsity via low-rank approximation.

## III. Nonlocal Low-rank Regularization for CS Recovery

In this section, we present a new model of nonlocal low-rank regularization for CS recovery. The proposed regularization model consists of two components: patch grouping for characterizing self-similarity of a signal and low-rank approximation for sparsity enforcement. The basic assumption underlying the proposed approach is that self-similarity is abundant in signals of our interest. Such assumption implies that a sufficient number of similar patches can be found for any exemplar patch of size $\sqrt{n} \times \sqrt{n}$ at position $i$ denoted by $\hat{\boldsymbol{x}}_i \in \mathbb{C}^n$. For each exemplar patch $\hat{\boldsymbol{x}}_i$, we perform a variant of k-nearest-neighbor search within a local window (e.g., $40 \times 40$) - i.e.,

$$G_i = \{i_j | \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{i_j}\| < T\}, \tag{4}$$

where $T$ is a pre-defined threshold, and $G_i$ denotes the collection of positions corresponding to those similar patches. After patch grouping, we obtain a data matrix $\boldsymbol{X}_i = [\boldsymbol{x}_{i_0}, \boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_{m-1}}], \boldsymbol{X}_i \in \mathbb{C}^{n \times m}$ for each exemplar patch $\boldsymbol{x}_i$, where each column of $\boldsymbol{X}_i$ denotes a patch similar to $\boldsymbol{x}_i$ (including $\boldsymbol{x}_i$ itself).

Under the assumption that these image patches have similar structures, the formed data matrix $\boldsymbol{X}_i$ has a low-rank property. In practice, $\boldsymbol{X}_i$ may be corrupted by some noise, which could lead to the deviation from the desirable low-rank constraint. One possible solution is to model the data matrix $\boldsymbol{X}_i$ as: $\boldsymbol{X}_i = \boldsymbol{L}_i + \boldsymbol{W}_i$, where $\boldsymbol{L}_i$ and $\boldsymbol{W}_i$ denote the low-rank matrix and the Gaussian noise matrix respectively. Then the low-rank matrix $\boldsymbol{L}_i$ can be recovered by solving the following optimization problem:

$$\boldsymbol{L}_i = \underset{\boldsymbol{L}_i}{\operatorname{argmin}} \operatorname{rank}(\boldsymbol{L}_i), \text{ s.t. } \|\boldsymbol{X}_i - \boldsymbol{L}_i\|_F^2 \leq \sigma_w^2, \tag{5}$$

where $\|\cdot\|_F^2$ denotes the Frobenious norm and $\sigma_w^2$ denotes the variance of additive Gaussian noise. In general, the rank-minimization is an NP-hard problem; hence we cannot solve Eq.(5) directly. To obtain an approximated solution, the nuclear norm $\|\cdot\|_*$ (sum of the singular values) can be used as a convex surrogate of the rank. Using the nuclear norm, the rank minimization problem can be efficiently solved by the technique of singular value thresholding (SVT) [24]. Despite good theoretical guarantee of the correctness [16], we conjecture that non-convex optimization toward rank minimization could lead to better recovery results just like what has been observed in the studies of $l_p$-optimization.

In this paper, we consider a smooth but *non-convex* surrogate of the rank rather than the nuclear norm. It has been shown in [17] that for a symmetric positive semidefinite

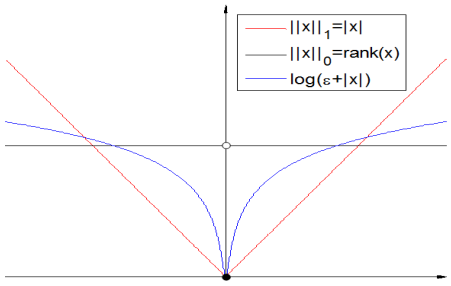



Fig. 1. Comparison of $L(x,\varepsilon)$, $\text{rank}(x) = \|x\|_0$ and the nuclear norm $= \|x\|_1$ in the case of a scalar.

matrix $\boldsymbol{X} \in \mathbb{R}^{n\times n}$, the rank minimization problem can be approximately solved by minimizing the following functional:

$$E(\boldsymbol{X},\varepsilon) \doteq \log\det(\boldsymbol{X}+\varepsilon\boldsymbol{I}), \tag{6}$$

where $\varepsilon$ denotes a small constant value. Note that this function $E(\boldsymbol{X},\varepsilon)$ approximates the sum of the logarithm of singular values (up to a scale). Therefore, the function $E(\boldsymbol{X},\varepsilon)$ is smooth yet non-convex. The logdet as a non-convex surrogate for the rank has also be more carefully justified from an information-theoretic perspective such as in [34]. Figure 1 shows the comparison of a non-convex surrogate function, the rank, and the nuclear norm in the scalar case. It can be seen from Fig. 1 that the surrogate function $E(\boldsymbol{X},\varepsilon)$ can better approximate the rank than the nuclear norm.

For a general matrix $\boldsymbol{L}_i \in \mathbb{C}^{n\times m}$, $n \le m$ that is neither square nor positive semidefinite, we slightly modify Eq.(6) into

$$\begin{aligned} L(\boldsymbol{L}_i,\varepsilon) &\doteq \log\ \det((\boldsymbol{L}_i\boldsymbol{L}_i^\top)^{1/2}+\varepsilon\boldsymbol{I}) \\ &= \log\ \det(\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}^{-1}+\varepsilon\boldsymbol{I}) \\ &= \log\ \det(\boldsymbol{\Sigma}^{1/2}+\varepsilon\boldsymbol{I}), \end{aligned} \tag{7}$$

where $\boldsymbol{\Sigma}$ is the diagonal matrix whose diagonal elements are eigenvalues of matrix $\boldsymbol{L}_i\boldsymbol{L}_i^\top$, i.e., $\boldsymbol{L}_i\boldsymbol{L}_i^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{-1}$, and $\boldsymbol{\Sigma}^{1/2}$ is the diagonal matrix whose diagonal elements are the singular values of the matrix $\boldsymbol{L}_i$. Therefore, we can see that $L(\boldsymbol{L}_i,\varepsilon)$ is a logdet$(\cdot)$ surrogate function of rank$(\boldsymbol{L}_i)$ obtained by setting $\boldsymbol{X} = (\boldsymbol{L}_i\boldsymbol{L}_i^\top)^{1/2}$. We then propose the following low-rank approximation problem for solving $\boldsymbol{L}_i$

$$\boldsymbol{L}_i = \underset{\boldsymbol{L}_i}{\text{argmin}}\, L(\boldsymbol{L}_i,\varepsilon) \text{ s.t. } \|\boldsymbol{X}_i-\boldsymbol{L}_i\|_F^2 \le \sigma_w^2. \tag{8}$$

In practice, this constrained minimization problem can be solved in its Lagrangian form, namely

$$\boldsymbol{L}_i = \underset{\boldsymbol{L}_i}{\text{argmin}}\, \|\boldsymbol{X}_i-\boldsymbol{L}_i\|_F^2 + \lambda L(\boldsymbol{L}_i,\varepsilon). \tag{9}$$

Eq.(9) is equivalent to Eq.(8) by selecting a proper $\lambda$. For each exemplar image patch, we can approximate the matrix $\boldsymbol{X}_i$ with a low-rank matrix $\boldsymbol{L}_i$ by solving Eq.(9).

How to use the patch-based low-rank regularization model for CS image recovery? The basic idea is to enforce the low-rank property over the sets of nonlocal similar patches for each extracted exemplar patch along with the constraint of linear measurements. With the proposed low-rank regularization term, we propose the following global objective functional for CS recovery:

$$\begin{aligned} (\hat{\boldsymbol{x}},\hat{\boldsymbol{L}}_i) = \underset{\boldsymbol{x},\boldsymbol{L}_i}{\text{argmin}}\, &\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \eta\sum_i\{\|\tilde{\boldsymbol{R}}_i\boldsymbol{x}-\boldsymbol{L}_i\|_F^2 + \\ &\lambda L(\boldsymbol{L}_i,\varepsilon)\}, \end{aligned} \tag{10}$$

where $\tilde{\boldsymbol{R}}_i\boldsymbol{x} \doteq [\boldsymbol{R}_{i_0}\boldsymbol{x}, \boldsymbol{R}_{i_1}\boldsymbol{x}, \ldots, \boldsymbol{R}_{i_{m-1}}\boldsymbol{x}]$ denotes the matrix formed by the set of similar patches for every exemplar patch $\boldsymbol{x}_i$. The proposed nonlocal low-rank regularization can exploit both the group sparsity of similar patches and nonconvexity of rank minimization; thus achieve better recovery than previous methods. In the next section, we will show that the proposed objective functional can be efficiently solved by the method of alternative minimization.

## IV. Optimization Algorithm for CS Image Recovery

The proposed objective functional can be solved by alternatively minimizing the objective functional with respect to the whole image $\boldsymbol{x}$ and low-rank data matrices $\boldsymbol{L}_i$.

### A. Low-rank Matrix Optimization via Iterative Single Value Thresholding

With an initial estimate of the unknown image, we first extract exemplar patches $\boldsymbol{x}_i$ at every $l$ pixels along each direction and group a set of similar patches for each $\boldsymbol{x}_i$, as described in section III. Then, we propose to solve the following minimization problem for each $\boldsymbol{L}_i$:

$$\boldsymbol{L}_i = \underset{\boldsymbol{L}_i}{\text{argmin}}\, \eta\|\tilde{\boldsymbol{R}}\boldsymbol{x}-\boldsymbol{L}_i\|_F^2 + \lambda L(\boldsymbol{L}_i,\varepsilon). \tag{11}$$

