

# A NEW FRAMEWORK FOR DESIGNING INCOHERENT SPARSIFYING DICTIONARIES

Gang Li,<sup>1</sup> Zhihui Zhu,<sup>2</sup> Huang Bai,<sup>3</sup> and Aihua Yu<sup>3</sup>

<sup>1</sup> School of Automation & EE, Zhejiang Univ. of Sci. & Tech., Hangzhou, Zhejiang, P.R. China

<sup>2</sup> Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO, USA

<sup>3</sup> College of Information Engr., Zhejiang University of Technology, Hangzhou, Zhejiang, P.R. China

## ABSTRACT

This paper deals with designing incoherent sparsifying dictionaries. A new framework is proposed, in which the sparse representation error and mutual coherence are embedded. An alternating minimization method is developed for solving the optimal dictionary problem. One of the significant features of the proposed approach is that the dictionary is directly updated with each atom being normalized. A gradient-based algorithm is derived for this purpose. Experiments are carried out and the results show that the proposed approach outperforms some prevailing ones in terms of minimizing sparse representation error and mutual coherence.

**Index Terms**— Dictionary learning, equiangular tight frame, incoherent dictionary, sparse representation

## 1. INTRODUCTION

Sparse and redundant representations look for the best approximation of a signal vector with a linear combination of few atoms from an over-complete set of well-designed vectors [1]. This topic is closely related to the *sparsifying dictionary learning* [2], [3] and the compressed sensing (CS) [4] - [6], in which the sparsity of the signals to be recovered is a prerequisite.

The  $l_p$ -norm of  $v \in \mathbb{R}^{N \times 1}$  is defined as<sup>1</sup>

$$\|v\|_p \triangleq \left( \sum_{n=1}^N |v(n)|^p \right)^{1/p}, \quad p \geq 1. \quad (1)$$

For convenience,  $\|v\|_0$  is used to denote the number of non-zero elements in  $v$ . Let  $y \in \mathbb{R}^{N \times 1}$  be the signal under consideration and assume

$$y = \sum_{l=1}^L s_l \psi_l \triangleq \Psi s \quad (2)$$

---

This work was supported by the Grants of NSFCs 61273195 & 61473262. Emails: ielgang@zjut.edu.cn, zhu@mines.edu.

<sup>1</sup>Throughout this paper, MATLAB notations are adopted:  $v(n)$  denotes the  $n$ th entry of the vector  $v$ ;  $Q(m, :)$ ,  $Q(:, k)$  and  $Q(i, j)$  respectively denote the  $m$ th row,  $k$ th column, and  $(i, j)$ th entry of the matrix  $Q$ .

where  $\Psi \in \mathbb{R}^{N \times L}$  is usually called a *dictionary* and is said *over-complete* if  $N < L$  and  $s$  is the coefficient vector. Such a model, referred to as *synthesis model*, has been vastly used in many applications such as denoising and compression.

The optimal vector  $s$  for a given  $y$  is determined with

$$s = \arg \min_{\tilde{s}} \|y - \Psi \tilde{s}\|_2^2 \text{ s.t. } \|\tilde{s}\|_0 \leq \kappa \quad (3)$$

where the dictionary  $\Psi$  is assumed to be given and  $\kappa$  is a prescribed sparsity level. The problem defined in (3) can be solved using the orthogonal matching pursuit (OMP) [7]-[9].

Let  $Y = [ y_1 \cdots y_m \cdots y_J ] \in \mathbb{R}^{N \times J}$  be the data matrix formed by a collection of  $J$  observed signal vectors. The traditional sparsifying dictionary learning can be stated mathematically as

$$\left\{ \begin{array}{l} (\Psi, S) \triangleq \arg \min_{\tilde{\Psi} \in \mathbb{R}^{N \times L}, \tilde{S} \in \mathbb{R}^{L \times J}} \|Y - \tilde{\Psi} \tilde{S}\|_F^2 \\ \text{s.t. } \|\tilde{S}(:, m)\|_0 \leq \kappa, \forall m \end{array} \right. \quad (4)$$

where  $\|\cdot\|_F$  denotes the *Frobenius norm*.

