

Research Article

Analysis Sparse Representation for Nonnegative Signals Based on Determinant Measure by DC Programming

Yujie Li ¹, Benying Tan,² Atsunori Kanemura ¹, Shuxue Ding,² and Wuhui Chen ³

¹National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

²School of Computer Science and Engineering, University of Aizu, Aizuwakamatsu, Japan

³School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

Correspondence should be addressed to Wuhui Chen; chenwuh@mail.sysu.edu.cn

Received 1 September 2017; Accepted 15 March 2018; Published 24 April 2018

Academic Editor: Tsendsuren Munkhdalai

Copyright © 2018 Yujie Li 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.

Analysis sparse representation has recently emerged as an alternative approach to the synthesis sparse model. Most existing algorithms typically employ the ℓ_0 -norm, which is generally NP-hard. Other existing algorithms employ the ℓ_1 -norm to relax the ℓ_0 -norm, which sometimes cannot promote adequate sparsity. Most of these existing algorithms focus on general signals and are not suitable for nonnegative signals. However, many signals are necessarily nonnegative such as spectral data. In this paper, we present a novel and efficient analysis dictionary learning algorithm for nonnegative signals with the determinant-type sparsity measure which is convex and differentiable. The analysis sparse representation can be cast in three subproblems, sparse coding, dictionary update, and signal update, because the determinant-type sparsity measure would result in a complex nonconvex optimization problem, which cannot be easily solved by standard convex optimization methods. Therefore, in the proposed algorithms, we use a difference of convex (DC) programming scheme for solving the nonconvex problem. According to our theoretical analysis and simulation study, the main advantage of the proposed algorithm is its greater dictionary learning efficiency, particularly compared with state-of-the-art algorithms. In addition, our proposed algorithm performs well in image denoising.

1. Introduction

Real signals around our daily life are always distributed in a high dimensional space; however low-dimensional structures are found in the signals, then we can represent the signals with a proper model by only a few parameters [1]. A proper model should be simple while matching the signals. In the past decades, the sparse and redundant representation model has been proven to be an efficient and beneficial model [2–5]. The theoretical background for sparse models is given by compressed sensing (CS) [6–8]. CS mathematically declares that if a signal is sparse or compressive, this original signal can be reconstructed by a few measurements, which are much fewer than the counts suggested by previous theories [6, 7, 9–11]. Sparse representation has also been described as an extraordinary powerful solution for a wide range of real-world applications, especially in image processing, such as image denoising, deblurring, inpainting, restoration,

superresolution, and also in the field of machine learning, computer vision, and so on [12–21].

Sparse representation can be formulated by either a synthesis model or an analysis model. The synthesis model is popular and mature. The analysis model has been less investigated for sparse representation, although several analysis dictionary learning algorithms have been proposed, such as the analysis K-SVD [12], Greedy Analysis Pursuit (GAP) [22], and the analysis thresholding algorithm [23].

In practice, some signals such as chemical concentrations in experimental results and pixels in video frames and images have inherent nonnegativity, and dedicated factorization methods have been proposed [24, 25]. The above analysis sparse representation algorithms are all for general signals which contain nonnegative and negative elements. The methods for general signals directly applied to nonnegative signals cannot achieve satisfying results. An existing analysis dictionary learning method for nonnegative signals [26],

which uses blocked determinants as the sparseness measure, is quite difficult or computationally expensive. The purpose of this paper is to address the problem on nonnegative analysis sparse representation.

In this paper, we present a novel algorithm of analysis sparse representation for nonnegative signals, which is parallel to the synthesis sparse representation in the principle and structure. Though this model has been studied in the past, there is still not a matured field for nonnegative analysis representation, while the algorithms designed for general signal cannot sufficiently be applied to the nonnegative signals. Thus, we focus on the nonnegative sparse representation with the analysis model. We cast the analysis sparse representation into three subproblems (analysis dictionary update, sparse coding, and signal recovery) and use an alternating scheme to obtain an optimization solution. We utilize the determinant-type of sparseness measure as the sparseness constraint, which is convex and differentiable. The objective function for sparse coding is nonconvex, and standard convex optimization methods cannot be employed. Fortunately the objective function is the difference of two convex functions, then we can introduce difference of convex (DC) programming to solve this nonconvex optimization problem.

The remainder of this paper is organized as follows. The conventional sparse representation problem is reviewed in Section 2. In Section 3, we introduce the analysis representation. In Section 4, we describe the problem formulation for analysis sparse representation and present the optimization framework. The experiments described in Section 5 demonstrate the practical advantages of the proposed algorithms compared with state-of-the-art algorithms both with artificial and real-world datasets. Finally, we present our conclusions in Section 6.

1.1. Notations. Here we list notations used in this paper. A boldface uppercase letter like \mathbf{X} is defined as a matrix, and a lowercase letter like x_{ij} is defined as the ij th entry of \mathbf{X} . A boldface lowercase letter, such as \mathbf{x} , is defined as a vector, and a lowercase letter x_j is defined as the j th entry of \mathbf{x} . Matrix slices $\mathbf{X}_{:,i}$ and $\mathbf{X}_{:,j}$ are defined as the i th row and the j th column of matrix \mathbf{X} , respectively. The Frobenius norm of matrix \mathbf{X} is defined as $\|\mathbf{X}\|_F = (\sum_{i,j} |x_{ij}|^2)^{1/2}$. The determinant value of a matrix is denoted by $\det(\cdot)$. Note that, in this paper, all parameters take real values.

2. Preliminaries

2.1. Sparse Representation. Sparse representation decomposes observed signals into a product of a dictionary matrix which contains signal bases and a sparse coefficient matrix [13–17], and there are two different structures: synthesis model and analysis model. The synthesis model is the first proposed sparse model and more popular. We first review it in this section.

Assume that we want to model the signals $\mathbf{X} \in \mathbb{R}^{m \times N}$, where m is the signal dimensionality and N is the number of measurements. The synthesis sparse model suggests that the signals could be expressed as

$$\mathbf{X} = \mathbf{D}\mathbf{H} \quad (1)$$

or

$$\mathbf{X} \approx \mathbf{D}\mathbf{H}, \quad (2)$$

$$\text{s.t. } \|\mathbf{X} - \mathbf{D}\mathbf{H}\|_F^2 \leq \varepsilon,$$

where $\mathbf{D} \in \mathbb{R}^{m \times n}$ refer to as a dictionary, $\mathbf{H} \in \mathbb{R}^{n \times N}$ is a representation coefficient matrix, and $\varepsilon \geq 0$ is a small residual [27]. Here, n is the number of bases, which are also called dictionary atoms. We further assume that the representation matrix \mathbf{H} is sparse (i.e., many zero entries) to obtain sparse representations of the signals. Equations (1) or (2) mean that each signal can be represented as a linear combination of a few atoms from the dictionary matrix \mathbf{D} .

A key issue in the sparse representation is the choice of the dictionary which the observed signals are used to decompose. One choice is a predefined dictionary, such as discrete Fourier transform (DFT), discrete cosine transform (DCT), and wavelets [28], which can be employed for learning a signal-specific dictionary from observed signals. Another choice, the learned dictionary, results in better matching to the contents of signals. In addition, the learned dictionary often exhibits better performance compared to predefined dictionaries in real-world applications [29, 30].

Intriguingly, there exists a “twin” of the synthesis model called the analysis model [31]. Assume that there is a matrix $\mathbf{\Omega} \in \mathbb{R}^{n \times m}$ that produces a sparse coefficient matrix \mathbf{H} by being multiplied to the signal matrix: $\mathbf{H} = \mathbf{\Omega}\mathbf{X}$. This equation can be obtained as the solution to a minimization problem of the error function $\|\mathbf{H} - \mathbf{\Omega}\mathbf{X}\|_F$. Remarkably, this error function is convex, and standard optimization methods can be employed. Since error functions in the synthesis model are nonconvex, optimization in the analysis model is often easier. We call $\mathbf{\Omega} \in \mathbb{R}^{n \times m}$ the analysis dictionary. Atoms in the analysis dictionary $\mathbf{\Omega}$ are its rows, rather columns in the synthesis dictionary \mathbf{D} . The term “analysis” means the dictionary analyzes the signal to produce a sparse result [32]. To emphasize the difference between the analysis and synthesis models, the term “cosparsity” has been introduced in the literature [31, 33], which counts the number of 0-valued elements of $\mathbf{\Omega}\mathbf{X}$, that is, zero elements coproduced by $\mathbf{\Omega}$ and \mathbf{X} [34]. The analysis sparse model is also called the cosparsity model, and then the analysis dictionary is also called the cosparsity dictionary.

