Research Article

An Efficient Algorithm for Learning Dictionary under Coherence Constraint

Huang Bai, Sheng Li, and Qianru Jiang

College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China

Correspondence should be addressed to Sheng Li; shengli@zjut.edu.cn

Received 30 March 2016; Revised 8 June 2016; Accepted 23 June 2016

Academic Editor: Srdjan Stankovic

Copyright © 2016 Huang Bai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dictionary learning problem has become an active topic for decades. Most existing learning methods train the dictionary to adapt to a particular class of signals. But as the number of the dictionary atoms is increased to represent the signals much more sparsely, the coherence between the atoms becomes higher. According to the greedy and compressed sensing theories, this goes against the implementation of sparse coding. In this paper, a novel approach is proposed to learn the dictionary that minimizes the sparse representation error according to the training signals with the coherence taken into consideration. The coherence is constrained by making the Gram matrix of the desired dictionary approximate to an identity matrix of proper dimension. The method for handling the proposed model is mainly based on the alternating minimization procedure and, in each step, the closed-form solution is derived. A series of experiments on synthetic data and audio signals is executed to demonstrate the promising performance of the learnt incoherent dictionary and the superiority of the learning method to the existing ones.

1. Introduction

Sparse representation (SR) theory [1, 2] indicates that a signal can be represented by certain linear combination of a few *atoms* of a prespecified *dictionary*. It is an evolving field, with state-of-the-art results in many signal processing tasks, such as coding, denoising, face recognition, deblurring, and compressed sensing [3–7].

A fundamental consideration in employing the above theory is the choice of the dictionary and this leads to the famous dictionary learning (DL) problem. DL has attracted a lot of attention since its introduction at the end of last century [8, 9]. Most of the research has been done to learn a data adaptive dictionary so that a particular class of signals can be sparsely represented in this dictionary with low approximation error.

Under the SR framework, a signal vector $\mathbf{y} \in \mathbb{R}^{N \times 1}$ can be expressed in the form of

$$\mathbf{y} \approx \sum_{k=1}^K \mathbf{x}(k) \mathbf{D}(:, k) \triangleq \mathbf{D}\mathbf{x}, \quad (1)$$

where $\mathbf{D} \in \mathbb{R}^{N \times K}$ is the dictionary with its columns $\{\mathbf{D}(:, k)\}$ referred to as atoms (throughout this paper, MATLAB

notations are used) and $\mathbf{x} \in \mathbb{R}^{K \times 1}$ is the corresponding sparse coefficient vector.

Let $\mathbf{v} \in \mathbb{R}^{K \times 1}$ with $\mathbf{v}(k)$ being its k th element. The l_p -norm of vector $\mathbf{v}$ is defined as

$$\|\mathbf{v}\|_p \triangleq \left(\sum_{k=1}^K |\mathbf{v}(k)|^p \right)^{1/p}, \quad p \geq 1. \quad (2)$$

Note that $\|\mathbf{v}\|_p$ is not a norm in a strict sense for $0 \leq p < 1$. For convenience, $\|\mathbf{v}\|_0$ is used to denote the number of nonzero elements in $\mathbf{v}$. A vector $\mathbf{y}$ given by (1) is said to be S -sparse in $\mathbf{D}$ if $\|\mathbf{x}\|_0 = S$.

Let $\{\mathbf{y}_m\}_{m=1}^M$ be a set of training samples from a class of signals to be considered. The basic problem of DL is to find a dictionary $\mathbf{D}$ such that, for each $\mathbf{y}_m$, there exists a vector $\mathbf{x}_m$ that is sparse. Such a problem has been widely investigated during the last decade or so [10–12] and can be formulated as

$$\min_{\mathbf{D}, \{\mathbf{x}_m\}_{m=1}^M} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \sum_{m=1}^M \tau_m \|\mathbf{x}_m\|_0, \quad (3)$$

where $\|\cdot\|_F$ denotes the *Frobenius* norm, $\{\tau_m\}$ are proper constants, $\|\mathbf{x}_m\|_0$ is the sparsity of sparse vector $\mathbf{x}_m$, and

$$\begin{aligned} \mathbf{Y} &\triangleq [\mathbf{y}_1 \cdots \mathbf{y}_m \cdots \mathbf{y}_M] \in \mathcal{R}^{N \times M}, \\ \mathbf{X} &\triangleq [\mathbf{x}_1 \cdots \mathbf{x}_m \cdots \mathbf{x}_M] \in \mathcal{R}^{K \times M}. \end{aligned} \quad (4)$$

Such a problem is difficult to be solved as it is nonconvex in $\mathbf{D}$ and $\mathbf{X}$, and $\|\cdot\|_0$ is nonsmooth and highly unstable. A popularly used approach is based on the alternating minimization strategy. A two-stage procedure is usually carried out for solving the above problem and also for avoiding the selections of $\{\tau_m\}$ [10–12]. The problem in the first stage is referred to as *sparse coding*, aiming at finding the (column) sparse matrix $\mathbf{X}$ with a given $\mathbf{D}$; that is

$$\begin{aligned} \hat{\mathbf{x}}_m &\triangleq \arg \min_{\mathbf{x}_m} \|\mathbf{y}_m - \mathbf{D}\mathbf{x}_m\|_2^2 \\ \text{s.t. } &\|\mathbf{x}_m\|_0 \leq S, \\ &\forall m. \end{aligned} \quad (5)$$

Note that the equivalent expression to (5) where the constraint is a fixed sparse representation error (SRE) level can also be formulated as

$$\begin{aligned} \hat{\mathbf{x}}_m &\triangleq \arg \min_{\mathbf{x}_m} \|\mathbf{x}_m\|_0 \\ \text{s.t. } &\|\mathbf{y}_m - \mathbf{D}\mathbf{x}_m\|_2^2 \leq \epsilon, \\ &\forall m \end{aligned} \quad (6)$$

with ϵ being the error threshold. Such a problem can be solved using the orthogonal matching pursuit (OMP) based methods [13, 14]. Furthermore, it can be shown that the solution of the above problem is the same as the one of the l_1 -based minimization below:

$$\begin{aligned} \hat{\mathbf{x}}_m &\triangleq \arg \min_{\mathbf{x}_m} \|\mathbf{x}_m\|_1 \\ \text{s.t. } &\|\mathbf{y}_m - \mathbf{D}\mathbf{x}_m\|_2^2 \leq \epsilon, \\ &\forall m, \end{aligned} \quad (7)$$

while the latter can be addressed using algorithms such as basis pursuit (BP) [15] and the l_1 -based optimization techniques [16].