Since $L(\boldsymbol{L}_i,\varepsilon)$ is approximately the sum of the logarithm of singular values (up to a scale), Eq.(11) can be rewritten as

$$\min_{\boldsymbol{L}_i} \|\boldsymbol{X}_i-\boldsymbol{L}_i\|_F^2 + \frac{\lambda}{\eta}\sum_{j=1}^{n_0}\log(\sigma_j(\boldsymbol{L}_i)+\varepsilon). \tag{12}$$

where $\boldsymbol{X}_i = \tilde{\boldsymbol{R}}_i\boldsymbol{x}$, $n_0 = \min\{n,m\}$, and $\sigma_j(\boldsymbol{L}_i)$ denotes the $j^{th}$ singular value of $\boldsymbol{L}_i$. For simplicity, we use $\sigma_j$ to denote the $j^{th}$ singular value of $\boldsymbol{L}_i$. Though $\sum_{j=1}^n \log(\sigma_j+\varepsilon)$ is non-convex, we can solve it efficiently using a local minimization method (a local minimum will be sufficient when an annealing strategy is used - please refer to Algorithm 1). Let $f(\boldsymbol{\sigma}) = \sum_{j=1}^n \log(\sigma_j+\varepsilon)$. Then $f(\boldsymbol{\sigma})$ can be approximated using first-order Taylor expansion as

$$f(\boldsymbol{\sigma}) = f(\boldsymbol{\sigma}^{(k)}) + \langle\nabla f(\boldsymbol{\sigma}^{(k)}), \boldsymbol{\sigma}-\boldsymbol{\sigma}^{(k)}\rangle, \tag{13}$$

where $\boldsymbol{\sigma}^{(k)}$ is the solution obtained in the $k^{th}$ iteration. Therefore, Eq.(12) can be solved by iteratively solving

$$\boldsymbol{L}_i^{(k+1)} = \underset{\boldsymbol{L}_i}{\text{argmin}}\, \|\boldsymbol{X}_i-\boldsymbol{L}_i\|_F^2 + \frac{\lambda}{\eta}\sum_{j=1}^{n_0}\frac{\sigma_j}{\sigma_j^{(k)}+\varepsilon}, \tag{14}$$

where we have used the fact that $\nabla f(\boldsymbol{\sigma}^{(k)}) = \sum_{j=1}^{n_0} \frac{1}{\sigma_j^{(k)}+\varepsilon}$ and ignored the constants in Eq.(13). For convenience, we can rewrite Eq. (14) into

$$\boldsymbol{L}_i^{(k+1)} = \underset{\boldsymbol{L}_i}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{X}_i - \boldsymbol{L}_i\|_F^2 + \tau\varphi(\boldsymbol{L}_i, \boldsymbol{w}^{(k)}), \quad (15)$$

where $\tau = \lambda/(2\eta)$ and $\varphi(\boldsymbol{L}_i, \boldsymbol{w}) = \sum_j^{n_0} w_j^{(k)}\sigma_j$ denotes the *weighted nuclear norm* with weights $w_j^{(k)} = 1/(\sigma_j^{(k)} + \varepsilon)$. Notice that since the singular values $\sigma_j$ are ordered in a descending order, the weights are ascending.

It is known that in the case of a real matrix, the weighted nuclear norm is a convex function only if the weights are descending, and the optimal solution to (15) is given by a weighted singular value thresholding operator, known as the proximal operator. In our case, the weights are ascending hence (15) is not convex. So we do not expect to find its global minimizer. In addition, we are dealing with a complex matrix. Nevertheless, one could still show that the weighted singular value thresholding gives one (possible local) minimizer to (15):

*Theorem 1 (Proximal Operator of Weighted Nuclear Norm):* For each $\boldsymbol{X} \in \mathbb{C}^{n\times m}$ and $0 \leq w_1 \leq \cdots \leq w_{n_0}$, $n_0 = \min\{m, n\}$, a minimizer to

$$\min_{\boldsymbol{L}} \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{L}\|_F^2 + \tau\varphi(\boldsymbol{L}, \boldsymbol{w}) \quad (16)$$

is given by the weighted singular value thresholding operator $\mathcal{S}_{\boldsymbol{w},\tau}(\boldsymbol{X})$:

$$\mathcal{S}_{\boldsymbol{w},\tau}(\boldsymbol{X}) := \boldsymbol{U}(\boldsymbol{\Sigma} - \tau\text{diag}(\boldsymbol{w}))_+\boldsymbol{V}^\top, \quad (17)$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the SVD of $\boldsymbol{X}$ and $(x)_+ = \max\{x, 0\}$.

The detailed proof of the Theorem 1 is given in the Appendix. Based on this theorem, the reconstructed matrix in the $(k+1)^{th}$ iteration is then obtained by

$$\boldsymbol{L}_i^{(k+1)} = \boldsymbol{U}(\tilde{\boldsymbol{\Sigma}} - \tau\text{diag}(\boldsymbol{w}^{(k)}))_+\boldsymbol{V}^\top, \quad (18)$$

where $\boldsymbol{U}\tilde{\boldsymbol{\Sigma}}\boldsymbol{V}^\top$ is the SVD of $\boldsymbol{X}_i$ and $w_j^{(k)} = 1/(\sigma_j^{(k)} + \varepsilon)$. Notice that even though the weighted thresholding is only a local minimizer, it always leads to a decreasing in the objective function value. In our implementation, we set $\boldsymbol{w}^{(0)} = [1, 1, \ldots, 1]^\top$ and the first iteration is equivalent to solving an unweighted nuclear norm minimization problem.

When $\boldsymbol{L}_i$ is a vector, the $\log\det(\cdot)$ leads to the well-known reweighted $\ell_1$-norm [28]. In [28] it has been shown that the reweighted $\ell_1$-norm performs much better than $\ell_1$-norm in approximating the $\ell_0$-norm and often leads to superior image recovery results. Similarly, our experimental results in the next section show that the $\log\det(\cdot)$ can lead to better CS recovery results than the nuclear norm.

### *B. Image Recovery via Alternative Direction Multiplier Method*

After solving for each $\boldsymbol{L}_i$, we can reconstruct the whole image by solving the following minimization problem:

$$\boldsymbol{x} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \eta\sum_i \|\tilde{\boldsymbol{R}}_i\boldsymbol{x} - \boldsymbol{L}_i\|_F^2. \quad (19)$$

Eq.(19) is a quadratic optimization problem admitting a closed-form solution - i.e.,

$$\boldsymbol{x} = (\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \eta\sum_i \tilde{\boldsymbol{R}}_i^\top\tilde{\boldsymbol{R}}_i)^{-1}(\boldsymbol{\Phi}^H\boldsymbol{y} + \eta\sum_i \tilde{\boldsymbol{R}}_i^\top\boldsymbol{L}_i), \quad (20)$$

where the superscript $H$ denotes the Hermitian transpose operation and $\tilde{\boldsymbol{R}}_i^\top\boldsymbol{L}_i \doteq \sum_{r=0}^{m-1}\boldsymbol{R}_r^\top\boldsymbol{x}_{i_r}$ and $\tilde{\boldsymbol{R}}_i^\top\tilde{\boldsymbol{R}}_i \doteq \sum_{r=0}^{m}\boldsymbol{R}_r^\top\boldsymbol{R}_r$. In Eq.(20), the matrix to be inverted is large. Hence, directly solving Eq.(20) is not possible. In practice, Eq.(20) can be computed by using a conjugate gradient (CG) algorithm.

When the measurement matrix $\boldsymbol{\Phi}$ is a partial Fourier transform matrix that has important applications in high speed MRI, we can derive a much faster algorithm for solving Eq.(19) by using the alternative direction multiplier method (ADMM) technique [13], [29]. The advantage of ADMM is that we can split Eq.(19) into two sub-problems that both admit closed-form solutions. By applying ADMM to Eq.(19), we obtain

$$(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\mu}) = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \beta\|\boldsymbol{x} - \boldsymbol{z} + \frac{\boldsymbol{\mu}}{2\beta}\|_2^2 + \eta\sum_i \|\tilde{\boldsymbol{R}}_i\boldsymbol{z} - \boldsymbol{L}_i\|_F^2, \quad (21)$$

where $\boldsymbol{z} \in \mathbb{C}^N$ is an auxiliary variable, $\boldsymbol{\mu} \in \mathbb{C}^N$ is the Lagrangian multiplier, and $\beta$ is a positive scalar. The optimization of Eq.(21) consists of the following iterations:

$$\begin{aligned}
\boldsymbol{z}^{(l+1)} &= \underset{\boldsymbol{z}}{\operatorname{argmin}}\, \beta^{(l)}\|\boldsymbol{x}^{(l)} - \boldsymbol{z} + \frac{\boldsymbol{\mu}^{(l)}}{2\beta^{(l)}}\|_2^2 + \eta\sum_i \|\tilde{\boldsymbol{R}}_i\boldsymbol{z} - \boldsymbol{L}_i\|_F^2,\\
\boldsymbol{x}^{(l+1)} &= \underset{\boldsymbol{x}}{\operatorname{argmin}}\, \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \beta^{(l)}\|\boldsymbol{x} - \boldsymbol{z}^{(l+1)} + \frac{\boldsymbol{\mu}^{(l)}}{2\beta^{(l)}}\|_2^2,\\
\boldsymbol{\mu}^{(l+1)} &= \boldsymbol{\mu}^{(l)} + \beta^{(l)}(\boldsymbol{x}^{(l+1)} - \boldsymbol{z}^{(l+1)}),\\
\beta^{(l+1)} &= \rho\beta^{(l)},
\end{aligned} \quad (22)$$

where $\rho > 1$ is a constant. For fixed $\boldsymbol{x}^{(l)}$, $\boldsymbol{\mu}^{(l)}$ and $\beta^{(l)}$, $\boldsymbol{z}^{(l+1)}$ admits a closed-form solution:

$$\boldsymbol{z}^{(l+1)} = (\eta\sum_i \tilde{\boldsymbol{R}}_i^\top\tilde{\boldsymbol{R}}_i + \beta^{(l)}\boldsymbol{I})^{-1}(\beta^{(l)}\boldsymbol{x}^{(l)} + \frac{\boldsymbol{\mu}^{(l)}}{2} + \eta\sum_i \tilde{\boldsymbol{R}}_i\boldsymbol{L}_i). \quad (23)$$

Note that the term $\sum_i \tilde{\boldsymbol{R}}_i^\top\tilde{\boldsymbol{R}}_i$ is a diagonal matrix. Each of the entries in the diagonal matrix corresponds to an image pixel location, and its value is the number of overlapping patches that cover the pixel location. The term $\sum_i \tilde{\boldsymbol{R}}_i\boldsymbol{L}_i$ denotes the patch average result - i.e., averaging all of the collected similar patches for each exemplar patch. Therefore, Eq.(23) can easily computed in one step. For fixed $\boldsymbol{z}^{(l+1)}$, $\boldsymbol{\mu}^{(l)}$ and $\beta^{(l)}$, the $\boldsymbol{x}-$subproblem can be solved by computing:

$$(\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \beta^{(l)}\boldsymbol{I})\boldsymbol{x} = (\boldsymbol{\Phi}^H\boldsymbol{y} + \beta^{(l)}\boldsymbol{z}^{(l+1)} - \frac{\boldsymbol{\mu}^{(l)}}{2}). \quad (24)$$

When $\boldsymbol{\Phi}$ is a partial Fourier transform matrix $\boldsymbol{\Phi} = \boldsymbol{D}\boldsymbol{F}$, where $\boldsymbol{D}$ and $\boldsymbol{F}$ denotes the down-sampling matrix and Fourier transform matrix respectively, Eq.(24) can be easily solved by transforming the problem from image space into Fourier space. By substituting $\boldsymbol{\Phi} = \boldsymbol{D}\boldsymbol{F}$ into Eq.(24) and

applying Fourier transform to each side of Eq.(24), we can obtain

$$\boldsymbol{F}((\boldsymbol{D}\boldsymbol{F})^H\boldsymbol{D}\boldsymbol{F}+\beta^{(l)}\boldsymbol{I})\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{D}\boldsymbol{F})^H\boldsymbol{y}+\boldsymbol{F}(\beta^{(l)}\boldsymbol{z}^{(l+1)}-\frac{\boldsymbol{\mu}^{(l)}}{2}) \quad (25)$$

The above equation can be simplified as

$$\boldsymbol{F}\boldsymbol{x} = (\boldsymbol{D}^\top\boldsymbol{D}+\beta^{(l)})^{-1}(\boldsymbol{D}^\top\boldsymbol{y}+\boldsymbol{F}(\beta^{(l)}\boldsymbol{z}^{(l+1)}-\frac{\boldsymbol{\mu}^{(l)}}{2})), \quad (26)$$

where the matrix to be inverted is a diagonal matrix (so it can be computed easily). Then, $\boldsymbol{x}^{(l+1)}$ can be computed by applying inverse Fourier transform to the right hand side of Eq. (26)- i.e.,

$$\boldsymbol{x}^{(l+1)} = \boldsymbol{F}^H\{(\boldsymbol{D}^\top\boldsymbol{D}+\beta^{(l)})^{-1}(\boldsymbol{D}^\top\boldsymbol{y}+\boldsymbol{F}(\beta^{(l)}\boldsymbol{z}^{(l+1)}-\frac{\boldsymbol{\mu}^{(l)}}{2}))\} \quad (27)$$

With updated $\boldsymbol{x}$ and $\boldsymbol{z}$, $\boldsymbol{\mu}$ and $\beta$ can be readily computed according to Eq.(22)

After obtaining an improved estimate of the unknown image, the low-rank matrixes $\boldsymbol{L}_i$ can be updated by Eq.(18). The updated $\boldsymbol{L}_i$ is then used to improve the estimate of $\boldsymbol{x}$ by solving Eq.(19). Such process is iterated until the convergence. The overall procedure is summarized below as **Algorithm 1**.