Now we look more closely at the analysis sparse model. The analysis model for one signal $\mathbf{x} \in \mathbb{R}^m$, which is a column in the signal matrix \mathbf{X} , can be represented using a proper analysis dictionary $\mathbf{\Omega} \in \mathbb{R}^{n \times m}$. The i th row, namely, i th atom, in $\mathbf{\Omega}$ is denoted by ω_i . We want to make the analysis representation vector $\mathbf{h} = \mathbf{\Omega}\mathbf{x}$ sparse. This is formulated by introducing a sparsity measure $M(\mathbf{h})$, so that it negatively behaves with the sparsity of \mathbf{h} , and minimizing $M(\mathbf{h})$ yields the sparsest solution:

$$\begin{aligned} \mathbf{\Omega} &= \underset{\mathbf{\Omega}}{\operatorname{argmin}} M(\mathbf{h}), \\ \text{s.t. } \mathbf{h} &= \mathbf{\Omega}\mathbf{x}. \end{aligned} \quad (3)$$

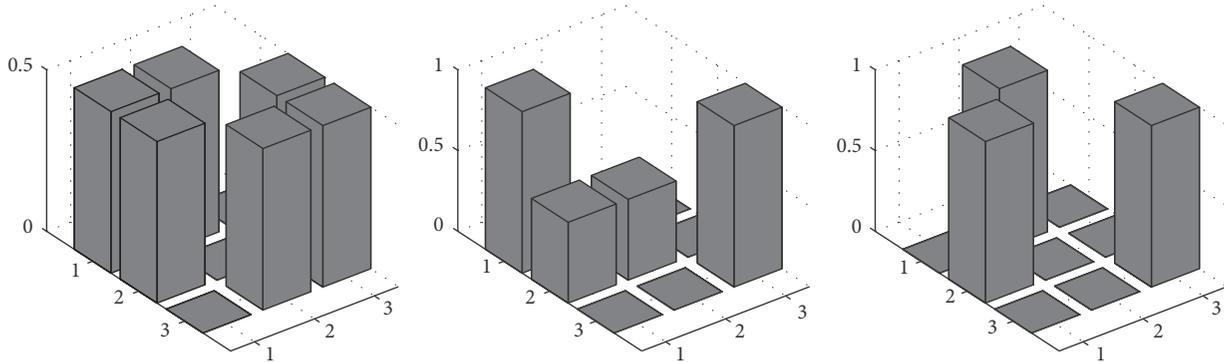


FIGURE 1: Illustration of various degrees of sparseness.

Although employing the ℓ_0 -norm, that is, setting $M(\mathbf{h}) = \|\mathbf{h}\|_0$, yields the sparsest solution [35], the optimization problem is combinatorial and often NP-hard. Therefore other sparsity measures such as the ℓ_1 -norm are employed to have easier optimization problems. Nevertheless, it is known that the ℓ_1 -norm often overpenalizes large elements and solutions are too sparse.

2.2. Sparseness Measure. The ℓ_p -norms, where $p = 0, 1$, or 2 , are popular measures for assessing the sparseness of a vector. Since the ℓ_0 -norm yields an NP-hard problem, its convex relaxation, the ℓ_1 -norm, is often preferred [36, 37]. The ℓ_1 -norm of a vector \mathbf{h} is defined to be the sum of the absolute values of \mathbf{h} ; namely, $\|\mathbf{h}\|_1 = \sum_i |\mathbf{h}_i|$. If the vector is nonnegative, that is, $\mathbf{h} \in \mathbb{R}_+$, the ℓ_1 -norm of \mathbf{h} is just $\|\mathbf{h}\|_1 = \sum_i h_i$. For nonnegative vectors, their ℓ_1 -norm is differentiable and smooth, and gradient methods can be used in optimization. Some authors introduce the ℓ_2 -norm with nonnegative matrix factorization since the nonnegative constraints yield sparse solutions. However, the results with the ℓ_2 -norm are not sparser than those with the ℓ_0 -norm or ℓ_1 -norm [38].

The sparsity measures mentioned above can reflect the instantaneous sparseness of one single signal [35], but they are not suitable for evaluating sparsity across different measurements [39]. In order to describe the joint sparseness of the nonnegative sources, we introduce determinant-type of sparsity measure. In spectral unmixing for remote sensing image interpretation, where signals are nonnegative, the determinant-type of sparsity measure [40] can get good results similar to the other sparseness-based methods from the results of numerical experiments [41]. Thus, the determinant-type measure can explicitly measure the sparseness of nonnegative matrices.

The determinant-type sparse measure has several good qualities. If a nonnegative matrix is normalized, we can find that the determinant value of the nonnegative matrix is well bounded and its value interpolates monotonously between two extremes of 0 and 1, along with the increasing of the sparsity. For instance, if the nonnegative matrix \mathbf{V} is nonsparse with its rows and satisfies sum-to-one, then the determinant of $\mathbf{V}\mathbf{V}^T$, $\det(\mathbf{V}\mathbf{V}^T)$, is close to 0. On the other hand, $\det(\mathbf{V}\mathbf{V}^T)$ approaches to 1 if and only if the matrix is

the most sparse [42]. Namely, the determinant value satisfies that $0 \leq \det(\mathbf{V}\mathbf{V}^T) \leq 1$, where $\det(\mathbf{V}\mathbf{V}^T) = 0$ if all entries of $\mathbf{V}\mathbf{V}^T$ are the same, and $\det(\mathbf{V}\mathbf{V}^T) = 1$ when the following two criteria are satisfied at the same time:

- (1) For all $i \in \{1, 2, \dots, m\}$, only one element in \mathbf{v}_i is nonzero.
- (2) For all $i, j \in \{1, 2, \dots, m\}$ and $i \neq j$, \mathbf{v}_i and \mathbf{v}_j are orthogonal, that is, $\mathbf{v}_i^T \mathbf{v}_j = 0$.

The detailed proof can be found in [40]. Thus we can use the determinant measure in the cost function. Figure 1 illustrates the sparseness degrees of three different matrices gauged by the determinant measure. The determinant values of the matrices from left to right are 0.0625, 0.5, and 1. We can see that the sparser the matrix is, the larger value the determinant measure is. Thus the sparse coding problem with determinant constraints can be expressed as an optimization problem

$$\max_{\mathbf{V}} \det(\mathbf{V}\mathbf{V}^T) = \min_{\mathbf{V}} -\det(\mathbf{V}\mathbf{V}^T). \quad (4)$$

2.3. Related Works. The existing methods for sparse representation with the analysis model employ various sparsity constraints. The analysis K-SVD algorithm [12] minimizes the error between the noisy (observed) signals and the estimated (reconstructed) signals with the ℓ_0 -norm as the sparsity constraint. The optimal backward greedy algorithm (OBG) is employed to estimate the analysis dictionary $\mathbf{\Omega}$, which has higher complexity. In [43–46], the ℓ_1 -norm is imposed as sparse constraint on the sparse coefficients and a projected subgradient-based algorithm is employed to learn the analysis dictionary. The analysis operator learning (AOL) algorithm [43] utilizing the ℓ_1 -norm constraint restricts the dictionary to a uniform normalized tight frame (UNTF) and learns the analysis dictionary as entire matrix by solving a cheap soft thresholding and then projecting the dictionary onto the UNTF set. Li et al. [26] propose a novel analysis sparse model based on the determinant-type measure, where the optimization problem is solved by an iterative sparseness maximization scheme. In their work, the problem can be cast into row-to-row optimization with respect to the analysis atoms and then utilize the quadratic programming (QP) technique to solve the optimization problem with respect

to each row. However this method has no signal recovery stage and cannot apply to denoising and other applications. In recently, DC programming is introduced to deal with the nonconvex optimization problems for sparse representation [47]. DC programming and DC algorithm (DCA) constitute the backbone of smoothness and nonsmoothness, nonconvex programming, and global optimization [48]. Thus it is considered to be a powerful tool for nonconvex optimizations [48, 49]. The exiting DC programming based sparse representation algorithms are with synthesis model, and we first propose DCA to analysis model.

