

# INCOHERENT DICTIONARY LEARNING FOR SPARSE REPRESENTATION BASED IMAGE DENOISING

*Jin Wang<sup>1</sup>, Jian-Feng Cai<sup>2</sup>, Yunhui Shi<sup>1</sup> and Baocai Yin<sup>1</sup>*

<sup>1</sup>Beijing Key Laboratory of Multimedia and Intelligent Software Technology,  
College of Metropolitan Transportation, Beijing University of Technology, Beijing, China, 100124

<sup>2</sup>Department of Mathematics, The University of Iowa, Iowa City, IA, USA, 52242

## ABSTRACT

Dictionary learning for sparse representation has been an active topic in the field of image processing. Most existing dictionary learning schemes focus on the representation ability of the learned dictionary. However, according to the theory of compressive sensing, the mutual incoherence of the dictionary is of crucial role in the sparse coding. Thus incoherent dictionary is desirable to improve the performance of sparse representation based image restoration. In this paper, we propose a new incoherent dictionary learning model that minimizes the representation error and the mutual incoherence by incorporating the constraint of mutual incoherence into the dictionary update model. The optimal incoherent dictionary is achieved by seeking an optimization solution. An efficient algorithm is developed to solve the optimization problem iteratively. Experimental results on image denoising demonstrate that the proposed scheme achieves better recovery quality and converges faster than K-SVD while keeping lower computation complexity.

**Index Terms**— Dictionary learning, incoherent, sparse representation, image denoising

## 1. INTRODUCTION

According to the sparse representation theory[1], a signal can be represented by the linear combination of a few typical atoms. The signal can be precisely represented by seeking the most sparse linear representation with a pre-specified dictionary. Such superiority made it widely applied in the area of image restoration and processing such as denoising[2], super-resolution[3], inpainting[4]. As the basis of sparse representation, the choice of dictionary plays an crucial role in the sparse representation problems. Many research has been made to learn a data adapted dictionary to make the signal well presented. [5] first introduces the dictionary learning problem and suggests the basis atoms in the dictionary are comparable with known image filters. [6] develops the first approach to learn an overcomplete dictionary by using probabilistic model of training data. [7] proposes a frame design technique called method of optimal directions(MOD)

and [8] presents the K-SVD algorithm which generalized the K-means clustering process. These methods share the same general idea of iterative alternations between sparse coding and dictionary update. Other related works include [9, 10, 11].

Most existing work on dictionary learning focus on the representation ability of the learned dictionary such as multi-scale representation and adaptation to the data. However, according to the research of [12, 13], the intrinsic property of a dictionary such as coherence has a direct influence on its performance. Thus incoherent dictionary is desirable to improve the performance of sparse representation. Most incoherent dictionary learning scheme attempt to decrease the coherence of current dictionary atoms in the dictionary update step or impose a new step. [14] proposes a scheme to enhance the mutual and cumulative coherence of learned dictionary with the Gram matrix norm in the modified dictionary update step. [15] develops the INK-SVD algorithm by adding a decorrelation step to the existing K-SVD iterative scheme. By decorrelating pairs of atoms in a dictionary in a greedy way, the desired mutual coherence is achieved. [16] improves the work in [15] by incorporating a decorrelation step and a dictionary rotation step into the update step. The target mutual coherence is reached by iterative projection and rotation of the dictionary. Different from the above methods, [17] proposes a sparsity-based orthogonal dictionary learning method to learn an orthogonal dictionary to minimize the mutual incoherence. The scheme greatly simplifies both the sparse coding and dictionary update step with comparable performance in image restoration tasks.

In this paper, we propose an incoherent dictionary learning scheme, which is desirable for sparse representation model of image. Existing incoherent dictionary learning schemes achieve the low mutual incoherence by adding an extra update step to the learned dictionary. Naturally we propose to incorporate the constraint on mutual coherence into the dictionary learning model to learn an incoherent dictionary more efficiently. We introduce a new representation model that minimizes both the representation error and the mutual incoherence. The optimal incoherent dictionary is

achieved by seeking an optimization solution. A Split Bregman and Augmented Lagrangian Method based algorithm is developed to efficiently solve the optimization problem iteratively. Experimental results on the image restoration applications like image denoising demonstrate that the proposed scheme achieves better recovery quality and converges faster than K-SVD while keeping lower computation complexity.

