



Article ID 1007-1202(2023)01-0053-08

DOI <https://doi.org/10.1051/wujns/2023281053>

Group Sparsity Residual Constraint Image Denoising Model with l_1/l_2 Regularization

□ WU Di¹, ZHANG Tao^{2†}, MO Xutao²

1. Key Laboratory of Multidisciplinary Management and Control of Complex Systems of Anhui Higher Education Institutes, Anhui University of Technology, Maanshan 243002, Anhui, China;

2. School of Mathematics and Physics, Anhui University of Technology, Maanshan 243032, Anhui, China

© Wuhan University 2023

Abstract: Group sparse residual constraint with non-local priors (GSRC) has achieved great success in image restoration producing state-of-the-art performance. In the GSRC model, the l_1 norm minimization is employed to reduce the group sparse residual. In recent years, non-convex regularization terms have been widely used in image denoising problems, which have achieved better results in denoising than convex regularization terms. In this paper, we use the ratio of the l_1 and l_2 norm instead of the l_1 norm to propose a new image denoising model, i.e., a group sparse residual constraint model with l_1/l_2 minimization (GSRC- l_1/l_2). Due to the computational difficulties arisen from the non-convexity and non-linearity, we focus on a constrained optimization problem that can be solved by alternative direction method of multipliers (ADMM). Experimental results of image denoising show that the pro-posed model outperforms several state-of-the-art image denoising methods both visually and quantitatively.

Key words: image denoising; l_1/l_2 minimization; group sparse representation

CLC number: TP 751

0 Introduction

The goal of image denoising is to estimate the latent image $\mathbf{x} \in \mathbb{R}^{m \times n}$ from its degraded observation $\mathbf{y} \in \mathbb{R}^{m \times n}$, which is typically an ill-posed inverse problem. To tackle the ill-posed nature of this problem, regularization methods based on various image prior information have to be incorporated into the image denoising process, which is usually realized by minimizing the ob-

jective function as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \quad (1)$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the original image and the degraded image, respectively. The first term is the data fidelity and the second term is the regularization term. λ is a regularization parameter that balances these two terms. During the past decades, a variety of image denoising methods have been proposed, such as total variation^[1,2],

Received date: 2022-10-18

Foundation item: Supported by the Open Fund of Key Laboratory of Anhui Higher Education Institutes (CS2021-07), the National Natural Science Foundation of China (61701004), the Outstanding Young Talents Support Program of Anhui Province (GXYQ 2021178) and University Natural Science Research Project of Anhui Province of China (KJ2020A0238)

Biography: WU Di, male, Master candidate, research direction: image processing. E-mail: wudi2020wudi@163.com

† To whom correspondence should be addressed. E-mail: zt9877@163.com

sparse representation^[3-5], non-local self-similarity^[6,7], and neural network models^[8-10].

Nowadays, sparse representation has become a mainstream direction in image processing. Aharon *et al*^[11] used singular value decomposition to update the basic atoms and corresponding coefficients in a minimum error criterion, and pioneered the K-SVD (singular value decomposition) algorithm with a K-time optimization process in 2006. In dictionary learning and sparse coding, each block is treated as an independent individual, while blocks are generally correlated with each other, which is usually ignored. Mairal *et al*^[12] assumed that the sparse coefficients of similar blocks had the same support mechanism and combined the image redundancy information to improve the K-SVD denoising performance by learning simultaneous sparse coding (LSSC). Dong *et al*^[13] trained principal component analysis (PCA) dictionaries by clustering all image blocks, and for each patch group, the most appropriate local PCA dictionary was selected to encode it, thus called the non-local centralized sparse representation (NCSR) model. This NCSR method introduces the concept of "coding residuals". Zhang *et al*^[14] formally proposed the concept of "group" as the basic processing unit of sparse coding, and proposed a group-based sparse representation (GSR) model to extend the mutually independent patch sparse to group sparse which is applicable to both non-local similarity and local sparsity of images. Zha *et al*^[15] proposed a group sparse residual constraint model with non-local priors (GSRC-NLP) and used soft thresholding to achieve image denoising, which essentially combines the NCSR method and the GSR model to upgrade the patch-based sparse coding residuals to the group-based sparse residuals.

Other than the popular l_1 norm, there are many regularization functions to promote sparsity. Bai *et al*^[16] proposed an adaptive correction term to alleviate the over constraint of the l_1 norm. The l_1/l_2 regularization term is also used to improve sparsity and can alleviate the over-constraint of l_1 norm to some extent^[17]. In addition to this, many non-convex regularization terms have been proposed, such as $l_{1/2}$ norm^[18], l_p norm^[19], and Logarithmic regularization term^[20].

In this paper, we study the ratio of the l_1 and l_2 norms, denoted as l_1/l_2 , to promote the sparsity of group sparse residual. The contributions of the paper are sum-

marized as follows:

First, the l_1/l_2 regularization term is used to reduce the group sparse residual, and then a group sparse residual constraint model with l_1/l_2 minimization is proposed. The proposed model is successfully solved by the alternative direction method of multipliers (ADMM) algorithm.

Second, experimental results on image denoising show that the proposed scheme is feasible and outperforms many state-of-the-art methods in both objective and perceptual quality.