In **Algorithm 1**, we use the global threshold $\tau$ to solve an unweighted nuclear norm minimization in the first $K_0$ ($K_0 = 45$ in our implementation) iterations to obtain a starting point (which we call "warm start") for the proposed nonconvex logdet based CS method. As shown in Fig. 2, we can see that the use of the warm start step improves the convergence speed and leads to better CS reconstruction. After the first $K_0$ iterations, we compute the adaptive weights $\boldsymbol{w}_i$ based on $\boldsymbol{\sigma}_i^{(k)}$ obtained in the previous iteration. We use ADMM technique to solve Eq.(19) when the measurement matrix is a partial Fourier transform matrix. For other CS measurement matrixes, we can use CG algorithm to solve Eq.(19). To save computational complexity, we update the patch grouping in every $T$ iterations. Empirically, we have found that **Algorithm 1** converges even when the inner loops only executes one iteration (larger $J$ values do not lead to noticeable PSNR gain). Hence, by setting $J = 1$ and $L = 1$ we can save much computational complexity of the proposed CS algorithm.

## V. Experimental Results

In this section, we report the experimental results of the proposed low-rank based CS recovery method. Here we generate the CS measurements by randomly sampling the Fourier transform coefficients of test images. However, the proposed CS recovery method can also be used for other CS sampling schemes. The number of compressive measurements $M$ is measured in terms of the percentages of total number of image pixels $N$ or Fourier coefficients. In our experiments, both the natural images and simulated MR images (complex-valued) are used to verify the performance of the proposed CS method. The main parameters of the proposed algorithms is set as follows: patch size - $6 \times 6$ (i.e., $n = 36$); total $m = 45$ similar patches are selected for each exemplar patch. For better CS recovery performance, the regularization parameter

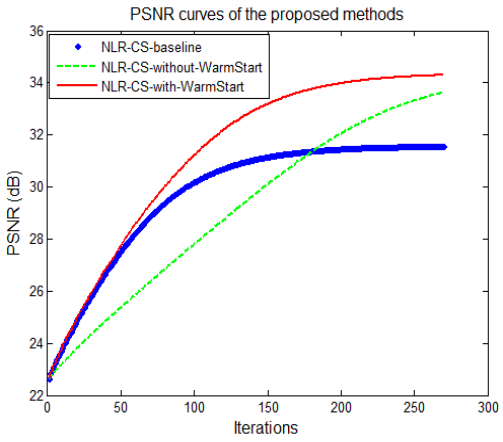



Fig. 2. PSNR curves of the three variants of the proposed CS methods on *Monarch* image at sensing rate $0.2N$ (random subsampling). **NLR-CS-baseline** denotes the proposed CS method using the standard nuclear norm; **NLR-CS-without-WarmStart** denotes the proposed logdet based CS method that does not include the warm start step (by setting $K_0 = 0$); **NLR-CS-with-WarmStart** denotes the proposed logdet based CS method that includes the warm start step (i.e., $K_0 = 45$).

$\lambda$ is tuned separately for each sensing rate. To reduce the computational complexity, we extract exemplar image patch in every 5 pixels along both horizontal and vertical directions. Both natural images and complex-valued MRI images are used in our experiments (the eight test natural images used in our experiment are shown in Fig. 3). Both source codes and test images accompanying this paper can be downloaded from the following website: `http://see.xidian.edu.cn/faculty/wsdong/NLR_Exps.htm`. We first present the experimental results for noiseless CS measurements and then report the results using noisy CS measurements.

### A. Experiments on Noiseless data

Let NLR-CS denote the proposed nonlocal low-rank regularization based CS method. To verify the effectiveness of the logdet function for rank minimization, we have implemented a variant of the proposed NLR-CS method that uses the standard nuclear norm, denoted as NLR-CS-baseline. NLR-CS-baseline can be implemented by slightly modifying **Algorithm 1**. To verify the performance of the proposed nonlocal low-rank regularization based CS methods, we compare them with several competitive CS recovery methods including the total variation (TV) method [30], the iterative reweighted TV method [28] that performs better than TV method, the BM3D based CS recovery method [25] (denoted as BM3D-CS) and a recently proposed MARX-PC method [27]. Note that BM3D is a well-known image restoration method that delivers state-of-the-art denoising results; MARX-PC exploits both the local and nonlocal image dependencies and is among the best CS methods so far. The source codes of all benchmark methods [25], [27], [30] were obtained from the authors' websites. To make a fair comparison among the competing methods, we have carefully tuned their parameters to achieve the best performance. The TV method [30] and ReTV method [28] are implemented based on the well-known l1-magic software. It

**Algorithm 1** CS via low-rank regularization

- **Initialization**:
  (a) Estimate an initial image $\hat{\boldsymbol{x}}$ using a standard CS recovery method (e.g., DCT/Wavelet based recovery method);
  (b) Set parameters $\lambda$, $\eta$, $\tau = \lambda/(2\eta)$, $\beta$, $K$, $J$, and $L$.
  (c) Initialize $\boldsymbol{w}_i = [1, 1, \ldots, 1]^\top$, $\boldsymbol{x}^{(1)} = \hat{\boldsymbol{x}}$, $\boldsymbol{\mu}^{(1)} = \mathbf{0}$;
  (d) Grouping a set of similar patches $G_i$ for each exemplar patch using $\boldsymbol{x}^{(1)}$;
- **Outer loop**: for $k = 1, 2, \ldots, K$ **do**
  (a) Patch dataset $\boldsymbol{X}_i$ construction: grouping a set of similar patches for each exemplar patch $\boldsymbol{x}_i$ using $\boldsymbol{x}^{(k)}$;
  (b) Set $\boldsymbol{L}_i^{(0)} = \boldsymbol{X}_i$;
  (c) **Inner loop** (Low-rank approximation, solving Eq. (12)): for $j = 1, 2, \ldots, J$ **do**
    (I) If ($k > K_0$), update the weights $\boldsymbol{w}_i = 1/(\boldsymbol{\sigma}(\boldsymbol{L}_i^{(j-1)}) + \varepsilon)$;
    (II) Singular values thresholding via Eq.(18): $\boldsymbol{L}_i^{(j)} = \mathcal{S}_{\boldsymbol{w}_i,\tau}(\boldsymbol{X}_i)$.
    (III) Output $\boldsymbol{L}_i = \boldsymbol{L}_i^{(j)}$ if $j = J$.
  **End for**
  (d) **Inner loop** (Solving Eq.(19)): for $l = 1, 2, \ldots, L$ **do**
    (I) Compute $\boldsymbol{z}^{(l+1)}$ via Eq.(23);
    (II) Compute $\boldsymbol{x}^{(l+1)}$ via Eq.(27);
    (III) $\boldsymbol{\mu}^{(l+1)} = \boldsymbol{\mu}^{(l)} + \beta^{(l)}(\boldsymbol{x}^{(l+1)} - \boldsymbol{z}^{(l+1)})$
    (IV) $\beta^{(l+1)} = \rho\beta^{(l)}$.
    (V) Output $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(l+1)}$ if $l = L$.
  **End for**
  (e) If $mod(k, T) = 0$, update the patch grouping.
  (f) Output the reconstructed image $\hat{\boldsymbol{x}} = \boldsymbol{x}^{(k)}$ if $k = K$.

**End for**

has been found that the CS reconstruction performances of TV and ReTV methods are not sensitive to their parameters. We have also carefully tuned the parameters of the BM3D-CS method for each CS sensing rates and subsampling scheme for the purpose of hopefully achieving the best possible performance. The MARX-PC method in [27] is our previous CS reconstruction method, whose parameters have already been optimized on the set of test images for each sensing rate.

In the first experiment, we generate CS measurements by randomly sampling the Fourier transform coefficients of input images [30], [31]. The PSNR comparison results of recovered images by competing CS recovery methods are shown in Table I. From Table I, one can see that 1) ReTV [28] performs better than conventional TV on almost all test images; 2) MARX-PC outperforms both TV and ReTV for most test images and sensing rates; 3) BM3D-CS only beats MARX-PC method on high sensing rates. On the average, the proposed NLR-CS-baseline method outperforms all previous benchmark methods (e.g., NLR-CS-baseline can outperform BM3D-CS by up to 1.5 dB). By using the nonconvex logdet surrogate of the rank, the proposed NLR-CS method performs much better than the NLR-CS-baseline method on all test images and sensing rates. The average gain of NLR-CS over NLR-CS-baseline is up to 2.24 dB; the PSNR gains of NLR-CS over TV, ReTV, MARX-PC, and BM3D-CS can be as much as 11.45 dB, 11.30 dB, 5.49 dB, and 6.36 dB respectively on *Barbara* image - the most favorable situation for nonlocal regularization. In many cases, the proposed NLR-CS can produce higher PSNRs than other competing methods using 0.1N-0.2N less CS measurements. To facilitate the evaluation of subjective qualities, parts of reconstructed images are shown in Figs.4-5. Apparently, the proposed NLR-CS achieves the best visual quality among all competing methods - it can not only perfectly reconstruct large-scale sharp edges but also well recover small-scale fine structures.