By nature, sparsifying dictionary learning is a matrix factorization problem [10]-[12]:  $Y \approx \Psi S$ . There exists an infinite number of solutions to (4). No analytical solutions have been found so far and given that (4) is highly non-convex, a practical approach used to attack such a problem is the alternating minimization strategy [13], leading to a class of algorithms [14] - [17] such as the method of optimal direction (MOD) [14] and the K-singular value decomposition (K-SVD) [15].

The approaches mentioned above focus on minimizing the sparse representation error  $\|Y - \tilde{\Psi} \tilde{S}\|_F^2$ . The concept of mutual coherence of matrix  $A$ , denoted as  $\mu(A)$ , plays an important role in numerical analysis. On the one hand, it was shown in [18] that a dictionary  $\Psi$  has well conditioned sub-matrices formed with columns of  $\Psi$  when  $\mu(\Psi)$  is small. On the other hand, as shown in [2], [7], any  $\kappa$ -sparse coefficient vector  $s_0$  can be exactly recovered from the observation/measurement  $y = \Psi s_0$  via (3) as long as

$$\kappa < \frac{1}{2} \left[ 1 + \frac{1}{\mu(\Psi)} \right]. \quad (5)$$

(5) suggests that a dictionary with small mutual coherence can enlarge the signal space in which the coefficient vector  $s$  can be achieved exactly. Therefore, it is desired to design the dictionary with mutual coherence taken into account. This leads to the so-call *incoherent sparsifying dictionary learning* (ISDL). Recently, an alternative formulation was proposed for ISDL in [19], where the two-stage alternating minimization approach is adopted, consisting of the sparse coding followed by a dictionary update. To the latter, two separated steps are employed: i) *atoms decorrelation*; ii) *dictionary rotation*. More details of this approach will be given in the next section.

Note that in the approach proposed in [19] the atoms decorrelation and minimization of sparse representation error are performed independently as they are under two different frameworks, making the algorithm converge slowly or may even diverge. Besides, simulations showed that such an approach leads to a quite large sparse representation error due to too much emphasis made on the mutual coherence reduction. In this paper, we propose a new framework for ISDL, under which both sparse representation error and mutual coherence of the dictionary are embedded in the same cost function. An algorithm for solving the optimal dictionary design is derived with guaranteed convergence. Simulations show that the dictionary obtained using the proposed approach outperforms those by the traditional MOD and K-SVD as well as the method proposed in [19] in terms of reducing mutual coherence and sparse representation error.

## 2. PROBLEM FORMULATION

The mutual coherence of a matrix  $A \in \mathbb{R}^{N \times L}$  is defined as

$$\mu(A) \triangleq \max_{1 \leq i \neq j \leq L} \left\{ \frac{|A(:, i)^T A(:, j)|}{\|A(:, i)\|_2 \|A(:, j)\|_2} \right\} \quad (6)$$

where  $\mathcal{T}$  represents the transpose. Roughly speaking,  $\mu(A)$  measures the maximum linear dependency possibly achieved by any two columns of matrix  $A$  and it can be shown [20] that

$$\underline{\mu} \triangleq \sqrt{\frac{L - N}{N(L - 1)}} \leq \mu(A) \leq 1 \quad (7)$$

where  $\underline{\mu}$  is the Welch bound of matrix  $A$ .

Let  $A$  be a unit-norm frame (i.e.,  $\|A(:, l)\|_2 = 1, \forall l$ ). Such a frame is said *equiangular* if  $|A(:, l)^T A(:, j)| = c, \forall l \neq j$ , where  $c$  is some positive constant. An equiangular tight frame (ETF) is a unit-norm frame which is tight and equiangular [20]. Furthermore, a unit-norm frame  $A$  achieves  $\mu(A) = \underline{\mu}$  if and only if  $A$  is an ETF [20].