1 Background and Related Work

1.1 Group Sparse Representation

In the framework of GSR method^[14], an image $\mathbf{x}$ with size N is divided into n overlapped patches $\mathbf{x}_i$ of size $\sqrt{b} \times \sqrt{b}$, and each patch is denoted by a vector $\mathbf{x}_i \in \mathbb{R}^{b \times m}$. Then for each patch $\mathbf{x}_i$, m similar matched patches are selected from a $W \times W$ sized searching window by K-nearest neighbor (KNN) algorithm^[21]. All patches in set $\mathcal{S}_i$ are stacked into a matrix $\mathbf{X}_i \in \mathbb{R}^{b \times m}$, i.e., $\mathbf{X}_i = \{\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,m}\}$. This matrix $\mathbf{X}_i$ consisting of patches with similar structures is called a patch group, where $\{\mathbf{x}_{i,j}\}$ denotes the j -th patch in the i -th patch group.

Give a dictionary $\mathbf{D}_i$, which is often learned from each group, such as PCA dictionary^[22]. Each group $\mathbf{X}_i$ can be sparsely represented by solving the following l_1 norm minimization problem:

$$\hat{\mathbf{A}}_i = \arg \min_{\mathbf{A}_i} \left\{ \frac{1}{2} \|\mathbf{Y}_i - \mathbf{D}_i \mathbf{A}_i\|_F^2 + \lambda \|\mathbf{A}_i\|_1 \right\} \quad (2)$$

where $\mathbf{A}_i$ represents the group sparse coefficient of each group $\mathbf{X}_i$, i.e., $\mathbf{X}_i = \mathbf{D}_i \mathbf{A}_i$. $\mathbf{D}_i$ is an adaptive PCA dictionary learned from each patch group.

1.2 Construction of Group Sparse Coefficient Residuals $\mathbf{A}_i - \mathbf{B}_i$

In Ref.[18], the authors pointed out that traditional GSR models fail to faithfully estimate the sparsity of each group due to the degradation of the observed image, and they showed that the quality of image restoration largely depends on the group sparsity residual, which is defined as the difference between each degraded group sparse coefficient $\mathbf{A}_i$ and the corresponding original group sparse coefficient $\mathbf{B}_i$:

$$\mathbf{R}_i = \mathbf{A}_i - \mathbf{B}_i \quad (3)$$

So, they proposed the group sparsity residual constraint model to boost up the accuracy of group sparse coefficient $\mathbf{A}_i$, which can be written as

$$\hat{\mathbf{A}}_i = \arg \min_{\mathbf{A}_i} \left\{ \frac{1}{2} \|\mathbf{Y}_i - \mathbf{D}_i \mathbf{A}_i\|_F^2 + \lambda \|\mathbf{A}_i - \mathbf{B}_i\|_1 \right\} \quad (4)$$

where $\mathbf{B}_i = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_m\}$ is the group sparse coefficient of each group corresponding to the original image. As the original image $\mathbf{x}$ is not available in image denoising, the non-local self-similarity prior is used to estimate $\mathbf{B}_i$. For each group including m non-local similar patches, a good estimation of $\boldsymbol{\beta}_k$ can be computed by the weighted average of each vector $\boldsymbol{\alpha}_j$ in $\mathbf{A}_i$. We have

$$\boldsymbol{\beta}_k = \sum_{j=1}^m \mathbf{w}_j \boldsymbol{\alpha}_j \quad (5)$$

where $\mathbf{A}_i = \mathbf{D}_i^T \mathbf{Y}_i$, $\boldsymbol{\beta}_k$ and $\boldsymbol{\alpha}_j$ represent the k -th and j -th vector of $\mathbf{B}_i$ and $\mathbf{A}_i$, respectively. After this, $\boldsymbol{\beta}_k$ is copied by m times to estimate $\mathbf{B}_i$. We can obtain $\mathbf{B}_i = \{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k, \dots, \boldsymbol{\beta}_k\}$. In Eq. (5), $\mathbf{w}_j$ is set to be inversely proportional to the distance between the target patch $\mathbf{y}_i$ and its similar patch $\mathbf{y}_{i,j}$, i.e.,

$$\mathbf{w}_j = \frac{1}{L} \exp(-\|\mathbf{y}_i - \mathbf{y}_{i,j}\|_2^2 / h)$$

where h is a predefined constant and L is a normalization factor [23].

1.3 l_1/l_2 Minimization

In Ref.[17], the authors employed the l_1/l_2 norm to promote the sparsity and proposed the following model for sparse recovery:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}, \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \quad (6)$$

Two main advantages of l_1/l_2 minimization are scale invariant and parameter free. As the ratio of l_1 and l_2 is non-convex and non-linear, solving model (6) and analyzing the convergence become theoretically difficult. The authors of Ref.[17] applied ADMM for minimizing model (11), and verified the convergence empirically by plotting residual errors and objective functions. In Ref.[17], the authors applied the l_1/l_2 norm on the image gradient for MRI reconstruction problem.

2 The GSRC- l_1/l_2 Model for Image Denoising

In this subsection, we employ the ratio of l_1 and l_2 norm as the regularizer to promote the sparsity of the

group sparse residual mentioned before. The proposed model is

$$\hat{\mathbf{A}}_i = \arg \min_{\mathbf{A}_i} \left\{ \frac{1}{2\lambda_0} \|\mathbf{y} - \mathbf{D}\mathbf{a}\|_2^2 + \sum_{i=1}^n \frac{\|\mathbf{A}_i - \mathbf{B}_i\|_1}{\|\mathbf{A}_i - \mathbf{B}_i\|_2} \right\} \quad (7)$$

where $\mathbf{D}$ represents the dictionary and $\mathbf{a}$ is the sparse coefficient.