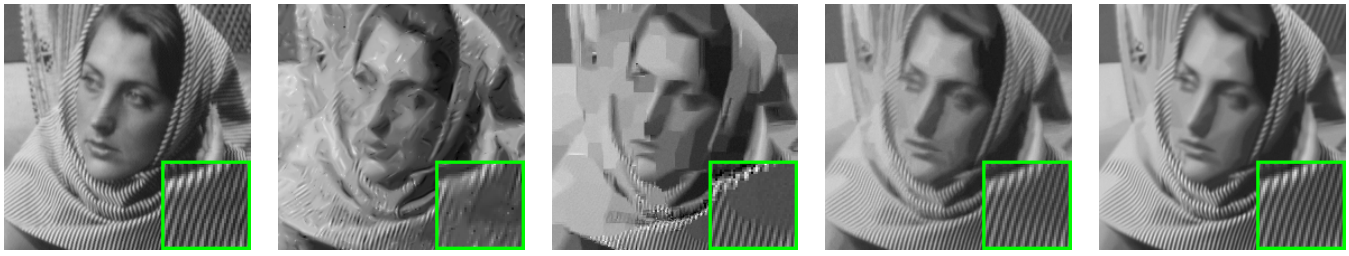
Next, we generate CS measurements by pseudo-radial subsampling of the Fourier transform coefficients of test images. An example of the radial subsampling mask is shown in Fig.8. Unlike random subsampling that generates incoherent aliasing artifacts, pseudo-radial subsampling produces streaking artifacts, which are more difficult to remove. The PSNR results of reconstructed images from the pseudo-radial CS measurements are included in Table II. One can see that the proposed NLR-CS method produces the highest PSNRs on all test images and CS measurement rates (except for three cases where NLR-CS slightly falls behind NLR-CS-baseline). The average PSNR improvements over BM3D-CS and NLR-CS-baseline are about 1.99 dB and 1.55 dB, respectively. Subjective justification about the superiority of the proposed NLR-CS method on the pseudo-radial subsampling case can be found in Figs.6-7.

Due to the potential applications of CS for MRI in reducing the scanning time, we have also conducted experiments on some complex-valued real-world MR images. Two sets of brain images with size of $256 \times 256$ acquired on a 1.5T Philips Achieva system are used in this experiment. The magnitude of these MR images are normalized to have the maximum value of 1. The $k-$space samples are simulated by applying 2D discrete Fourier transform to those MR images. The CS-MRI acquisition processes are simulated by sub-sampling the $k-$space data. Two sub-sampling strategies including random

(a) Barbara (b) boats (c) Girl (d) foreman

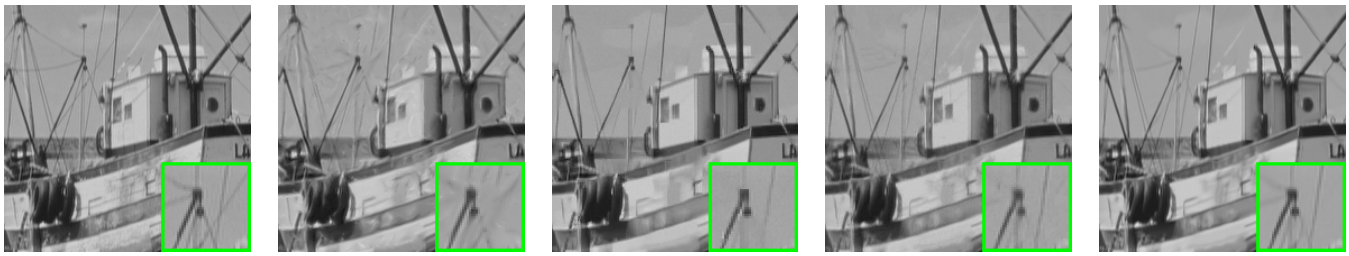
(e) house (f) lena256 (g) Monarch (h) Parrots

Fig. 3. Test photographic images used for compressive sensing experiments.



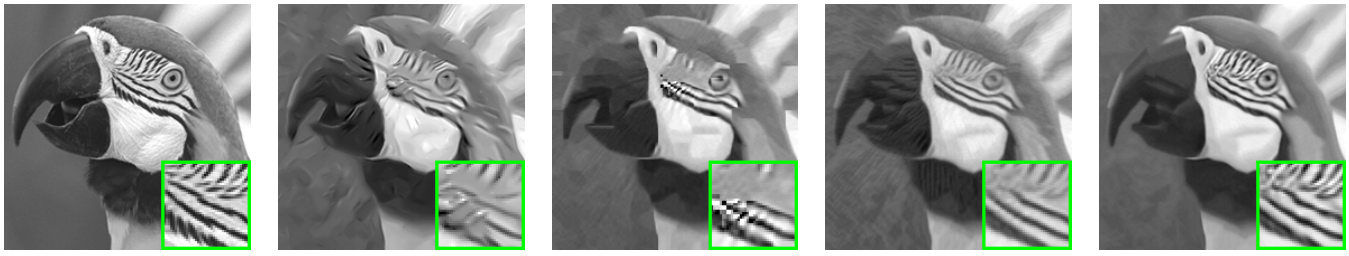
(a) (b) (c) (d) (e)

Fig. 4. CS recovered *Barbara* images with $0.05N$ measurements (random sampling). (a) Original image; (b) MARX-PC recovery [27] (24.11 dB); (c) BM3D-CS recovery [25] (24.34 dB); (d) Proposed **NLR-CS-baseline** recovery (27.59 dB) (e) Proposed **NLR-CS** recovery (**29.79** dB).



(a) (b) (c) (d) (e)

Fig. 5. CS recovered *Boats* images with $0.1N$ measurements (random sampling). (a) Original image; (b) MARX-PC recovery [27] (32.12 dB); (c) BM3D-CS recovery [25] (32.09 dB); (d) Proposed **NLR-CS-baseline** recovery (32.68 dB); (e) Proposed **NLR-CS** recovery (**35.33** dB).



(a) (b) (c) (d) (e)

Fig. 6. CS recovered *Parrots* images with pseudo radial subsampling (20 radial lines, i.e., $0.08N$ measurements). (a) Original image; (b) MARX-PC recovery [27] (24.80 dB); (c) BM3D-CS recovery [25] (25.04 dB); (d) Proposed **NLR-CS-baseline** recovery (25.79 dB); (e) Proposed **NLR-CS** recovery (**28.05** dB).

sub-sampling [23] and pseudo-radial sampling are used here. Two examples of test MR images and associated sub-sampling masks are shown in Fig. 8. We have compared the proposed method against the conventional CS-MRI method of [32]

TABLE I

THE PSNR (dB) RESULTS OF TEST CS RECOVERY METHODS WITH RANDOM SUBSAMPLING SCHEME [31].

| Image | Method | Number of measurements | | | | | |
|---|---|---|---|---|---|---|---|
| | | $M = 0.05N$ | $M = 0.1N$ | $M = 0.15N$ | $M = 0.2N$ | $M = 0.25N$ | $M = 0.3N$ |
| Barbara | TV | 22.79 | 24.78 | 26.72 | 28.87 | 31.23 | 33.80 |
| | ReTV | 22.58 | 24.74 | 26.87 | 29.33 | 31.82 | 34.50 |
| | MARX-PC | 24.11 | 30.24 | 33.76 | 36.00 | 37.90 | 39.62 |
| | BM3D-CS | 24.34 | 28.99 | 33.29 | 36.07 | 38.46 | 40.60 |
| | **NLR-CS-Baseline** | 27.59 | 32.66 | 35.65 | 37.93 | 39.83 | 41.80 |
| | **NLR-CS** | **29.79** | **35.47** | **38.17** | **40.00** | **41.52** | **43.02** |
| Boats | TV | 25.58 | 29.01 | 31.52 | 33.77 | 35.98 | 38.07 |
| | ReTV | 25.57 | 28.96 | 31.80 | 34.30 | 36.58 | 38.70 |
| | MARX-PC | 27.88 | 32.12 | 34.75 | 37.07 | 39.03 | 40.90 |
| | BM3D-CS | 27.31 | 32.09 | 35.28 | 37.70 | 39.81 | 41.67 |
| | **NLR-CS-Baseline** | 28.37 | 32.68 | 35.65 | 38.05 | 40.11 | 42.14 |
| | **NLR-CS** | **30.23** | **35.33** | **38.50** | **40.63** | **42.27** | **43.82** |
| Cameraman | TV | 25.09 | 28.63 | 31.48 | 34.20 | 36.84 | 39.65 |
| | ReTV | 25.49 | 29.29 | 32.01 | 34.88 | 37.84 | 40.65 |
| | MARX-PC | 26.38 | 29.32 | 31.82 | 33.94 | 36.56 | 39.08 |
| | BM3D-CS | 27.12 | 30.27 | 33.88 | 36.91 | 39.49 | 41.79 |
| | **NLR-CS-Baseline** | 27.13 | 30.73 | 33.48 | 36.17 | 38.64 | 41.15 |
| | **NLR-CS** | **28.36** | **32.30** | **36.10** | **39.16** | **41.59** | **43.79** |
| Foreman | TV | 32.50 | 36.02 | 38.34 | 40.49 | 42.53 | 44.61 |
| | ReTV | 32.63 | 36.29 | 38.67 | 40.99 | 42.90 | 45.01 |
| | MARX-PC | 35.26 | 38.29 | 40.26 | 42.00 | 43.66 | 45.41 |
| | BM3D-CS | 33.10 | 36.85 | 39.00 | 40.73 | 42.36 | 44.02 |
| | **NLR-CS-Baseline** | 34.84 | 38.38 | 40.76 | 42.67 | 44.51 | 46.49 |
| | **NLR-CS** | **35.70** | **39.39** | **41.98** | **44.02** | **46.12** | **48.46** |
| House | TV | 30.50 | 33.63 | 35.54 | 37.20 | 38.96 | 40.94 |
| | ReTV | 30.91 | 33.79 | 35.77 | 37.29 | 39.13 | 41.18 |
| | MARX-PC | 32.83 | 35.27 | 36.87 | 38.65 | 40.48 | 42.40 |
| | BM3D-CS | 32.56 | 35.96 | 38.10 | 39.76 | 41.42 | 42.99 |
| | **NLR-CS-Baseline** | 33.42 | 37.20 | 39.39 | 41.14 | 42.89 | 44.56 |
| | **NLR-CS** | **34.80** | **38.39** | **40.62** | **42.49** | **44.30** | **45.94** |
| Lena | TV | 26.48 | 29.63 | 32.32 | 34.74 | 37.33 | 39.79 |
| | ReTV | 26.55 | 29.95 | 32.78 | 35.17 | 37.82 | 40.27 |
| | MARX-PC | 29.20 | 33.17 | 36.00 | 38.16 | 40.41 | 42.26 |
| | BM3D-CS | 27.02 | 31.84 | 35.51 | 38.22 | 40.39 | 42.51 |
| | **NLR-CS-Baseline** | 28.75 | 33.18 | 36.41 | 38.88 | 41.07 | 43.25 |
| | **NLR-CS** | **30.69** | **35.75** | **38.95** | **41.38** | **43.43** | **45.36** |
| Monarch | TV | 24.21 | 28.78 | 31.93 | 34.84 | 37.37 | 39.63 |
| | ReTV | 24.39 | 29.17 | 32.52 | 35.38 | 37.91 | 40.17 |
| | MARX-PC | 27.01 | 31.17 | 34.03 | 36.53 | 39.00 | 41.20 |
| | BM3D-CS | 24.73 | 29.79 | 33.99 | 37.00 | 39.46 | 41.84 |
| | **NLR-CS-Baseline** | 27.15 | 31.56 | 34.62 | 37.24 | 39.58 | 41.84 |
| | **NLR-CS** | **28.85** | **34.30** | **37.78** | **40.31** | **42.38** | **44.16** |
| Parrots | TV | 27.65 | 31.84 | 34.76 | 37.00 | 39.45 | 41.65 |
| | ReTV | 28.44 | 32.53 | 35.44 | 37.71 | 39.87 | 42.02 |
| | MARX-PC | 27.93 | 34.22 | 36.51 | 38.49 | 40.35 | 42.27 |
| | BM3D-CS | 29.13 | 33.63 | 36.62 | 38.79 | 40.83 | 42.68 |
| | **NLR-CS-Baseline** | 30.34 | 34.92 | 37.78 | 39.83 | 41.73 | 43.56 |
| | **NLR-CS** | **32.18** | **36.56** | **39.56** | **41.44** | **43.19** | **44.72** |
| Average | TV | 26.85 | 30.29 | 32.83 | 35.14 | 37.46 | 39.77 |
| | ReTV | 27.07 | 30.59 | 33.23 | 35.63 | 37.98 | 40.31 |
| | MARX-PC | 28.83 | 32.98 | 35.50 | 37.61 | 39.67 | 41.64 |
| | BM3D-CS | 28.16 | 32.43 | 35.71 | 38.15 | 40.28 | 42.26 |
| | **NLR-CS-Baseline** | 29.70 | 33.91 | 36.72 | 38.99 | 41.05 | 43.10 |
| | **NLR-CS** | **31.33** | **35.93** | **38.96** | **41.18** | **43.10** | **44.91** |