The rest of this paper is organized as follows. Section 2 describes the problem and introduces the optimization model. Section 3 presents the proposed numerical algorithm. Section 4 describes the application of proposed scheme in the image restoration task of image denoising. Implementation details and experimental results are shown in Section 5. Section 6 concludes this paper.

## 2. PROBLEM FORMULATION

Suppose  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \in \mathbb{R}^{m \times n}$  denotes the training image set, and each column of  $\mathbf{Y}$  represents the vector form one image patch in the training set.  $\Phi \in \mathbb{R}^{m \times k}$  is the dictionary to be learned,  $\mathbf{X} \in \mathbb{R}^{k \times n}$  is the matrix of sparse coefficient with each column  $\mathbf{x}_i$  denoting the sparse coefficient for one sample image patch  $\mathbf{y}_i$ . The dictionary learning for sparse representation is usually expressed as:

$$\min_{\Phi, \mathbf{X}} \|\mathbf{Y} - \Phi \mathbf{X}\|_F^2, s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T_0 \quad (1)$$

where  $T_0$  is the threshold of sparse coefficients. The commonly applied strategy to solve this problem is to start with an initial dictionary and alternate between the following two stages until convergence: sparse coding stage where the sparse coefficient  $\mathbf{X}$  is calculated with a given dictionary  $\mathbf{Y}$  and dictionary update stage where the dictionary is updated to minimize the overall representation error with given  $\mathbf{X}$ . In the following sections, we'll follow this commonly deployed method by applying a standard sparse coding step and focus on the dictionary update step.

The mutual coherence  $\mu(\Phi)$  measures the correlation of different atoms in a dictionary, which is defined as[18]:

$$\mu(\Phi) = \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|, \|\phi_s\|_2 = 1, s = 1, \dots, k \quad (2)$$

where  $\langle \phi_i, \phi_j \rangle$  is the inner product of two normalized vectors  $\phi_i$  and  $\phi_j$ . Rather than adding a decorrelation step to the update step in the previous work, we incorporate the mutual incoherence into the objective function as a constraint in the dictionary update step. The optimal dictionary is achieved by seeking an optimization solution that both minimize the representation error and the mutual incoherence. The Gram matrix  $\mathbf{G} = \Phi^T \Phi$  is introduced, as the mutual incoherence  $\mu(\Phi)$  can also be defined as the maximal off-diagonal amplitude:  $\mu = \max_{i \neq j} |G_{ij}|$ . Thus the absolute value of matrix element  $G_{ij}$  represents the coherence between the  $i_{th}$  and  $j_{th}$  columns of dictionary  $\Phi$ , then the constraint on mutual incoherence of  $\Phi$  can be imposed on the matrix  $\mathbf{G}$  as

$|G_{ij}| \leq \mu, i \neq j$  and  $diag(\mathbf{G}) = \mathbf{1}$ . It can be reformed as  $\mathbf{B} \leq \mathbf{G} \leq \mathbf{A}$ , where  $\mathbf{B}$  and  $\mathbf{A}$  are the lower bound and upper bound for  $\mathbf{G}$  respectively. The matrix  $\mathbf{G}$  should be low rank and semi positive definite as  $\mathbf{G} = \Phi^T \Phi$ , which leads to the following optimization model:

$$\min_{\mathbf{G}} \alpha \|\mathbf{G}\|_* + \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2, \quad (3)$$

$$s.t. \mathbf{B} \leq \mathbf{G} \leq \mathbf{A}, \mathbf{G} \succeq 0$$

where  $\|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2$  is the data fidelity term, notice the term is equivalent to  $\|\mathbf{Y} - \Phi \mathbf{X}\|_F^2$ . Once  $\mathbf{G}$  is obtained,  $\Phi$  can be solved by decomposition of  $\mathbf{G}$ .

## 3. NUMERICAL ALGORITHM

As we've mentioned in the previous section, the commonly applied alternative iterate scheme is adopted and the standard sparse coding step is applied. To solve the problem (3), auxiliary variables  $\mathbf{P}$  and  $\mathbf{Q}$  are introduced:

$$\min_{\mathbf{G}, \mathbf{P}, \mathbf{Q}} \alpha \|\mathbf{G}\|_* + \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2, \quad (4)$$

$$s.t. \mathbf{B} \leq \mathbf{P} \leq \mathbf{A}, \mathbf{Q} \succeq 0, \mathbf{P} = \mathbf{G}, \mathbf{Q} = \mathbf{G}$$