3. Formulation

Now we turn to describe the problem of analysis sparse representation. We consider the following assumption: given an observed signal vector $\mathbf{y} \in \mathbb{R}_+^m$, we assume that \mathbf{y} is a noisy version of a signal $\mathbf{x} \in \mathbb{R}_+^m$. Thus, $\mathbf{y} = \mathbf{x} + \mathbf{u}$, where \mathbf{u} is additive positive white Gaussian noise. Using an analysis dictionary $\mathbf{\Omega} \in \mathbb{R}^{n \times m}$, each row of which defines $1 \times m$ analysis atom, we assume that \mathbf{x} satisfies $\|\mathbf{\Omega}\mathbf{x}\|_0 = m - l$, where l is the cosparsity of the signal defied to be the number of zero elements. Then we extend to signals matrix, and we use matrix \mathbf{X} to define signals with each column as one signal. Here we use $M(\cdot)$ as the sparse measure. Taking into account the noise in the measured signals, we formulate an optimization task for analysis dictionary learning by

$$\begin{aligned} \min_{\mathbf{\Omega}, \mathbf{X}} \quad & M(\mathbf{\Omega}\mathbf{X}), \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{X}\|_F^2 \leq \sigma, \\ & \mathbf{\Omega} \in \mathcal{L}, \end{aligned} \quad (5)$$

where σ is a noise-level parameter, \mathcal{L} is the constraint on the analysis dictionary, and M is the sparse regularization. We prefer to use an alternative, regularized version of above equation with penalty multipliers. Here we employ \mathbf{Z} as an approximation of $\mathbf{\Omega}\mathbf{X}$, which makes the learning easier and faster in practice. Thus, the analysis sparse coding is cheap and can be obtained exactly by thresholding the product of $\mathbf{\Omega}\mathbf{X}$ and the sparsity measure on \mathbf{Z} . Thus the analysis sparse representation can be rewritten as

$$\begin{aligned} \min_{\mathbf{\Omega}, \mathbf{X}, \mathbf{Z}} \quad & M(\mathbf{Z}) + \lambda \|\mathbf{Y} - \mathbf{X}\|_F^2 + \beta \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2, \\ \text{s.t.} \quad & \|\omega_i\|_2 = 1, \quad \forall i, \end{aligned} \quad (6)$$

where $\mathbf{Z} \in \mathbb{R}^{n \times N}$ is the representation coefficient matrix. Having a sparse representation means the representation matrix \mathbf{Z} is sparse. We introduce normalization constraint ($\forall i \|\omega_i\|_2 = 1$) to analysis dictionary which helps remove a scale ambiguity.

We can cast the entire problem into three variable-wise optimization subproblems. The first one is the analysis dictionary update subproblem and the next two problems are sparse coefficient matrix and signal update subproblems.

A pseudocode for such an analysis dictionary learning is presented in Algorithm 1 (ADLA).

(1) Initialization: $\mathbf{\Omega}_0, \mathbf{Z}_0, \mathbf{X}_0 = \mathbf{Y}, i = 0$,
(2) **while** not converged **do**
(3) $\mathbf{\Omega}_{i+1} = \min_{\mathbf{\Omega}} \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2$, s.t. $\forall i \|\omega_i\|_2 = 1$
(4) $\mathbf{Z}_{i+1} = \min_{\mathbf{Z}} M(\mathbf{Z}) + \beta \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2$
(5) $\mathbf{X}_{i+1} = \min_{\mathbf{X}} \lambda \|\mathbf{Y} - \mathbf{X}\|_F^2 + \beta \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2$
(6) $i = i + 1$
(7) **end while**

ALGORITHM 1: Analysis dictionary learning algorithm (ADLA).

3.1. Analysis Sparse Representation with Determinant Constraint. In this section, we introduce a novel determinant-type constrained sparse method to learn an analysis dictionary for sparse representation of nonnegative signals. What is more, we set the parameters λ and β in (6) as positive constants. Considering these factors, our minimization problem for analysis sparse dictionary learning with the determinant-type constraint can be expressed as

$$\begin{aligned} \min_{\mathbf{\Omega}, \mathbf{X}, \mathbf{Z}} \quad & -\det(\mathbf{Z}\mathbf{Z}^T) + \lambda \|\mathbf{Y} - \mathbf{X}\|_F^2 + \beta \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2, \\ \text{s.t.} \quad & \mathbf{Z} > 0; \\ & \|\omega_i\|_2 = 1, \\ & \|\mathbf{z}_i\|_2 = 1, \end{aligned} \quad (7)$$

$\forall i$.

The above objective function is nonconvex, which is hard to solve by traditional convex optimization methods.

4. Proposed Algorithm

The analysis representation with determinant-type constraint problem can be cast into three subproblems: an analysis dictionary update stage, a sparse coefficient coding stage, and a signal recovery stage. The corresponding formulations are as follows.

Analysis Dictionary Update Stage

$$\begin{aligned} \min_{\mathbf{\Omega}} \quad & \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2, \\ \text{s.t.} \quad & \|\omega_i\|_2 = 1, \quad \forall i. \end{aligned} \quad (8)$$

Sparse Coefficient Coding Stage

$$\begin{aligned} \min_{\mathbf{Z}} \quad & -\det(\mathbf{Z}\mathbf{Z}^T) + \beta \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2, \\ \text{s.t.} \quad & \mathbf{Z} > 0; \end{aligned} \quad (9)$$

$$\|\mathbf{z}_i\|_2 = 1, \quad \forall i. \quad (10)$$

Signal Recovery Stage

$$\min_{\mathbf{X}} \quad \lambda \|\mathbf{Y} - \mathbf{X}\|_F^2 + \beta \|\mathbf{\Omega}\mathbf{X} - \mathbf{Z}\|_F^2. \quad (11)$$

We will describe each stage in detail in the following subsections.

4.1. Analysis Dictionary Update. We use a projected subgradient type algorithm to solve the analysis dictionary update subproblem as follows:

$$\begin{aligned} \min_{\Omega} \quad & \|\Omega \mathbf{X} - \mathbf{Z}\|_F^2, \\ \text{s.t.} \quad & \|\omega_i\|_2 = 1, \quad \forall i. \end{aligned} \quad (12)$$

Since the objective function above is a quadratic function of Ω , the solution can be analytically found as the zero-gradient point. The gradient of the objective function is

$$\begin{aligned} \frac{\partial f(\Omega)}{\partial \Omega} &= (\Omega \mathbf{X} - \mathbf{Z}) \mathbf{X}^T \\ &= \Omega \mathbf{X} \mathbf{X}^T - \mathbf{Z} \mathbf{X}^T. \end{aligned} \quad (13)$$

By solving $\partial f(\Omega)/\partial \Omega = 0$, we obtain

$$\Omega = \mathbf{Z} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1}. \quad (14)$$

The projection of a dictionary onto a space with fixed row norms can be easily done for nonzero rows by scaling each row to have a unit norm. We use P_{UN} to denote this projection. If a row is zero, we set the row to a normalized random vector. Namely, P_{UN} is not uniquely defined. This is due to the fact that the set of uniformly normalized vectors is not convex. The projection can be found by

$$\begin{aligned} P_{\text{UN}}(\Omega) &= [P_{\text{UN}}(\omega_i)]_i, \\ P_{\text{UN}}(\omega) &= \begin{cases} \frac{\omega}{\|\omega\|_2} & \|\omega\|_2 \neq 0 \\ \nu & \text{otherwise,} \end{cases} \end{aligned} \quad (15)$$

where ν is a random vector on the unit sphere.

