CrossMark

ORIGINAL ARTICLE

Compressed sensing image reconstruction via adaptive sparse nonlocal regularization

Zhiyuan Zha[1] · **Xin Liu**[2] · **Xinggan Zhang**[1] · **Yang Chen**[1] · **Lan Tang**[3] ·
Yechao Bai[1] · **Qiong Wang**[1] · **Zhenhong Shang**[4]

Abstract Compressed sensing (CS) has been successfully utilized by many computer vision applications. However, the task of signal reconstruction is still challenging, especially when we only have the CS measurements of an image (CS image reconstruction). Compared with the task of traditional image restoration (e.g., image denosing, debluring and inpainting, etc.), CS image reconstruction has partly structure or local features. It is difficult to build a dictionary for CS image reconstruction from itself. Few studies have shown promising reconstruction performance since most of the existing methods employed a fixed set of bases (e.g., wavelets, DCT, and gradient spaces) as the dictionary, which lack the adaptivity to fit image local structures. In this paper, we propose an adaptive sparse nonlocal regularization (ASNR) approach for CS image reconstruction. In ASNR, an effective self-adaptive learning dictionary is used to greatly reduce artifacts and the loss of fine details. The dictionary is compact and learned from the reconstructed image itself rather than natural image dataset. Furthermore, the image sparse nonlocal (or nonlocal self-similarity) priors are integrated into the regularization term, thus ASNR can effectively enhance the quality of the CS image reconstruction. To improve the computational efficiency of the ASNR, the split Bregman iteration based technique is also developed, which can exhibit better convergence performance than iterative shrinkage/thresholding method. Extensive experimental results demonstrate that the proposed ASNR method can effectively reconstruct fine structures and suppress visual artifacts, outperforming state-of-the-art performance in terms of both the PSNR and visual measurements.

Keywords Compressed sensing · Nonlocal self-similarity · Dictionary learning · Split Bregman iteration

✉ Zhiyuan Zha
zhazhiyuan.mmd@gmail.com

✉ Xinggan Zhang
zhxg@nju.edu.cn

1 School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China

2 The Center for Machine Vision and Signal Analysis, University of Oulu, Finland and School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, Shanxi, China

3 Department of Electronic Science and Engineering, Nanjing University, and National Mobile Commun. Research Lab., Southeast University, Nanjing 210023, China

4 School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

1 Introduction

The recent development of the compressed sensing (CS) theory [1,2] has drawn much attention in signal/image communities. The CS theory has uncovered that by achieving the possibility of reconstructing a signal from a smaller number of random measurements, as long as the signal has a sparse representation in some domain. The CS theory has the potential of dramatically improving the energy efficiency of sensors, leading to extensive applications in the real world. For instance, in recent years, several CS-based imaging systems have been established, including the compressed spectral imaging system [3], the single-pixel camera [4], and the high-speed video camera [5].

In the CS theory, if a signal is sparse in some transform domain, it is usually sampled by a random projection which is independent and reconstructed by minimizing the ℓ_0 norm or the ℓ_1 norm optimization problem with some prior information which usually constitutes regularization terms. However,

Springer

since the ℓ_0 norm minimization is a difficult combinatorial optimization problem, solving this problem directly is both NP-hard and unstable in the presence of noise. For this reason, it has been proposed to replace the non-convex ℓ_0 norm by its convex ℓ_1 norm, and provably it is probable that the ℓ_1 norm is equivalent to the ℓ_0 norm in many practical problems. Therefore, under these circumstances, one is allowed to solve the problem easily by the ℓ_1 norm, rather than the ℓ_0 norm. The optimization problem by the ℓ_1 norm can be solved by linear programming [6] and several CS restoration algorithms, known as the gradient projection sparse reconstruction [7], the orthogonal matching pursuit [8], and the iterative thresholding [9], which have been proposed recently.

As a basic image inverse problem in the field of image restoration, the purpose of CS image reconstruction is to reconstruct high quality image from considerably fewer measurements than suggested by the Nyquist sampling theory. In addressing this issue, the most effective method now is to utilize image prior knowledge or some constraints, which plays a key role in the CS image reconstruction problems to recover the missing information. Accordingly, effective regularization terms are designed to reflect the image priors, which is vital for the CS image reconstruction.

The classic regularization models, such as the quadratic Tikhonov regularization model [10], the Mumford-Shah (MS) model [11], and the TV regularization model [12], utilize the local structure patterns of an image and high effectiveness to preserve image edges and recover the smooth regions of an image. Nevertheless, they usually destroy the image details and cannot handle fine structures well, since they have only exploited the image local statistics, while ignoring the nonlocal statistics.

Besides the local structural schema, perhaps the most remarkable nonlocal statistics are the nonlocal self-similarity models demonstrated by natural images in the image processing, which was initially proposed as a solution to image denoising, called the non-local means (NLM) denoising filtering [13]. The basic idea of the nonlocal self-similarity method is very simple: the image patches that have similar patterns can be spatially far from each other and thus we can collect them from the whole image. The nonlocal self-similarity characterizes the repetitiveness of textures and structures reflected by natural images within nonlocal regions, which can be applied to retain the edges and the sharpness effectively to maintain image nonlocal consistency [13–16]. Later, inspired by the success of the NLM denoising filtering, a flurry of nonlocal regularization terms were proposed to solve the image inverse problems and to utilize the nonlocal self-similarity properties of natural images as exhibited in [14–19]. Note that the NLM-based regularization term in [17] is implemented at the pixel level, i.e., from one pixel to another pixel. The block-level NLM-based method is introduced in [14] to address the image deblurring

and super-resolution problems. The connection between the spectral bases of the nonlocal graph Laplacian operator and the filtering methods was proposed by Peyr in [15]. Jung et al. [18] proposed a nonlocal form of the approximation of the MS regularizer for color image restoration instead of the traditional local MS regularizer. Recently, Dong et al. [19] proposed a nonlocally centralized sparse representation model for image restoration, which first obtains a good estimation of the sparse coding coefficients of the original image, and then centralizes the sparse coding coefficients of the observed image to those estimations to improve the performance of sparse representation-based image restoration.

In recent years, the patch-based sparse representation along with dictionary learning (DL) from natural images, has led to very impressive reconstruction results [14,19–23]. The sparse representation model assumes that each patch of an image can be accurately represented by few elements from a basis set called a dictionary, which is usually learned from a natural image dataset [20]. Therefore, one important issue of the sparsity-based image restoration is the selection of the dictionary. The conventional analytically designed dictionaries, such as the DCT, wavelet, bandlet and curvelet, etc., share the superiority of fast conduct, yet they usually lack the adaptivity to fit the image local structures, resulting in poor reconstruction performance. Compared with these predefined sparse transforms, the DL method eliminates the disadvantage of a fixed sparse transform. For example, it is impossible for one predefined sparse transform to be universally optimal and sufficient to characterize the many complex structures of natural images. The purpose of many DL methods is to learn a universal and over-complete dictionary which can represent various image structures. However, it has been shown that sparse decomposition over an over-complete dictionary is potentially unstable, and it tends to generate visual artifacts [24,25]. Recently, the PCA sub-dictionary has been an illustrious DL method, which clusters the training patches extracted from a set of natural images into some clusters, and finally learns a compact PCA sub-dictionary for each cluster [14].

To reconstruct an image accurately from its compressed sensing measurements, Wu et al. [26] proposed a model-assisted adaptive recovery of compressed sensing for exploiting the 2D piecewise autoregressive model and, in turn, for making the CS recovery adaptive to spatially transform the second order statistics of an image. Ravishankar et al. [27] began with a rough estimate for the compressed sensing measurement, and they simultaneously updated the sparse dictionary and the sparse coefficients of all overlapped patches, and finally averaged all the reconstructed patches to estimate the image iteratively. Many recent image CS reconstruction works integrated some prior knowledge into the image CS reconstruction framework, such as the tree-structured wavelet [28], tree-structured DCT [29], and

Gaussian mixture models [30] have integrated some prior knowledge into the CS image reconstruction framework. Mun et al. [31] proposed a projection-driven CS reconstruction incorporated with block-based random image sampling, which aimed at encouraging sparsity in the domain of directional transforms.

Different from the task of traditional image restoration (e.g, image denoising, deblurring, inpainting, etc.), which can directly use the image nonlocal self-similarity (NSS) property and due to the encouragement by the recent excellent sparse representation methods in the image processing. In this paper, we propose a unified variational framework for adaptive sparse nonlocal regularization (ASNR) model of CS image reconstruction. Specifically, an effective self-adaptive dictionary learning as sparse domain is used to greatly reduce visual artifacts and the loss of fine details. The dictionary is compact and learned from the reconstructed image itself rather than natural image dataset. In addition, the image sparse nonlocal (or nonlocal self-similarity) priors are integrated into the regularization term to enhance the quality of CS image reconstruction. To improve the computational efficiency of the proposed method, we have developed a convenient implementation utilizing the split Bregman iteration technique, which can exhibit good convergence performance. Extensive experimental results validate that the proposed method can effectively reconstruct fine structures and suppress visual artifacts, outperforming many current state-of-the-art schemes in terms of both the PSNR and visual measurements.

The rest of this paper is organized as follows: Sect. 2 briefly states the prior work in the CS theory and the patch-based sparse representation. Section 3 presents the modeling of the ASNR and introduces the implementation details. Section 4 presents the extensive experimental results, and Sect. 5 concludes this paper.

2 Background and related work

2.1 Compressed sensing

The CS theory demonstrates that the signal can be recovered by linear projection from its measurements, which may even be far below the traditional Nyquist sampling rate, as long as the signal has a sparse representation in some domain. In the CS theory, a signal x of size N is viewed as sparse in domain Ψ, if its transform θ coefficients over Ψ is expressed by $x = \Psi\theta$, where θ are mostly zeros or nearly sparse if the vast majority of coefficients θ are either zeros or very close to zero. The number of significant elements within the coefficient vector θ is usually quantifying the sparsity of x in Ψ.

More specifically, one can seek the perfect reconstruction of a signal x from $y \in \Re^{M}$ with M randomized linear measurements, which is formulated as the following constrained optimization problem, namely,

$$\arg \min_{\theta} ||\theta||_0, \quad s.t. \quad y = \Phi\Psi\theta \tag{1}$$

where $\Phi \in \Re^{M \times N}$ represents the random projection matrix and such that M is much smaller than N. $|| \bullet ||_0$ represents the ℓ_0 norm, counting the number of nonzero entries of a vector θ.