The Augmented Lagrangian function of (4) is:

$$\min_{\mathbf{G}, \mathbf{P}, \mathbf{Q}} \alpha \|\mathbf{G}\|_* + \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2 + \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle$$

$$+ \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2 + \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2 \quad (5)$$

$$+ \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle, s.t. \mathbf{B} \leq \mathbf{P} \leq \mathbf{A}, \mathbf{Q} \succeq 0$$

And (5) can be solved by the following iterations:

$$\min_{\mathbf{G}} \alpha \|\mathbf{G}\|_* + \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2 + \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle$$

$$+ \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2 + \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2 \quad (6)$$

$$+ \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle$$

$$\min_{\mathbf{P}} \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2 + \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle, s.t. \mathbf{B} \leq \mathbf{P} \leq \mathbf{A} \quad (7)$$

$$\min_{\mathbf{Q}} \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2 + \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle, s.t. \mathbf{Q} \succeq 0 \quad (8)$$

$$\mathbf{S}_{k+1} = \mathbf{S}_k + \mu_1 (\mathbf{G} - \mathbf{P}) \quad (9)$$

$$\mathbf{T}_{k+1} = \mathbf{T}_k + \mu_2 (\mathbf{G} - \mathbf{Q}) \quad (10)$$

(6) can be solved by splitting, after introducing auxiliary variable  $\mathbf{N}$  it can be expressed as:

$$\min_{\mathbf{G}, \mathbf{N}} \alpha \|\mathbf{N}\|_* + \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2 + \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle$$

$$+ \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2 + \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2 \quad (11)$$

$$+ \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle + \mu_3 \|\mathbf{G} - (\mathbf{N} + \mathbf{L})\|_F^2$$

which can be splitted into the following two sub-problems:

$$\min_{\mathbf{G}} \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2 + \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2$$

$$+ \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle + \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2 \quad (12)$$

$$+ \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle + \mu_3 \|\mathbf{G} - (\mathbf{N} + \mathbf{L})\|_F^2$$

---

**Algorithm 1** INCOHERENT DICTIONARY UPDATE

---

**Input:** Training image data  $\mathbf{Y}$  and sparse coefficients  $\mathbf{X}$ .**Output:** Learned dictionary  $\Phi$ .

- 1: Initialization:  $\mathbf{S}; \mathbf{T}; \mathbf{N}; \mathbf{L}; \mathbf{P}; \mathbf{Q}; \alpha; \beta; \mu_1; \mu_2; \mu_3; \epsilon; k = 0$ .
  - 2: **while**  $\frac{\|\mathbf{G}_{k+1} - \mathbf{G}_k\|_F}{\|\mathbf{G}_k\|_F} > \epsilon$  **do**
  - 3:
$$\mathbf{G}_{k+1} = \min_{\mathbf{G}} \beta \|\mathbf{Y}^T \mathbf{Y} - \mathbf{X}^T \mathbf{G} \mathbf{X}\|_F^2 + \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2$$
$$+ \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle + \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2$$
$$+ \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle + \mu_3 \|\mathbf{G} - (\mathbf{N} + \mathbf{L})\|_F^2$$
  - 4:
$$\mathbf{N}_{k+1} = \min_{\mathbf{N}} \alpha \|\mathbf{N}\|_* + \mu_3 \|\mathbf{G} - (\mathbf{N} + \mathbf{L})\|_F^2$$
  - 5:
$$\mathbf{L}_{k+1} = \mathbf{L}_k - (\mathbf{G} - \mathbf{N})$$
  - 6:
$$\mathbf{P}_{k+1} = \min_{\mathbf{P}} \mu_1 \|\mathbf{G} - \mathbf{P}\|_F^2 + \langle \mathbf{S}, \mathbf{G} - \mathbf{P} \rangle, s.t. \mathbf{B} \leq \mathbf{P} \leq \mathbf{A}$$
  - 7:
$$\mathbf{Q}_{k+1} = \min_{\mathbf{Q}} \mu_2 \|\mathbf{G} - \mathbf{Q}\|_F^2 + \langle \mathbf{T}, \mathbf{G} - \mathbf{Q} \rangle, s.t. \mathbf{Q} \succeq 0$$
  - 8:
$$\mathbf{S}_{k+1} = \mathbf{S}_{k+1} = \mathbf{S}_k + \mu_1 (\mathbf{G} - \mathbf{P})$$
  - 9:
$$\mathbf{T}_{k+1} = \mathbf{T}_{k+1} = \mathbf{T}_k + \mu_2 (\mathbf{G} - \mathbf{Q})$$
  - 10: **end while**
  - 12: Decomposition of  $\mathbf{G}: \Phi^T \Phi = \mathbf{G}$ .
  - $$\mathbf{W} = \min_{\mathbf{W}} \|\mathbf{Y} - \mathbf{W} \Phi \mathbf{X}\|_F^2, s.t. \mathbf{W}^T \mathbf{W} = \mathbf{I}$$
  - 13: Post processing:  $\Phi = \mathbf{W} \Phi$ .
- 