$$\mathbf{Z}^{(k+1)} = \underset{\mathbf{Z}}{\operatorname{argmin}} \left\{ \sum_{j=1}^N (\beta (\Omega \mathbf{X})_{:,j}^T (\Omega \mathbf{X})_{:,j} - 2\beta (\Omega \mathbf{X})_{:,j}^T \mathbf{Z}_{:,j} + \beta \mathbf{Z}_{:,j}^T \mathbf{Z}_{:,j} - \bar{\mathbf{Z}}_{:,j}^{(k)T} \mathbf{Z}_{:,j}) : \mathbf{Z} \in \mathbb{R}_+^n \right\}. \quad (21)$$

For the sparse coding stage in iteration scheme, we fix the dictionary Ω and signals \mathbf{X} and update only \mathbf{Z} in (16) and (17). To keep the coefficient matrix \mathbf{Z} nonnegative, the update procedure of sparse coefficient \mathbf{Z} can be briefly described as follows:

$$\mathbf{Z}_{:,j}^{(k+1)} = \left((\Omega \mathbf{X})_{:,j} + \frac{1}{2\beta} \bar{\mathbf{Z}}_{:,j}^{(k)} \right)_+, \quad \forall j, \quad (22)$$

where $(\cdot)_+ = \max(0, \cdot)$ takes the positive part. Algorithm 2 shows the procedures of the proposed algorithm for sparse coding based on DC programming.

4.2. Sparse Coding by DC Programming. Optimization for estimating sparse coefficient \mathbf{Z} is a core problem. The formulation is the subtraction of two convex functions, making the objective function not generally nonconvex. Thus we cannot use traditional convex optimization methods to solve this problem.

In this paper, to solve the objective function, we introduce the DC programming scheme to translate the minimization optimization problem into DC programming problem. According to the theory of DC programming [48, 49], we construct a DC function for our problem. Then the reformulated DC objective function F can be expressed as

$$F(\mathbf{Z}) = f_1(\mathbf{Z}) - f_2(\mathbf{Z}), \quad (16)$$

where

$$\begin{aligned} f_1(\mathbf{Z}) &= \beta \|\Omega \mathbf{X} - \mathbf{Z}\|_F^2, \\ f_2(\mathbf{Z}) &= \det(\mathbf{Z} \mathbf{Z}^T). \end{aligned} \quad (17)$$

Here $f_1(\mathbf{Z})$ and $f_2(\mathbf{Z})$ are two convex functions. Then, the sparse coding problem can be reformulated as a DC program:

$$\min \left\{ f_1(\mathbf{Z}) + \chi_{\mathbb{R}_+^{n \times N}}(\mathbf{Z}) - f_2(\mathbf{Z}) : \mathbf{Z} \in \mathbb{R}_+^{n \times N} \right\}, \quad (18)$$

where $\chi_{\mathbb{R}_+^n}(\mathbf{Z})$ is an indicator function, which is defined by $\chi_{\mathbb{R}_+^n}(\mathbf{Z}) = 0$ if $\mathbf{Z} \in \mathbb{R}_+^{n \times N}$ and $+\infty$ otherwise.

We utilize the scheme of DCA [48]. At each iteration k ,

$$\bar{\mathbf{Z}}^{(k)} = \nabla f_2(\mathbf{Z}^{(k)}) = (\det(\mathbf{Z} \mathbf{Z}^T)) (\mathbf{Z} \mathbf{Z}^T)^{-1} \mathbf{Z}. \quad (19)$$

At the iteration $k+1$,

$$\begin{aligned} \mathbf{Z}^{(k+1)} &= \underset{\mathbf{Z}}{\operatorname{argmin}} \left\{ f_1(\mathbf{Z}) - \langle \bar{\mathbf{Z}}^{(k)}, \mathbf{Z} \rangle : \mathbf{Z} \in \mathbb{R}_+^n \right\} \\ &= \underset{\mathbf{Z}}{\operatorname{argmin}} \left\{ \beta \|\Omega \mathbf{X} - \mathbf{Z}\|_F^2 - \langle \bar{\mathbf{Z}}^{(k)}, \mathbf{Z} \rangle : \mathbf{Z} \in \mathbb{R}_+^n \right\}. \end{aligned} \quad (20)$$

Consider that $\mathbf{Z}^{(k+1)}$ is separable according to the columns of \mathbf{Z} , the above equation can be rewritten as

4.3. Signal Recovery. Then we consider the signal recovery stage. In this stage the analysis dictionary Ω and the coefficients \mathbf{Z} are fixed. The optimization problem for signal recovery is formulated as

$$\min_{\mathbf{X}} f(\mathbf{X}) = \lambda \|\mathbf{Y} - \mathbf{X}\|_F^2 + \beta \|\Omega \mathbf{X} - \mathbf{Z}\|_F^2. \quad (23)$$

The above objective function (23) is a quadratic function of signals \mathbf{X} . Thus, the optimal solution of above objective function can be analytically obtained by setting its

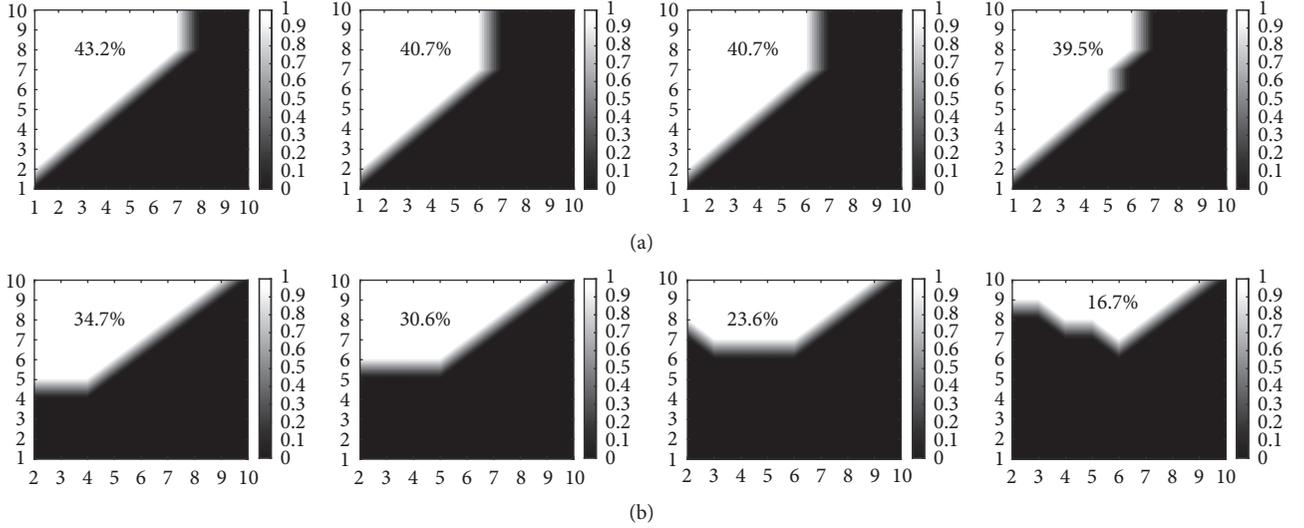


FIGURE 2: Illustration of phase transition: (a) determinant-type measure, $N = 50, 100, 150, 200$; (b) ℓ_1 -norm, $N = 50, 100, 150, 200$ (from left to right).

Require: $m \times N$ matrix \mathbf{X} , $n \times m$ matrix $\mathbf{\Omega}$, and $m < n < N$.