As we mentioned before, patch group is the unite of the proposed sparse representation. Supposing $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the corresponding n patch group obtained by block matching is denoted by $\mathbf{X}_i, \mathbf{Y}_i \in \mathbb{R}^{b \times m}$. Let $\mathbf{x} = \mathbf{D}\mathbf{a}$, then Eq. (7) can be rewritten as

$$\hat{\mathbf{A}}_i = \arg \min_{\mathbf{A}_i} \left\{ \frac{1}{2\lambda_0} \|\mathbf{y} - \mathbf{x}\|_2^2 + \sum_{i=1}^n \frac{\|\mathbf{A}_i - \mathbf{B}_i\|_1}{\|\mathbf{A}_i - \mathbf{B}_i\|_2} \right\} \quad (8)$$

By means of Theorem I in Ref.[15], the following equation holds with a very large probability at each iteration.

$$\frac{1}{N} \|\mathbf{x} - \mathbf{y}\|_2^2 = \frac{1}{K} \sum_{i=1}^n \|\mathbf{X}_i - \mathbf{Y}_i\|_F^2 \quad (9)$$

Based on Eq. (9), minimization problem (8) is equivalent to

$$\hat{\mathbf{A}}_i = \arg \min_{\mathbf{A}_i} \sum_{i=1}^n \left\{ \frac{1}{2\lambda} \|\mathbf{Y}_i - \mathbf{X}_i\|_F^2 + \frac{\|\mathbf{A}_i - \mathbf{B}_i\|_1}{\|\mathbf{A}_i - \mathbf{B}_i\|_2} \right\} \quad (10)$$

where $\lambda = \frac{\lambda_0 K}{N}$, $K = b \times m \times n$. Let $\mathbf{X}_i = \mathbf{D}_i \mathbf{A}_i$ and $\mathbf{Y}_i = \mathbf{D}_i \mathbf{S}_i$, $\mathbf{D}_i$ is a PCA based dictionary learned from each group $\mathbf{Y}_i$, so it is orthogonal. By utilizing ADMM and introducing two auxiliary variables $\mathbf{P}_i = \mathbf{A}_i - \mathbf{B}_i$ and $\mathbf{Q}_i = \mathbf{A}_i - \mathbf{B}_i$, Eq. (10) can be transformed into another equivalent constraint form:

$$\begin{cases} \arg \min_{\mathbf{A}_i, \mathbf{P}_i, \mathbf{Q}_i} \sum_{i=1}^n \left(\frac{1}{2\lambda} \|\mathbf{A}_i - \mathbf{S}_i\|_F^2 + \frac{\|\mathbf{P}_i\|_1}{\|\mathbf{Q}_i\|_2} \right) \\ \text{s.t. } \mathbf{P}_i = \mathbf{A}_i - \mathbf{B}_i, \mathbf{Q}_i = \mathbf{A}_i - \mathbf{B}_i \end{cases} \quad (11)$$

The augmented Lagrange function plays a central role in the ADMM method. For the problem (11), the augmented Lagrange function can be written as

$$\begin{aligned} L_{\rho_1, \rho_2}(\mathbf{A}_i, \mathbf{P}_i, \mathbf{Q}_i, \mu_1, \mu_2) &= \frac{\|\mathbf{P}_i\|_1}{\|\mathbf{Q}_i\|_2} + \frac{1}{2\lambda} \|\mathbf{A}_i - \mathbf{S}_i\|_F^2 + \langle \mu_1, \mathbf{P}_i - \mathbf{A}_i + \mathbf{B}_i \rangle \\ &+ \frac{\rho_1}{2} \|\mathbf{P}_i - \mathbf{A}_i + \mathbf{B}_i\|_F^2 \\ &+ \langle \mu_2, \mathbf{Q}_i - \mathbf{A}_i + \mathbf{B}_i \rangle + \frac{\rho_2}{2} \|\mathbf{Q}_i - \mathbf{A}_i + \mathbf{B}_i\|_F^2 \end{aligned} \quad (12)$$

Then, the ADMM iteration goes as follows

$$\mathbf{A}_i^{k+1} = \arg \min_{\mathbf{A}_i} \frac{1}{2\lambda} \|\mathbf{A}_i - \mathbf{S}_i\|_F^2$$

$$+ \frac{\rho_1}{2} \left\| A_i - (P_i^k + B_i + \frac{\mu_1}{\rho_1}) \right\|_F^2 + \frac{\rho_2}{2} \left\| A_i - (Q_i^k + B_i + \frac{\mu_2}{\rho_2}) \right\|_F^2 \quad (13)$$

$$P_i^{k+1} = \arg \min_{P_i} \frac{\|P_i\|_1}{\|Q_i^k\|_2} + \frac{\rho_1}{2} \left\| A_i^{k+1} - (P_i + B_i + \frac{\mu_1}{\rho_1}) \right\|_F^2 \quad (14)$$

$$Q_i^{k+1} = \arg \min_{Q_i} \frac{\|P_i^{k+1}\|_1}{\|Q_i\|_2} + \frac{\rho_2}{2} \left\| A_i^{k+1} - (Q_i + B_i + \frac{\mu_2}{\rho_2}) \right\|_F^2 \quad (15)$$

$$\mu_1^{k+1} = \mu_1^k + \rho_1 (P_i^{k+1} - A_i^{k+1} + B_i) \quad (16)$$

$$\mu_2^{k+1} = \mu_2^k + \rho_2 (Q_i^{k+1} - A_i^{k+1} + B_i) \quad (17)$$

Next, we will discuss the optimization strategies for solving each sub-problem one by one.