$$\min_{\mathbf{N}} \alpha \|\mathbf{N}\|_* + \mu_3 \|\mathbf{G} - (\mathbf{N} + \mathbf{L})\|_F^2 \quad (13)$$

And  $\mathbf{L}_{k+1} = \mathbf{L}_k - (\mathbf{G} - \mathbf{N})$ .

Since the objective function of (12) is strict convex, the global optimum of  $\mathbf{G}$  can be achieved by taking the derivative and setting it as zero. And the low rank recovery problem (13) can be efficiently solved by the singular value thresholding algorithm [19]. (7) can be transformed into a quadratic program (QP) which can be solved with a standard QP solver. (8) is equivalent to the following problem:

$$\min_{\mathbf{Q}} \mu_2 \|\mathbf{G} - \mathbf{Q} + 1/(2\mu_2) \mathbf{T}\|_F^2, s.t. \mathbf{Q} \succeq 0 \quad (14)$$

which can be solved by thresholding the eigenvalue of matrix  $\mathbf{G} + 1/(2\mu_2) \mathbf{T}$  and keeping the positive ones.

When the Gram matrix  $\mathbf{G}$  is obtained,  $\Phi$  can be got by applying matrix decomposition to  $\mathbf{G}$ . Since the decomposition is not unique, a post processing step is introduced as called rotation in [16]. Denote  $\mathbf{W}$  as an orthogonal matrix, we have  $(\mathbf{W} \Phi)^T (\mathbf{W} \Phi) = \Phi^T \Phi = \mathbf{G}$ . Thus an orthogonal matrix  $\mathbf{W}$  can be applied on  $\Phi$  to minimize the representation error, which can be represented as:

$$\min_{\mathbf{W}} \|\mathbf{Y} - \mathbf{W} \Phi \mathbf{X}\|_F^2, s.t. \mathbf{W}^T \mathbf{W} = \mathbf{I} \quad (15)$$

where  $\mathbf{I}$  is a identity matrix. And  $\mathbf{W}$  can be solved by using SVD on  $\Phi \mathbf{X} \mathbf{Y}^T$ , that is  $\Phi \mathbf{X} \mathbf{Y}^T = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ , and the orthogonal matrix to minimize (15) is  $\mathbf{W} = \mathbf{V} \mathbf{U}^T$ . And the whole algorithm is as described in algorithm 1. To accelerate the convergence, a continuation strategy is applied:

$\mu_{k+1} = \rho * \mu_k$ , where  $\rho$  is a pre-defined constant. Thus the whole incoherent dictionary update can be solved with mainly several SVD operations and one EVD operation. It's more efficient than K-SVD, which requires an SVD operation when update each column of the dictionary, and SVD operation is time consuming especially for large scale matrix. So the proposed scheme is more computational efficient than K-SVD, especially for large scale dictionary learning problems.

#### 4. APPLICATION IN IMAGE DENOISING

To evaluate the performance of the learned incoherent dictionary for sparse representation based image restoration, the incoherent dictionary learning scheme is applied in image denoising. As indicated in [2], the dictionary can be learned from noisy images by extracting overlapping image patches with size  $\sqrt{m} \times \sqrt{m}$  uniformly at random as training data  $\mathbf{Y}$ . Applying the Algorithm 1 on the training data will produce an incoherent dictionary  $\Phi$  of size  $m \times k$ , with which the noiseless version of each noisy image patch  $\mathbf{p}$  can be sparsely represented. For a noisy patch  $\mathbf{p}$ , the sparse coefficient  $\hat{\alpha}$  can be obtained by solving a  $l_0$  minimization problem. Many existing algorithms can be used to solve the sparse coding problem. The corresponding clean patch can be represented as:  $\hat{\mathbf{p}} = \Phi \hat{\alpha}$ . And the final denoised image can be reconstructed by averaging the overlapping patches to avoid visual artifacts on block boundaries. See Algorithm 2 for more details.