- (1) Initialize \mathbf{Z} , let $k = 1$
- (2) **while** not converged **do**
- (3) Compute $\bar{\mathbf{Z}}^{(k)}$ using (19).
- (4) Update $\mathbf{Z}^{(k+1)}$ using (22).
- (5) **end while**

ALGORITHM 2: Sparse coding based on DC programming (SCDC).

gradient equal to zero. The gradient of the objective function is

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \lambda(\mathbf{X} - \mathbf{Y}) + \beta \mathbf{\Omega}^T (\mathbf{\Omega} \mathbf{X} - \mathbf{Z}). \quad (24)$$

By solving $\partial f(\mathbf{X})/\partial \mathbf{X} = 0$, we obtain

$$\mathbf{X} = \frac{\lambda \mathbf{Y} + \beta \mathbf{\Omega}^T \mathbf{Z}}{\lambda \mathbf{I} + \beta \mathbf{\Omega}^T \mathbf{\Omega}}. \quad (25)$$

5. Experiments and Discussion

This section presents the results of numerical experiments to evaluate the performance of the proposed algorithms. The programs were coded in Matlab (R2016b) and run on a machine with a 3.3 GHz Intel Core i7 CPU and 16 GB of memory under Microsoft Windows 10.

5.1. Phase Transition. The goal of this phase transition experiment is to study how the proposed algorithm with the determinant-type sparsity measure can identify nonzero elements in a sparse matrix $\mathbf{Z} \in \mathbb{R}^{m \times N}$. Specifically, we examine when the success or failure of nonzero element identification switches according to the sparsity of the matrix.

We varied the number s of nonzero elements in sparse matrix \mathbf{Z} , the number m of the dimensionality of the dictionary atoms, and the number of measurements N . The range for s was from 1 to 10, m from 1 to 10, and N from 50 to 200.

Figure 2 compares the phase transition diagrams of the determinant-type measure and the ℓ_1 -norm. The white regions indicate success of identifying nonzero elements, whereas the black regions indicate failure. It can be seen that the area of white for the determinant-type measure is larger than that of the ℓ_1 -norm. It demonstrates that the determinant-type measure outperforms the ℓ_1 -norm especially when the number of measurements N increases.

5.2. Dictionary Recovery. This subsection examines if the proposed algorithm can recover the true analysis dictionary. To quantitatively evaluate the performance, we used synthetic signals with a known ground-truth dictionary.

Then, to evaluate the sparsity of the coefficient, we use the Hoyer sparsity [50], which can be normalized to satisfy more of the sparsity constraints [51]:

$$\text{Hoyer}_{\text{Sparsity}}(\mathbf{x}) = \frac{\sqrt{n} - (\sum |x_i|) / \sqrt{\sum x_i^2}}{\sqrt{n} - 1}, \quad (26)$$

where \mathbf{x} is a column of the coefficient matrix. The Hoyer sparsity for a matrix is defined as

$$\text{Hoyer}_{\text{Sparsity}}(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \text{Hoyer}_{\text{Sparsity}}(\mathbf{x}_i). \quad (27)$$

The larger the Hoyer sparsity measure, the sparser the coefficient matrix. Note that the sparse representation optimization problem with Hoyer sparsity constraint is hard to solve; thus it is generally used for evaluation.

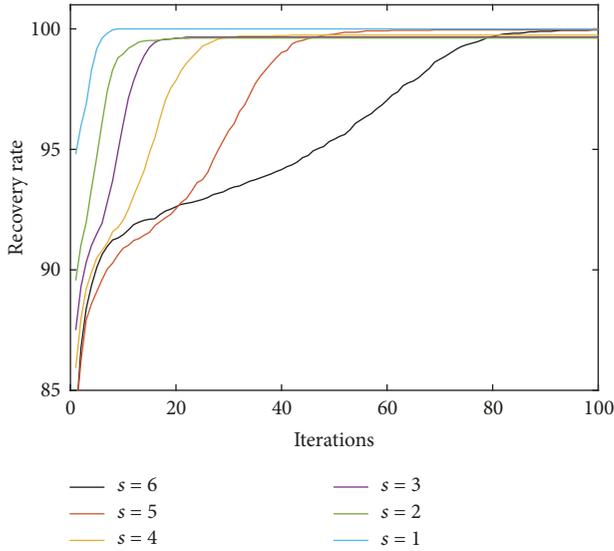


FIGURE 3: The recovery curves.

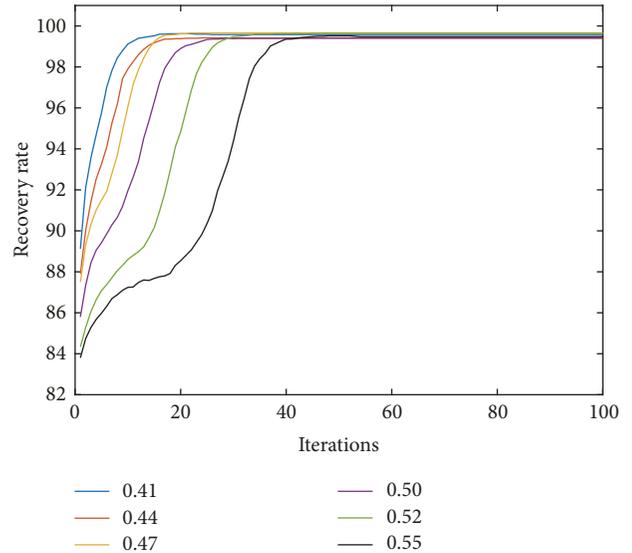


FIGURE 5: The recovery curves.

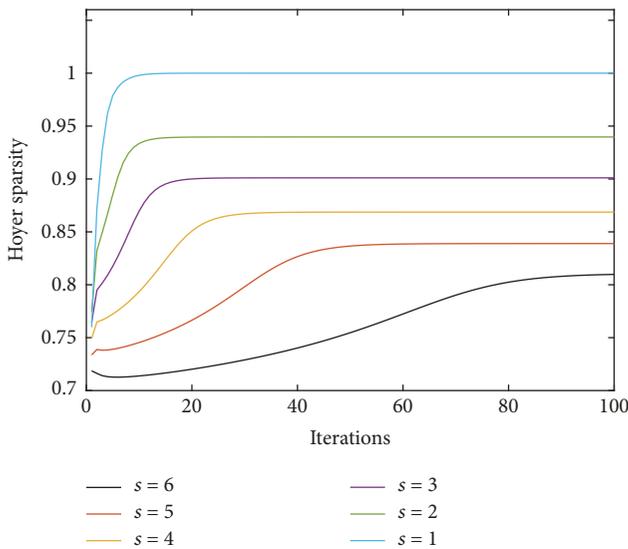


FIGURE 4: The Hoyer sparsity.

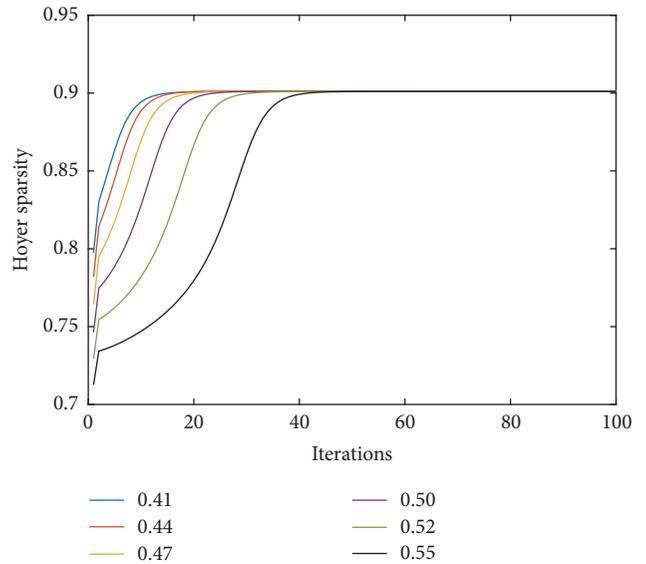
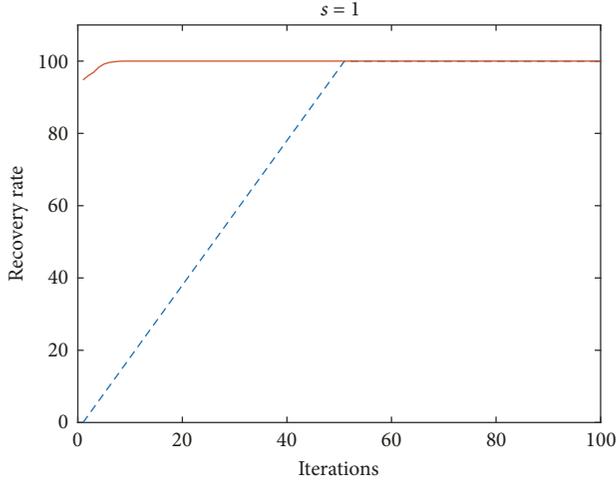


FIGURE 6: The Hoyer sparsity.