2.1 A_i Sub-Problem

The sub-problem of A_i is a least square problem. From the optimal conditions, we can obtain

$$\frac{1}{\lambda} (A_i - S_i) + \rho_1 (A_i - (P_i^k + B_i + \frac{\mu_1}{\rho_1})) + \rho_2 (A_i - (Q_i^k + B_i + \frac{\mu_2}{\rho_2})) = 0$$

The solution of A_i is

$$A_i^{k+1} = \frac{S_i + \lambda \rho_1 (P_i^k + B_i + \frac{\mu_1}{\rho_1}) + \lambda \rho_2 (Q_i^k + B_i + \frac{\mu_2}{\rho_2})}{1 + \lambda \rho_1 + \lambda \rho_2} \quad (18)$$

2.2 P_i Sub-Problem

The P_i sub-problem is equivalent to

$$P_i^{k+1} = \arg \min_{P_i} \|P_i\|_1 + \frac{\rho_1 \|Q_i^k\|_2}{2} \left\| P_i - (A_i^{k+1} - B_i - \frac{\mu_1}{\rho_1}) \right\|_F^2$$

It has a closed form solution via soft shrinkage, i.e.,

$$P_i^{k+1} = \max \left\{ \left| A_i^{k+1} - B_i - \frac{\mu_1}{\rho_1} \right| - \frac{1}{\|Q_i^k\|_2 \rho_1}, 0 \right\} \circ \text{sgn}(A_i^{k+1} - B_i - \frac{\mu_1}{\rho_1}) \quad (19)$$

where $\circ$ represents point-wise product.

2.3 Q_i Sub-Problem

As for the Q_i sub-problem, let $c^k = \|P_i^{k+1}\|_1$, $d^k =$

$A_i^{k+1} - B_i - \frac{\mu_2}{\rho_2}$ and the minimization problem reduces to

$$Q_i^{k+1} = \arg \min_{Q_i} \frac{c^k}{\|Q_i\|_2} + \frac{\rho_2}{2} \|Q_i - d^k\|_F^2 \quad (20)$$

If $d^k = 0$, then any vector Q_i with $\|Q_i\|_2 = \sqrt{\frac{c^k}{\rho_2}}$ is a

solution to the minimization problem. If $c^k = 0$, then $Q_i = d^k$ is the solution. If $c^k \neq 0$, $d^k \neq 0$, by taking derivative of

the objective function with respect to Q_i , we obtain

$$(-\frac{c^k}{\|Q_i\|_2^3} + \rho_2) Q_i = \rho_2 d^k \quad (21)$$

As a result, there exists a positive number τ^k such that $Q_i = \tau^k d^k$. Thus, Eq. (21) reduces to

$$(-\frac{c^k}{(\tau^k)^3 \|d^k\|_2^3} + \rho_2) \tau^k d^k = \rho_2 d^k$$

Given d^k , we denote

$$\eta^k = \|d^k\|_2, M^k = \frac{c^k}{\rho_2 (\eta^k)^3}.$$

Then τ^k is the root of

$$\tau^3 - \tau^2 - M^k = 0$$

Let $F(\tau) = \tau^3 - \tau^2 - M^k$, we can find $F(\tau) = 0$ has only one real root, which has the following closed-form:

$$\tau = \frac{1}{3} + \frac{1}{3} (C^k + \frac{1}{C^k})$$

$$\text{where } C^k = \sqrt[3]{\frac{27M^k + 2 + \sqrt{(27M^k + 2)^2 - 4}}{2}}.$$

In summary, we have

$$Q_i^{k+1} = \begin{cases} e^k, & \text{if } d^k = 0 \\ \tau^k d^k, & \text{if } d^k \neq 0 \end{cases} \quad (22)$$

where e^k is a random vector with the l_2 norm to be $\sqrt[3]{\frac{c^k}{\rho_2}}$.

Finally, the final estimated group patch X_i is computed by $\hat{X}_i = D_i \hat{A}_i$. We obtain the complete image $\hat{x}$ by putting the group back to its original position and averaging the overlapping pixels.

2.4 Settings of Parameters and Iterations

In addition, to obtain better results, we can perform the following several iterations of the denoising process. In the t -th iteration, an iterative regularization strategy is used to update the estimation of the noise variance, and therefore updating y^t . The standard deviation of the noise σ in the t -th iteration is adjusted to

$$\sigma_n^t = \rho \sqrt{(\sigma_n^2 - \|y - \hat{x}^t\|_2^2)}$$

where $\rho > 0$ is a constant.