(denoted as SparseMRI), the TV method, the reweighted TV (denoted as ReTV) method [28], and the baseline zero-filling reconstruction method. The source code of [32] was downloaded from the authors' website. The TV and ReTV method are implemented using the iterative reweighted least square (IRLS) approach. For a fair comparison, we have tried our best to tune the parameters of the SparseMRI method to achieve the highest PSNR results.

In Figs.9-10, we compare the reconstructed MR magnitude images by the test CS-MRI recovery methods for variable density 2D random sampling. In Figs.9 and 10, the sampling rate is $0.2N$, i.e., 5-fold undersampling of the $k-$space data. We can see that the SparseMRI method cannot preserve the sharp edges and fine image details. The ReTV method outperforms the SparseMRI method in terms of PSNR; however, visual artifacts can still be clearly observed. By contrast, the proposed NLR-CS-baseline preserves the edges and local structures better than ReTV method; and the proposed NLR-CS can further dramatically outperforms the NLR-CS-baseline. Figs.11-12 show the results for the pseudo-radial sampling scheme. The sub-sampling rates are $0.16N$ and $0.13N$ (i.e., 6.0 fold and 7.62 fold undersampling), respectively. We can see that

TABLE II

THE PSNR (dB) RESULTS OF DIFFERENT CS RECOVERY METHODS WITH PSEUDO RADIAL SUBSAMPLING SCHEME.

| Image | Method | Number of measurements | | | | | |
|---|---|---|---|---|---|---|---|
| | | $M = 0.08N$ (20 lines) | $M = 0.13N$ (35 lines) | $M = 0.14N$ (50 lines) | $M = 0.24N$ (65 lines) | $M = 0.29N$ (80 lines) | $M = 0.34N$ (95 lines) |
| Barbara | TV | 21.65 | 23.19 | 23.93 | 24.38 | 25.43 | 26.52 |
| | ReTV | 21.42 | 22.98 | 23.71 | 24.06 | 25.15 | 26.42 |
| | MARX-PC | 21.17 | 22.99 | 24.25 | 25.41 | 29.14 | 32.35 |
| | BM3D-CS | 21.85 | 24.38 | 28.45 | 31.29 | 33.67 | 34.83 |
| | **NLR-CS-Baseline** | **23.36** | 26.99 | 30.89 | 33.12 | 35.19 | 36.60 |
| | **NLR-CS** | 23.16 | **28.07** | **33.08** | **35.48** | **37.28** | **38.75** |
| Boats | TV | 22.53 | 25.48 | 28.00 | 29.97 | 31.87 | 33.38 |
| | ReTV | 22.52 | 25.60 | 28.13 | 30.14 | 31.94 | 33.44 |
| | MARX-PC | 22.63 | 26.63 | 29.63 | 31.84 | 33.85 | 35.29 |
| | BM3D-CS | 23.21 | 27.45 | 30.72 | 32.98 | 34.83 | 35.68 |
| | **NLR-CS-Baseline** | 23.85 | 27.62 | 30.46 | 32.57 | 34.55 | 35.99 |
| | **NLR-CS** | **24.44** | **29.00** | **31.83** | **33.98** | **35.76** | **37.57** |
| Cameraman | TV | 22.22 | 25.28 | 27.87 | 28.86 | 31.33 | 32.47 |
| | ReTV | 22.41 | 25.62 | 28.40 | 29.41 | 31.98 | 33.19 |
| | MARX-PC | 22.39 | 26.18 | 28.14 | 29.78 | 31.41 | 32.66 |
| | BM3D-CS | 22.32 | 26.50 | 29.62 | 31.83 | 33.28 | 34.23 |
| | **NLR-CS-Baseline** | 22.60 | 26.36 | 29.37 | 30.88 | 33.12 | 34.27 |
| | **NLR-CS** | **24.06** | **28.22** | **30.59** | **32.68** | **35.00** | **36.39** |
| Foreman | TV | 28.74 | 32.58 | 35.49 | 37.18 | 38.57 | 39.97 |
| | ReTV | 29.19 | 33.02 | 35.74 | 37.56 | 38.90 | 40.16 |
| | MARX-PC | 29.50 | 34.02 | 37.32 | 38.46 | 39.98 | 41.27 |
| | BM3D-CS | 29.80 | 34.52 | 37.42 | 38.91 | 40.13 | 40.74 |
| | **NLR-CS-Baseline** | 30.53 | 34.57 | 37.67 | 39.29 | 40.72 | 42.02 |
| | **NLR-CS** | **32.11** | **35.80** | **38.78** | **40.30** | **41.70** | **43.02** |
| House | TV | 28.11 | 30.61 | 33.30 | 34.81 | 35.44 | 36.83 |
| | ReTV | 28.69 | 31.04 | 33.51 | 34.91 | 35.49 | 36.76 |
| | MARX-PC | 29.27 | 31.99 | 34.89 | 36.00 | 36.76 | 37.83 |
| | BM3D-CS | 30.78 | 32.65 | 35.91 | 36.65 | 37.26 | 38.02 |
| | **NLR-CS-Baseline** | 29.25 | 32.73 | 35.69 | 37.53 | **37.52** | **39.20** |
| | **NLR-CS** | **32.09** | **34.06** | **36.24** | **37.78** | 37.49 | 39.14 |
| Lena | TV | 23.22 | 26.19 | 28.93 | 30.68 | 32.39 | 33.78 |
| | ReTV | 23.37 | 26.44 | 29.15 | 30.91 | 32.61 | 33.91 |
| | MARX-PC | 23.86 | 27.41 | 31.18 | 32.96 | 35.25 | 36.52 |
| | BM3D-CS | 23.55 | 27.50 | 30.87 | 33.17 | 35.24 | 36.47 |
| | **NLR-CS-Baseline** | 24.34 | 28.01 | 31.17 | 33.58 | 35.50 | 36.88 |
| | **NLR-CS** | **25.52** | **29.65** | **33.08** | **35.67** | **37.84** | **38.95** |
| Monarch | TV | 18.95 | 24.32 | 28.03 | 30.68 | 32.89 | 34.35 |
| | ReTV | 19.19 | 24.99 | 28.67 | 31.15 | 33.24 | 34.67 |
| | MARX-PC | 19.62 | 25.94 | 29.88 | 31.92 | 33.93 | 35.30 |
| | BM3D-CS | 19.75 | 26.24 | 30.34 | 32.81 | 35.02 | 35.97 |
| | **NLR-CS-Baseline** | 20.96 | 27.00 | 30.47 | 32.89 | 35.06 | 36.41 |
| | **NLR-CS** | **21.78** | **29.12** | **33.22** | **35.71** | **37.79** | **39.08** |
| Parrots | TV | 24.13 | 28.48 | 31.58 | 33.66 | 35.30 | 36.64 |
| | ReTV | 24.20 | 28.93 | 32.00 | 33.92 | 35.56 | 36.82 |
| | MARX-PC | 24.80 | 30.90 | 33.32 | 35.10 | 36.35 | 37.56 |
| | BM3D-CS | 25.04 | 30.43 | 34.17 | 36.02 | 37.39 | 38.35 |
| | **NLR-CS-Baseline** | 25.79 | 30.49 | 33.90 | 36.13 | 37.64 | 39.03 |
| | **NLR-CS** | **28.05** | **32.59** | **35.67** | **37.51** | **38.93** | **40.09** |
| Average | TV | 23.69 | 27.02 | 29.64 | 31.28 | 32.90 | 34.24 |
| | ReTV | 23.87 | 27.33 | 29.91 | 31.51 | 33.11 | 34.42 |
| | MARX-PC | 24.16 | 28.26 | 30.96 | 32.68 | 34.58 | 36.10 |
| | BM3D-CS | 24.54 | 28.71 | 32.19 | 34.21 | 35.85 | 36.79 |
| | **NLR-CS-Baseline** | 25.08 | 29.22 | 32.45 | 34.50 | 36.16 | 37.55 |
| | **NLR-CS** | **26.40** | **30.81** | **34.06** | **36.14** | **37.72** | **39.12** |

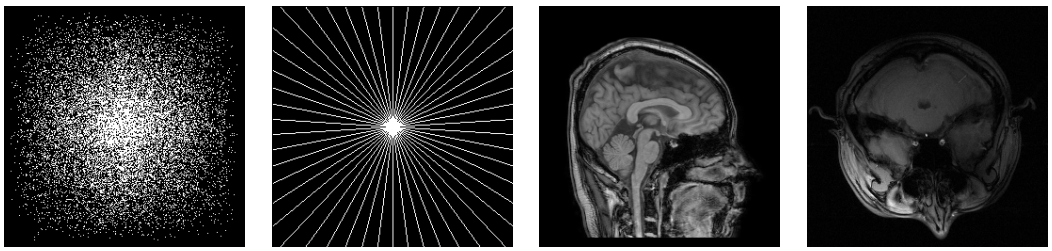
the proposed NLR-CS-baseline and NLR-CS methods both significantly outperform the SparseMRI method. In Fig.13 we plot the PSNR curves as a function of sensing rates for both random sampling and pseudo-radial sampling schemes. It can be seen from Fig.13 that the PSNR results of the proposed NLR-CS are much higher than all other competing methods at very low sensing rates, which implies that the proposed NLR-CS can better remove the artifacts and preserve important image structures more effectively even when the undersampling factor is high (i.e., large speed-up factor).