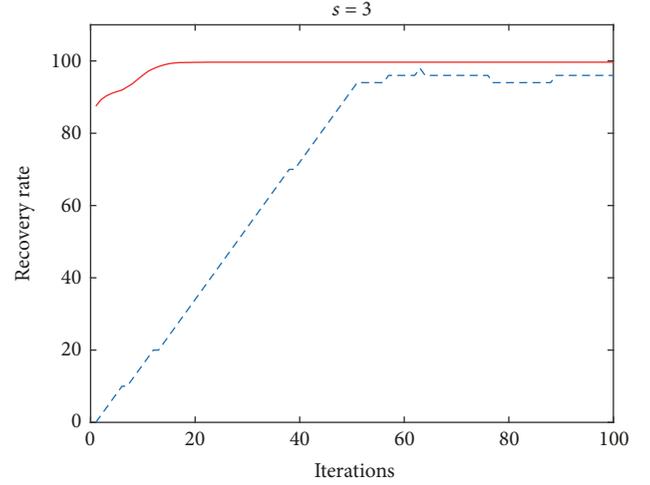
5.2.1. *Experimental Setup.* A ground-truth dictionary and observed signals were built as follows. A random synthesis dictionary $\mathbf{W} \in \mathbb{R}_+^{m \times m}$ was generated by taking the absolute values of i.i.d. zero mean, unit variance normal random variables. The analysis dictionary was set as $\mathbf{\Omega} = \mathbf{W}^{-1}$. A set of observation signals $\mathbf{X} \in \mathbb{R}_+^{m \times N}$ was generated by $\mathbf{X} = \mathbf{W}\mathbf{Z}$; namely, each signal column \mathbf{x} was generated by a linear combination of s different atoms in the dictionary, with s nonnegative corresponding coefficients in unity random and independent locations. Naturally, the elements of the observed signals became nonnegative with different sparsities. In this experiment, we chose a set of size $N = 1000$ of such observational signal matrices.

To initialize the proposed algorithm, we used a linear model to generate the initial estimate by combining the ground-truth dictionary $\mathbf{\Omega}_0$ and a normalized random matrix $\mathbf{\epsilon}$, that is, $\mathbf{\Omega}_{\text{init}} = (1-\rho)\mathbf{\Omega}_0 + \rho\mathbf{\epsilon}$, and then projecting it to satisfy the uniformly normalized and orthogonal constraints. When ρ is zero, we actually initialize $\mathbf{\Omega}_{\text{init}}$ with the ground-truth dictionary $\mathbf{\Omega}_0$ and when $\rho = 1$, the initial analysis dictionary will be random.

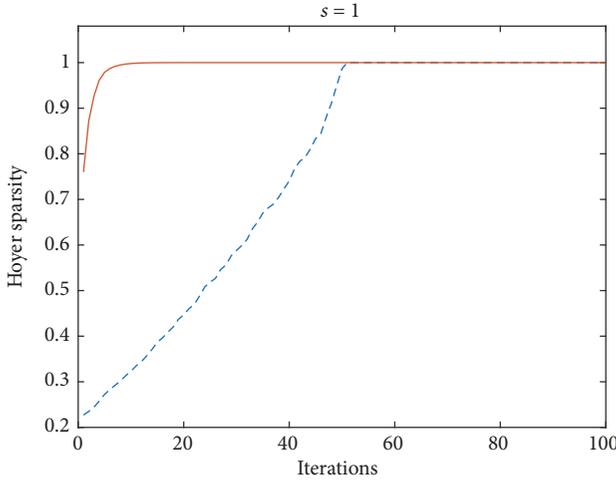
The learned analysis dictionary by our algorithm was compared with the ground-truth dictionary. Since there is a row-shuffle indeterminacy, we find corresponding rows by sweeping all rows of the learned dictionary and ground-truth dictionary and finding the closest rows between the two



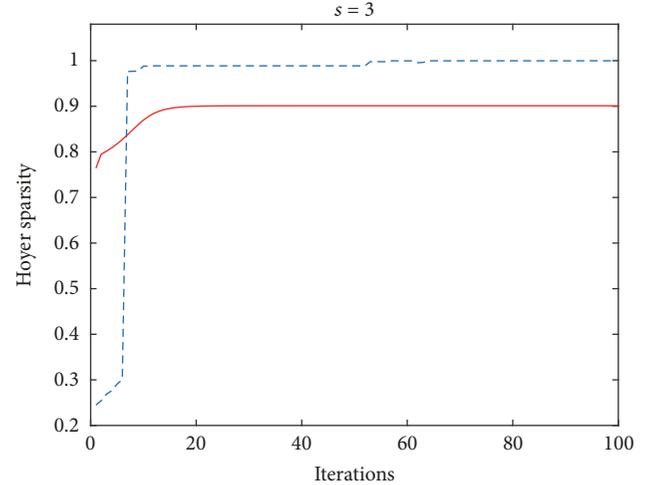
--- SADL-DET
— Proposed



— Proposed
--- SADL-DET



--- SADL-DET
— Proposed



— Proposed
--- SADL-DET

FIGURE 7: The comparison with SADL-DET when $s = 1$.

FIGURE 8: The comparison with SADL-DET when $s = 3$.

dictionaries. The distance less than 0.01 was considered as a success:

$$\max_j \min_i \left(1 - \left| \mathbf{\Omega}'_i{}^T \mathbf{\Omega}_{0j} \right| \right) < 0.01, \quad (28)$$

where $\mathbf{\Omega}'_i$ was an atom from the learned dictionary, and $\mathbf{\Omega}_j$ was an atom from the ground-truth dictionary.

First we use different number of nonzeros s . We set $s = 1, 2, 3, 4, 5, 6$ and $\rho = 0.9$ to see the performance between different sparsity. The Hoyer sparsities of ground-truth coefficient matrices are 1 ($s = 1$), 0.95 ($s = 2$), 0.91 ($s = 3$), 0.87 ($s = 4$), 0.84 ($s = 5$), and 0.81 ($s = 6$). Figure 3 presents the recovery curves of the analysis dictionary $\mathbf{\Omega}_0$ of size 50×50 , from which we can see that the results can reach to 100% in the cases of $s = 1, 5$, and 6 and more than 99% in the cases of $s = 2, 3$, and 4. Figure 4 presents the Hoyer sparsity of the learned sparse coefficients, which shows

the convergence of Hoyer sparsity value of different s and they nearly converge to the Hoyer sparsities of ground-truth coefficient matrices.

Next, we investigated the role of ρ . We chose $\rho = 0.7, 0.8, 0.9, 1.0, 1.1, 1.2$ with $s = 3$ and drew the recovery curves and the Hoyer sparsity. Figure 5 shows the recovery curves of the analysis dictionary with different ρ . From this figure, we can see that the ratio of recovery can reach about 100% in many different situations. Figure 6 presents the Hoyer sparsity of the learned sparse coefficient, which shows the Hoyer sparsity converges to the same value which nearly equal to the ground-truth Hoyer sparsity 0.91 ($s = 3$).