Since the noise variance changes after each regularization process, the parameter λ , which is used to balance the fidelity and regularization terms, needs to change along with it. After each regularization, the parameter λ of group Y_i is set to

$$\lambda = \frac{\bar{c} 2 \sqrt{2} \sigma_n^2}{\delta_i + \kappa} \quad (23)$$

where δ_i denotes the estimated standard deviation of $A_i - B_i$ and $\bar{c}, \kappa$ are constants. The complete description of the

proposed GSRC- l_1/l_2 model for image denoising employing ADMM framework is presented in Algorithm 1.

Algorithm 1 The proposed GSRC- l_1/l_2 model for image denoising

1. Require: Noisy image $\mathbf{y}$.
 2. Initialization: Input $\hat{\mathbf{x}}^0 = \mathbf{y}$, $\mathbf{y}^0 = \mathbf{y}$, $\mathbf{P}_i^0$, $\mathbf{Q}_i^0$, μ_1 , μ_2 , δ_i , ρ_1 , ρ_2 , σ_n , c , m , L , h , ρ , μ and ε .
 3. For $t = 1$ to Max-Iter
 4. Iterative Regularization $\mathbf{y}' = \hat{\mathbf{x}}^{t-1} + \gamma(\mathbf{y} - \mathbf{y}^{t-1})$
 5. For each patch $\mathbf{y}_i$ in $\mathbf{y}'$ do
 6. Find non-local similar patches from group $\mathbf{Y}_i$.
 7. Obtain dictionary $\mathbf{D}_i$ by $\mathbf{Y}_i$ using PCA.
 8. Update $\mathbf{A}_i$ by computing $\mathbf{A}_i = \mathbf{D}_i^\top \mathbf{Y}_i$.
 9. Update $\mathbf{B}_i$ by computing Eq. (5).
 10. Update λ by computing Eq. (23).
 11. While $k < \text{ADMM-Iter}$ or $\frac{\|\mathbf{A}_i^k - \mathbf{A}_i^{k-1}\|_2}{\|\mathbf{A}_i^k\|_2} > \varepsilon$
 12. Update $\mathbf{A}_i^{k+1}$ by computing Eq. (18).
 13. Update $\mathbf{P}_i^{k+1}$ by computing Eq. (19).
 14. Compute $\mathbf{Q}_i^{k+1}$ by computing Eq. (22).
 15. Update μ_1^{k+1} via (16).
 16. Update μ_2^{k+1} via (17).
 17. Get the estimation: $\hat{\mathbf{X}}_i = \mathbf{D}_i \hat{\mathbf{A}}_i$.
 18. End while
 19. End for
 20. Aggregate $\hat{\mathbf{X}}_i$ to the denoised image $\hat{\mathbf{x}}'$.
 21. End for
 22. Output: The final denoised image $\hat{\mathbf{x}}$.
-

3 Experiments

In this section, we will present extensive experimental results to evaluate the denoising performance of the proposed GSRC- l_1/l_2 model. To evaluate the quality of the recovered images, the peak signal to noise ratio (PSNR) and structural similarity (SSIM) metrics are used. The source codes of all comparison methods were obtained from the original authors, and we used the default parameter settings. For color images, image restoration operators are only performed for the luminance component in this paper.

3.1 Setting Adaptive Parameters of the GSRC- l_1/l_2 Model

In this subsection, we report the performance of the proposed GSRC- l_1/l_2 for image denoising and compare

them with some leading denoising methods, including BM3D^[24], NCSR^[13], PGPD^[25], SAIST^[26] and GSRC-NLP^[15]. The parameter setting of the proposed GSRC- l_1/l_2 model for image denoising is as follows. The size of searching window for block matching is set to 25×25 . All experiments are conducted under the MATLAB 2018a environment on a computer with Intel (R) Core (TM) i7-10750H with 2.60 GHz CPU and NVIDIA GeForce GTX 1650 Ti and Windows 10 platform. In the GSRC- l_1/l_2 model, the size of patch $\sqrt{b} \times \sqrt{b}$ is set to 7×7 for $20 < \sigma \leq 50$. In addition, the parameter settings of the model are different. The parameters $(c, \gamma, \rho_1, \rho_2)$ are $(0.6, 0.1, 0.023, 0.001)$, $(0.7, 0.1, 0.0034, 0.001)$ and $(0.7, 0.1, 0.0034, 0.001)$ set to for $\sigma_n = 30$, $\sigma_n = 40$ and $\sigma_n = 50$.

3.2 Analysis of Results

Note that BM3D, PGPD and SAIST are the GSR-based or NSS-based image denoising methods. NCSR and GSRC-NLP are group sparsity residual based image denoising methods. We test the denoising performance on the presented test images and the highest PSNR is recorded in this paper. The PSNR results for all methods are shown in Table 1.

The proposed GSRC- l_1/l_2 has an average PSNR increase of 0.26dB, 0.32dB, 0.27dB, 0.07dB and 0.03dB compared with BM3D, NCSR, PGPD, SAIST and GSRC-NLP, respectively. In addition, the proposed GSRC- l_1/l_2 has an average SSIM increase of 0.013 3, 0.012 9, 0.014 4, 0.007 2 and 0.003 3 compared with BM3D, NCSR, PGPD, SAIST and GSRC-NLP, respectively. We can observe that the proposed GSRC- l_1/l_2 can obtain better results than other competing methods.

Figures 1 and 2 show the visual comparison of the images Barbara and Mural, respectively. To better compare the details of the reconstructed images, we zoom in the same parts of each image to twice the original size. It can be seen that BM3D, NCSR, SAIST and GSRC-NLP are prone to over smooth the images. PGPD does not protect image details. In contrast, the proposed GSRC- l_1/l_2 is able to preserve local structure of the image more effectively than other competing approaches.