### *B. Experiments on noisy data*

In this subsection, we conduct similar experiments with noisy CS measurements to demonstrate the robustness of the proposed NLR-CS to noise. A significant amount of complex-valued additive Gaussian noise was added to the CS measurements. For natural images the subsampling ratios of the Fourier coefficients are fixed with 0.2 (randomly subsampling) and 0.24 (65 radial lines). The standard derivations of additive noise vary to generate the signal-to-noise ratio (SNR) between 11.5 dB and 35.0 dB. In this experiment, the MARX method

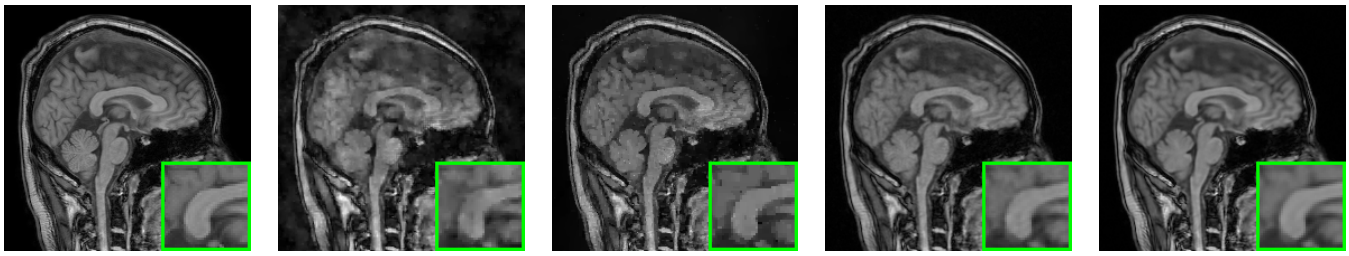
(a) (b) (c) (d) (e)

Fig. 7. CS recovered *Barbara* images with pseudo radial subsampling (35 radial lines, i.e., $0.13N$ measurements). (a) Original image; (b) MARX-PC recovery [27] (22.99 dB); (c) BM3D-CS recovery [25] (24.38 dB); (d) Proposed **NLR-CS-baseline** recovery (26.99 dB); (e) Proposed **NLR-CS** recovery (**28.07** dB).



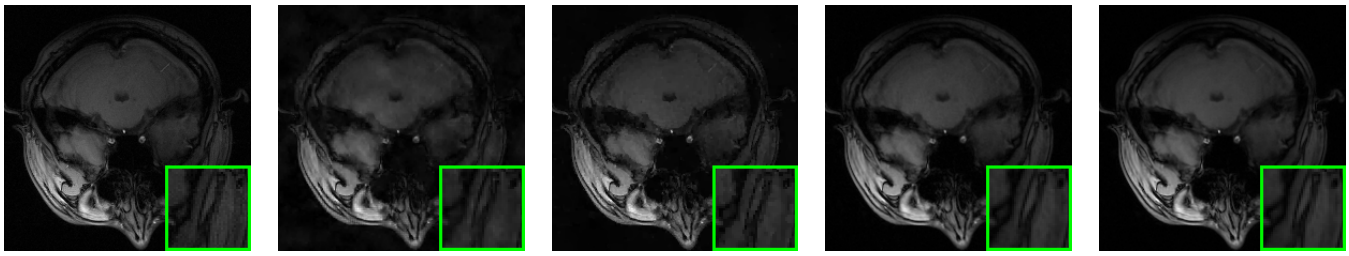
(a) (b) (c) (d)

Fig. 8. Sub-sampling masks and test MR images. (a) random sub-sampling mask; (b) pseudo-radial sub-sampling mask; (c) *Head* MR image; (d) *Brain* MR image.



(a) (b) (c) (d) (e)

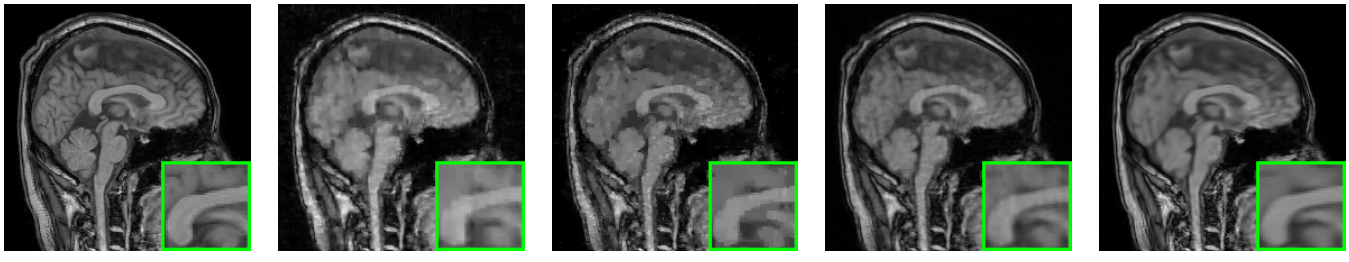
Fig. 9. Reconstructed *Head* MR images using $0.2N$ $k-$space samples (5 fold under-sampling, random sampling). (a) original MR image (magnitude); (b) SparseMRI method [32] (22.45 dB); (c) ReTV method [28] (27.31 dB) (d) proposed **NLR-CS-baseline** method (30.84 dB); (e) proposed **NLR-CS** method (**33.31** dB).



(a) (b) (c) (d) (e)

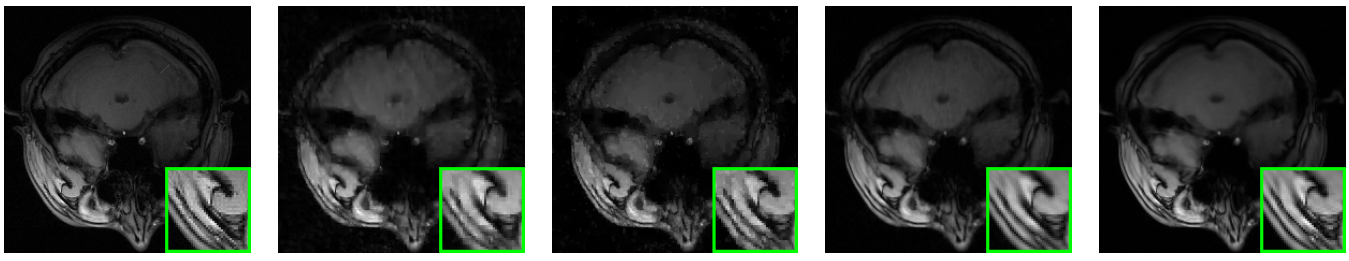
Fig. 10. Reconstructed *Brain* MR images using $0.2N$ $k-$space samples (5 fold under-sampling, random sampling). (a) original MR image (magnitude); (b) SparseMRI method [32] (30.20 dB); (c) ReTV method [28] (33.49 dB) (d) proposed **NLR-CS-baseline** (35.39 dB); (e) proposed **NLR-CS** method (**36.34** dB).

is not included since it is sensitive to noise. The PSNR comparison of the reconstructed images are shown in Fig. 14. One can observe that 1) TV and ReTV outperform BM3D-CS in the cases of low SNRs while it goes the other way round when $SNR > 21dB$; 2) the proposed NLR-CS outperforms other competing methods in all situations. Cropped portions of the reconstructed *Monarch* image from $0.2N$ noisy CS measurements (SNR=17.5 dB) are shown in Fig. 15. For complex-valued MR images, we fix the subsampling ratio of the k-space data to be 0.4 (randomsubsampling) and 0.24 (65 radial lines). The SNRs of the noisy CS measurements are still between 11.9 dB and 39.5 dB. The PSNR and subjective



(a) (b) (c) (d) (e)

Fig. 11. Reconstruction of *Head* MR images from 45 radial lines (pseudo-radial sampling, 6.0 fold under-sampling). (a) original MR image (magnitude); (b) SparseMRI method [32] (24.02 dB); (c) ReTV method [28] (27.09 dB) (d) proposed **NLR-CS-baseline** (27.97 dB) (e) proposed **NLR-CS** method (**29.67** dB).



(a) (b) (c) (d) (e)

Fig. 12. CS-MRI reconstruction of *Brain* MR images from 35 radial lines (pseudo-radial sampling, 7.62 fold under-sampling). (a) original MR image; (b) SparseMRI method [32] (29.03 dB); (c) ReTV method [28] (30.48 dB) (d) proposed **NLR-CS-baseline** (32.05 dB) (e) proposed **NLR-CS** method (**33.46** dB).

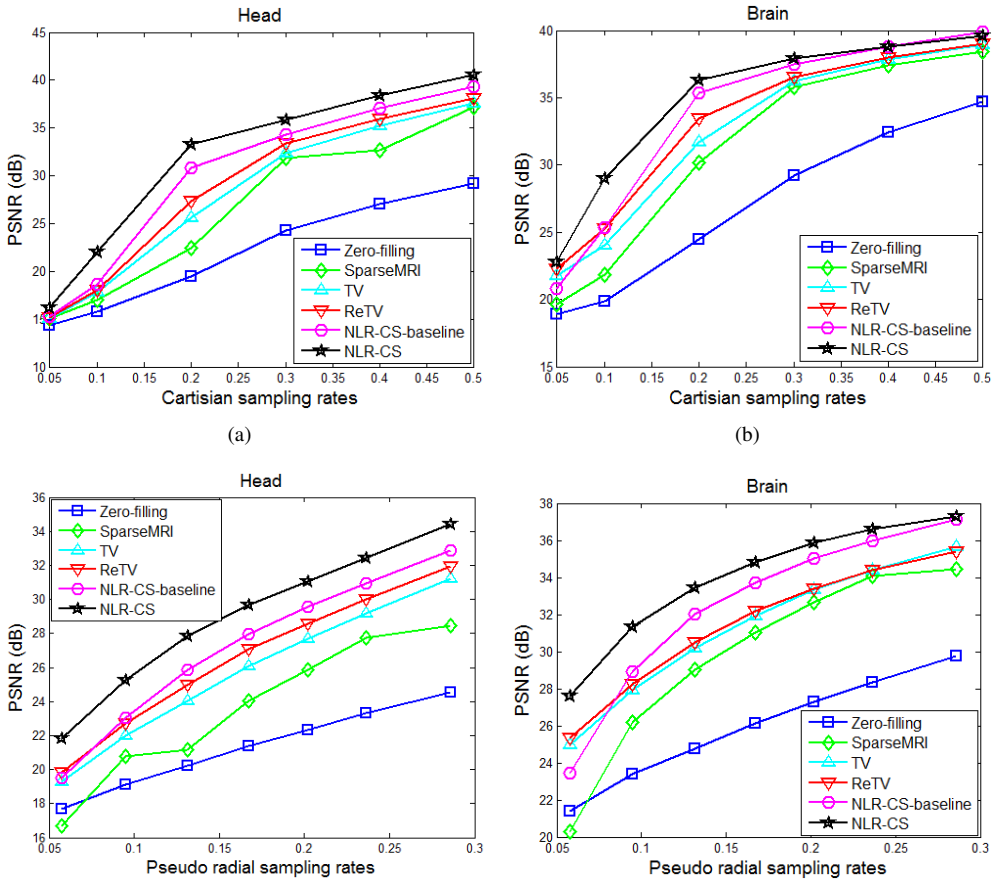



Fig. 13. PSNR results as sampling rates varies. (a)-(b) Cartesian random sampling; (c)-(d) pseudo-radial sampling.

quality comparison results are shown in Fig. 16 and Fig. 17 respectively, which clearly shows the proposed NLR-CS is the best among all competing methods.