Many algorithms for solving (3) are different from each other mainly in the 2nd stage, that is, *dictionary updating*. For the dictionary $\mathbf{D}$, in order to code the signals of interest more sparsely, we usually set $K > N$ which means that $\mathbf{D}$ is overcomplete. However, this redundancy increases the pairwise similarity of dictionary atoms. According to the work in [13], such a similarity has a direct influence on the dictionary's performance, especially for the accuracy in sparse coding stage. If any two atoms degenerate to the same vector, this will lead to overfitting to the training data. Thus, incoherent dictionary is expected to improve the performance of the SR model.

Yaghoobi et al. proposed a design method for parametric dictionary [17]. The authors attempted to optimize the dictionary to make the corresponding Gram matrix approximate

to the Gram of an equiangular tight frame (ETF), which possesses good coherence behavior. However, this method relies extremely on a priori knowledge of appropriate parameters choosing criterion that is related to a given class of signals. A new algorithm was developed in [18] named INK-SVD. In each iteration of K-SVD algorithm [11], the dictionary updating stage is followed by an additional decorrelation step. Each pair of atoms which has coherence above the threshold should have its inner angle increased symmetrically so as to reduce the coherence. But this procedure will implicitly destroy the original SR result from the K-SVD algorithm. To compensate this problem, the authors of [19] improved the work of [18] by incorporating a new decorrelation step (also related to the ETF according to its low coherence) and a dictionary rotation operation to the update stage. In [20], a weighting model was formulated to balance the coherence of the dictionary and the sparse representation ability, and a gradient-based method was carried out for solving the corresponding problem.

The main objective of this paper is to propose a new incoherent dictionary learning (IDL) method that constrains the coherence of the dictionary and minimizes the SRE and the contributions are threefold:

- (i) A novel model is proposed for learning the incoherent dictionary. The main contribution is also located in the dictionary updating procedure. When minimizing the SRE, that is, $\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2$, $\mathbf{D}$ is under the coherence constraint by making the corresponding Gram matrix approximate to an identity matrix of proper dimension.
- (ii) An iterative algorithm that updates the sparse coefficients and the components of the dictionary alternately is put forward to solve the design problem. In every step of dictionary updating, the solution of each component of dictionary is derived analytically.
- (iii) A series of experiments on synthetic data and audio signals is carried out to demonstrate the performance of each compared algorithm.

The remainder of this paper is arranged as follows. In Section 2, some preliminaries are provided and the main issue of learning incoherent dictionary is also formulated in this part. The algorithm proposed for addressing the corresponding design problem is investigated in Section 3. Simulations are carried out in Section 4 to examine the performance of the proposed algorithm and to compare with the existing ones. Some concluding remarks are given in Section 5.

2. Preliminaries and Problem Formulation

In this section, some preliminaries will be introduced and two main comparisons of this paper are also reviewed in detail. Based on these, we formulate the problem of incoherent dictionary learning with the purpose of increasing the approximation performance of the dictionary to a particular class of signals under the coherence constraint.

The most fundamental quality associated with a dictionary is the *mutual coherence* (MC) [21]. MC indicates the degree of similarity between different dictionary columns. It equals the maximum absolute inner product between two distinct atoms:

$$\mu(\mathbf{D}) \triangleq \max_{1 \leq i \neq j \leq K} \frac{|\mathbf{D}(:, i)^{\mathcal{T}} \mathbf{D}(:, j)|}{\|\mathbf{D}(:, i)\|_2 \|\mathbf{D}(:, j)\|_2}, \quad (8)$$

where $\mathcal{T}$ denotes the transpose operator. As shown in [21], a S -sparse signal generated according to (1) can be exactly recovered with OMP as long as

$$S < \frac{1}{2} \left[1 + \frac{1}{\mu(\mathbf{D})} \right]. \quad (9)$$

Roughly speaking, MC measures how two atoms can look alike. Equation (9) is just a worst-case bound and only reflects the most extreme correlations in the dictionary. Nevertheless, MC is easy to be manipulated and it captures well the behaviors of some dictionaries. Generally, a dictionary is called incoherent if the corresponding MC is small [18, 19]. Besides, as pointed out in [19], the coherence of a dictionary is related to the condition number of its subdictionaries. This implies that achieving a low MC value results in well-conditioned subdictionaries.

Define the Gram matrix of $\mathbf{D}$ as

$$\mathbf{G} \triangleq \mathbf{D}^{\mathcal{T}} \mathbf{D}. \quad (10)$$

It is common to study MC in (8) via the Gram matrix. Let $\mathbf{D}_{\text{sc}}$ be the diagonal matrix whose k th element is given by $1/\sqrt{\mathbf{G}(k, k)}$ for $k = 1, \dots, K$. The Gram matrix of $\bar{\mathbf{D}} \triangleq \mathbf{D} \mathbf{D}_{\text{sc}}$, denoted as $\bar{\mathbf{G}}$, is then normalized, such that $\bar{\mathbf{G}}(k, k) = 1, \forall k$. Obviously, $\mu(\mathbf{D}) = \max_{i \neq j} |\bar{\mathbf{G}}(i, j)|$.

For $\mathbf{D} \in \mathfrak{R}^{N \times K}$, it has been shown in [22] that $\mu(\mathbf{D})$ is bounded with

$$\underline{\mu} \triangleq \sqrt{\frac{K-N}{N(K-1)}} \leq \mu(\mathbf{D}) \leq 1 \quad (11)$$

with $\underline{\mu}$ being the Welch bound. If each atomic inner product meets this bound, the dictionary is called an ETF. An ETF has a very nice MC behavior and has been considered to be utilized in optimal dictionary design [17, 19].

2.1. Related Works. It is worth noting that ETFs only exist for those matrices $\mathbf{D} \in \mathfrak{R}^{N \times K}$ with dimensionality constrained with $K \leq N(N+1)/2$ if the atoms are real. So, one usually replaces the set of ETF Grams with a relaxed version [17, 19] that is defined as

$$\begin{aligned} \mathcal{S}_{\eta} &\triangleq \left\{ \mathbf{G} \in \mathfrak{R}^{K \times K} : \mathbf{G} = \mathbf{G}^{\mathcal{T}}, \mathbf{G}(k, k) \right. \\ &= 1, \forall k; \max_{i \neq j} |\mathbf{G}(i, j)| \leq \eta \left. \right\}, \end{aligned} \quad (12)$$

where $0 < \eta < 1$ is a constant to control the searching space. Clearly, when $\eta \geq \underline{\mu}$, $\mathcal{S}_{\eta}$ contains all the ETF Grams.