---

**Algorithm 2** DENOISING VIA LEARNED DICTIONARY

---

**Input:** Noisy image  $\mathbf{I}$  with additive Gaussian white noise, standard deviation  $\sigma$  of noise, patch size  $n$ .**Output:** Denoised image  $\hat{\mathbf{I}}$ .

- 1: Initialization: extract image patches to form training data  $\mathbf{Y}$  in Algorithm 1, error term  $\epsilon = C\sigma$ ,
  - 2: Learn a incoherent dictionary  $\Phi$  using Algorithm 1 with  $\mathbf{Y}$ .
  - 4: Denoise each patch  $\mathbf{p}$  in  $\mathbf{I}$ :
$$\hat{\alpha} = \min_{\alpha} \|\alpha\|_0, s.t. \|\Phi \alpha - \mathbf{p}\|_2^2 \leq \epsilon^2$$
  - 5: Denoised patch  $\hat{\mathbf{p}} = \Phi \hat{\alpha}$ .
  - 6: Averaging overlapping patches to synthesize denoised image  $\hat{\mathbf{I}}$ .
- 

#### 5. EXPERIMENTAL RESULTS

Extensive experiments have been carried out to test the performance of the incoherent dictionary learning scheme on a set of standard test images as shown in Fig.1. In our implementation, the image patches for training are of size  $8 \times 8$ . About  $6 \times 10^4$  patches are extracted from a  $512 \times 512$  image uniformly at random as training data. The DCT transform is adopted as the initial dictionary. For the sparse coding problem both in dictionary update and image denoising as in Algorithm 1 and Algorithm 2, the orthonormal matching pursuit (OMP) [20] is applied for its efficiency. The error term  $\epsilon$  in Algorithm 2 depends on the standard deviation of noise,

**Table 1.** PSNR values comparison of denoised images with different methods

Noise Level	Methods	Barbara	Lena	House	Peppers	Fingerprint	Boat	Average
$\sigma = 10$	Overcomplete DCT	33.99	35.29	35.43	33.91	32.15	33.45	34.04
	Global trained dictionary	33.08	35.42	35.66	34.33	32.26	33.54	34.05
	K-SVD $8 \times 8$	34.48	35.52	35.99	34.27	32.33	33.67	34.38
	Proposed $8 \times 8$	34.59	35.59	36.19	34.49	32.36	33.74	<b>34.49</b>
$\sigma = 20$	Overcomplete DCT	29.95	31.99	32.20	30.15	28.01	29.87	30.36
	Global trained dictionary	28.88	32.26	32.91	30.86	28.22	30.21	30.56
	K-SVD $8 \times 8$	30.89	32.45	33.40	30.84	28.49	30.37	31.07
	Proposed $8 \times 8$	30.94	32.47	33.50	30.95	28.50	30.42	<b>31.13</b>
$\sigma = 30$	Overcomplete DCT	27.51	30.04	30.18	27.96	25.51	27.91	28.19
	Global trained dictionary	26.49	30.39	31.05	28.84	25.88	28.32	28.50
	K-SVD $8 \times 8$	28.51	30.57	31.56	28.91	26.37	28.50	29.07
	Proposed $8 \times 8$	28.56	30.61	31.62	28.95	26.38	28.54	<b>29.11</b>
$\sigma = 40$	Overcomplete DCT	25.98	28.50	28.7	26.38	23.53	26.54	26.61
	Global trained dictionary	25.10	28.87	29.37	27.27	24.11	26.92	26.94
	K-SVD $8 \times 8$	26.93	28.97	29.63	27.37	24.73	27.06	27.45
	Proposed $8 \times 8$	26.97	29.01	29.66	27.41	24.76	27.10	<b>27.49</b>
$\sigma = 50$	Overcomplete DCT	24.74	27.50	27.53	25.35	22.04	25.60	25.46
	Global trained dictionary	24.09	27.82	28.05	26.19	22.72	25.92	25.80
	K-SVD $8 \times 8$	25.42	27.89	28.09	26.07	23.31	26.02	26.13
	Proposed $8 \times 8$	25.49	27.91	28.10	26.10	23.34	26.06	<b>26.17</b>