However, since the $|| \bullet ||_0$ norm minimization is discontinuous and an NP-hard problem, it is often relaxed to the convex ℓ_1 norm minimization. By selecting an appropriate regularization parameter, Eq. (1) is equivalent to the following unconstrained optimization problem, i.e.,

$$\arg \min \frac{1}{2}||y - \Phi\Psi\theta||_2^2 + \lambda||\theta||_1 \tag{2}$$

According to [32], CS is capable of recovering K-sparse signal x with a high probability from $M = O(K\log(N/K))$ random measurements by solving this ℓ_1 norm optimization problem.

2.2 Patch-based sparse representation

Recently, the sparse representation based modeling has been proved to be a promising model for the image restoration studies [14,19–23], which assumes that a natural image is sparse in some domain spanned by a set of bases or the atoms of the dictionary. According to [21], the basic unit of sparse representation for images is patch. Mathematically, for an image $x \in \Re^N$, let $x_i = R_i x, i = 1, 2, \ldots, n$ denote an image patch of size $\sqrt{bc} \times \sqrt{bc}$ extracted at location i, where R_i is the matrix extracting patch x_i from x at location i. Given a dictionary $D \in \Re^{bc \times M}, bc \leq M$, the sparse coding processing of each patch x_i is to find a sparse vector α_i to satisfy $x_i = D\alpha_i$ over dictionary D, where α_i is a sparse vector whose coefficients are mostly zero or close to zero. Obviously, the image patches can be overlapped to better suppress noise and visual artifacts, and we can obtain a redundant patch-based representation. Then the entire image x can be reconstructed by averaging all the reconstructed patches x_i, which can be mathematically expressed as

$$x \approx D\alpha = \left(\sum_{i=1}^{n} R_i^T R_i\right)^{-1} \left(\sum_{i=1}^{n} R_i^T D\alpha_i\right) \tag{3}$$

where α denotes the concatenation of all α_i, that is, $\alpha = [\alpha_1^T, \alpha_2^T, \ldots, \alpha_n^T]^T$. Note that the above equation is nothing but signifying that the overall image is reconstructed by averaging each reconstructed patch of x_i.

After this, we merged Eq. (3) into Eq. (1). The CS recovery scheme for sparse coding by the ℓ_0 norm is formulated as

$$\arg \min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_0, \quad s.t., \quad \boldsymbol{y} = \boldsymbol{\Phi D\alpha} \tag{4}$$

where $\boldsymbol{D}$ replaces Ψ in Eq. (1), standing for a learning dictionary, and $\boldsymbol{\alpha}$ represents the patch-based sparse coding coefficients for the entire image over dictionary $\boldsymbol{D}$.

However, since the ℓ_0 norm minimization is discontinuous and NP-hard, solving Eq. (4) is a difficult combinatorial optimization problem. On this account, a straightforward method has been proposed for solving the non-convex ℓ_0 norm by its convex ℓ_1 counterpart, and Eq. (4) can be turned into the following unconstrained optimization problem:

$$\arg \min_{\boldsymbol{\alpha}} \frac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi D\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1 \tag{5}$$

The above ℓ_1-minimization problem is a classic sparse coding problem and can be efficiently solved by various methods, such as the iterative shrinkage algorithm [9], the split Bregman algorithm [33,34], and the alternative direction multiplier method [35].

3 Modeling of compressed sensing image reconstruction via adaptive sparse nonlocal regularization

Now, the goal is to reconstruct the desired image $\boldsymbol{x}$ from CS measurements $\boldsymbol{y}$. In the scenario of CS image reconstruction, this problem can be converted to decoding the CS measurements $\boldsymbol{y}$ so as to obtain the desired $\boldsymbol{\alpha}$

$$\arg \min_{\boldsymbol{\alpha}} \frac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi D\alpha}||_2^2 + \lambda \boldsymbol{R(\alpha)} \tag{6}$$

where $\boldsymbol{R(\alpha)}$ represents some regularization terms and λ represents the regularization parameter. However, in this paper, due to not knowing whether $\boldsymbol{R(\alpha)}$ is a convex or a non-convex minimization, we have developed a convenient implementation using the split Bregman iteration algorithm to solve Eq. (6) without considering the property of $\boldsymbol{R(\alpha)}$ in advance.

3.1 Split Bregman based iterative algorithm for the proposed ASNR model

In this section, we exploit the algorithmic framework of the split Bregman iteration [33,34] to solve Eq. (6) and introduce the implementation details of the proposed method.

The split Bregman iteration (SBI) has recently been proposed by Goldstein [33] for solving a very broad class of ℓ_1 norm related minimization problems. By exploiting the variable splitting technique, the SBI translates the unconstrained

Table 1 Split Bregman iteration

Split Bregman Iteration (SBI)
1.   **set** $t = 0, choose\ \mu > 0, \boldsymbol{b}_0 = 0, \boldsymbol{z}_0 = 0, \boldsymbol{\alpha}_0 = 0.$
2.   **Repeat**
3. $\boldsymbol{z}^{(t+1)} = \arg \min_z(z) + \frac{\mu}{2}||\boldsymbol{z} - \boldsymbol{G\alpha}^{(t)} - \boldsymbol{b}^{(t)}||_2^2$
4. $\boldsymbol{\alpha}^{(t+1)} = \arg \min_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) + \frac{\mu}{2}||\boldsymbol{z}^{(t+1)} - \boldsymbol{G\alpha} - \boldsymbol{b}^{(t)}||_2^2$
5. $\boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} - (\boldsymbol{z}^{(t+1)} - \boldsymbol{G\alpha}^{(t+1)})$
6. $t \leftarrow t+1;$
7.   **Until**   stoping criterion is satisfied.

minimization problem into a constrained one, after which the constrained minimization problem can be solved by invoking the Bregman iteration [33,34]. Next, we will briefly introduce the SBI, by considering a constrained optimization problem, i.e.,

$$\min_{\boldsymbol{z}\in\Re^{\boldsymbol{N}}, \boldsymbol{\alpha}\in\Re^{\boldsymbol{M}}} f(\boldsymbol{z}) + g(\boldsymbol{\alpha}), \quad s.t. \quad \boldsymbol{z} = \boldsymbol{G\alpha} \tag{7}$$

where $\boldsymbol{G} \in \Re^{\boldsymbol{M}\times\boldsymbol{N}}$ and $f : \Re^{\boldsymbol{N}} \to \Re$, $g : \Re^{\boldsymbol{M}} \to \Re$ are convex functions. Note that although $g(\boldsymbol{\alpha})$ is a convex function, it does not imply that it is infeasible when $g(\boldsymbol{\alpha})$ is a non-convex function, which has been done in [36]. The split Bregman iteration algorithm is described in Table 1.

In the SBI, the problem sequence approaching infinity can be avoided with the fixed parameter μ, which has been done in [37]. According to the SBI, each sub-problem minimization may be simpler than solving the original problem directly if the original minimization is split into two sub-problems.

Now, let us come back to Eq. (6) and show how to exploit the framework of the SBI to solve it. We first convert Eq. (6) into another equivalent constrained form by introducing a variable $\boldsymbol{z}$, namely,

$$\arg \min_{\boldsymbol{z},\boldsymbol{\alpha}} \frac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi z}||_2^2 + \lambda \boldsymbol{R(\alpha)}, \quad s.t. \quad \boldsymbol{z} = \boldsymbol{D\alpha} \tag{8}$$

By defining $f(\boldsymbol{z}) = ||\boldsymbol{y} - \boldsymbol{\Phi z}||_2^2$, $g(\boldsymbol{\alpha}) = \lambda \boldsymbol{R(\alpha)}$, we have

$$\boldsymbol{z}^{(t+1)} = \arg \min_{\boldsymbol{z}} \frac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi z}||_2^2 + \frac{\mu}{2}||\boldsymbol{z} - \boldsymbol{D\alpha}^{(t)} - \boldsymbol{b}^{(t)}||_2^2 \tag{9}$$

and

$$\boldsymbol{\alpha}^{(t+1)} = \arg \min_{\boldsymbol{\alpha}} \lambda \boldsymbol{R(\alpha)} + \frac{\mu}{2}||\boldsymbol{z}^{(t+1)} - \boldsymbol{D\alpha} - \boldsymbol{b}^{(t)}||_2^2 \tag{10}$$

We can then update $\boldsymbol{b}^{(t)}$ by line 5 in Table 1:

$$\boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} - (\boldsymbol{z}^{(t+1)} - \boldsymbol{D\alpha}^{(t+1)}) \tag{11}$$

Therefore, according to the SBI, the minimization for Eq. (6) is equivalent to solve two sub-problems, that is, z and $\boldsymbol{\alpha}$ sub-problems. Next, we will expound that every split sub-problem is suitable for an efficient solution. To avoid confusion, the subscribe t may be omitted for conciseness.

1. z sub-problem

Given $\boldsymbol{\alpha}$, the z sub-problem denoted by Eq. (9) becomes

$$\arg \min_z \boldsymbol{Q}_1(z) = \arg \min_z \frac{1}{2}||\boldsymbol{y} - \boldsymbol{\Phi} z||_2^2 + \frac{\mu}{2}||z - \boldsymbol{D}\boldsymbol{\alpha} - \boldsymbol{b}||_2^2 \tag{12}$$

Clearly, z is actually a closed form of Eq. (12); $\boldsymbol{Q}_1(z)$ is a strictly convex quadratic function and its gradient can be expressed as

$$\boldsymbol{d}(\boldsymbol{Q}_1(z)) = \boldsymbol{\Phi}^T \boldsymbol{\Phi} z - \boldsymbol{\Phi}^T \boldsymbol{y} + \mu(z - \boldsymbol{D}\boldsymbol{\alpha} - \boldsymbol{b}) \tag{13}$$

Setting $\boldsymbol{d}(\boldsymbol{Q}_1(z))$ to be zero gives the exact minimization of Eq. (12), namely

$$\hat{z} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mu \boldsymbol{I})^{-1}(\boldsymbol{\Phi}^T \boldsymbol{y} + \mu(\boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{b})) \tag{14}$$

where $\boldsymbol{I}$ represents the identity matrix.