Based on the fact that an ETF achieves the minimal mutual coherence, the optimal ISDL was formulated in [19] as (4) but with an extra constraint  $\mu(\tilde{\Psi}) \leq \mu_0$ , where  $\mu_0 \geq \underline{\mu}$  is a fixed target mutual coherence level and such a problem was

addressed in [19] with the algorithm outlined below:

### **Alg <sub>$\Psi$</sub> <sup>IPR</sup>- ISDL using iterative projections and rotations**

*Initialization:* Set the training data matrix  $Y$ , the number of iterations  $N_{ite}$  and the parameter  $\mu_0$ ; Initiate  $\tilde{\Psi}$  with a randomly generated  $\Psi_0 \in \mathbb{R}^{N \times L}$ .

*Begin:*  $k = 1, 2, \dots, N_{ite}$

- *Step I: Sparse coding* - to update  $S$  via

$$\begin{cases} S_k \triangleq \arg \min_{\tilde{S}} \|Y - \Psi_{k-1} \tilde{S}\|_F^2 \\ \text{s.t. } \|\tilde{S}(:, m)\|_0 \leq \kappa, \forall m \end{cases} \quad (8)$$

which can be solved by a greedy algorithm like OMP.

- *Step II: Dictionary update* - to update  $\Psi$  via

$$\tilde{\Psi}_k \triangleq \arg \min_{\tilde{\Psi} \in \mathbb{R}^{N \times L}} \|Y - \tilde{\Psi} S_k\|_F^2.$$

- *Step III: Atoms decorrelation* using

$$\hat{\Psi}_k \triangleq \arg \min_{\tilde{\Psi} \in \mathbb{R}^{N \times L}} \|\tilde{H} - \tilde{\Psi}^T \tilde{\Psi}\|_F^2 \quad (9)$$

subject to  $\tilde{H} = \arg \min_{H \in \mathcal{S}_H^{etf}} \|H - \tilde{\Psi}_k^T \tilde{\Psi}_k\|_F^2$  with  $\mathcal{S}_H^{etf}$  being the set of relaxed ETF Grams:

$$\mathcal{S}_H^{etf} \triangleq \{H \in \mathbb{R}^{L \times L} : H = H^T, H(l, l) = 1, \forall l, \max_{i \neq j} |H(i, j)| \leq \xi\}. \quad (10)$$

- *Step VI: Dictionary rotation* to complete the dictionary update with  $\Psi_k = V \hat{\Psi}_k$ , where

$$V \triangleq \arg \min_{\tilde{V}} \|Y - \tilde{V} \hat{\Psi}_k S_k\|_F^2 \text{ s.t. } \tilde{V}^T \tilde{V} = I_N. \quad (11)$$

*End*

*Outputs:*  $\Psi = \Psi_{N_{ite}}$  and  $S = S_{N_{ite}}$ .

As mentioned in the previous section, **Alg <sub>$\Psi$</sub> <sup>IPR</sup>** has some weak points. The main problem in this algorithm is that the atoms decorrelation and sparse representation error reduction are not considered in the same framework. This motivates us to consider an alternative approach in which the cost function takes the two into account in the same framework.

## 3. AN ALTERNATIVE FRAMEWORK FOR ISDL

The problem we encounter here is actually to design such a dictionary  $\Psi$  that yields a small sparse representation error and has its atoms incoherent. To deal with such a multi-target problem, we consider the following measure for ISDL:

$$\varrho(\Psi, S, H) \triangleq (1 - \beta) \|Y - \Psi S\|_F^2 + \beta \|H - \Psi^T \Psi\|_F^2 \quad (12)$$

where  $H \in \mathcal{S}_H^{etf}$  and  $0 \leq \beta \leq 1$  is a weighting factor to balance the importance of the two terms.

The corresponding optimal ISDL is then formulated as

$$\begin{aligned} (\Psi, S, H) &\triangleq \arg \min_{\tilde{\Psi}, \tilde{S}, \tilde{H}} \varrho(\tilde{\Psi}, \tilde{S}, \tilde{H}) \\ \text{s.t. } &\tilde{H} \in \mathcal{S}_H^{etf}, \|\tilde{S}(:, m)\|_0 \leq \kappa, \forall m \\ &\|\tilde{\Psi}(:, l)\|_2 = 1, \forall l. \end{aligned} \quad (13)$$

**Remark 3.1:** It should be pointed out that the  $l_2$ -based normalization in (13) is used for the term  $\|H - \Psi^\top \Psi\|_F^2$  to have the intended physical meaning of coherence difference between the target Gram and that of the dictionary.