Besides the space $\mathcal{S}_{\eta}$, the authors of [19] define a spectral constraint set as

$$\begin{aligned} \mathcal{F} &\triangleq \left\{ \mathbf{G} \in \mathfrak{R}^{K \times K} : \mathbf{G} = \mathbf{G}^{\mathcal{T}}, \text{eig}(\mathbf{G}) \geq 0, \text{rank}(\mathbf{G}) \right. \\ &\leq N \left. \right\}. \end{aligned} \quad (13)$$

Here $\text{eig}(\cdot)$ returns the vector of eigenvalues and $\text{rank}(\cdot)$ is the rank operator. The algorithm for learning incoherent dictionary proposed in [19] can be outlined as follows:

- (i) *Sparse coding* with OMP.
- (ii) *Dictionary updating* employing K-SVD.
- (iii) *Atoms decorrelation* through an iterative projection procedure.
- (iv) *Dictionary rotation* to minimize the approximate error while keeping the MC unchanged.

The main contributions of [19] lie in the last two steps. The atoms decorrelation is executed by iteratively projecting the Gram of the output dictionary of K-SVD between the sets $\mathcal{S}_{\eta}$ and $\mathcal{F}$ until a stopping criterion is met. With the singular value decomposition (SVD) of the resulting positive semidefinite Gram matrix $\hat{\mathbf{G}}$ being expressed as

$$\hat{\mathbf{G}} = \mathbf{V}_{\mathbf{G}} \begin{bmatrix} \Sigma_{\mathbf{G}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}_{\mathbf{G}}^{\mathcal{T}}, \quad (14)$$

where $\mathbf{V}_{\mathbf{G}} \in \mathfrak{R}^{K \times K}$ is orthonormal and $\Sigma_{\mathbf{G}} \in \mathfrak{R}^{N \times N}$ is the diagonal singular value matrix with all its elements being nonnegative, the incoherent dictionary can be obtained as

$$\hat{\mathbf{D}} = \mathbf{U} [\sqrt{\Sigma_{\mathbf{G}}} \quad \mathbf{0}] \mathbf{V}_{\mathbf{G}}^{\mathcal{T}} \quad (15)$$

with $\mathbf{U}$ being an *arbitrary* orthonormal matrix. Finally, the authors consider this degree of freedom to further reduce the SRE by solving

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U} \in O(N)} \|\mathbf{Y} - \mathbf{U} [\sqrt{\Sigma_{\mathbf{G}}} \quad \mathbf{0}] \mathbf{V}_{\mathbf{G}}^{\mathcal{T}} \mathbf{X}\|_F^2, \quad (16)$$

where $O(N)$ is the set of $N \times N$ orthonormal matrices. This is the rotation procedure.

Remark 1. Compared with the decorrelation operation in [18], the above-mentioned atoms decorrelation can achieve a much smaller MC value. Besides, the additional rotation procedure can slightly redeem SR ability. However, the approximation performance of the dictionary is highly damaged by the iterative projections. Though the dictionary rotation procedure is carried out for compensation, the effect of the sole degree of freedom $\mathbf{U}$ on the SR ability is quite limited.

In [20], the authors consider another strategy for IDL, where the dictionary's coherence is minimized along with the SRE. The cost function can be expressed as

$$\min_{\mathbf{D}} \left\{ \|\mathbf{D}\mathbf{X} - \mathbf{Y}\|_F^2 + \gamma \|\mathbf{D}^{\mathcal{T}} \mathbf{D} - \mathbf{I}_K\|_F^2 \triangleq \varrho(\mathbf{D}) \right\} \quad (17)$$

with $\mathbf{I}_K$ denoting the identity matrix of dimension K . It is clear that $\mathbf{I}_K$ is the simplest ETF Gram (with $N = K$). The Lagrange multiplier γ controls the trade-off between minimizing the SRE and minimizing the dictionary's coherence. With the gradient of $\varrho(\mathbf{D})$ being calculated as

$$\frac{d\varrho(\mathbf{D})}{d\mathbf{D}} = 2(\mathbf{D}\mathbf{X}\mathbf{X}^\mathcal{T} - \mathbf{Y}\mathbf{X}^\mathcal{T}) + 4\gamma(\mathbf{D}\mathbf{D}^\mathcal{T}\mathbf{D} - \mathbf{D}), \quad (18)$$

the update of the dictionary is then executed by the steepest descent algorithm [20].

Remark 2. (i) The choice of γ remains open-ended and there is no selection criterion introduced in [20]. From the simulation results of [20], larger γ introduces better performance.

(ii) It is well known that the gradient-based algorithms may easily fall into a local minimum if the initialization is not properly set [23, 24]. As gradient-based method is carried out for solving (17), the efficiency and accuracy can be further improved.

2.2. Problem Formulation. Let $\mathbf{D} \in \mathfrak{R}^{N \times K}$ be the dictionary, let $\mathbf{Y} = \{\mathbf{y}_m\}_{m=1}^M$ be the signal set with $\mathbf{y}_m \in \mathfrak{R}^{N \times 1}$, and let $\mathbf{X} = \{\mathbf{x}_m\}_{m=1}^M$ be the corresponding sparse coefficient matrix in $\mathbf{D}$ with $\mathbf{x}_m \in \mathfrak{R}^{K \times 1}$ as defined previously. For the problem indicated in (3), we update $\mathbf{X}$ and $\mathbf{D}$ alternately. For a fixed $\mathbf{D}$, $\mathbf{X}$ can be calculated by the greedy algorithms or the l_1 -based convex optimization methods. In the following, we focus our discussion on the dictionary updating stage. For the traditional case, that is,

$$\min_{\mathbf{D}} \|\mathbf{D}\mathbf{X} - \mathbf{Y}\|_F^2, \quad (19)$$

the authors of [10] simply update the dictionary as $\hat{\mathbf{D}} = \mathbf{Y}\mathbf{X}^\mathcal{T}(\mathbf{X}\mathbf{X}^\mathcal{T})^{-1}$. But when $\mathbf{X}$ is not full rank, this method fails to work. The K-SVD algorithm [11] minimizes (19) for each atom separately. When updating the dictionary, the coefficients are also renewed simultaneously. In every iteration between coefficients and dictionary, it needs K times SVD operations. It is a time-consuming algorithm and not well suited to enforcing a coherence constraint which is important for the implementation of sparse coding.