3.3 Empirical Convergence Analysis of GSRC- l_1/l_2 Model

The existing literature on the ADMM convergence requires the existence of one separable function in the objective function, whose gradient is Lipschitz continu-

Table 1 PSNR and SSIM result by denoising methods

Image	Noise level	BM3D		NCSR		PGPD		SAIST		GSRC-NLP		GSRC- l_1/l_2	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
Mural	30	26.48	0.756 6	26.27	0.747 2	26.27	0.751 7	26.36	0.749 1	26.36	0.756 3	26.39	0.759 7
	40	25.21	0.709 3	24.89	0.692 7	25.05	0.710 7	25.09	0.700 1	25.12	0.703 4	25.18	0.707 5
	50	24.25	0.674 1	23.92	0.653 2	24.17	0.665 7	24.15	0.657 9	24.22	0.656 1	24.27	0.667 4
Foreman	30	33.10	0.883 0	32.60	0.883 0	32.73	0.880 3	33.05	0.889 9	33.19	0.894 5	32.91	0.886 3
	40	31.72	0.862 1	31.57	0.873 6	31.43	0.859 9	31.78	0.874 4	31.86	0.876 8	31.86	0.871 9
	50	30.53	0.840 3	30.60	0.856 3	30.32	0.838 2	30.60	0.855 4	30.78	0.860 9	30.90	0.858 5
Kodim	30	27.31	0.797 1	27.37	0.791 5	27.11	0.777 1	27.33	0.784 4	27.48	0.798 6	27.50	0.810 8
	40	25.95	0.669 1	25.79	0.725 5	25.76	0.717 7	25.90	0.722 3	25.98	0.732 1	26.06	0.739 5
	50	25.01	0.682 8	24.92	0.679 1	24.76	0.668 0	24.84	0.669 6	24.94	0.673 7	25.03	0.685 5
Barbara	30	29.05	0.868 5	28.85	0.859 6	28.96	0.856 8	29.39	0.869 5	29.39	0.872 2	29.40	0.875 8
	40	27.43	0.817 9	27.26	0.799 5	27.42	0.807 3	27.95	0.828 2	27.86	0.828 2	27.90	0.832 5
	50	26.05	0.765 9	25.99	0.751 4	26.28	0.761 9	26.69	0.781 0	26.59	0.779 8	26.65	0.784 7
Lena	30	29.04	0.859 3	29.36	0.859 7	29.56	0.860 8	29.61	0.865 4	29.66	0.869 2	29.61	0.870 1
	40	27.62	0.824 7	28.01	0.828 1	28.16	0.827 0	28.23	0.835 5	28.16	0.835 7	28.26	0.837 7
	50	26.41	0.784 1	26.95	0.801 0	27.08	0.795 8	27.11	0.804 8	27.10	0.807 4	27.19	0.806 9
Average		27.68	0.786 3	27.62	0.786 8	27.67	0.785 3	27.87	0.7925	27.91	0.796 3	27.94	0.799 7

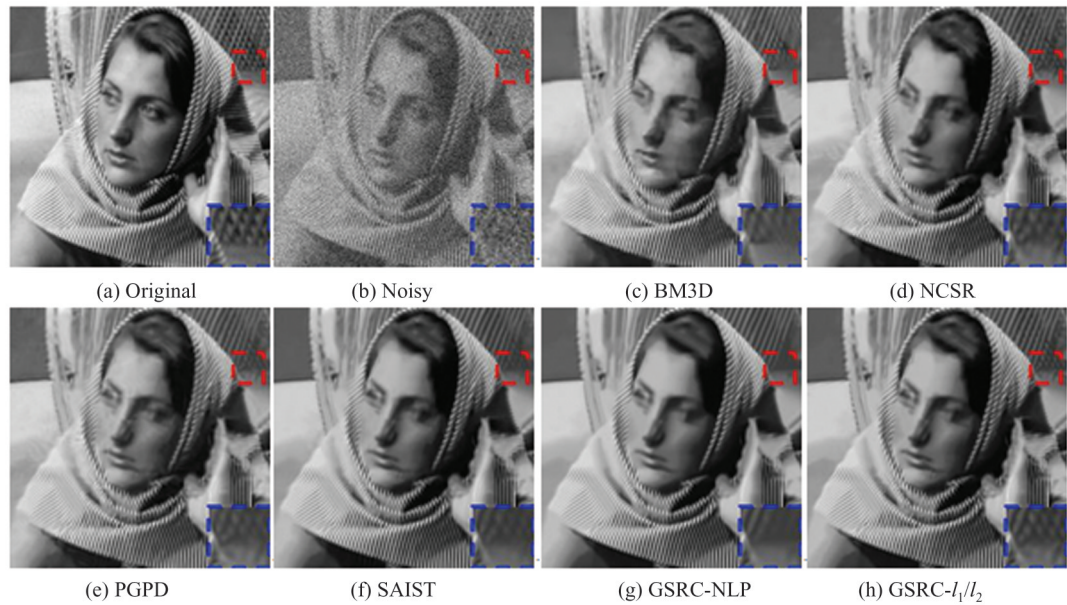


Fig. 1 Denoising results on image Barbara by different methods (noise level $\sigma_n=30$)

ous. However, the proposed GSRC- l_1/l_2 model does not satisfy this assumption. Therefore, it is difficult to analyze the convergence theoretically. Instead, in this subsection we will show the convergence empirically by plotting residual errors and objective functions, which provides strong supports for the convergence validation.