Now, let us consider the optimal dictionary design problem (13). As realized,  $\varrho(\tilde{\Psi}, \tilde{S}, \tilde{H})$  is a highly non-convex function of  $\tilde{\Psi}$ ,  $\tilde{S}$ , and  $\tilde{H}$ . To address such a problem, we propose the following alternating minimization-based algorithm that has been popularly adopted in sparse dictionary learning:

#### **Alg <sub>$\Psi$</sub> <sup>GSD</sup> - Gradient-based ISDL**

*Initialization:* With a normalized dictionary  $\Psi_0$  randomly generated, set a prescribed iteration number  $N_{ite}$ .

*Begin For*  $k = 1 : N_{ite}$

- *Step I:* With  $\tilde{\Psi} = \Psi_{k-1}$ , update  $H$  with the solution of (13) which is equivalent to

$$H_k \triangleq \arg \min_{\tilde{H}} \|\tilde{H} - G_{k-1}\|_F^2 \quad \text{s.t. } \tilde{H} \in \mathcal{S}_H^{etf}$$

where  $G_{k-1} \triangleq \Psi_{k-1}^\top \Psi_{k-1}$ . The solution to the above is obtained by applying the following *shrinkage* operation [13] to  $G_{k-1}$ :

$$H_k(i, j) = \begin{cases} \tau, & i \neq j, |\tau| \leq \xi \\ \text{sign}(\tau)\xi, & i \neq j, |\tau| \geq \xi \end{cases} \quad (14)$$

where  $\tau = G_{k-1}(i, j)$ ,  $\text{sign}(\cdot)$  is the sign function, and  $\xi$  is the parameter characterizing the space of relaxed ETF Grams via (10).

- *Step II:* With  $\tilde{\Psi} = \Psi_{k-1}$  and the obtained  $H = H_k$ ,  $S$  is updated with the solution of (13), which is actually equivalent to the standard sparse coding (8).
- *Step III:* With  $H_k$  and  $S_k$  obtained above, update the dictionary  $\Psi$  with  $\Psi_k$ , the solution of

$$\min_{\tilde{\Psi}} \varrho(\tilde{\Psi}, S_k, H_k) \quad \text{s.t. } \|\tilde{\Psi}(:, l)\|_2 = 1, \forall l. \quad (15)$$

*End For loop*

*Output:*  $\Psi = \Psi_{N_{ite}}$ .

Here, we present an approach to attack (15). The key to the success of this approach is the following parametrization of the dictionary

$$\tilde{\Psi} = XD_X \quad (16)$$

where  $D_X \triangleq \text{diag}(d_1, \dots, d_l, \dots, d_L)$  with

$$D_X(l, l) = d_l \triangleq \|X(:, l)\|_2^{-1}, \quad \forall l. \quad (17)$$

Clearly, the columns of such  $\tilde{\Psi}$  are inherently  $l_2$ -normalized as long as  $X$  has no zero columns. It then follows from (16) - (17) that the problem of updating dictionary using (15) is converted to a unconstraint minimization of form

$$X_{opt} \triangleq \arg \min_X \varrho(X D_X, \tilde{S}, \tilde{H}) \quad (18)$$

where  $\tilde{S} = S_k$ ,  $\tilde{H} = H_k$ .

The problem defined by (18) can then be addressed using the following iterative procedure:

$$X_n = X_{n-1} - \lambda \frac{\partial \varrho}{\partial X}|_{X=X_{n-1}} \quad (19)$$

where  $\lambda > 0$  is the step-size and with an initial  $X_0 \in \mathbb{R}^{N \times L}$  (of no zero-columns) given, say the  $X_0$  such that  $\Psi_{k-1} = X_0 D_{X_0}$ , run (19) and  $\lim_{n \rightarrow +\infty} X_n$  yields an estimate of  $X_{opt}$ . Clearly, the updated dictionary at the  $k$ th iteration of  $\text{Alg}_{\Psi}^{GSD}$  is given by  $\Psi_k = X_{opt} D_{X_{opt}}$ .

The expression for  $\frac{\partial \varrho}{\partial X}$  has been derived and will not be presented due to the limited space.