In the CS image reconstruction, since $\boldsymbol{\Phi}$ is a random projection matrix without a special structure, computing the inverse by Eq. (14) at each iteration is too costly to implement numerically. Therefore, to avoid computing the matrix inverse, an iterative method is highly desirable for solving Eq. (12). In this paper, we exploit the gradient descent method with the optimal step to solve Eq. (12), namely

$$\hat{z} = z - \eta \boldsymbol{d} \tag{15}$$

where $\boldsymbol{d}$ is the gradient direction of the objective function $\boldsymbol{Q}_1(z)$, $\eta = \mathbf{abs}(\boldsymbol{d}^T \boldsymbol{d}/\boldsymbol{d}^T \partial \boldsymbol{d})$ is the optimal step, $\partial = \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mu \boldsymbol{I}$, and $\boldsymbol{I}$ is the identity matrix.

Finally, it only requires an iterative calculation of the following equation to solve the z sub-problem, namely,

$$\hat{z} = z - \eta(\boldsymbol{\Phi}^T \boldsymbol{\Phi} z - \boldsymbol{\Phi}^T \boldsymbol{y} + \mu(z - \boldsymbol{D}\boldsymbol{\alpha} - \boldsymbol{b})) \tag{16}$$

where $\boldsymbol{\Phi}^T \boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^T \boldsymbol{y}$ can be calculated in advance, making the solving of the above equation more efficient.

2. $\boldsymbol{\alpha}$ sub-problem

Given z, similarly, according to Eq. (10), the $\boldsymbol{\alpha}$ sub-problem can be written as

$$\arg \min_{\boldsymbol{\alpha}} \boldsymbol{Q}_2(\boldsymbol{\alpha}) = \arg \min_{\boldsymbol{\alpha}} \frac{1}{2}||\boldsymbol{D}\boldsymbol{\alpha} - \boldsymbol{r}||_2^2 + \frac{\lambda}{\mu}\boldsymbol{R}(\boldsymbol{\alpha}) \tag{17}$$

where $\boldsymbol{r} = z - \boldsymbol{b}$.

To make Eq. (17) as tractable as possible, we will make a conversion: let $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}$, and then Eq. (17) can be rewritten as:

$$\arg \min_{\boldsymbol{\alpha}} \boldsymbol{Q}_2(\boldsymbol{\alpha}) = \arg \min_{\boldsymbol{\alpha}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{r}||_2^2 + \frac{\lambda}{\mu}\boldsymbol{R}(\boldsymbol{\alpha}) \tag{18}$$

Note that it is difficult to solve Eq. (18) due to the unknown structure of $\boldsymbol{R}(\boldsymbol{\alpha})$. In this paper, a general assumption is made, with which even a closed form can be achieved. Specifically, $\boldsymbol{r}$ can be regarded as some type of noisy observation of $\boldsymbol{x}$, and then the assumption that each element of $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{r}$ follows an independent zero-mean distribution with variance σ^2 is made. The universality of the above assumption merits attention, which means that it needs to be, for example, neither Gaussian nor Laplacian nor a generalized Gaussian distribution (GGD) [38] process, but something more general. The following conclusion can be proved with this assumption.

Theorem 1 *Defining $\boldsymbol{x}, \boldsymbol{r} \in \Re^N$, $\boldsymbol{x}_i, \boldsymbol{r}_i \in \Re^n$, $\boldsymbol{e}(j)$ is every element of the error vector $\boldsymbol{e}$, where $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{r}$, $j = 1, \ldots, N$. Assume that $\boldsymbol{e}(j)$ follows an independent zero mean distribution with variance σ^2. Therefore, for any $\varepsilon > 0$, we can represent the relationship between $\frac{1}{N}||\boldsymbol{x} - \boldsymbol{r}||_2^2$ and $\frac{1}{K}\sum_{i=1}^n ||\boldsymbol{x}_i - \boldsymbol{r}_i||_2^2$ by the following property, namely,*

$$\lim_{\substack{N \to \infty \\ K \to \infty}} \boldsymbol{P}\left\{\left|\frac{1}{N}||\boldsymbol{x} - \boldsymbol{r}||_2^2 - \frac{1}{K}\sum_{i=1}^n ||\boldsymbol{x}_i - \boldsymbol{r}_i||_2^2\right| < \varepsilon\right\} = 1 \tag{19}$$

where $\boldsymbol{P}(\bullet)$ represents the probability and $\mathrm{K} = \mathrm{n} \times \mathrm{k}$. The detailed proof of the Theorem 1 is given in the Appendix.

On the basis of Theorem 1, we have the following equation with very large probability (restricted to 1) at each iteration t, namely,

$$\frac{1}{N}||\boldsymbol{x} - \boldsymbol{r}||_2^2 = \frac{1}{K}\sum_{i=1}^n ||\boldsymbol{x}_i - \boldsymbol{r}_i||_2^2 \tag{20}$$

Next, the following equation can be achieved by merging Eq. (20) into Eq. (18), that is

$$\arg \min_{\boldsymbol{\alpha}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{r}||_2^2 + \frac{\lambda}{\mu}\boldsymbol{R}(\boldsymbol{\alpha})$$
$$= \arg \min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^n ||\boldsymbol{x}_i - \boldsymbol{r}_i||_2^2 + \frac{\lambda K}{\mu N}\sum_{i=1}^n \boldsymbol{R}_i(\boldsymbol{\alpha}) \tag{21}$$
$$= \arg \min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^n ||\boldsymbol{x}_i - \boldsymbol{r}_i||_2^2 + \tau \sum_{i=1}^n \boldsymbol{R}_i(\boldsymbol{\alpha})$$

where $\tau = \frac{\lambda K}{\mu N}$.

It is obvious that the minimization of Eq. (21) can be efficient by solving n sub-problem for all the image patches x_i. Each patch-based sub-problem can be calculated by:

$$\alpha_i = \arg\min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} ||x_i - r_i||_2^2 + \tau \sum_{i=1}^{n} R_i(\alpha)$$
$$= \arg\min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} ||D\alpha_i - r_i||_2^2 + \tau \sum_{i=1}^{n} R_i(\alpha) \quad (22)$$

Clearly, Eq. (22) can also be viewed as a sparse coding problem. Now, the question how to determine the specific structure of $R(\alpha)$ is quite essential. It has been widely recognized that image prior knowledge plays a significant role in the performance of CS image reconstruction, and thus designing an effective regularization term to reflect image priors is at the core of CS image reconstruction. Therefore, several regularization terms have been developed in recent years. For instance, when $R_i(\alpha)$ is $||\alpha_i||_1$, Eq. (22) becomes a convex ℓ_1 norm minimization problem and can be efficiently solved by iterative shrinkage algorithm [9]. Meanwhile, when $R_i(\alpha)$ is $||\alpha_i||_0$, Eq. (22) becomes a non-convex ℓ_0 norm minimization problem, and apparently its constrained optimization form can be efficiently solved by the orthogonal matching pursuit (OMP) algorithm [20].

From the above analysis, it is clear to see that the SBI method solves not only the convex optimization, but also the non-convex optimization problem. However, since r is regarded as a good approximation of the original image x at each iteration, it is challenging to recover the true sparse coding coefficients α from r. Only using the local sparsity constraint $||\alpha_i||_1$ or $||\alpha_i||_0$ in Eq. (22) may not lead to a sufficiently accurate CS image reconstruction, and even $||\alpha_i||_0$ is a non-convex ℓ_0 norm minimization. Therefore, good priors of natural images are crucial to improve the CS image reconstruction performance. It is well-known that the image sparse coding coefficients α are not randomly distributed due to the local and nonlocal correlations existing in natural images. To achieve a sufficiently accurate CS reconstruction performance, we first use the NCSR model [19] as a regularization term instead of a conventional local sparsity constraint term for the CS image reconstruction, that is,

$$\Sigma_i ||\alpha_i - u_i||_1 \quad (23)$$

where u_i represent some reasonably good approximation of the sparse-coded patch-based original image x.

In our CS reconstruction method, although the original image x is unknown, for each patch x_i in image r, we search the nonlocal similar patches for it from within a large enough window in image r, because r can be regarded as a good approximation of the original image x. We exploit the weighted average of the first x_i^q patches closest to L to pre-

dict x_i', that is, $x_i' = \Sigma_{q=1}^{L} b_i^q x_i^q$. The weight b_i^q is set to be inversely proportional to the distance between patches x_i and $x_i^q : b_i^q = \exp(-||x_i - x_i^q||_2^2/h)/W$, where h is a predetermined scalar and W is the normalization factor. In the case of an orthogonal dictionary (e.g., the local PCA dictionary used in this work, see Sect. 3.2 for more details), each patch x_i of r and its nonlocal prediction patch x_i' are easily coded by a dictionary D_i, that is, $x_i = D_i\alpha_i$ and $x_i' = D_i u_i$. Since the dictionary is orthogonal, the sparse coding coefficients α_i and u_i should be similar, and,

$$u_i = D_i^{-1} x_i' \quad (24)$$

Accordingly, we can exploit $R_i(\alpha) = \Sigma_i||\alpha_i - u_i||_p$ to regularize the solution of Eq. (22), namely,

$$\alpha_i = \arg\min_{\alpha_i} \frac{1}{2} \sum_{i=1}^{n} ||D\alpha_i - r_i||_2^2 + \tau \Sigma_i ||\alpha_i - u_i||_p \quad (25)$$

where p denotes the ℓ_p norm. In general, p is equal to 1, which leads to the Laplacian distribution and p is equal to 2, which follows the Gaussian distribution. If $0 < p < 1$, which follows the hyper-Laplacian distribution.