## VI. Concluding Remarks

In this paper, we have presented a new approach toward CS based on nonlocal low-rank regularization and alternative direction multiplier method. Nonlocal low-rank regularization enables us to exploit both the group sparsity of similar patches and the nonconvexity of rank minimization; alternative direction multiplier method offers a principled and computationally efficient solution to image reconstruction from recovered patches. When compared against existing CS techniques in the open literature, our NLR-CS is mostly favored in terms of both subjective and objective qualities. Moreover, it shows significant performance improvements over a wide range of images including photographic and real-world complex-valued MR ones. One of the open issues remaining to be addressed is the modeling of data term (or likelihood function) in the CS-MRI. How the proposed nonlocal low-rank regularization works with the real-world (not simulated) k-space data seems a natural next step in our effort of pushing CS from theory to practice. Meantime, how nonlocally-regularized image reconstruction algorithm jointly works with parallel MRI - such as SENSE and SMASH - deserves further study.

## Acknowledgement

The authors would like to thank Prof. Jianfeng Cai of University of Iowa for his kind help with the proof of Theorem 1. They also thank two anonymous reviewers for their valuable comments and constructive suggestions that have significantly improved the presentation of this paper.

## Appendix

In order to prove Theorem 1, we first show the following theorem for the proximal operator of weighted nuclear norm for real-valued matrices. Then, we extend the result to prove Theorem 1 for complex-valued matrices. It has been shown in [37], [38] that if the weights $\boldsymbol{w}$ are in a non-descending order, the weighted nuclear norm is concave and a fixed-point solution can be obtained by weighted singular value thresholding operator [37], [38]. The following theorem is an extension of Theorem 2.1 of [24].

*Theorem 2 (Proximal Operator for the Real Case):* For $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ and $0 \leq w_1 \leq \cdots \leq w_{n_0}$, where $n_0 = \min\{m, n\}$, one minimizer to

$$\min_{\boldsymbol{L}} \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{L}\|_F^2 + \tau\varphi(\boldsymbol{L}, \boldsymbol{w}) \tag{28}$$

is given by the weighted singular value thresholding operator $\mathcal{S}_{\boldsymbol{w},\tau}(\boldsymbol{X})$:

$$\mathcal{S}_{\boldsymbol{w},\tau}(\boldsymbol{X}) := \boldsymbol{U}(\boldsymbol{\Sigma} - \tau\text{diag}(\boldsymbol{w}))_+\boldsymbol{V}^\top, \tag{29}$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the SVD of $\boldsymbol{X}$ and $(x)_+ = \max\{x, 0\}$.

*Proof:* For fixed weights $\boldsymbol{w}$, $h(\boldsymbol{L}) := \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{L}\|_F^2 + \tau\varphi(\boldsymbol{L}, \boldsymbol{w})$, $\boldsymbol{L}^*$ minimizes $h$ if and only if it satisfies the following optimal condition, i.e.,

$$\boldsymbol{0} \in \boldsymbol{X} - \boldsymbol{L}^* + \tau\partial\varphi(\boldsymbol{L}^*, \boldsymbol{w}) \tag{30}$$

where $\partial\varphi(\boldsymbol{L}^*, \boldsymbol{w})$ is the set of subgradients of the weighted nuclear norm. Let matrix $\boldsymbol{L} \in \mathbb{R}^{n\times m}$ and its SVD be $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$. It is known from [36] that the subgradient of $\varphi(\boldsymbol{L}, \boldsymbol{w})$ can be derived as

$$\begin{aligned}\partial\varphi(\boldsymbol{L}, \boldsymbol{w}) =&\{\boldsymbol{U}\boldsymbol{W}_r\boldsymbol{V}^\top + \boldsymbol{Z} : \boldsymbol{Z} \in \mathbb{R}^{n\times m}, \boldsymbol{U}^\top\boldsymbol{Z} = 0,\\ &\boldsymbol{Z}\boldsymbol{V} = 0, \sigma_j(\boldsymbol{Z}) \leq w_{r+j}, j = 1, \ldots, n_0 - r\},\end{aligned} \tag{31}$$

where $r$ is the rank of $\boldsymbol{L}$ and $\boldsymbol{W}_r$ is the diagonal matrix composed by the first $r$ rows and $r$ columns of the diagonal matrix $\text{diag}(\boldsymbol{w})$. To show that $\boldsymbol{L}^*$ satisfies Eq. (31), we rewritten the SVD of $\boldsymbol{X}$ as $\boldsymbol{X} = \boldsymbol{U}_0\boldsymbol{\Sigma}_0\boldsymbol{V}^\top + \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^\top$, where $\boldsymbol{U}_0$, $\boldsymbol{V}_0$ (resp. $\boldsymbol{U}_1$, $\boldsymbol{V}_1$) are the singular vectors associated with the singular values greater than $\tau w_j$ (resp. smaller than or equal to $\tau w_j$). Let $\boldsymbol{L}^* = \mathcal{S}_{\boldsymbol{w},\tau}(\boldsymbol{X})$. Then, we have

$$\boldsymbol{L}^* = \boldsymbol{U}_0(\boldsymbol{\Sigma}_0 - \tau\boldsymbol{W}_r)\boldsymbol{V}_0^\top. \tag{32}$$

Therefore,

$$\begin{aligned}\boldsymbol{X} - \boldsymbol{L}^* &= \boldsymbol{U}_0(\tau\boldsymbol{W}_r)\boldsymbol{V}_0^\top + \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}^\top\\ &= \tau(\boldsymbol{U}_0\boldsymbol{W}_r\boldsymbol{V}_0^\top + \boldsymbol{U}_1(\tau^{-1}\boldsymbol{\Sigma}_1)\boldsymbol{V}_1^\top)\\ &= \tau(\boldsymbol{U}_0\boldsymbol{W}_r\boldsymbol{V}_0^\top + \boldsymbol{Z}),\end{aligned} \tag{33}$$

where $\boldsymbol{Z} = \boldsymbol{U}_1(\tau^{-1}\boldsymbol{\Sigma}_1)\boldsymbol{V}_1^\top$. By definition, $\boldsymbol{U}_0^\top\boldsymbol{Z} = \boldsymbol{0}$, $\boldsymbol{Z}\boldsymbol{V}_0 = \boldsymbol{0}$. Since the diagonal elements of $\boldsymbol{\Sigma}_1$ are smaller than $\tau w_{j+r}$, it is easy to verify that $\sigma_j(\boldsymbol{Z}) \leq w_{r+j}$, $j = 1, 2, \ldots, n_0 - r$. Therefore, $\boldsymbol{X} - \boldsymbol{L}^* \in \tau\partial\varphi(\boldsymbol{L}^*, \boldsymbol{w})$, which concludes the proof. ■

*Proof of Theorem 1:* Let $\boldsymbol{X} = (\boldsymbol{X}_1 + i\boldsymbol{X}_2) \in \mathbb{C}^{n\times m}$ be an arbitrary complex-valued matrix and its SVD be $\boldsymbol{X} = (\boldsymbol{U} + i\boldsymbol{P})\boldsymbol{\Sigma}(\boldsymbol{V} + i\boldsymbol{Q})^\top$, where $\boldsymbol{X}_1 \in \mathbb{R}^{n\times m}$, $\boldsymbol{X}_2 \in \mathbb{R}^{n\times m}$, $\boldsymbol{U} \in \mathbb{R}^{n\times n}$, $\boldsymbol{P} \in \mathbb{R}^{n\times n}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n_0\times n_0}$, $\boldsymbol{V} \in \mathbb{R}^{m\times m}$, $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$, and $n_0 = \min\{n, m\}$. Then, it is easy to verify that the SVD of the matrix

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{X}_1 & \boldsymbol{X}_2\\ -\boldsymbol{X}_2 & \boldsymbol{X}_1\end{bmatrix}, \quad \boldsymbol{A} \in \mathbb{R}^{2n\times 2m} \tag{34}$$

can be expressed as

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{U} & \boldsymbol{P}\\ -\boldsymbol{P} & \boldsymbol{U}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma} & \\ & \boldsymbol{\Sigma}\end{bmatrix}\begin{bmatrix}\boldsymbol{V} & -\boldsymbol{Q}\\ \boldsymbol{Q} & \boldsymbol{V}\end{bmatrix}^\top. \tag{35}$$

Let $\boldsymbol{L} = (\boldsymbol{Y}_1 + i\boldsymbol{Y}_2) \in \mathbb{C}^{n\times m}$ and the functions $h_1$ and $h_2$ be

$$\begin{aligned}h_1(\boldsymbol{L}) &= \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{L}\|_F^2 + \tau\varphi(\boldsymbol{L}, \boldsymbol{w}),\\ h_2(\boldsymbol{B}) &= \frac{1}{4}\|\boldsymbol{A}-\boldsymbol{B}\|_F^2 + \frac{\tau}{2}\varphi(\boldsymbol{A}, \tilde{\boldsymbol{w}}).\end{aligned} \tag{36}$$

where $\tilde{\boldsymbol{w}} = [\boldsymbol{w}^\top, \boldsymbol{w}^\top]^\top \in \mathbb{R}_+^{2n_0}$ and the matrix $\boldsymbol{B} \in \mathbb{R}^{2n\times 2m}$ is defined as

$$\boldsymbol{B} = \begin{bmatrix}\boldsymbol{Y}_1 & \boldsymbol{Y}_2\\ -\boldsymbol{Y}_2 & \boldsymbol{Y}_1\end{bmatrix}. \tag{37}$$

Using the equality expressed in Eq.(35), it is can be verified that $h_1(\boldsymbol{L}) = h_2(\boldsymbol{B})$. According to Theorem 2, one minimizer to the following minimization problem

$$\min_{\boldsymbol{B}} \frac{1}{4}\|\boldsymbol{A}-\boldsymbol{B}\|_F^2 + \frac{\tau}{2}\varphi(\boldsymbol{A}, \tilde{\boldsymbol{w}}) \tag{38}$$

is given by $\boldsymbol{B}^* = \mathcal{S}_{\tilde{\boldsymbol{w}},\tau}(\boldsymbol{A})$. Since $h_1(\boldsymbol{L}) = h_2(\boldsymbol{B})$, the function $h_2(\boldsymbol{B}^*)$ has the same minimum value as $h_1(\boldsymbol{L}^*)$ with $\boldsymbol{L}^* = \mathcal{S}_{\boldsymbol{w},\tau}(\boldsymbol{X})$. ■