As an ETF can achieve small MC value, this motivates us to design a dictionary that is as close as possible to an ETF [17–20]. So the following constrained model is proposed:

$$\begin{aligned} \min_{\mathbf{D}, \{\mathbf{x}_m\}_{m=1}^M} \quad & \|\mathbf{D}\mathbf{X} - \mathbf{Y}\|_F^2 + \sum_{m=1}^M \tau_m \|\mathbf{x}_m\|_0 \\ \text{s.t.} \quad & \min_{\mathbf{D}} \|\mathbf{D}^\mathcal{T}\mathbf{D} - \mathbf{I}_K\|_F^2. \end{aligned} \quad (20)$$

The closed-form solution set of $\{\min_{\mathbf{D}} \|\mathbf{D}^\mathcal{T}\mathbf{D} - \mathbf{I}_K\|_F^2\}$ has been derived in [7, 24] as

$$\mathbf{D} = \mathbf{U} [\mathbf{I}_N \quad \mathbf{0}] \mathbf{V}^\mathcal{T}, \quad (21)$$

where $\mathbf{U}$ and $\mathbf{V}$ are both *arbitrary* orthonormal matrices of dimensions N and K , respectively. So (20) can be rewritten as

$$\begin{aligned} \min_{\mathbf{U} \in O(N), \mathbf{V} \in O(K), \{\mathbf{x}_m\}_{m=1}^M} \quad & \|\mathbf{D}\mathbf{X} - \mathbf{Y}\|_F^2 + \sum_{m=1}^M \tau_m \|\mathbf{x}_m\|_0 \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{U} [\mathbf{I}_N \quad \mathbf{0}] \mathbf{V}^\mathcal{T}. \end{aligned} \quad (22)$$

Remark 3. (i) Here we choose the identity matrix $\mathbf{I}_K$ as the target Gram for the following reasons: it is easy to handle (avoiding the iterative projection between $\mathcal{S}_\eta$ and $\mathcal{F}$ as carried out in [19]) and expression (21) contains more degrees of freedom than (15) for further minimizing the SRE.

(ii) As pointed out in [20], a flatter singular value spectrum of the dictionary indicates a less coherent dictionary. Our design strategy is under constraint (21) which means that the nonzero singular values of our designed dictionary are all equal (the same as (17) with $\gamma \rightarrow \infty$). Hence, better coherence performance can be expected.

3. Coherence Constrained Dictionary Learning

In this section, an alternating minimization algorithm is developed to address the dictionary learning problem (22).

3.1. Algorithm for IDL. To solve the above multivariate problem and also avoid the selections of $\{\tau_m\}$, the alternating minimization strategy as introduced for addressing (3) seems a natural choice. The pseudocode of the proposed algorithm (named CCDL, standing for Coherence Constrained Dictionary Learning) is summarized as follows:

Initialization

$\mathbf{D}^{(0)}$: initial random $N \times K$ dictionary;

$\mathbf{Y}$: training data;

Iter₁: number of iterations between sparse coding and dictionary updating;

Iter₂: number of iterations between updating $\mathbf{U}$ and $\mathbf{V}$.

Calculate the SVD of $\mathbf{D}^{(0)}$ as $\mathbf{D}^{(0)} = \mathbf{U}^{(0)} \mathbf{\Sigma}^{(0)} (\mathbf{V}^{(0)})^\mathcal{T}$ and set $i = 1$.

For $1 \leq i \leq \text{Iter}_1$, do the following.

Step 1. Set $\mathbf{D}^{(i-1)} = \mathbf{U}^{(i-1)} [\mathbf{I}_N \quad \mathbf{0}] (\mathbf{V}^{(i-1)})^\mathcal{T}$, and update $\mathbf{X}$ column by column by solving

$$\hat{\mathbf{x}}_m^{(i)} = \arg \min_{\mathbf{x}_m} \|\mathbf{D}^{(i-1)} \mathbf{x}_m - \mathbf{y}_m\|_F^2 + \tau_m \|\mathbf{x}_m\|_0, \quad \forall m \quad (23)$$

with an OMP based algorithm and get the approximate solution $\mathbf{X}^{(i)} = \{\hat{\mathbf{x}}_m^{(i)}\}_{m=1}^M$.

Set $j = 1$ and $\mathbf{U}_{(0)}^{(i-1)} = \mathbf{U}^{(i-1)}$ and $\mathbf{V}_{(0)}^{(i-1)} = \mathbf{V}^{(i-1)}$.

Step 2. While $1 \leq j \leq \text{Iter}_2$, do the following:

- (a) For fixed $\mathbf{V}_{(j-1)}^{(i-1)}$, update $\mathbf{U}_{(j)}^{(i-1)}$ by solving

$$\min_{\mathbf{U} \in \mathcal{O}(N)} \left\| \mathbf{U} [\mathbf{I}_N \ \mathbf{0}] \left(\mathbf{V}_{(j-1)}^{(i-1)} \right)^{\mathcal{T}} \mathbf{X}^{(i)} - \mathbf{Y} \right\|_F^2. \quad (24)$$

The analytical solution will be given in the next subsection.

- (b) For fixed $\mathbf{U}_{(j)}^{(i-1)}$, update $\mathbf{V}_{(j)}^{(i-1)}$ by solving

$$\min_{\mathbf{V} \in \mathcal{O}(K)} \left\| \mathbf{U}_{(j)}^{(i-1)} [\mathbf{I}_N \ \mathbf{0}] \mathbf{V}^{\mathcal{T}} \mathbf{X}^{(i)} - \mathbf{Y} \right\|_F^2. \quad (25)$$

The solution will be derived in the next subsection.

- (c) Return to Step 2 with $j \rightarrow j + 1$.

Step 3. Set $\mathbf{U}^{(i)} = \mathbf{U}_{(j-1)}^{(i-1)}$ and $\mathbf{V}^{(i)} = \mathbf{V}_{(j-1)}^{(i-1)}$. End for if $i > \text{Iter}_1$.

End. Output $\mathbf{D} = \mathbf{U}^{(i-1)} [\mathbf{I}_N \ \mathbf{0}] \left(\mathbf{V}^{(i-1)} \right)^{\mathcal{T}}$.