Let

$$\chi_i = \alpha_i - u_i \quad (26)$$

Here we perform some experiments to investigate the statistical property of χ, where χ represents the set of χ_i. We have tested the fidelity term in Eq. (25) on the image *Leaves* with $0.2N$ measurements, *Monarch* with $0.3N$ measurements and *boats* with $0.4N$ measurements, respectively. We plot the histogram of χ as well as the fitting Gaussian, Laplacian and hyper-Laplacian distributions of χ in Figs. 1a, 2a and 3a, respectively. To better observe the fitting of the tails, we also plot these distributions in the log domain in Figs. 1b, 2b and 3b, respectively. It can be observed that the histogram of χ can be characterized well by the Laplacian distribution. Therefore, the ℓ_1 norm is adopted to regularize χ, and the proposed model can be formulated as:

$$\alpha_i = \arg\min_{\alpha_i} \frac{1}{2} \sum_{i=1}^{n} ||D\alpha_i - r_i||_2^2 + \tau \Sigma_i ||\alpha_i - u_i||_1 \quad (27)$$

Note that for fixed τ, u_i, Eq. (27) is convex and can be achieved efficiently using an iterative thresholding algorithm. We adopt the surrogate algorithm in [9] to solve it. In the $(t+1)$-iteration, the proposed shrinkage operator can be calculated as

$$\alpha_i^{(t+1)} = S_\rho(D^{-1} x_i^{(t)} - u_i^{(t)}) + u_i^{(t)} \quad (28)$$

Fig. 1 The distribution of χ for image *Leaves* with $0.2N$ measurements and the fitting Gaussian, Laplacian distribution, hyper-Laplacian distribution in **a** linear and **b** log domain, respectively
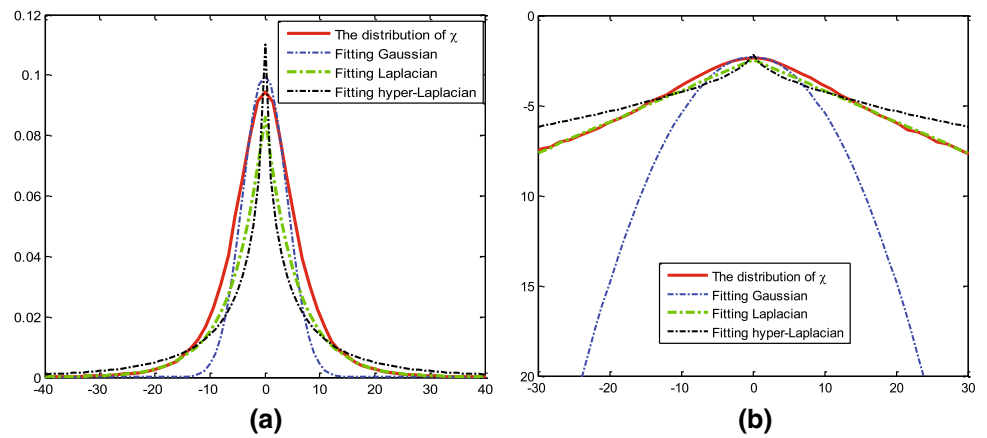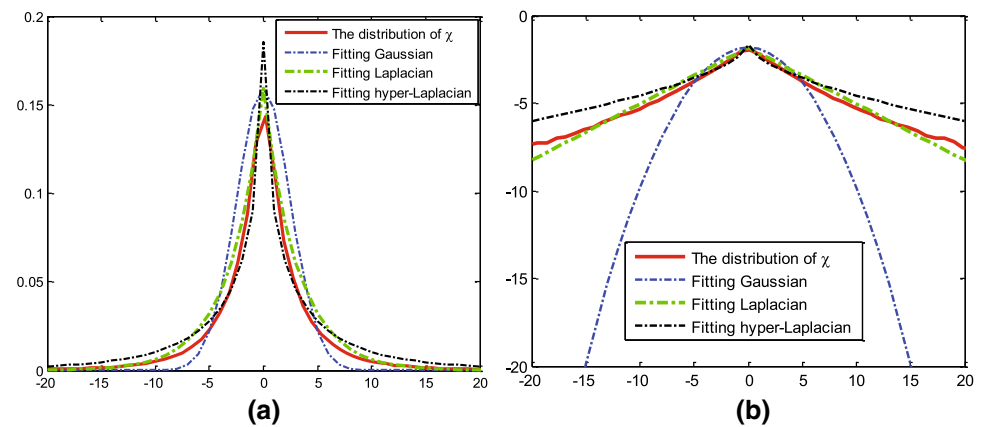

(a)

(b)

Fig. 2 The distribution of χ for image *Monarch* with $0.3N$ measurements and the fitting Gaussian, Laplacian distribution, hyper-Laplacian distribution in **a** linear and **b** log domain, respectively

(a)

(b)

Fig. 3 The distribution of χ for image *Boats* with $0.4N$ measurements and the fitting Gaussian, Laplacian distribution, hyper-Laplacian distribution in **a** linear and **b** log domain, respectively

(a)

(b)
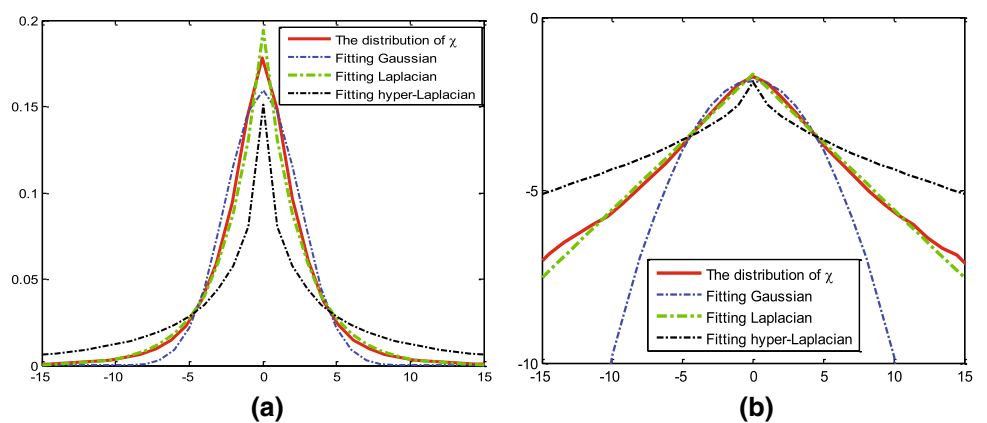
where $S_\rho(\bullet)$ is the soft thresholding operator, $\rho = \tau/c$ and c is an auxiliary parameter. The above shrinkage operator follows the standard surrogate algorithm, from which more details can be seen in [9]. Hence, if we know the dictionary D, α can be obtained by sparse coding all the overlapped patches, which is the final solution for α sub-problem in Eq. (17). In the next sub-section, we will discuss how to learn a self-adaptive dictionary D.

According to the above analysis, the main difference between [19] and this paper for sparse coding coefficient α is that we utilize the spit Bregman iteration method to solve a sub-problem to achieve α, while [19] solves α by a direct method. The rationale behind is that each sub-problem minimization may be more effective and robust than the original problem. In addition, the iterative shrinkage (IS) algorithm [9] is applied to solve the original problem in [19]. However,

Fig. 4 The distribution of $e^{(t)}$ for image *Leaves* in the context of CS image reconstruction at different iteration. **a** $(t) = 3$; **b** $(t) = 9$;
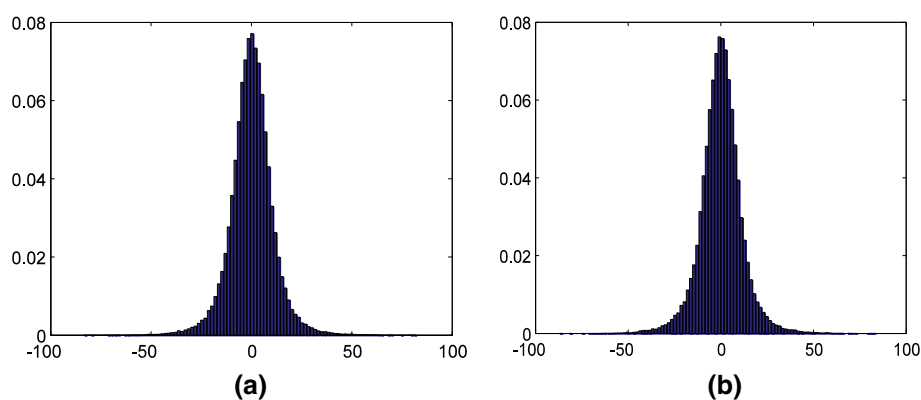
the major drawback of the IS algorithm is its low convergence speed and instability [39].

Note that the main difference between [39] and our method is that we utilizing the local and nonlocal redundancies to constraint sparse coding coefficient, while method in [39] only using the local sparsity constraint. More specifically, we not only exploited the image nonlocal self-similarity to better estimate the sparse coding coefficients, but also centralized the sparse coding coefficients of the observed image to those estimates. However, [39] only utilized the image nonlocal self-similarity to estimate the local sparse coefficients, which may not an inaccurate CS reconstruction results.

In addition, based on the above analysis and due to the unknown structure of $\boldsymbol{R}(\boldsymbol{\alpha})$ in Eq. (18), we assume that the error vector $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{r}$ follows an independent zero-mean distribution with variance σ^2 and achieves a conclusion of Theorem 1. Later, we deduce that Eq. (18) can convert into Eq. (21) which can efficiently solve n sub-problems for all the image patches $\boldsymbol{x}_i$. In other words, Eq. (21) can also be viewed as the sparse coding problem.

Because we know the detailed structure of $\boldsymbol{R}(\boldsymbol{\alpha})$ in Eq. (27), namely, $\boldsymbol{R}(\boldsymbol{\alpha}) = ||\boldsymbol{\alpha} - \boldsymbol{u}||_1$, a series of experiments can be made. Specifically, because we regard $\boldsymbol{r}$ as the approximation of the original image $\boldsymbol{x}$, we achieve some experiments to survey the statistics of $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{r}$. In these experiments, a gray image *Leaves* is used as an example in the context of compressed sensing image reconstruction, where the original image is undersampled by a random Gaussian matrix with $0.2N$ measurements. At each iteration t, $\boldsymbol{r}^{(t)}$ can be obtained by $\boldsymbol{r}^{(t)} = \boldsymbol{z}^{(t)} - \boldsymbol{b}^{(t-1)}$. Since it is difficult to achieve the exact minimization of Eq. (27), we approximate $\boldsymbol{x}^{(t)}$ by $\boldsymbol{r}^{(t)}$ without the loss of generality. Thus, it is possible to achieve the histogram of residual $\boldsymbol{e}^{(t)} = \boldsymbol{x}^{(t)} - \boldsymbol{r}^{(t)}$ at each iteration t. Figure 4 shows the distributions of residual $\boldsymbol{e}^{(t)}$ when t is equal to 3 and 9, respectively.