**Remark 3.2:** As the OMP can yield a solution almost equal to the true one of (8) and the gradient-based algorithm can ensure a solution  $\Psi_k$  better than  $\Psi_{k-1}$  even if it may be different from the one defined by (15), the proposed algorithm can ensure  $\varrho(\Psi_k, S_k, H_k) \leq \varrho(\Psi_{k-1}, S_{k-1}, H_{k-1})$  and hence the convergence of the iterative procedure is guaranteed given that the cost function is positive. So, our proposed algorithm is numerically much more stable than that in [19]. More interesting features of this algorithm will be demonstrated in the next section.

## 4. SIMULATION RESULTS

The setup is as follows. We generate a dictionary  $\Psi \in \mathbb{R}^{30 \times 60}$  randomly, then  $\{y_m^*\}_{m=1}^{4000}$  with each  $y_m^*$  being sparse with  $\kappa = 4$  in  $\Psi$  and  $\|y_m^*\|_2 = 1$ . The actual signal  $y_m$  is given by  $y_m = y_m^* + \epsilon_m$ ,  $\forall m$ , where  $\epsilon_m$  is an additive Gaussian noise with standard variance  $\sigma$ . This setting is the same as used in [1]. Note that  $\mu = \sqrt{1/59} = 0.1302$  and  $\mu(\Psi) = 0.55$ .

We compare our proposed  $\text{Alg}_{\Psi}^{GSD}$  with  $\text{Alg}_{\Psi}^{MOD}$  [14],  $\text{Alg}_{\Psi}^{KSVD}$  [15], and  $\text{Alg}_{\Psi}^{JPR}$  [19] in terms of average representation error  $\mathcal{E}_k \triangleq \|Y - \Psi_k S_k\|_F / \sqrt{JN}$ , percentage of recovered atoms  $r(\Psi_k)$  to be defined below, mutual coherence, and average mutual coherence  $\bar{\mu}_{av}(\Psi_k)$ , defined as

$$\bar{\mu}_{av}(\Psi_k) \triangleq \sqrt{\frac{\|I_L - \bar{\Psi}_k^\top \bar{\Psi}_k\|_F^2}{L(L-1)}}$$

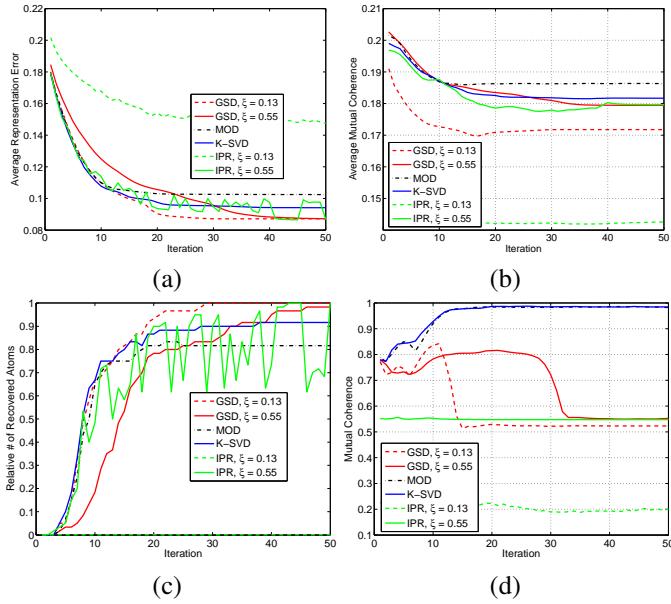
where  $\bar{\Psi}_k$  is the normalized version of  $\Psi_k$ . We say an atom  $\Psi(:, l)$  in the true dictionary  $\Psi$  is recovered in  $\Psi_k$  if

$$\min_n (1 - |\Psi_k(:, n)^\top \Psi(:, l)|) < 0.01. \quad (20)$$

Denote  $\mathcal{L}_k$  as the number of recovered atoms in  $\Psi_k$ . The percentage of the recovered atoms is defined as  $r(\Psi_k) \triangleq \mathcal{L}_k / L$ .