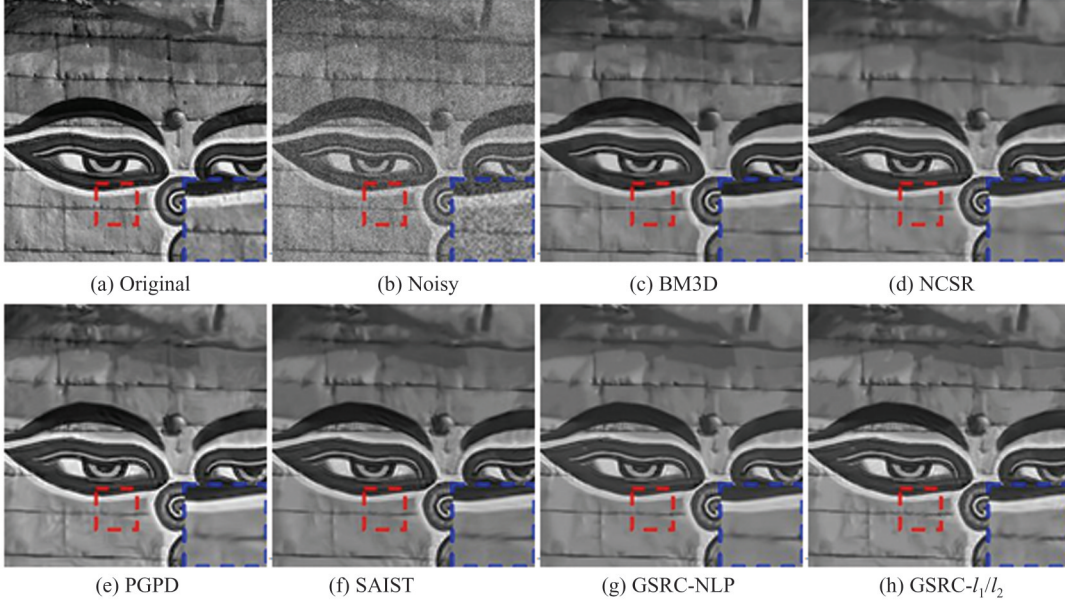


Fig. 2 Denoising results on image Mural by different methods (noise level $\sigma_n = 40$)

In Fig. 3, we empirically demonstrate the convergence of the proposed GSRC- l_1/l_2 algorithm. There are two auxiliary variables P_i and Q_i in l_1/l_2 , i.e., $P_i = A_i - B_i$ and $Q_i = A_i - B_i$. Figure 3(a) shows the values of $\sum_{i=1}^n \|P_i^k - A_i^k + B_i^k\|_2$ and $\sum_{i=1}^n \|Q_i^k - A_i^k + B_i^k\|_2$ with respect to

iteration counter k . Figure 3(b) shows the values of the objective function (11) versus iteration counter k . The constraints and objective functions in Fig. 3 decrease rapidly with respect to the iteration counters, which is the heuristic evidence of the convergence of the Algorithm 1.

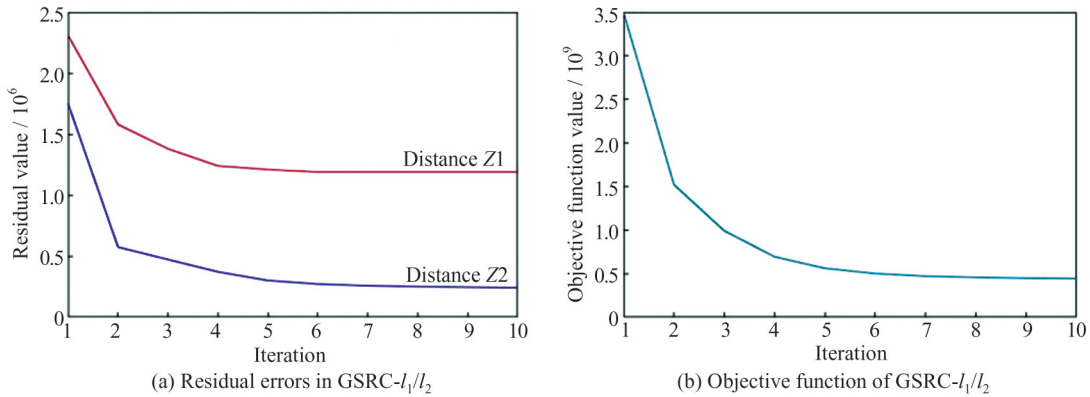


Fig. 3 Plots of residual errors and objective function for empirically demonstrating the convergence

4 Conclusion

To improve the performance of group sparse based image restoration approaches, we introduce an effective sparsity promoting strategies, i.e., l_1/l_2 minimization to promote the sparsity of the group sparse residual. We use the non-local self-similar prior of images along with a self-supervised learning scheme to obtain good estimates of the group sparsity coefficients for each original

group. Then we adapt the l_1/l_2 minimization to reduce the group sparse residual. Compared with l_1 minimization in GSRC-NLP method, the proposed GSRC- l_1/l_2 can lead to better performance in image denoising. We apply the ADMM algorithm to solve the models and explore the convergence of the proposed algorithm. Experimental results show that the model is comparable with the tested denoising methods, and outperforms many state-of-the-art image denoising methods in terms of

both objective and perceptual quality metrics.