From Fig. 4, it is clear that the distribution of $\boldsymbol{e}^{(t)}$ at each iteration is very similar to the generalized Gaussian distribution (GGD) [38] with zero-mean and variance σ^2. Therefore, we assume that the error vector $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{r}$ follows an inde-

pendent zero-mean distribution with variance σ^2 available in Eq. (18).

3.2 Self-adaptive PCA dictionary learning

Recently, there have been many works about learning dictionaries [14,40–43] from natural image patches. It is well-known that the KSVD [20,21] can represent various image local structures by learning a universal dictionary from natural image dataset. However, it has been shown that sparse coding over a highly redundant dictionary is potentially unstable and tends to generate visual artifacts [24,25]. In our method, we adopt the strategy in [14], that is, the K-means clustering combined with the local PCA are used to learn a set of PCA sub-dictionaries.

To obtain the self-adaptive PCA dictionary, the training image patches are usually derived from the natural image dataset or the original image. However, in many practical situations the training example images are simply not available, and, even in the CS image reconstruction, the original image is also not available; we can only use the CS measurements from Eq. (1). Although the original image $\boldsymbol{x}$ is unknown, we can still extract all the image patches from $\boldsymbol{r}$ to achieve a self-adaptive PCA dictionary, because $\boldsymbol{r}$ can be regarded as a good approximation of the original image $\boldsymbol{x}$ at each iteration in Eq. (27). We divide $\boldsymbol{r}$ into patches and cluster the patches into multiple clusters. To remove the redundancy from each cluster, we use the PCA to handle it to learn a local sub-dictionary. Meanwhile, the centroid of each cluster is calculated. For each given image patch, the Euclidian distance between the patch and the centroid of each sub-dictionary is calculated to adaptively choose the proper sub-dictionary. More details about the PCA dictionary can be seen in [14].

3.3 Summary of the proposed method

So far, the above two sub-problems have been solved. In fact, we achieve an efficient solution by solving each sub-problem separately, which can guarantee the whole method

Table 2 Complete distribution of proposed ASNR model for CS image reconstruction

Input: the observed measurement y and the measurement matrix Φ.
Initialization: $t = 0$, $b^{(0)} = 0$, $z^{(0)} = 0$, $\alpha^{(0)} = 0$, L, bc, c, λ, η, μ;
Repeat
Update $z^{(t+1)}$ by Eq.(16);
$r^{(t+1)} = z^{(t+1)} - b^{(t)}$; $\tau = \lambda K/\mu N$; $\rho = \tau/c$;
Update $D^{(t+1)}$ by K-means and PCA from $r^{(t+1)}$;
For   Each patch x_i;
Update $\hat{\alpha}_i^{(t+1)}$ by computing by Eq.(28);
End For
Update $\hat{\alpha}^{(t+1)}$ by concatenating all $\hat{\alpha}_i^{(t+1)}$;
Update $b^{(t+1)}$ by computing by Eq.(11) ;
$t \leftarrow t + 1$;
Until
maximum iteration number is reached.
Output:
The final restored image $\hat{x} = D \circ \hat{\alpha}$.

to be more efficient and effective. Based on the above derivations, the complete description of the proposed method for the CS image reconstruction via the ASNR model is provided in Table 2.

4 Experimental results

4.1 Experimental framework and parameters setting

In this section, we will report the experimental results of the proposed ASNR-based CS image reconstruction method. We generated the CS measurements at the block level by utilizing the Gaussian random projection matrix to test images, i.e., the block-based CS recovery with block size of 32 × 32. In our experiments, both the natural images and the MRI images were used to verify the performance of the proposed ASNR method. The main parameters of the proposed method were set as follows: the size of each patch, i.e., $\sqrt{bc} \times \sqrt{bc}$ was 6 × 6, μ was set to be 0.0025, η was set to be 1, and the total $L = 40$ matched patches were selected for each exemplar patch. The parameter λ that balances the fidelity

term and the regularization term should be tuned separately for each sensing rate. The test images which were used in our experiments consisted of eight natural images and two MRI images of size 256 × 256 as shown in Fig. 5. The platform for the comparison was an Intel(R) Core(TM) i3-4150 with a 3.50GHz CPU and 4GB memory, and Windows 7 operating system.

4.2 Numerical performance comparisons

To verify the performance of the proposed ASNR method, we compared it with several competitive CS recovery methods including the BCS method [44], the NESTA method [45], the COS method [46], the BM3D based CS recovery method [47] (denoted as the BM3D-CS), the ASD-CS method [39], the NCSR-CS method [19], the adaptively learned sparsifying based CS recovery via minimization [48] (denoted as the ALSB), and the MRK-CS method [49]. The source codes of all bench methods [19,39,46–49] were obtained from the authors' websites. To have a fair comparison among the competing methods, we used the built-in parameter settings in each implementation which performed optimally.

In the first experiment, the numerical performance comparison of the proposed ASNR method was as follows. The PSNR comparison results of recovered images by competing CS recovery methods are shown in Table 3. The ASNR method consistently outperformed other methods on most of the test images over different numbers of CS measurements, which further increases the performance advantage of the proposed ASNR method (e.g., the ASNR can outperform eight comparative methods by up to 6.75, 5.16, 3.90, 2.52, 0.66, 0.58, 1.77 and 1.22 dB, respectively). On average, the proposed ASNR method outperforms all previous benchmark methods. The PSNR gains of the ASNR over the BCS, NESTA, COS, BM3D-CS, ASD-CS, NCSR-CS, ALSB, and MRK-CS can be as much as 9.70, 9.83, 5.74, 4.15, 1.06, 0.71, 1.32 and 3.34 dB, respectively on *Barbara* image— the most

Fig. 5 Test images used for compressed sensing reconstruction experiments. **a** House. **b** Barbara. **c** Leaves. **d** Boats. **e** Foreman. **f** Goldhill. **g** starfish. **h** Monarch. **i** Brain MRI image. **j** Bone MRI image
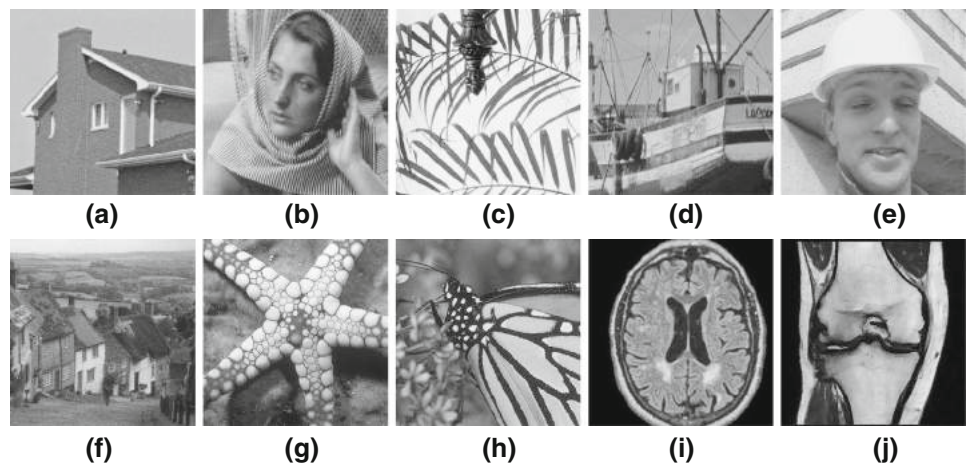
(a)   (b)   (c)   (d)   (e)

(f)   (g)   (h)   (i)   (j)

Table 3 continued

Images	Method	Number of measurements				
		$M = 0.1N$	$M = 0.2N$	$M = 0.3N$	$M = 0.4N$	$M = 0.75N$
Average	BCS [44]	24.07	27.03	29.13	30.98	37.82
	NESTA [45]	23.36	28.03	31.09	33.65	40.84
	COS [46]	23.42	30.35	32.58	34.83	41.10
	BM3D-CS [47]	25.08	30.57	34.06	36.42	43.98
	ASD-CS [39]	27.91	32.62	35.73	38.21	45.00
	NCSR-CS [19]	28.31	33.26	35.88	38.00	44.35
	ALSB [48]	26.95	31.28	34.16	36.64	44.88
	MRK-CS [49]	27.88	31.92	34.88	37.32	44.64
	ASNR	**28.69**	**33.27**	**36.15**	**38.50**	**46.16**