Then we compared our proposed algorithm with the existing determinant-type based algorithm SADL-DET [26], which is also applied for nonnegative signals. We compared the recovery rate and the Hoyer sparsity for these algorithms in the case of $s = 1, 3, 5$. Figure 7 shows the results in the case of $s = 1$, which shows our proposed algorithm converges

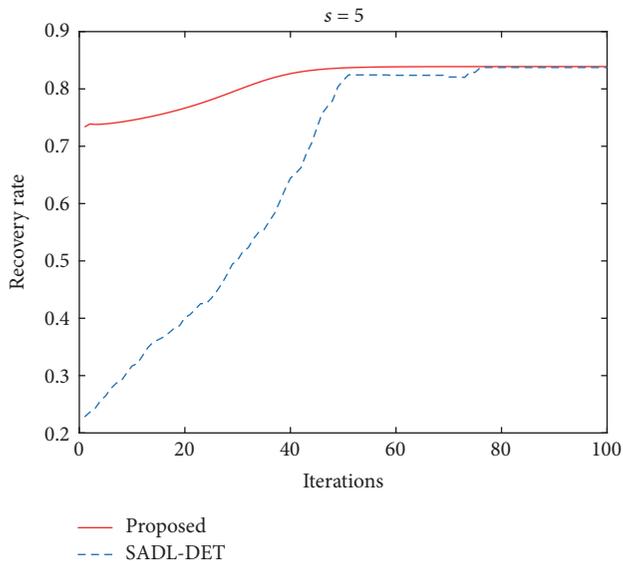
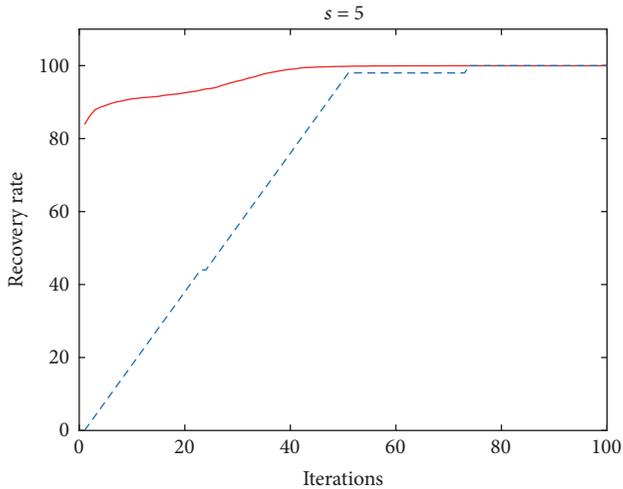


FIGURE 9: The comparison with SADL-DET when $s = 5$.

faster than SADL-DET. Figure 8 shows the results in the case of $s = 3$, which shows our proposed algorithm also converge faster in the recovery curve and can obtain a better recovery rate. Figure 9 shows the results in the case of $s = 5$, which shows that our proposed algorithm can achieve good results.

We compared the computational consumption. Figure 10 indicates that our proposed algorithm is nearly 10^3 times faster compared with SADL-DET.

5.3. Image Denoising. We tested our proposed method with a practical problem of image denoising [18]. The noise level and the quality of denoising were evaluated by the peak signal-to-noise ratios (PSNRs (PSNR (dB) is defined as $10\log_{10}(255^2/\|x - y\|)$, where x and y denote the original signal and the signal polluted by noise, resp.)). The tested images, as also the tested noise levels, are from the Cropped Yale face database (from <http://vision.ucsd.edu/extyaleb/CroppedYaleBZip/CroppedYale.zip>). Note that, to make sure

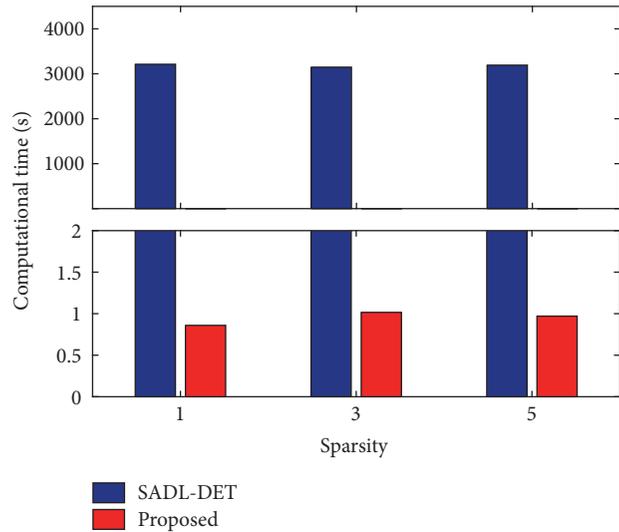


FIGURE 10: The computational consumption.

TABLE 1: Image denoising performance of PSNR (dB).

Face ID	Noisy level	COAOL	Proposed	FD
01	10.7662	11.0553	19.2743	11.0553
02	10.7675	11.0686	19.0550	11.0152
03	10.7381	11.0324	18.7899	10.9849
04	10.7542	11.0452	18.9949	11.0025
05	10.7918	11.0881	18.4613	11.0410
06	10.7502	11.0408	18.9840	10.9959
07	10.7844	11.0838	19.2423	11.0335
08	10.8030	11.0926	18.7126	11.0541
09	10.7905	11.0792	18.5088	11.0400
10	10.7528	11.0408	18.3358	11.0000
11	10.7764	11.0641	18.7527	11.0242
12	10.7623	11.0429	18.6075	11.0092
13	10.7268	11.0181	17.4434	10.9735
15	10.8005	11.0786	18.9708	11.0487
16	10.7784	11.0660	18.6266	11.0261
17	10.7745	11.0626	19.1770	11.0224
18	10.7546	11.0398	19.0336	11.0031
19	10.7442	11.0307	19.2140	10.9920
20	10.7765	11.0703	18.9491	11.0255

that the observed signals are nonnegative, the noises we added to the original images are nonnegative uniformly distributed pseudorandom values.

Figure 11 shows original, noisy, and denoised images by our proposed algorithm. Then, we compare our proposed algorithm with the analysis model based algorithm, Constrained Overcomplete Analysis Operator Learning (COAOL), and we also represent the denoised results using fixed Finite Difference (FD) operator. The average PSNRs of the denoised results for different face cases are presented in Table 1, which shows that our algorithm performs well

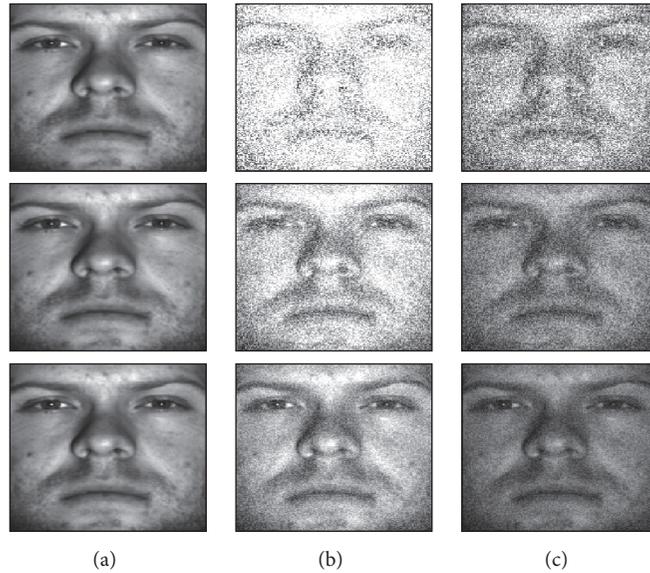


FIGURE 11: Face image denoising. (a) Original face images. (b) Noisy images. (c) Denoised faces using our proposed algorithm.