3.2. Update the Components of Dictionary. Now, let us focus on solving (24) and (25). For convenience, we omit the superscript (i) in the expressions. As the sparse coefficient matrix $\mathbf{X}$ is assumed to be fixed in Step 2, we can rewrite the cost function of (22) as

$$\min_{\mathbf{U} \in \mathcal{O}(N), \mathbf{V} \in \mathcal{O}(K)} \left\{ \left\| \mathbf{U} [\mathbf{I}_N \ \mathbf{0}] \mathbf{V}^{\mathcal{T}} \mathbf{X} - \mathbf{Y} \right\|_F^2 \triangleq \varrho(\mathbf{U}, \mathbf{V}) \right\}, \quad (26)$$

where $\mathbf{U}$ and $\mathbf{V}$ can be updated alternately. Let $\mathbf{X}$ have the following SVD (for arbitrary matrix $\mathbf{M}$, the general SVD form can be expressed as $\mathbf{M} = \mathbf{U}^m \mathbf{\Sigma}^m (\mathbf{V}^m)^{\mathcal{T}}$):

$$\mathbf{X} = \mathbf{U}^x \mathbf{\Sigma}^x (\mathbf{V}^x)^{\mathcal{T}}. \quad (27)$$

Define

$$\begin{aligned} \mathbf{Q}(\mathbf{U}) &\triangleq \mathbf{U}^{\mathcal{T}} \mathbf{Y} \mathbf{V}^x \\ \mathbf{P}(\mathbf{V}) &\triangleq [\mathbf{I}_N \ \mathbf{0}] \mathbf{V}^{\mathcal{T}} \mathbf{X} \\ \tilde{\mathbf{V}} &\triangleq (\mathbf{U}^x)^{\mathcal{T}} \mathbf{V} = [\tilde{\mathbf{V}}_1 \ \tilde{\mathbf{V}}_2] \end{aligned} \quad (28)$$

with $\tilde{\mathbf{V}}_1 \in \mathfrak{R}^{K \times N}$. We then have two alternative expressions for $\varrho(\mathbf{U}, \mathbf{V})$:

$$\varrho(\mathbf{U}, \mathbf{V}) = \left\| \mathbf{U} \mathbf{P}(\mathbf{V}) - \mathbf{Y} \right\|_F^2 = \left\| \tilde{\mathbf{V}}_1^{\mathcal{T}} \mathbf{\Sigma}^x - \mathbf{Q}(\mathbf{U}) \right\|_F^2. \quad (29)$$

Assume that $\mathbf{U}_{(j-1)}$ and $\mathbf{V}_{(j-1)}$ be given. In what follows, we derive a procedure for updating $(\mathbf{U}, \mathbf{V})$ such that

$$\varrho(\mathbf{U}_{(j)}, \mathbf{V}_{(j)}) \leq \varrho(\mathbf{U}_{(j-1)}, \mathbf{V}_{(j-1)}). \quad (30)$$

3.2.1. Update $\mathbf{U}$. First of all, consider

$$\mathbf{U}_{(j)} \triangleq \arg \min_{\mathbf{U} \in \mathcal{O}(N)} \varrho(\mathbf{U}, \mathbf{V}_{(j-1)}). \quad (31)$$

This model can be solved by the following theorem [19].

Theorem 4. For both $\mathbf{B}$ and $\mathbf{C}$ belonging to $\mathfrak{R}^{N \times M}$, the solution of

$$\min_{\mathbf{A} \in \mathcal{O}(N)} \left\| \mathbf{A} \mathbf{B} - \mathbf{C} \right\|_F^2 \quad (32)$$

can be characterized as

$$\mathbf{A} = \mathbf{V}^t (\mathbf{U}^t)^{\mathcal{T}}, \quad (33)$$

where $\mathbf{V}^t$ and $\mathbf{U}^t$ are both orthonormal matrices given by the following SVD:

$$\mathbf{B} \mathbf{C}^{\mathcal{T}} \triangleq \mathbf{T} = \mathbf{U}^t \mathbf{\Sigma}^t (\mathbf{V}^t)^{\mathcal{T}}. \quad (34)$$

Let

$$\mathbf{P}(\mathbf{V}_{(j-1)}) \mathbf{Y}^{\mathcal{T}} = \mathbf{U}_{(j-1)}^u \mathbf{\Sigma}_{(j-1)}^u (\mathbf{V}_{(j-1)}^u)^{\mathcal{T}}. \quad (35)$$

The solution to (31) can be derived as

$$\mathbf{U}_{(j)} = \mathbf{V}_{(j-1)}^u (\mathbf{U}_{(j-1)}^u)^{\mathcal{T}}. \quad (36)$$

3.2.2. Update $\mathbf{V}$. Now, let us consider

$$\mathbf{V}_{(j)} \triangleq \arg \min_{\mathbf{V} \in \mathcal{O}(K)} \varrho(\mathbf{U}_{(j)}, \mathbf{V}). \quad (37)$$

Obviously, (30) is satisfied with $\mathbf{U}_{(j)}$ and $\mathbf{V}_{(j)}$ obtained as the solutions of (31) and (37), respectively.

Note that $\varrho(\mathbf{U}_{(j)}, \mathbf{V}) = \left\| \tilde{\mathbf{V}}_1^{\mathcal{T}} \mathbf{\Sigma}^x - \mathbf{Q}(\mathbf{U}_{(j)}) \right\|_F^2$. Define

$$\begin{aligned} \tilde{\varrho}(\tilde{\mathbf{V}}, \mathbf{W}) &\triangleq \left\| \tilde{\mathbf{V}}^{\mathcal{T}} \mathbf{\Sigma}^x - \begin{bmatrix} \mathbf{Q}(\mathbf{U}_{(j)}) \\ \mathbf{W} \end{bmatrix} \right\|_F^2 \\ &= \left\| \begin{bmatrix} \tilde{\mathbf{V}}_1^{\mathcal{T}} \mathbf{\Sigma}^x \\ \tilde{\mathbf{V}}_2^{\mathcal{T}} \mathbf{\Sigma}^x \end{bmatrix} - \begin{bmatrix} \mathbf{Q}(\mathbf{U}_{(j)}) \\ \mathbf{W} \end{bmatrix} \right\|_F^2 \\ &= \left\| \begin{bmatrix} \tilde{\mathbf{V}}_1^{\mathcal{T}} \mathbf{\Sigma}^x - \mathbf{Q}(\mathbf{U}_{(j)}) \\ \tilde{\mathbf{V}}_2^{\mathcal{T}} \mathbf{\Sigma}^x - \mathbf{W} \end{bmatrix} \right\|_F^2. \end{aligned} \quad (38)$$

It is clear that such a function has the following properties:

$$\begin{aligned} \tilde{\varrho}(\tilde{\mathbf{V}}, \tilde{\mathbf{V}}_2^{\mathcal{T}} \mathbf{\Sigma}^x) &= \varrho(\mathbf{U}_{(j)}, \mathbf{V}) \\ \tilde{\varrho}(\tilde{\mathbf{V}}, \mathbf{W}) &\geq \varrho(\mathbf{U}_{(j)}, \mathbf{V}), \end{aligned} \quad (39)$$

where $\tilde{\mathbf{V}}$ is defined in (28).