Bold represents the best or optimal meaning

Fig. 6 CS reconstruction *foreman* image with 0.1*N* measurements. **a** Original image; **b** BCS reconstruction [44] (29.76 dB); **c** NESTA reconstruction [45] (29.94 dB); **d** COS reconstruction [46] (24.66 dB); **e** BM3D-CS reconstruction [47] (32.68 dB); **f** ASD-CS reconstruction [39] (35.03 dB); **g** NCSR-CS reconstruction [19] (34.18 dB); **h** ALSB reconstruction [48] (34.34 dB); **i** MRK-CS reconstruction [49] (35.29 dB); **j** Proposed ASNR reconstruction (36.11 dB)
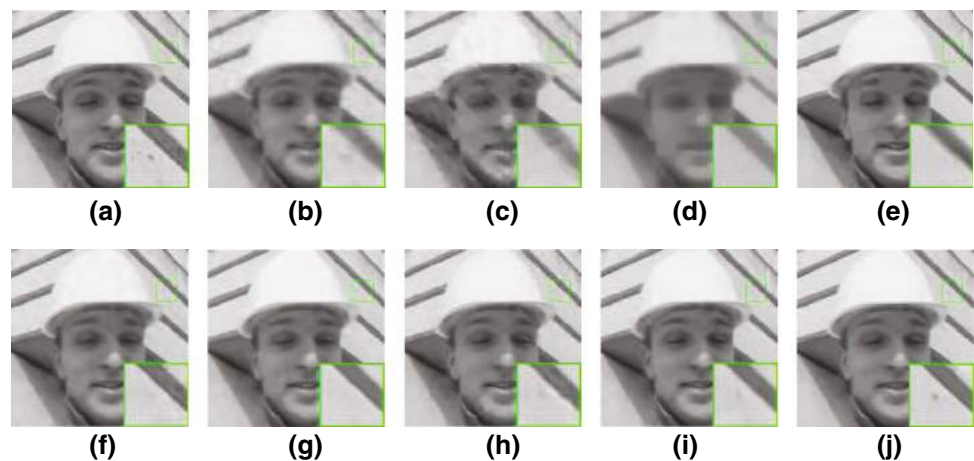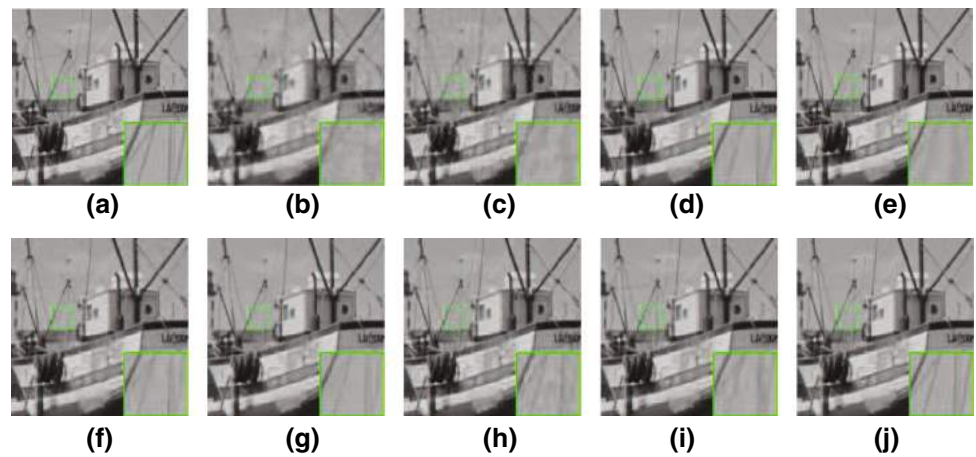
Fig. 7 CS reconstruction *Bone* image with 0.1*N* measurements. **a** Original image; **b** BCS reconstruction [44] (24.93 dB); **c** NESTA reconstruction [45] (24.82 dB); **d** COS reconstruction [46] (24.49 dB); **e** BM3D-CS reconstruction [47] (27.39 dB); **f** ASD-CS reconstruction [39] (29.23 dB); **g** NCSR-CS reconstruction [19] (29.14 dB); **h** ALSB reconstruction [48] (27.93 dB); **i** MRK-CS reconstruction [49] (29.09 dB); **j** Proposed ASNR reconstruction (29.38 dB)
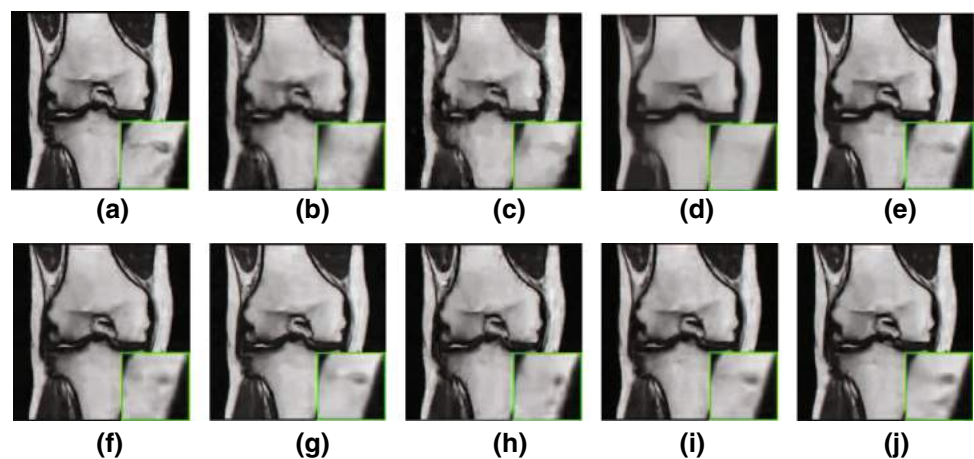
favorable situation for an adaptive sparse nonlocal regularization and the split Bregman iteration algorithm is exploited. In many cases, it has to be pointed out that the gain of the proposed ASNR method with low sampling ratios is larger than with high sampling ratios. The reason for this is that the self-adaptive PCA dictionary learning method can capture the local image feature better and greatly reduce visual artifacts in the case of low sampling ratios than some prede-fined transform and some over-complete dictionary learning methods.

4.3 Visual quality comparisons

To facilitate the evaluation of subjective qualities, some of the reconstructed images are shown in Figs. 6, 7, 8, 9, 10, 11. From them, we can see that the BCS and NESTA methods

Fig. 8 CS reconstruction *boats* image with 0.2*N* measurements. **a** Original image; **b** BCS reconstruction [44] (27.05 dB); **c** NESTA reconstruction [45] (28.30 dB); **d** COS reconstruction [46] (31.43 dB); **e** BM3D-CS reconstruction [47] (30.99 dB); **f** ASD-CS reconstruction [39] (33.15 dB); **g** NCSR-CS reconstruction [19] (33.56 dB); **h** ALSB reconstruction [48] (31.92 dB); **i** MRK-CS reconstruction [49] (32.37 dB); **j** Proposed **ASNR** reconstruction (33.63 dB)

Fig. 9 CS reconstruction *Barbara* image with 0.2*N* measurements. **a** Original image; **b** BCS reconstruction [44] (24.31 dB); **c** NESTA reconstruction [45] (23.96 dB); **d** COS reconstruction [46] (26.54 dB); **e** BM3D-CS reconstruction [47] (28.85 dB); **f** ASD-CS reconstruction [39] (32.27 dB); **g** NCSR-CS reconstruction [19] (33.27 dB); **h** ALSB reconstruction [48] (31.86 dB); **i** MRK-CS reconstruction [49] (27.99 dB); **j** Proposed ASNR reconstruction (33.40 dB)

Fig. 10 CS reconstruction *Leaves* image with 0.3*N* measurements. **a** Original image; **b** BCS reconstruction [44] (23.31 dB); **c** NESTA reconstruction [45] (27.39 dB); **d** COS reconstruction [46] (30.54 dB); **e** BM3D-CS reconstruction [47] (32.50 dB); **f** ASD-CS reconstruction [39] (32.53 dB); **g** NCSR-CS reconstruction [19] (33.86 dB); **h** ALSB reconstruction [48] (30.84 dB); **i** MRK-CS reconstruction [49] (32.37 dB); **j** Proposed ASNR reconstruction (34.78 dB)
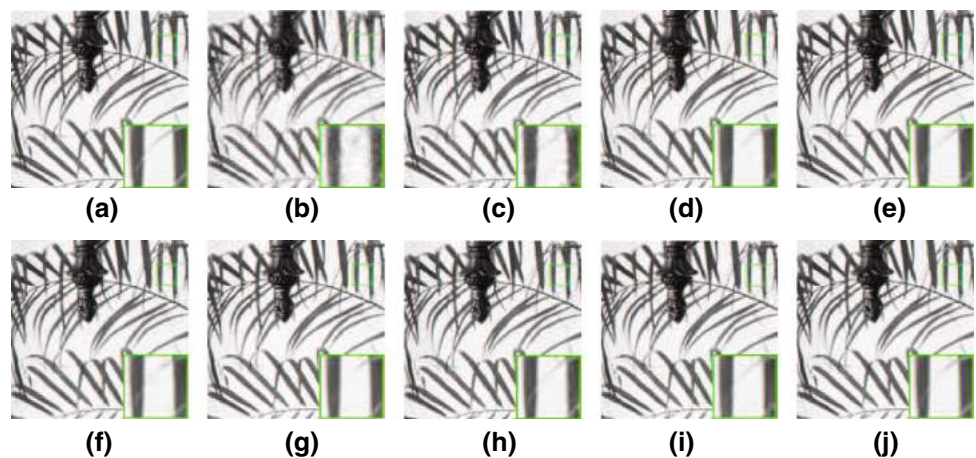
cannot reconstruct sharp edges and fine image details. The COS, BM3D-CS, ASD-CS, NCSR-CS, ALSB and MRK-CS methods can generate better visual quality than the BCS and NESTA methods. However, the COS, BM3D-CS and NCSR-CS still suffer from some undesirable over-smooth phenomena, such as the loss of some fine image details and visual artifacts that can still be clearly observed from the ASD-CS, ALSB and MRK-CS. By contrast, the proposed ASNR approach not only removes most of the visual artifacts and ring effects, but also effectively preserves large-scale sharp edges and small-scale fine image details in comparison with the other competing methods. The high performance of the proposed ASNR approach is attributed to the following facts: First, to exploit the self-adaptive PCA sub-dictionary can effectively capture the local image feature to greatly reduce visual artifacts. Second, an adaptive sparse nonlocal

Fig. 11 CS reconstruction *starfish* image with 0.3N measurements. **a** Original image; **b** BCS reconstruction [44] (27.17 dB); **c** NESTA reconstruction [45] (28.77 dB); **d** COS reconstruction [46] (30.24 dB); **e** BM3D-CS reconstruction [47] (30.45 dB); **f** ASD-CS reconstruction [39] (32.79 dB); **g** NCSR-CS reconstruction [19] (33.18 dB); **h** ALSB reconstruction [48] (29.58 dB); **i** MRK-CS reconstruction [49] (32.01 dB); **j** Proposed ASNR reconstruction (33.28 dB)
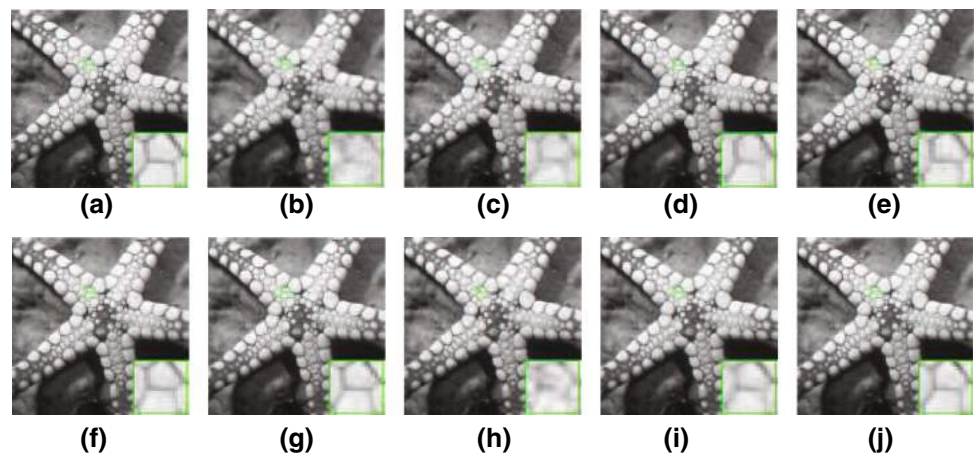
Fig. 12 Three different types of dictionaries. **a** SDCT. **b** ODCT. **c** GDIC
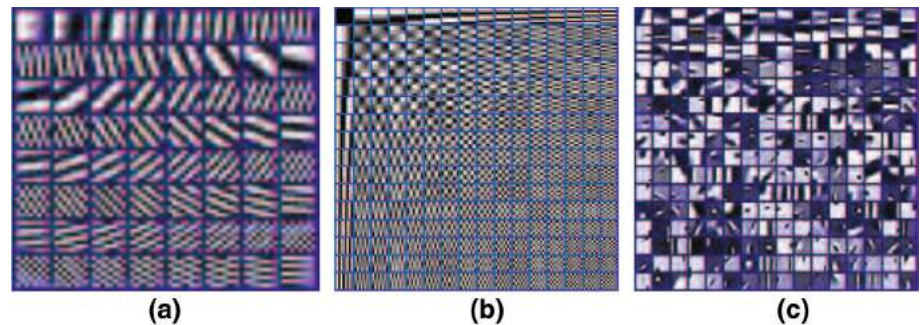
regularization can be introduced to further enhance the CS image reconstruction performance. Third, the split Bregman iteration algorithm is developed to improve the computational efficiency of the proposed ASNR method.