References

- [1] Blomgren P, Chan T F. Color TV: Total variation methods for restoration of vector-valued images[J]. *IEEE Transactions on Image Processing*, 1998, **7**(3):304-309.
- [2] Bertalmio M, Caselles V, Rouge B, *et al.* A TV based restoration model with local constraints. [J]. *Journal of Scientific Computing*, 2003, **19**(1):95-122.
- [3] Li H, Liu F. Image denoising via sparse and redundant representations over learned dictionaries in wavelet domain [C]// 2009 *Fifth International Conference on Image and Graphics*. Washington D C: IEEE, 2009: 754-758.
- [4] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration[J]. *IEEE Transactions on Image Processing*, 2007, **17**(1):53-69.
- [5] Aharon M, Elad M, Bruckstein A M. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them[J]. *Linear Algebra & Its Applications*, 2006, **416**(1): 48-67.
- [6] Zha Z, Yuan X, Zhu C, *et al.* Image restoration via simultaneous nonlocal selfsimilarity priors[J]. *IEEE Transactions on Image Processing*, 2020, **29**: 8561-8576.
- [7] Gu S H, Zhang L, Zuo W M, *et al.* Weighted nuclear norm minimization with application to image denoising[C]// 2014 *IEEE Conference on Computer Vision and Pattern Recognition*. Washington D C : IEEE, 2014: 2862-2869.
- [8] Zhang K, Zuo W M, Chen Y J, *et al.* Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising [J]. *IEEE Transactions on Image Processing*, 2016, **26**(7): 3142-3155.
- [9] Dong W S, Wang P Y, Yin W T, *et al.* Denoising prior driven deep neural network for image restoration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41** (10): 2305-2318.
- [10] Zhang K, Zuo W M, Gu S H, *et al.* Learning deep CNN denoiser prior for image restoration[C]// 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington D C : IEEE, 2017: 2808-2817.
- [11] Aharon M, Elad M, Bruckstein. A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Processing*, 2006, **54**(11): 4311-4322.
- [12] Mairal J, Bach F, Ponce J, *et al.* Non-local sparse models for image restoration[C]// 2009 *IEEE 12th International Conference on Computer Vision*. Washington D C : IEEE, 2009: 2272-2279
- [13] Dong W S, Zhang L, Shi G M, *et al.* Nonlocally centralized sparse representation for image restoration[J]. *IEEE Transactions on Image Processing*, 2013, **22**(4):1620-1630.
- [14] Zhang J, Zhao D B, Gao W. Group-based sparse representation for image restoration[J]. *IEEE Transactions on Image Processing*, 2014, **23**(8): 3336-3351.
- [15] Zha Z Y, Yuan X, Wen B H, *et al.* Group sparsity residual constraint with non-local priors for image restoration[J]. *IEEE Transactions on Image Processing*, 2020, **29**: 8960-8975.
- [16] Bai M, Zhang X, Shao Q. Adaptive correction procedure for TVL1 image deblurring under impulse noise[J]. *Inverse Problems*, 2016, **32**(8): 289-338.
- [17] Rahimi Y, Wang C, Dong H B, *et al.* A scale-invariant approach for sparse signal recovery[J]. *SIAM Journal on Scientific Computing*, 2019, **41**(6):A3649-A3672.
- [18] Xu Z B, Chang X Y, Xu F M, *et al.* L-1/2 regularization: A thresholding representation theory and a fast solver[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(7):1013-1027.
- [19] Zuo W M, Meng D Y, Zhang L, *et al.* A generalized iterated shrinkage algorithm for non-convex sparse coding[C]// 2013 *IEEE International Conference on Computer Vision*. Washington D C: IEEE, 2013: 217-224.
- [20] Jia X, Feng X. Bayesian inference for adaptive low rank and sparse matrix estimation[J]. *Neurocomputing*, 2018(5):71-83.
- [21] Keller J M, Gray M R, Givens J A. A fuzzy *k*-nearest neighbor algorithm[J]. *IEEE Transactions on Systems Man & Cybernetics*, 1985, **SMC-15**(4): 580-585.
- [22] Dong W S, Lei Z, Shi G M, *et al.* Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization[J]. *IEEE Transactions on Image Processing*, 2011, **20**(7):1838-1857.
- [23] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]// 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Washington D C: IEEE, 2005, **2**: 60-65.
- [24] Dabov K, Foi A, Katkovnik V, *et al.* Image denoising by sparse 3-D transform-domain collaborative filtering[J]. *IEEE Transactions on Image Processing*, 2007, **16**(8):2080-2095.
- [25] Xu J, Zhang L, Zuo W M, *et al.* Patch group based nonlocal self-similarity prior learning for image denoising[C]// 2015 *IEEE International Conference on Computer Vision (ICCV)*. Washington D C: IEEE, 2016: 244-252.
- [26] Dong W S, Shi G M, Li X. Nonlocal image restoration with bilateral variance estimation: A low-rank approach[J]. *IEEE Transactions on Image Processing*, 2013, **22**(2):700-711.

□