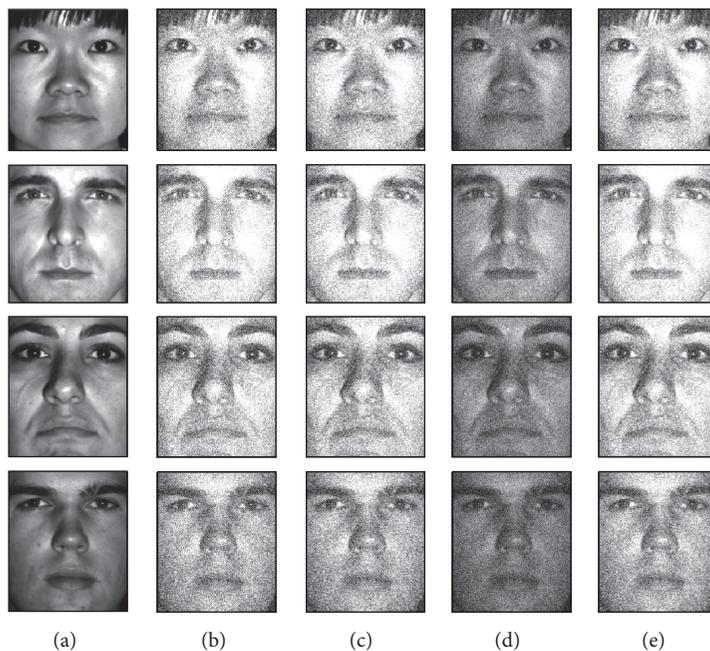


FIGURE 12: Face image denoising. (a) Original face images. (b) Noisy images. (c) Denoised faces using COAOL. (d) Denoised faces using our proposed algorithm. (e) Denoised faces using FD.

in image denoising. To intuitively see the denoising performance with different methods, we take faces 05, 10, 15, and 20, for instance. Figure 12 shows the original faces, noisy faces, and denoised faces with different methods, COAOL, ours, and FD. Although COAOL has a good performance for general signals, it cannot achieve good results for nonnegative noise. Our proposed algorithm has outstanding advantage in nonnegative image denoising.

6. Conclusion

In this study, we have proposed a novel and efficient analysis sparse representation algorithm with the determinant-type sparse measure, which focuses on nonnegative signals processing. We separated the whole problem into three subproblems: analysis dictionary updates, sparse coding, and signals recovery. In the sparse coding stage, we employ DC

programming to solve the nonconvex minimization problems. The experimental results verify the effectiveness of the proposed algorithm and the determinant measure of sparsity. Remarkable advantages of the proposed algorithm include the faster running time and good performance in image denoising. Moving forward, we can employ the proposed algorithm for more applications, such as image inpainting, superresolution, and other applications.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by NEDO and JSPS KAKENHI 26730130 and 16K00335.

References

- [1] R. G. Baraniuk, V. Cevher, and M. B. Wakin, "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.
- [2] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [3] M. Huang, W. Yang, J. Jiang et al., "Brain extraction based on locally linear representation-based classification," *NeuroImage*, vol. 92, pp. 322–339, 2014.
- [4] 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.
- [5] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.
- [6] D. L. Donoho, "Compressed sensing," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [7] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.
- [8] Y. Wang, J. Zeng, Z. Peng, X. Chang, and Z. Xu, "Linear convergence of adaptively iterative thresholding algorithms for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2957–2971, 2015.
- [9] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [10] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, 2006.
- [11] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
- [12] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 661–677, 2013.
- [13] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.
- [14] M. G. Jafari and M. D. Plumbley, "Fast dictionary learning for sparse representations of speech signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1025–1031, 2011.
- [15] M. J. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, vol. 52, no. 1, pp. 64–79, 2009.
- [16] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 3, pp. 922–932, 2012.
- [17] C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using Kronecker bases," *Neural Computation*, vol. 25, no. 1, pp. 186–220, 2013.
- [18] 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.
- [19] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proceedings of the 2011 IEEE International Conference on Computer Vision, ICCV 2011*, pp. 1259–1266, esp, November 2011.
- [20] Z. Li, S. Ding, Y. Li, Z. Yang, S. Xie, and W. Chen, "Manifold optimization-based analysis dictionary learning with an $\ell_{1/2}$ -norm regularizer," *Neural Networks*, vol. 98, pp. 212–222, 2018.
- [21] Z. Li, T. Hayashi, S. Ding, and Y. Li, "Dictionary learning with the $\ell_{1/2}$ -regularizer and the coherence penalty and its convergence analysis," *International Journal of Machine Learning and Cybernetics*, pp. 1–14, 2017.
- [22] R. Giryas, S. Nam, M. Elad, R. Gribonval, and M. E. Davies, "Greedy-like algorithms for the cosparsity analysis model," *Linear Algebra and its Applications*, vol. 441, pp. 22–60, 2014.
- [23] T. Peleg and M. Elad, "Performance guarantees of the thresholding algorithm for the cosparsity analysis model," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 59, no. 3, pp. 1832–1845, 2013.
- [24] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 29–47, 2006.
- [25] A. B. Hamza and D. J. Brady, "Reconstruction of reflectance spectra using robust nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3637–3642, 2006.
- [26] Y. Li, S. Ding, and Z. Li, "Dictionary learning with the cosparsity analysis model based on summation of blocked determinants as the sparseness measure," *Digital Signal Processing*, vol. 48, pp. 298–309, 2016.
- [27] Y. Li, S. Ding, and Z. Li, "A dictionary-learning algorithm for the analysis sparse model with a determinant-type of sparsity measure," in *Proceedings of the 2014 19th International Conference on Digital Signal Processing, DSP 2014*, pp. 152–156, Hong Kong, August 2014.
- [28] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.
- [29] M. Elad, *Sparse and Redundant Representations, From Theory to Applications in Signal and Image Processing*, Springer, New York, NY, USA, 2010.

- [30] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse image and signal processing*, Cambridge University Press, Cambridge, 2010.
- [31] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparsity analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.
- [32] S. Ravishanker and Y. Bresler, "Learning sparsifying transforms," *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1072–1086, 2013.
- [33] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "Cosparsity analysis modeling—uniqueness and algorithms," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP '11)*, pp. 5804–5807, Prague, Czech Republic, May 2011.
- [34] S. Hawe, M. Kleinstueber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2138–2150, 2013.
- [35] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, 2009.
- [36] 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.
- [37] C. A. Zala, I. Barrodale, and J. S. Kennedy, "High-resolution signal and noise field estimation using the L_1 (least absolute values) norm," *IEEE Journal of Oceanic Engineering*, vol. 12, no. 1, pp. 253–264, 1987.
- [38] Z. Li, S. Ding, Y. Li, Z. Tang, and W. Chen, "Improving dictionary learning using the Itakura-Saito divergence," in *Proceedings of the 2nd IEEE China Summit and International Conference on Signal and Information Processing, IEEE ChinaSIP 2014*, pp. 733–737, chn, July 2014.
- [39] B. Willmore and D. J. Tolhurst, "Characterizing the sparseness of neural codes," *Network: Computation in Neural Systems*, vol. 12, no. 3, pp. 255–270, 2001.
- [40] Z. Yang, Y. Xiang, S. Xie, S. Ding, and Y. Rong, "Nonnegative blind source separation by sparse component analysis based on determinant measure," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 10, pp. 1601–1610, 2012.
- [41] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.
- [42] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1112–1125, 2011.
- [43] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Constrained overcomplete analysis operator learning for cosparsity signal modelling," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2341–2355, 2013.
- [44] M. Yaghoobi and M. E. Davies, "Relaxed analysis operator learning," in *NIPS Workshop on Analysis Operator Learning vs. Dictionary Learning: Fraternal Twins in Sparse Modeling*, 2012.
- [45] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Noise aware analysis operator learning for approximately cosparsity signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 5409–5412, March 2012.
- [46] G. Peyré and Fadili. J. M., "Learning analysis sparsity priors," in *Proceedings of the International Conference on Sampling Theory and Applications (SampTA)*, 4 pages, 2011.
- [47] B. Tan, Y. Li, S. Ding, and X. Li, "Recovering nonnegative sparse signals with a determinant-type of sparse measure and dc programming," in *Proceedings of International Conference on Applied Computer and Communication Technologies (ComCom 2017)*, 2017.
- [48] H. A. Le Thi, X. T. Vo, and T. Pham Dinh, "Efficient nonnegative matrix factorization by DC programming and DCA," *Neural Computation*, vol. 28, no. 6, pp. 1163–1216, 2016.
- [49] L. T. An and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Annals of Operations Research*, vol. 133, pp. 23–46, 2005.
- [50] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [51] N. Hurley and S. Rickard, "Comparing measures of sparsity," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.