4.4 Effect of the dictionary selection

For this sub-section, another experiment was conducted to further evaluate the influence of different dictionary on the proposed ASNR reconstruction method. All the parameters were the same as specified in Sect. 4.1, except for the usage of the proposed dictionaries, including the standard DCT dictionary (SDCT), over-complete DCT dictionary (ODCT), and the globally trained dictionary (GDIC), as shown in Fig. 12. For different dictionaries, the CS reconstruction results for the image *Starfish* with 0.3N measurements are shown in Fig. 13. It is obvious that the reconstruction error images of the proposed self-adaptive PCA dictionary have less texture features than those of other dictionary learning methods, as shown in Fig. 13g–j, l–o, respectively. In other words, the proposed self-adaptive PCA learning dictionary method generates better results than other competing dictionary learning methods. The reason is that overcomplete redundant dictionary is potentially unstable and tends to generate visual artifacts, while the proposed PCA sub-dictionary can capture the local image feature better and greatly reduce artifacts. In addition, we compare the proposed PCA sub-dictionary,

which is learned from natural image dataset and the reconstructed image itself. We use the training dataset 2 as natural image dataset in [14]. From Figs. 14 and 15, we can see that the dictionary learned from natural image dataset loses of the fine details, while the dictionary learned from the reconstructed image itself can obtain fine details and produce more effective results than the dictionary learned from natural image dataset. The reason is that the natural image dataset can not always give fine description for various matching textures. However, the similar texture can be found in the reconstructed image without any doubt.

4.5 Effect of the number of best matched patches

In this sub-section, we will discuss how to select the best-matching patch numbers L for enhancing the performance of the proposed ASNR method. Specifically, to investigate the sensitivity of L, we conducted two experiments on three test images with respect to different L, ranging from 20 to 140, in the case of 0.1N measurement and 0.3N measurement, respectively. The performance comparison with different L is shown in Fig. 16. It is clear to see that all the curves are almost flat in Fig. 16, in other words, the performance of the proposed ASNR method is not too sensitive to L. The best performance of each case was usually achieved with L in the range [20, 60]. Thus, L was empirically set to be 40.

Fig. 13 CS reconstruction *Starfish* image with 0.3*N* measurements by different dictionaries. **a** Original image; **b** SDCT reconstruction (28.22 dB); **c** DCT reconstruction (29.21 dB); **d** GDIC reconstruction (30.19 dB); **e** proposed self-adaptive PCA dictionary reconstruction (33.28 dB); **f** and **k** different local magnification patch of the original image, respectively. **g**–**j** and **l**–**o** are the reconstruction error images by **f** and **k** with different dictionaries, respectively
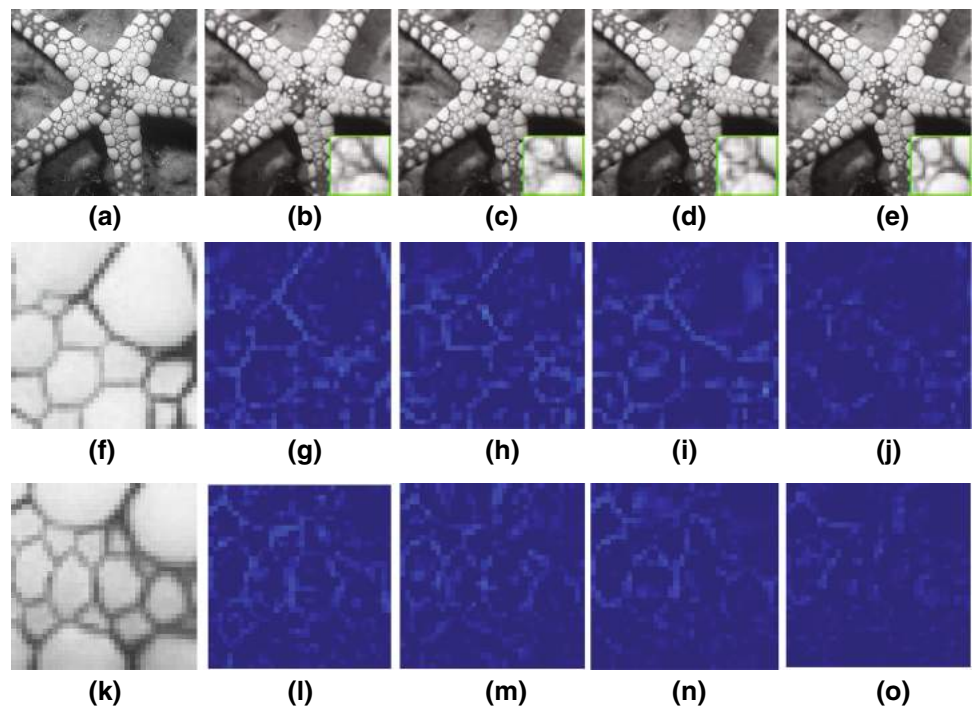
Fig. 14 CS reconstruction *House* image with 0.2*N* measurements by different dictionaries. **a** Original image; **b** dictionary learning from natural image dataset (35.56 dB); **c** dictionary learning from the reconstructed image itself (37.16 dB)
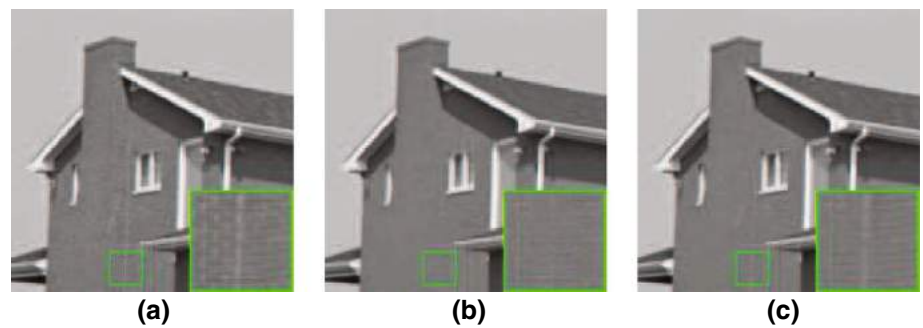
Fig. 15 CS reconstruction *Leaves* image with 0.2*N* measurements by different dictionaries. **a** Original image; **b** dictionary learning from natural image dataset (28.14 dB); **c** dictionary learning from the reconstructed image itself (28.89 dB)
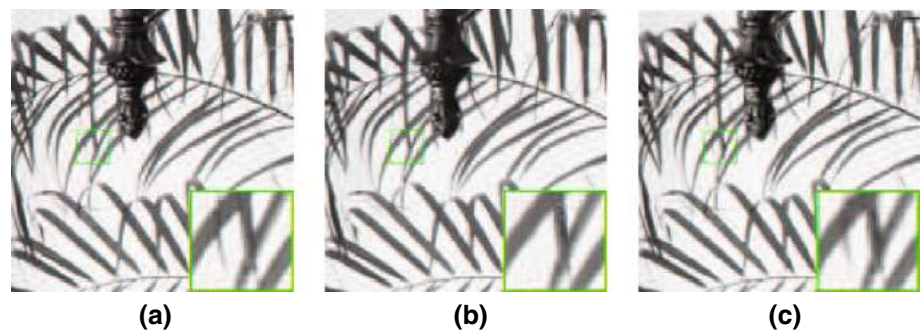
4.6 Comparison of the noise level for different methods

To test the noise sensitivity of the proposed ASNR method, we first conducted experiments with zero-mean Gaussian white noise added to the compressed sensing measurements. The small noise level was 1e−3, 2e−3, 3e−3, 4e−3 and 5e−3. Figure 17 shows the PSNR versus noise level in test images depicting *boats* and *Brain* under five under-sampling ratios (0.1, 0.2, 0.3, 0.4, and 0.75). It is obvious that the PSNR

of the reconstruction image remains stable for all specified noise levels, in other words, the proposed ASNR method can adequately remove the aliasing and noise in the reconstructed image since sparse coding based on an self-adaptive PCA dictionary and an adaptive sparse nonlocal regularization has the intrinsic denoising property. In addition, we conduct similar experiments with noisy CS measurements to demonstrate the robustness of the proposed ASNR under large noise situation. A significant amount of additive white Gaussian noise was