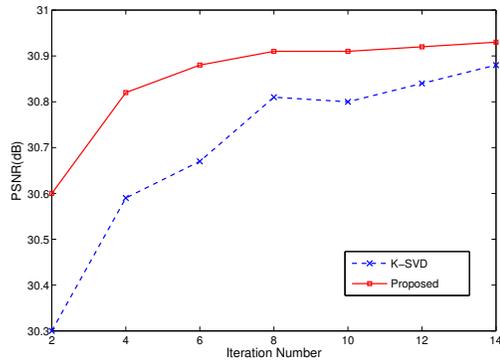
which can be described as  $\epsilon = C\sigma$ . And  $C$  is empirically chosen to be 1.15 as in [2]. The dictionary size is  $64 \times 256$  in our experiments to handle the  $8 \times 8$  image patch. The parameters in Algorithm 1 are empirically chosen as  $\alpha = 0.02$ ,  $\beta = 0.2$ ,  $\mu_1 = 0.1$ ,  $\mu_2 = 0.1$  and  $\mu_3 = 0.1$ . Different levels of white Gaussian noise with standard deviation  $\sigma$  is added to the test images, and the denoised result of our scheme is compared with other schemes based on overcomplete DCT, a global trained dictionary and K-SVD. The iteration number is 15 as more iterations won't improve the PSNR value. The PSNR values is listed in Table 1. Also the convergence speed is compared between our scheme and K-SVD. As shown in Fig.2, our scheme converges much faster than K-SVD, especially at the beginning of iterations.



**Fig. 1.** Test image set

**6. CONCLUSION**

In this paper, we propose a new incoherent dictionary learning scheme for sparse representation based image restoration.



**Fig. 2.** Convergence speed comparison between our scheme and K-SVD (Barbara,  $\sigma = 20$ )

A new representation model is introduced by incorporating the constraint on mutual incoherence into the dictionary update. And an efficient algorithm is developed to solve this optimization problem. The evaluation of its application in image denoising validate its efficiency. Experimental results on the standard test images demonstrate that the proposed scheme has better recovery quality than K-SVD while have lower computation complexity, which is desirable in sparse representation based image restoration applications.

**7. ACKNOWLEDGEMENT**

This work was supported by the Natural Science Foundation of China under Grant No. 61390510, 61033004, 61170103, 61370118, 61227004, 61370120, and Doctoral Fund of Innovation of Beijing University of Technology.

## 8. REFERENCES

- [1] Stéphane G Mallat and Zhifeng Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [2] Michael Elad and Michal Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [3] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, “Image super-resolution via sparse representation,” *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [4] Cewu Lu, Jiaping Shi, and Jiaya Jia, “Online robust dictionary learning,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 415–422.
- [5] Bruno A Olshausen and David J Field, “Sparse coding with an overcomplete basis set: A strategy employed by v1?,” *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [6] Michael S Lewicki and Terrence J Sejnowski, “Learning overcomplete representations,” *Neural computation*, vol. 12, no. 2, pp. 337–365, 2000.
- [7] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy, “Method of optimal directions for frame design,” in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1999, vol. 5, pp. 2443–2446.
- [8] Michal Aharon, Michael Elad, and Alfred Bruckstein, “K-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [9] Sylvain Lesage, Rémi Gribonval, Frédéric Bimbot, and Laurent Benaroya, “Learning unions of orthonormal bases with thresholded singular value decomposition,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’05)*. IEEE, 2005, vol. 5, pp. v–293.
- [10] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad, “Dictionaries for sparse representation modeling,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.
- [11] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, “Online learning for matrix factorization and sparse coding,” *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.
- [12] Joel A Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [13] Alfred M Bruckstein, David L Donoho, and Michael Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.
- [14] Ignacio Ramirez, Federico Lecumberry, and Guillermo Sapiro, “Sparse modeling with universal priors and learned incoherent dictionaries,” Tech. Rep., Tech Report, IMA, University of Minnesota, 2009.
- [15] Boris Mailhé, Daniele Barchiesi, and Mark D Plumbley, “Ink-svd: Learning incoherent dictionaries for sparse representations,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 3573–3576.
- [16] Daniele Barchiesi and Mark D Plumbley, “Learning incoherent dictionaries for sparse approximation using iterative projections and rotations,” *IEEE Transactions on Signal Processing*, vol. 61, pp. 2055–2065, 2013.
- [17] Chenglong Bao, Jian-Feng Cai, and Hui Ji, “Fast sparsity-based orthogonal dictionary learning for image restoration,” in *ICCV*, 2013.
- [18] David L Donoho, Michael Elad, and Vladimir N Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [19] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [20] Yagyensh Chandra Pati, Ramin Rezaifar, and PS Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*. IEEE, 1993, pp. 40–44.