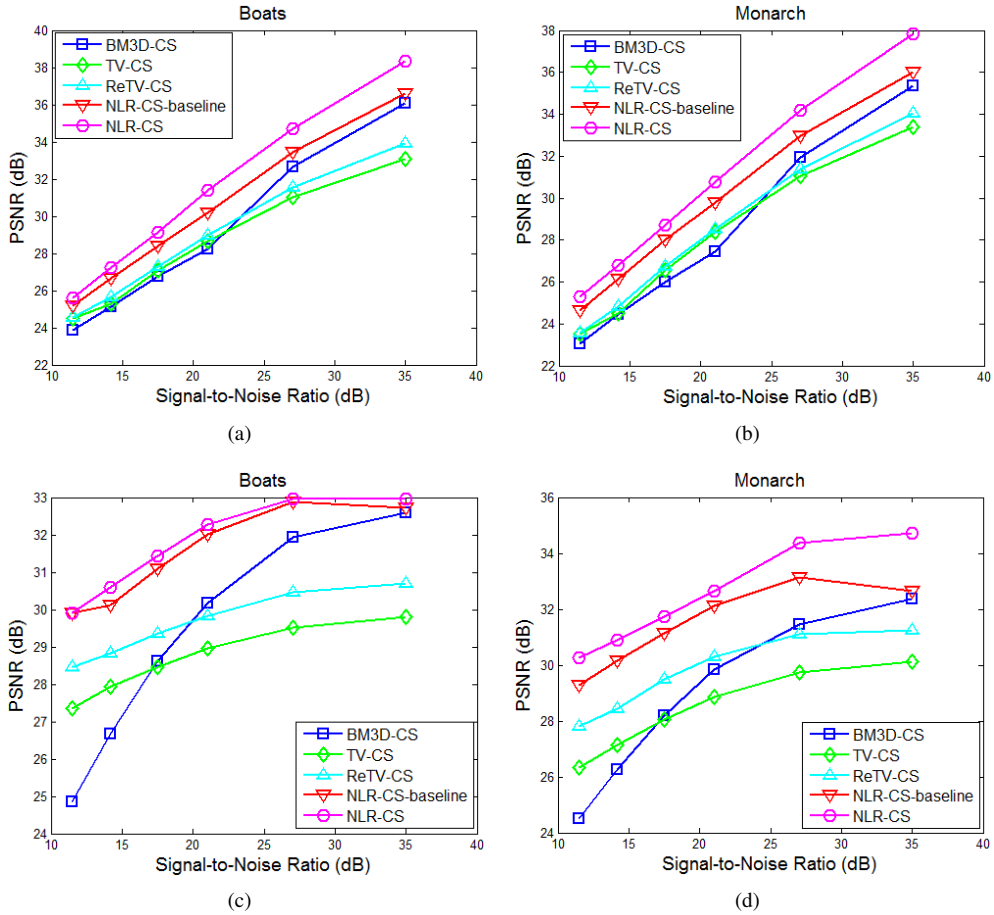
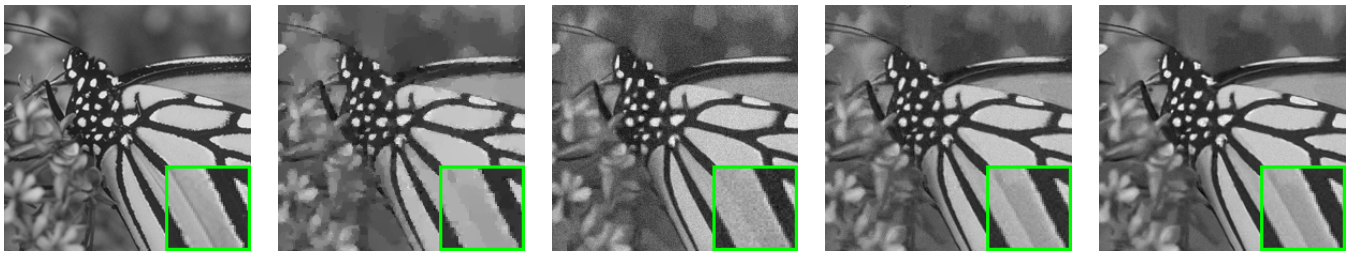



Fig. 14. PSNRs of the reconstructed images from noisy measurements. (a)-(b) Random subsampling ($0.2N$); (c)-(d) pseudo-radial sampling (65 lines).



(a) (b) (c) (d) (e)

Fig. 15. Reconstruction of *Monarch* images from $0.2N$ noisy CS measurements (SNR=17.5 dB). (a) original image; (b) ReTV recovery [28] (26.73 dB); (c) BM3D-CS recovery [25] (26.00 dB); (c) proposed **NLR-CS-baseline** recovery (27.99 dB); (d) proposed **NLR-CS** recovery (**28.70** dB).

## REFERENCES


[1] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information." *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. L. Donoho, "Compressed sensing" *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Sep. 2006.

[3] M. Duarte, M. Davenport, D. Takbar, J. Laska, T. Sun, K. Kelly, R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[4] M. E. Gehm, R. John, D. Brady, R. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14013–14027, Oct. 2007.

[5] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *IEEE International Conference on Computer Vision*, 2011, pp. 287–294.

[6] M. Lustig, D. .L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, March 2008.

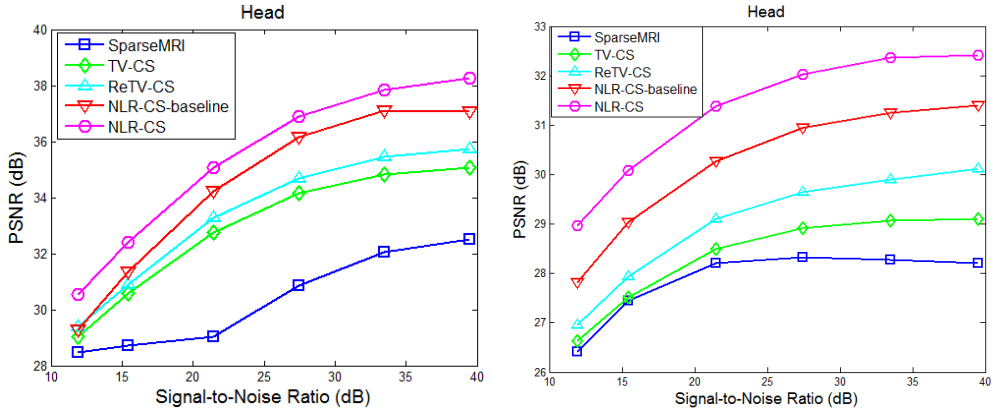
[7] I. Daubechies, M. Defriese, and C. DeMol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[8] M. Zibulevsky and M. Elad, "l1-l2 optimization in signal and image processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 76–88, May 2010.

[9] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 253–276, 2010.

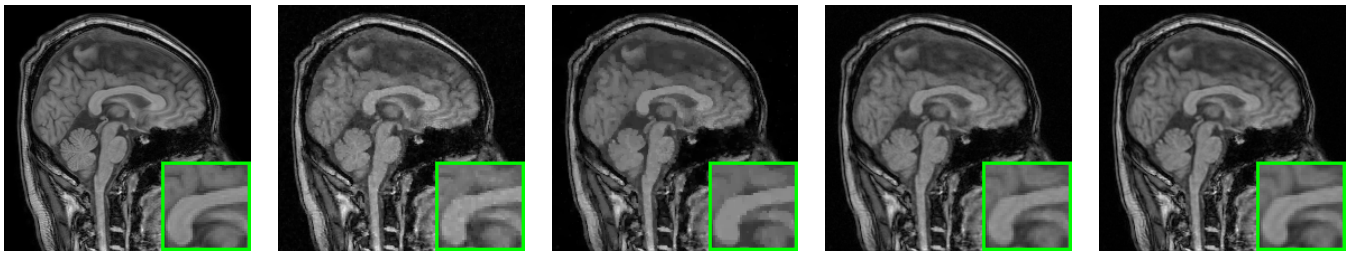
[10] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithm, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[11] W. Dong, G. Shi, X. Li, L. Zhang, and X. Wu, "Image reconstruction with locally adaptive sparsity and nonlocal robust regularization," *Signal Processing: Image Communication*, vol. 27, pp. 1109–1122, 2012.




(a) (b)

Fig. 16. PSNR of the reconstructed MR image *Head* from noisy measurements. (a) random subsampling (ratio 0.4); (b) pseudo-radial sampling (65 lines).



(a) (b) (c) (d) (e)

Fig. 17. Reconstruction of *Head* MR images from 65 noisy radial lines (SNR=15.42 dB). (a) original MR image (magnitude); (b) SparseMRI method [32] (28.73 dB); (c) ReTV method [28] (30.88 dB) (d) proposed **NLR-CS-baseline** (31.34 dB) (e) proposed **NLR-CS** method (**32.41** dB).


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

[13] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Dept. Electr. Comput. Eng., UIUC, Urbana, Tech. Rep. UILU-ENG-092215*, Oct. 2009.

[14] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *Statist. Sci.*, vol. 27, no. 4, pp. 450–468, 2012.

[15] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 457–464, 2011.

[16] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis," *Journal of the ACM*, vol. 58, no. 3, 2011.

[17] M. Fazel, H. Hindi, and S. Boyd, "Log-Det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *Proc. of Am. Control Conf.*, pp. 2156–2162, 2003.

[18] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[19] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 417–424.

[20] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 2272–2279.

[21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[22] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, oct 2007.

[23] J. Trzasko and A. Manduca, "Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic $l_{-}\{0\}$-Minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.

[24] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, p. 1956, 2010.

[25] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *IEEE International Conference on Image Processing*, vol. 1, San Antonio, TX, USA, Sep 2007.

[26] E. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[27] X. Wu, W. Dong, X. Zhang, and G. Shi, "Model-assisted adaptive recovery of compressed sensing with imaging applications," *IEEE Trans. on Image Processing*, vol. 21, no. 2, pp. 451–458, Feb. 2012.

[28] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted $l_1$ minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Treads in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[30] http://www.acm.caltech.edu./l1magic

[31] E. Candes and J. Romberg, "Practical signal recovery from random projections," in *Proc. SPIE Comput. Imaging*, vol. 5674, pp.76–86, 2005.

[32] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson, Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[33] Klaas P. Pruessmann et al., "SENSE: sensitivity encoding for fast MRI," *Magnetic resonance in medicine*, vol. 42, no.5, pp. 952–962, 1999.

[34] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.

[35] Daniel K. Sodickson and Warren J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays," *Magnetic Resonance in Medicine*, vol. 38, no. 4, pp. 591–603, 1997.

[36] A. S. Lewis, "The convex analysis of unitarily invariant matrix functions," *Journal of Convex Analysis*, vol. 2, no. 1/2, pp. 173-183, 1995.

[37] S. Gaiffas and G. Lecue, "Weighted algorithms for compressed sensing and matrix completion," *arXiv: 1107.1638v1*.

[38] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2014.