Denote

$$\mathbf{W}(k) \triangleq (\tilde{\mathbf{V}}_2(k))^{\mathcal{T}} \mathbf{\Sigma}^x, \quad k = 0, 1, 2, \dots, \quad (40)$$

where each $\tilde{\mathbf{V}}(k) = (\mathbf{U}^x)^{\mathcal{T}} \mathbf{V}(k)$ with $\mathbf{V}(0) \triangleq \mathbf{V}_{(j-1)}$ is constructed using

$$\tilde{\mathbf{V}}(k) = \arg \min_{\tilde{\mathbf{V}} \in \mathcal{O}(K)} \tilde{\varrho}(\tilde{\mathbf{V}}, \mathbf{W}(k-1)), \quad \forall k = 1, 2, \dots \quad (41)$$

It follows from (38) and Theorem 4 that the solution to (41), as understood, is given by

$$\tilde{\mathbf{V}}(k) = \mathbf{U}^v(k-1) (\mathbf{V}^v(k-1))^{\mathcal{T}}, \quad (42)$$

where

$$\begin{aligned} \Sigma^x \begin{bmatrix} \mathbf{Q}(\mathbf{U}_{(j)}) \\ \mathbf{W}(k-1) \end{bmatrix}^{\mathcal{T}} \\ = \mathbf{U}^v(k-1) \Sigma^v(k-1) (\mathbf{V}^v(k-1))^{\mathcal{T}}. \end{aligned} \quad (43)$$

It turns out from (39) and (40) that

$$\begin{aligned} \tilde{\varrho}(\tilde{\mathbf{V}}(k), \mathbf{W}(k-1)) &\geq \tilde{\varrho}(\tilde{\mathbf{V}}(k), \mathbf{W}(k)) \\ &= \varrho(\mathbf{U}_{(j)}, \mathbf{V}(k)), \end{aligned} \quad (44)$$

while (41) indicates that

$$\begin{aligned} \tilde{\varrho}(\tilde{\mathbf{V}}(k), \mathbf{W}(k-1)) &\leq \tilde{\varrho}(\tilde{\mathbf{V}}(k-1), \mathbf{W}(k-1)) \\ &= \varrho(\mathbf{U}_{(j)}, \mathbf{V}(k-1)). \end{aligned} \quad (45)$$

This implies that constructing $\{\tilde{\mathbf{V}}(k)\}$ using (42) and hence $\{\mathbf{V}(k)\} = \{\mathbf{U}^x \tilde{\mathbf{V}}(k)\}$ makes $\varrho(\mathbf{U}_{(j)}, \mathbf{V}(k))$ a decreasing sequence and, therefore, the solution to (37) can be estimated with

$$\hat{\mathbf{V}}_{(j)} \triangleq \lim_{k \rightarrow \infty} \mathbf{V}(k) = \mathbf{U}^x \lim_{k \rightarrow \infty} \tilde{\mathbf{V}}(k). \quad (46)$$

Remark 5. (i) It may be possible that $\mathbf{V}_{(j)} \neq \hat{\mathbf{V}}_{(j)}$, but $\varrho(\mathbf{U}_{(j)}, \hat{\mathbf{V}}_{(j)}) \leq \varrho(\mathbf{U}_{(j-1)}, \mathbf{V}_{(j-1)})$ is always true. Therefore, $(\mathbf{U}, \mathbf{V})$ can be updated with $(\mathbf{U}_{(j)}, \hat{\mathbf{V}}_{(j)})$.

(ii) For the whole CCDL, there actually exist three loops (indexed by i , j , and k , resp.). For the loop indexed by k , (44) and (45) indicate that the procedure of updating $\mathbf{V}$ makes $\varrho(\mathbf{U}_{(j)}, \mathbf{V}(k))$ decrease as k increases. So the solution (or an approximate one) of (37) can be gotten. Besides, the solution of (31) is derived analytically as (36). All these result in the convergence of the second loop indexed by j , that is, dictionary updating. Assuming that OMP performs perfectly in the sparse coding stage, the nonincreasing trend of Step 1 is ensured. To sum up, the cost function (22) decreases in every step and hence the convergence of CCDL is guaranteed.

4. Experiment Results

In this section, we evaluate the performance of the proposed model and algorithm with synthetic data and audio signals.

FIGURE 1: Convergence performance of loop indexed by k with synthetic $\tilde{\mathbf{D}}$.

4.1. Convergence Performance. Firstly, several simulations will be carried out to verify the convergence performance of the proposed CCDL. As the main contributions of the new method lie in the dictionary updating stage, we focus on the performance of designing the orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$, that is, solving (26).

Set $N = 20$, $K = 80$, and number of signals $M = 1000$. There exist two loops in dictionary updating as introduced in the second point of Remark 5 indexed by k and j , respectively. The maximum iteration numbers for k and j are both fixed to 100.

4.1.1. For Synthetic Dictionary. $\mathbf{X}$ is taken as a $K \times M$ Gaussian random matrix. Two orthonormal matrices $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are generated to form the authentic dictionary $\tilde{\mathbf{D}}$:

$$\tilde{\mathbf{D}} = \tilde{\mathbf{U}} [\mathbf{I}_N \quad \mathbf{0}] \tilde{\mathbf{V}}^{\mathcal{T}}. \quad (47)$$

Then $\mathbf{Y}$ is produced as $\mathbf{Y} = \tilde{\mathbf{D}}\mathbf{X}$. The performance is evaluated by

$$\|\tilde{\mathbf{D}} - \hat{\mathbf{D}}\|_F^2 \triangleq e \quad (48)$$

with $\hat{\mathbf{D}}$ being the learnt dictionary.

Starting from an initial random dictionary, Figures 1 and 2 show the convergence performance of the loops indexed by k (with $j = 1$) and j , respectively.

Remark 6. (i) Seen from the minimum values of Figures 1 and 2 (very close to zero), they indicate that the design processes of $\mathbf{U}$ and $\mathbf{V}$ can result in a dictionary which is almost the same as the authentic one, $\tilde{\mathbf{D}}$.