We perform 30 iterations of (19) for solving (15) in  $\text{Alg}_{\Psi}^{GSD}$ . In the legends of Fig.s 1,2, we refer to the algorithms  $\text{Alg}_{\Psi}^{GSD}$ ,  $\text{Alg}_{\Psi}^{MOD}$ ,  $\text{Alg}_{\Psi}^{KSVD}$ , and  $\text{Alg}_{\Psi}^{IPR}$  as GSD, MOD, KSVD and IPR, respectively.

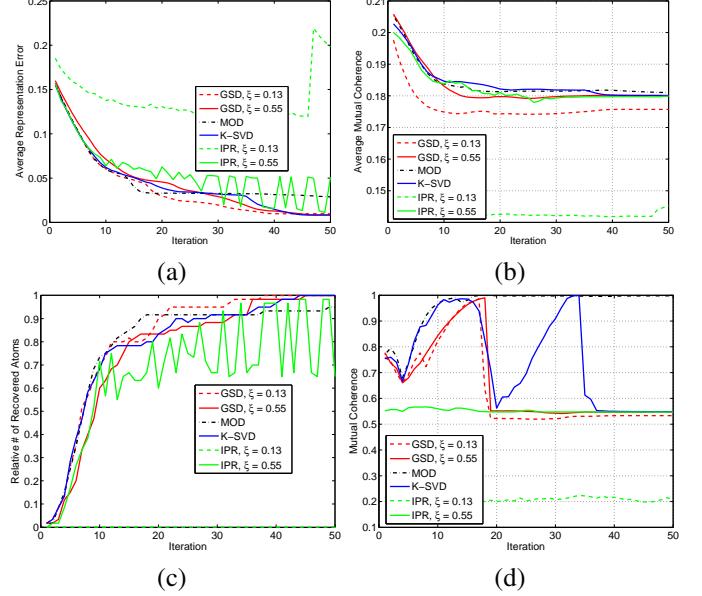
Fig. 1 shows the performance of each of  $\text{Alg}_{\Psi}^{GSD}$ ,  $\text{Alg}_{\Psi}^{IPR}$ ,  $\text{Alg}_{\Psi}^{KSVD}$  and  $\text{Alg}_{\Psi}^{MOD}$  with  $\sigma = 0.1$  and for different values of  $\xi$  (that is used in  $\text{Alg}_{\Psi}^{GSD}$  and  $\text{Alg}_{\Psi}^{IPR}$ ). While Fig. 2 presents the same things but for  $\sigma = 0$ .



**Fig. 1.** Evolution of the four indicators: (a) -  $\mathcal{E}_k$ ; (b) -  $\bar{\mu}_{av}(\Psi_k)$ ; (c) -  $r(\Psi_k)$ ; (d) -  $\mu(\Psi_k)$ . Here  $\sigma = 0.1$  and  $\xi = \underline{\mu}, \mu(\Psi)$  for both  $\text{Alg}_{\Psi}^{GSD}$  and  $\text{Alg}_{\Psi}^{IPR}$ .

#### Remark 4.1

- As observed from Figs. 1,2(a-b), the indicators  $\mathcal{E}_k$  and  $\bar{\mu}_{av}(\Psi_k)$  decay steadily for  $\text{Alg}_{\Psi}^{GSD}$ , which implies the convergence of our proposed algorithm. In  $\text{Alg}_{\Psi}^{GSD}$ , there is a tradeoff between sparse representation error and mutual coherence. By choosing  $\beta$ , we can balance this tradeoff depending on the particular application. In the experiments, for  $\xi = \underline{\mu}$ , we set  $\beta = 0.8$  when  $\sigma = 0$  and  $\beta = 0.9$  when  $\sigma = 0.1$ .
- As seen, our  $\text{Alg}_{\Psi}^{GSD}$  outperforms the others in terms of  $\mathcal{E}_k$  and  $r(\Psi_k)$  when there exists an additive noise, while when the signal is purely sparse (i.e.,  $\sigma = 0$ ),  $\text{Alg}_{\Psi}^{KSVD}$  and  $\text{Alg}_{\Psi}^{GSD}$  yield the best performance in terms of these two indicators.