Fig. 16 Performance comparison with different matched patch numbers L for three test images. **a** The PSNR (dB) results achieved by different L in the case of $0.1N$ measurement. **b** The PSNR (dB) results achieved by different L in the case of $0.3N$ measurement
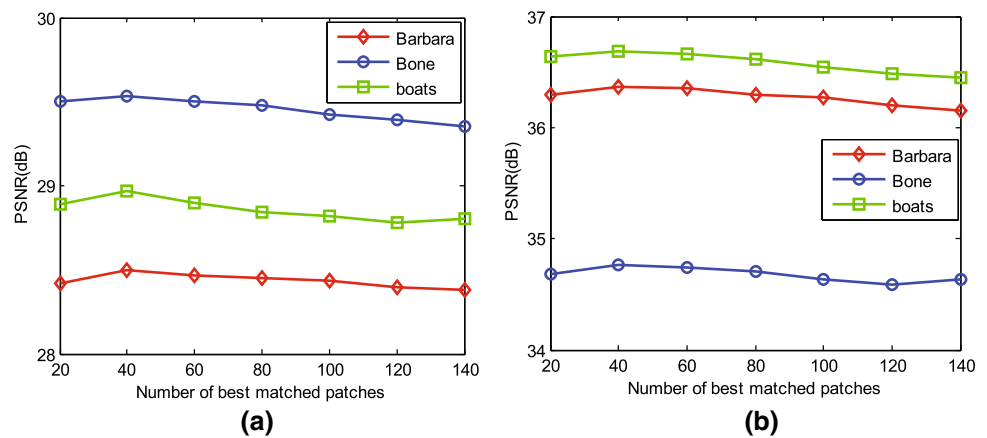

Fig. 17 Noise sensitivity analysis of the proposed ASNR method. **a** PSNR results versus measurement noise level for *boats*. **b** PSNR results versus measurement noise level for *Brain*
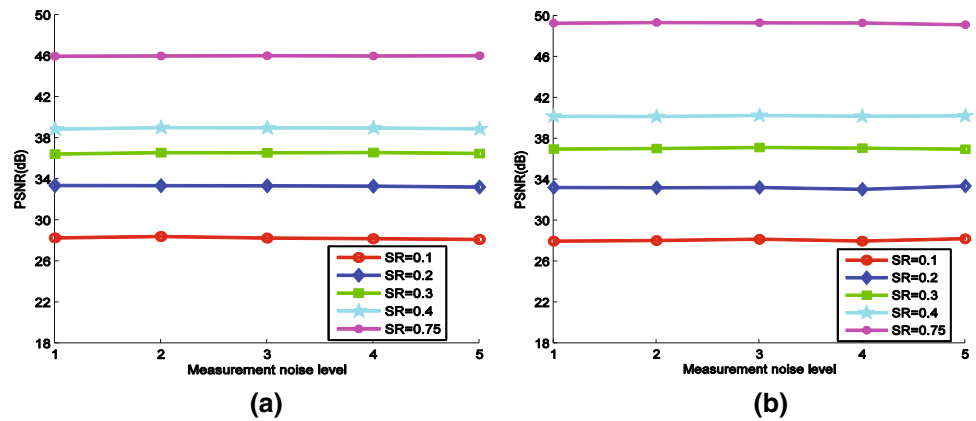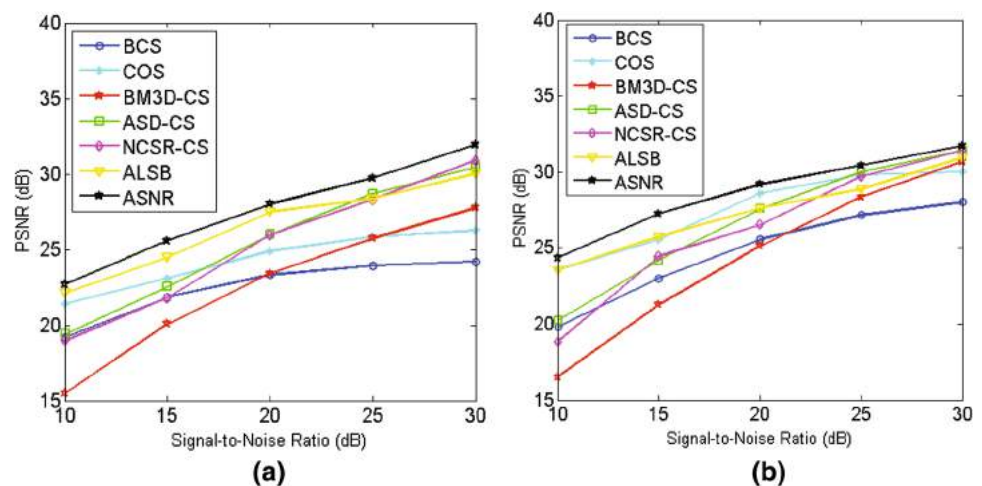

Fig. 18 PSNR of the different methods from noisy levels. **a** PSNR results versus SNR with $0.2N$ measurements for image *Barbara*. **b** PSNR results versus SNR with $0.2N$ measurements for image *Bone*

added to the CS measurements. The standard derivations of additive noise vary to generate signal-to-noise-ratio (SNR) between 10, 15, 20, 25 and 30 dB. All the parameters were the same as specified in Sect. 4.1, except $\mu = 19, 9, 3.5, 2.5$ and 0.25 corresponding to SNR $= 10, 15, 20, 25$ and 30 dB, respectively. In these experiments, the NESTA and MRK-CS methods have not included since they are sensi-

tive to noise. The PSNR comparison of the reconstructed images are shown in Fig. 18. One can observe that the proposed method outperforms other competing methods in all situations and subjective quality comparison results with SNR $= 20, 15, 10$ dB for image *Barbara*, *Bone*, *starfish* are shown in Figs. 19, 20, 21, respectively. It is obvious that the proposed ASNR method achieves the best perfor-

Fig. 22 Comparison between SBI and IST. **a** PSNR results achieved by SBI and IST with ratio = 0.2 for image *foreman*. **b** PSNR results achieved by SBI and IST with ratio = 0.3 for image *Barbara*
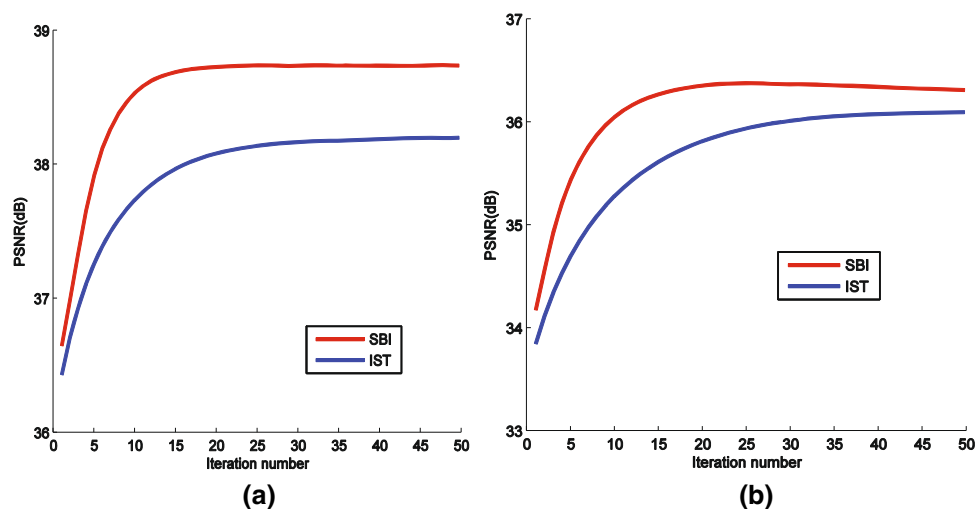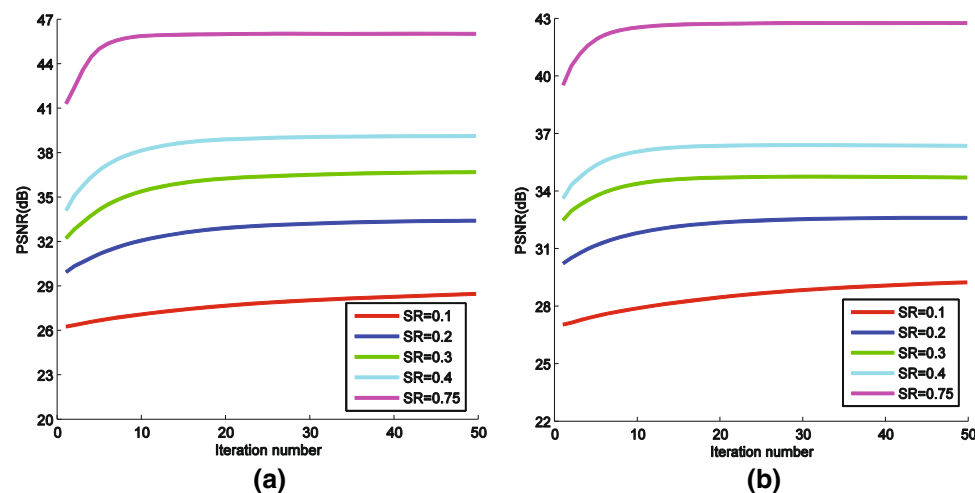

Fig. 23 Convergence analysis of the proposed ASNR method. **a** PSNR results versus iteration numbers for *boats*. **b** PSNR results versus iteration numbers for *Bone*

mance to suppress noise among all the competing methods. Therefore, the proposed ASNR method is robust to noise measurement.

4.7 Comparsion between SBI and IST

In this sub-section, another classical optimization method, the iterative shrinkage/thresholding (IST) [50] is exploited to solve our proposed ASNR model. We compared SBI with IST by setting ratio = 0.2 and ratio = 0.3 on test image *foreman* and *Barabar*. Figure 22 shows progression curves of the PSNR (dB) results. It can be seen that SBI method only requires 15 iterations (about 59 s per iteration) to achieve convergence, while IST needs 30 iterations (about 58 s per iteration). Furthermore, the SBI method can achieve better PSNR results than IST method. Therefore, compared to IST method, the proposed SBI method can achieve better convergence speed with fewer computations.

4.8 Convergence analysis

To elaborate the convergence performance of the proposed ASNR method, Fig. 23 shows the evolutions of the PSNR versus the iteration numbers for the test images depicting *boats* and *Bone* with various sampling ratios. It can be observed that with the growth of the iteration number, the PSNR curve of the reconstructed images gradually increases and ultimately tends to be flat and stable. Therefore, the good convergence performance of the proposed ASNR method is exhibited in Fig. 23.

5 Conclusion

In this paper, we have presented a new method toward CS image reconstruction based on adaptive sparse nonlocal regularization. The approach utilizes the nonlocal self-similarity prior integrated into an adaptive regularization to improve the