(ii) The loop indexed by k is embedded in that of j . When $j = 1$, that is, the case of Figure 1, e already achieves the minimum value that is very close to zero. It manifests that the orthonormal matrix $\mathbf{V}$ plays a more important role in minimizing e . The result of Figure 2 also verifies this

FIGURE 2: Convergence performance of loop indexed by j with synthetic $\tilde{\mathbf{D}}$.

FIGURE 3: Convergence performance of loop indexed by k with random $\tilde{\mathbf{D}}$.

conclusion as the value of e converges in one iteration of the loop indexed by j . Review Remarks 1 and 3 that state that one of the main differences between the proposed algorithm and the method in [19] is the extra degree of freedom $\mathbf{V}$. So better sparse representation ability of the proposed algorithm can be expected.

Figures 1 and 2 show the efficiency of the proposed CCDL, where the authentic dictionary is generated with an ideal format as (47). In what follows, random generated dictionary will be considered.

4.1.2. For Random Dictionary. In this case, the matrix $\mathbf{X}$, the authentic dictionary $\tilde{\mathbf{D}}$, and the initial dictionary are all chosen randomly of proper dimensions without any correlation. $\mathbf{Y}$ is produced as $\mathbf{Y} = \tilde{\mathbf{D}}\mathbf{X}$. Figures 3 and 4 depict

FIGURE 4: Convergence performance of loop indexed by j with random $\tilde{\mathbf{D}}$.

the convergence performance of the loops indexed by k (with $j = 1$) and j , respectively.

The phenomena observed in this case are similar to the previous one, except for the fact that the value of e is larger compared to the case with synthetic $\tilde{\mathbf{D}}$. It should be pointed out that, for random $\tilde{\mathbf{D}} = \tilde{\mathbf{U}} [\tilde{\Sigma} \ \mathbf{0}] \tilde{\mathbf{V}}^T$,

$$e \geq \|\tilde{\Sigma} - \mathbf{I}_N\|_F^2 \quad (49)$$

always holds [25]. This explains why the minimum of e cannot approach zero in this case.

4.2. Simulations with Synthetic Data. We now carry out experiments to illustrate the performance of dictionaries learnt with different approaches. As comparisons, algorithms in [11, 19, 20] will be performed. For convenience, the learning systems are denoted as Dict_{new} , $\text{Dict}_{\text{ksvd}}$, Dict_{BP} , and Dict_{SDB} for the proposed CCDL and the methods in the references just mentioned, respectively.

We generate two $N \times K$ dictionaries $\mathbf{D}^{(0)}$ and $\tilde{\mathbf{D}}$, both with normally distributed entries. $\mathbf{D}^{(0)}$ is used as the initial condition for executing different learning algorithms, and $\tilde{\mathbf{D}}$ is the authentic dictionary. A set of M S -sparse $K \times 1$ vectors $\{\mathbf{x}_m\}_{m=1}^M$ is produced, where each nonzero element of $\mathbf{x}_m$ is randomly positioned with a Gaussian distribution of i.i.d that has zero mean and unit variance. With the authentic dictionary $\tilde{\mathbf{D}}$, the set of signal vectors $\{\mathbf{y}_m\}_{m=1}^M$ is generated by $\mathbf{y}_m = \tilde{\mathbf{D}}\mathbf{x}_m, \forall m$, for training the dictionaries.

Set $N = 20$, $K = 80$, $S = 5$, and $M = 5000$, and the number of iterations for dictionary learning is fixed $\text{Iter}_1 = 100$ for all the four methods. Besides, the number of iterative projections and rotations in [19] is set to 100, and the gradient descent is executed 100 times with step size equaling to 0.1 in [20]. For CCDL, the maximum iteration numbers for the loops indexed by k and j are fixed as 100 and 10, respectively.

The mutual coherence performances of different dictionaries are compared and the results are shown in Figure 5. For

FIGURE 5: The mutual coherence performance. (a) Results of Dict_{BP} versus η . (b) Results of Dict_{SDB} versus $\gamma = 2^p$.

Dict_{BP}, the horizontal axis refers to the constant η which controls the searching space $\mathcal{S}_\eta$ in (12), while for Dict_{SDB} the horizontal axis indicates the weighting factor γ for $\gamma = 2^p$ with p being an integer varying within $[-10, 10]$. In order to have clear comparisons, some results beyond certain ranges have been omitted, mainly concerning Dict_{BP} with too small η .

With synthetic data, we test the sparse representation abilities of the learnt dictionaries. The representation accuracy is usually quantified with the mean square error (MSE) defined as [11]

$$\sigma_{\text{mse}} \triangleq \frac{1}{N \times M} \sum_{m=1}^M \|\hat{\mathbf{y}}_m - \mathbf{y}_m\|_2^2, \quad (50)$$

where $\hat{\mathbf{y}}_m = \hat{\mathbf{D}}\hat{\mathbf{x}}_m$ is the reconstructed signal with $\hat{\mathbf{D}}$ being the output dictionary of Dict_{new}, Dict_{ksvd}, Dict_{BP}, or Dict_{SDB} and $\{\hat{\mathbf{x}}_m\}_{m=1}^M$ being the corresponding coefficients of $\{\mathbf{y}_m\}_{m=1}^M$ in different dictionaries calculated by OMP. Figure 6 depicts the MSE results of different systems.

Remark 7. (i) As known, for Dict_{BP}, when η approximates to 1, the system regresses to Dict_{ksvd}. The results in Figures 5 and 6 have confirmed this conclusion with the cases where $\eta = 1$. Besides, if too small η is chosen, the MSE performance of Dict_{BP} degenerates, though small mutual coherence values would be achieved.

(ii) The results of Dict_{SDB} fluctuate a lot for both tests. Though some surprisingly good performance is achieved, this superiority is highly sensitive to the data. As will be seen in the next experiment, when musical audio signals are tested, the fluctuations of Dict_{SDB} become gentle.

(iii) Seen from the recovery accuracy indicator that is most crucial for evaluating the systems' performance, that is, Figure 6, the results of Dict_{new} are superior to those of the other three methods in most of the cases.

4.3. Experiments with Musical Audio Signals. The effectiveness of all the algorithms will be evaluated for audio signal coding task which is popularly used for testing the performance of incoherent dictionaries [19, 20].