**Fig. 2.** Evolution of the four indicators: (a) -  $\mathcal{E}_k$ ; (b) -  $\bar{\mu}_{av}(\Psi_k)$ ; (c) -  $r(\Psi_k)$ ; (d) -  $\mu(\Psi_k)$ . Here  $\sigma = 0$  and  $\xi = \underline{\mu}, \mu(\Psi)$  for both  $\text{Alg}_{\Psi}^{GSD}$  and  $\text{Alg}_{\Psi}^{IPR}$ .

- It should be pointed out that in the KSVD code we use, a new atom will be generated randomly to replace the present one if the latter has coherence larger than 0.99 with another atom. The sharp change caused by this action can be observed from Fig 2(d). This step is enough for KSVD (but not enough for MOD algorithm) to escape the coherent solution in the noiseless case.
- It is interesting to note that no matter there is an additive noise or not,  $\text{Alg}_{\Psi}^{IPR}$  performs poorly when  $\xi$  is set to a small value. This is suspected due to the fact that much effort has to be made to reduce the mutual coherence and hence no much room left for reducing the sparse representation error. We can observe when  $\xi$  is set to be close to  $\mu(\Psi)$ , which is practically not available,  $\text{Alg}_{\Psi}^{IPR}$  yields a better performance, while our  $\text{Alg}_{\Psi}^{GSD}$  always has better results than  $\text{Alg}_{\Psi}^{IPR}$  does when we just choose  $\xi = \underline{\mu}$ .

## 5. CONCLUSIONS

A novel framework has been proposed for ISDL and an alternative minimization-based algorithm has been derived for solving the optimal dictionary design. Experiments have shown that the proposed approach outperforms the prevailing ones for ISDL.

In the proposed algorithm, the dictionary update is done using a gradient-based method. More efficient algorithms are needed for speeding up the convergence and further enhancing the performance of the dictionary.

## 6. REFERENCES

- [1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer New York Dordrecht Heidelberg London, 2010.
- [2] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthonormal) dictionaries via  $l_1$  minimization,” *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197 - 2202, Mar. 2003.
- [3] I. Tošić and P. Frossard, “Dictionary learning: What is the right representation for my signal?”, *IEEE Signal Process. Mag.*, pp. 27 - 38, Mar. 2011.
- [4] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Sept. 2006.
- [5] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.
- [6] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21 - 30, Mar. 2008.
- [7] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231 - 2242, Oct. 2004.
- [8] J. A. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
- [9] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230 - 2249, 2009.
- [10] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online learning for matrix factorization and sparse coding,” *J. Mach. Learn. Res.*, vol. 11, pp. 19-60, Jan. 2010.
- [11] R. Gribonval and K. Schnass, “Dictionary identification: sparse matrix-factorization via  $l_1$ -minimization,” *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3523 - 3539, 2010.
- [12] K. Schnass, “On the identifiability of overcomplete dictionaries via the minimisation principle underlying K-SVD”, *Appl. Comput. Harmon. Anal.*, vol. 37, pp. 464-491, 2014.
- [13] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr., and T. Strohmer, “Designing structured tight frame via alternating projection,” *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188-209, 2005.
- [14] K. Engan, S. O. Aase, and J. H. Hakon-Housoy, “Method of optimal direction for frame design,” in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, pp. 2443-2446, 1999.
- [15] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311-4322, 2006.
- [16] M. Yaghoobi, L. Daudet, and M. E. Davies, “Parametric dictionary design for sparse coding,” *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4800-4810, 2009.
- [17] S. K. Sahoo and A. Makur, “Dictionary training for sparse representation as generalization of K-means clustering,” *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 587-590, 2013.
- [18] J. A. Tropp, “On the conditioning of random subdictionaries,” *Appl. Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 1-24, Jul. 2008.
- [19] D. Barchiesi and M.D. Plumbe, “Learning incoherent dictionaries for sparse approximation using iterative projections and rotations,” *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2055-2065, Apr. 2013.
- [20] T. Strohmer and R. W. Heath, “Grassmannian frames with applications to coding and communication,” *Appl. Comp. Harmonic Anal.*, vol. 14, no. 3, pp. 257-275, May 2003.