The audio signals are selected from the "testMusic16kHz" set of SMALLbox [26]. Just like the operations in [19], we divide the recording into 50% overlapping blocks of 256 samples with rectangular windows and arrange the resulting time-domain signals as columns of the training data matrix $\mathbf{Y}$. For each musical excerpt, the resulting $\mathbf{Y} \in \mathbb{R}^{256 \times 624}$; that is, $N = 256$ and $M = 624$. An overcomplete Gabor dictionary of size 256×512 is used as the initialization which means $K = 512$. The sparsity level is fixed as $S = 12$ for all the tests. The numbers of iterations are kept with the same settings as used for that of synthetic data. When learning the dictionaries, OMP algorithm is applied for the sparse coding stage.

The recovery accuracy is quantified with the signal-to-noise ratio (SNR) defined as [19]

$$\sigma_{\text{snr}} \triangleq 20 \lg \frac{\|\mathbf{Y}\|_F}{\|\mathbf{Y} - \mathbf{D}\hat{\mathbf{X}}\|_F} \quad (51)$$

with $\mathbf{D}$ being the output dictionary of each learning method and $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_m\}_{m=1}^M$ being the sparse coefficient matrix corresponding to $\mathbf{Y} = \{\mathbf{y}_m\}_{m=1}^M$ calculated from

$$\begin{aligned} \hat{\mathbf{x}}_m &\triangleq \arg \min_{\mathbf{x}_m} \|\mathbf{x}_m\|_1 \\ \text{s.t. } &\|\mathbf{y}_m - \mathbf{D}\mathbf{x}_m\|_2^2 \leq \epsilon, \\ &\forall m \end{aligned} \quad (52)$$

by the l_1 -Homotopy algorithm [16] with $\epsilon = 0.001$.

FIGURE 6: The MSE performance. (a) Results of Dict_{BP} versus η . (b) Results of Dict_{SDB} versus $\gamma = 2^p$.

FIGURE 7: The mutual coherence performance for each of the learnt dictionaries. (a) Results of Dict_{BP} versus η . (b) Results of Dict_{SDB} versus $\gamma = 2^p$.

For all Dict_{new}, Dict_{ksvd}, Dict_{BP}, and Dict_{SDB}, we test each of the ten musical excerpts in “testMusic16kHz” set and keep the average results for comparison. The mutual coherence behavior for each of the learnt dictionaries is depicted in Figure 7.

When recovering with l_1 -Homotopy, the SNR performance is shown in Figure 8.

Remark 8. In this case, where musical audio signals are tested and l_1 -Homotopy algorithm is applied for signal reconstruction, the fluctuations of Dict_{SDB} become gentle for both SNR and mutual coherence performance versus the weighting factor γ . For Dict_{BP}, the results of mutual coherence and SNR are more uniform; that is, smaller mutual coherence leads to higher SNR value. The performance of

FIGURE 8: The recovery SNR performance with l_1 -Homotopy used for reconstruction. (a) Results of Dict_{BP} versus η . (b) Results of Dict_{SDB} versus $\gamma = 2^p$.

Dict_{new} is superior to the others as these results are obtained by averaging ten musical excerpts.

5. Conclusion

In this paper, we have investigated the problem of learning incoherent dictionary. The contributions are threefold. The first one is to have proposed a novel model for IDL which considers minimizing the sparse representation error according to the training signals under the coherence constraint by making the Gram matrix of the dictionary approximate to the identity matrix. An alternating minimization algorithm named CCDL has been presented for solving the learning problem and the solution of each component of the optimum dictionary is derived analytically as the second contribution. The last one is to carry out experiments on synthetic data and musical audio signals to demonstrate the superiority of the proposed model and algorithm.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work was supported by Grants of NSFCs 61273195, 61473262, and 61503339, ZJNSF LQ14F030008, and Zhejiang Hua Yue Institute of Information and Data Processing.

References

- [1] A. M. Bruckstein, D. L. Donoho, and M. 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.
- [2] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, NY, USA, 2010.
- [3] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [4] A. Li and H. Shouno, "Dictionary-based image denoising by fused-lasso atom selection," *Mathematical Problems in Engineering*, vol. 2014, Article ID 368602, 10 pages, 2014.
- [5] Z. Jian, H. Luxi, J. Jian, and X. Yu, "A fast iterative pursuit algorithm in robust face recognition based on sparse representation," *Mathematical Problems in Engineering*, vol. 2014, Article ID 683494, 11 pages, 2014.
- [6] X. Zhu and F. Kui, "Image recovery algorithm based on learned dictionary," *Mathematical Problems in Engineering*, vol. 2014, Article ID 964835, 6 pages, 2014.
- [7] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, and L. Chang, "Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1581–1594, 2015.
- [8] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, no. 2, pp. 337–365, 2000.
- [9] I. Tosic and P. Frossard, "Dictionary learning: what is the right representation for my signals," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.
- [10] K. Engan, S. O. Aase, and J. H. Hakon-Housoy, "Method of optimal direction for frame design," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2443–2446, Phoenix, Ariz, USA, 1999.
- [11] M. Aharon, M. Elad, and A. 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.

- [12] S. K. Sahoo and A. Makur, "Dictionary training for sparse representation as generalization of K-means clustering," *IEEE Signal Processing Letters*, vol. 20, no. 6, pp. 587–590, 2013.
- [13] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [14] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [15] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [16] M. S. Asif and J. Romberg, "Sparse recovery of streaming signals using $\mathbf{I}_1$ -homotopy," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4209–4223, 2014.
- [17] M. Yaghoobi, L. Daudet, and M. E. Davies, "Parametric dictionary design for sparse coding," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4800–4810, 2009.
- [18] B. Maill, D. Barchiesi, and M. D. Plumbley, "INK-SVD: learning incoherent dictionaries for sparse representations," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 3573–3576, IEEE, Kyoto, Japan, March 2012.
- [19] D. Barchiesi and M. D. Plumbley, "Learning incoherent dictionaries for sparse approximation using iterative projections and rotations," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2055–2065, 2013.
- [20] C. D. Sigg, T. Dikk, and J. M. Buhmann, "Learning dictionaries with bounded self-coherence," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 861–864, 2012.
- [21] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [22] T. Strohmer and J. Heath, "Grassmannian frames with applications to coding and communication," *Applied and Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257–275, 2003.
- [23] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing," *Signal Processing*, vol. 92, no. 4, pp. 999–1009, 2012.
- [24] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2887–2898, 2013.
- [25] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 2nd edition, 2013.
- [26] I. Damnjanovic, M. E. P. Davies, and M. D. Plumbley, "SMALLbox-an evaluation framework for sparse representations and dictionary learning algorithms," in *Proceedings of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA '10)*, pp. 418–425, St. Malo, France, 2010.

Hindawi

Submit your manuscripts at
<http://www.hindawi.com>