Information Engineering Express
International Institute of Applied Informatics
2015, Vol.1, No.4, 1 – 10

Nonnegative Dictionary-Learning Algorithm Based on L_1 Norm with the Sparse Analysis Model

Yujie Li* , Shuxue Ding* , Zhenni Li* , Wuhui Chen*

Abstract

Sparse representation has been proven to be a powerful tool for analysis and processing of signals and images. Most of existing methods for sparse representation are based on the synthesis model. This paper presents a method for dictionary learning and sparse representation with the so-called analysis model. Different from the synthesis sparse model, in this analysis model, the analysis dictionary multiplying the signal can lead to a sparse outcome. The analysis dictionary learning problem has received less attention with and only a few algorithms has been proposed recently. What is more, there have still been few investigations in the context of dictionary learning for nonnegative signal representation. So, in this paper we focus on the nonnegative dictionary learning for signal representation. We use ℓ_1-norm as the sparsity measure to learn an analysis dictionary from signals in analysis sparse model. In addition, we adopt the Euclidean distance as the error measure in the formulation. Numerical experiments on recovery of analysis dictionary show that the proposed analysis dictionary learning algorithm performs well for nonnegative signal representation.

Keywords: nonnegative dictionary learning, sparse representation, ℓ_1-norm, analysis model.

1 Introduction

Situated at the heart of signal and image processing, data models are fundamental for stabilizing the solution, and enabling other tasks, such as signal compression, denoising, sampling, and so on [1]. A model is a set of mathematical relations that the data is believed to satisfy. Models are central in signal and image processing. How to choose a highly simple and reliable model is an essential task. Among the many ways, the sparse-based model has been proved to be important and useful. Most of existing methods for sparse representation are based on the synthesis sparse model. The signal can be sparsely represented using transform-domain methods, namely, the sparse representation for signals. This model can be described as $\mathbf{X} = \mathbf{WH}$, or $\mathbf{X} \approx \mathbf{WH}$ satisfying $\|\mathbf{X} - \mathbf{WH}\|_2 \leq \varepsilon$ [2]. Here $\mathbf{W} \in \mathbb{R}^{m\times n}$ is an overcomplete $(m < n)$ dictionary, implying that the dictionary is redundant. $\mathbf{X} \in \mathbb{R}^{m\times N}$ is the observed signal matrix, which can be represented as a linear combination of atom-

* School of Computer Science and Engineering, The University of Aizu, Aizu-Wakamatsu City, Fukushima, Japan

s of the dictionary. The matrix $\mathbf{H} \in \mathbb{R}^{n\times N}$ is the sparse representation coefficients of the signals $\mathbf{X}$ [3]. The number of non-zeros $||\mathbf{H}||_0$ is very small, and we say that $\mathbf{X}$ has a sparse representation in $\mathbf{W}$. The name "synthesis" comes from the relation $\mathbf{X} = \mathbf{WH}$, with the interpretation that the model describes a way to synthesize signal $\mathbf{X}$ [3]. The synthesis model has been the focus of many papers; it is safely to say that the synthesis model is a mature and stable field [3]. And this model has been heavily investigated in the last decade and many remarkable results achieved in almost all signal processing applications, the examples in [4][5].

Interestingly, the synthesis model has a "twin" model that takes an analysis point of view [3]. Assume that there is $\Omega \in \mathbb{R}^{n\times m}$ so that Ω can be used to find sparse $\mathbf{H}$ so that $\mathbf{H} = \Omega\mathbf{X}$ [6]. This analysis model (also called as cosparse model) can be converted as a minimization problem of the error function $||\mathbf{H} - \Omega\mathbf{X}||_F$, where the operator F denotes the Frobenius norm. Based on a series of experiments, it seems that the cosparse analysis model is very successful at accurately reconstructing signals from signals. They offer the advantages of low computational complexity and of being universally applicable to a wide set of signals. The analysis model relies on a matrix $\Omega \in \mathbb{R}^{n\times m}$, which we refer to as the analysis dictionary. Here "analysis" means the dictionary analyzes the signal to produce a sparse result [7]. Different from the synthesis model, the rows of analysis dictionary represent analysis atoms. The key point of this analysis model is the expectation that the analysis representation vector $\mathbf{H} = \Omega\mathbf{X}$ is sparse with many zeros, so that the computation will be easy.

Now we focus on the analysis model mentioned above. Here $\Omega \in \mathbb{R}^{n\times m}$ be a signal transformation or an analysis operator. Its rows in this matrix are the row vectors which refer to as atoms. To capture various aspects of the information in the signal, we typically have $(m \leq n)$. Since there is a relationship between the synthesis sparse dictionary and the analysis sparse dictionary by the equation $\mathbf{D} = \Omega^{\dagger}$. Then we can use analysis sparse representation to solve synthesis sparse problems. The analysis dictionary learning problem has received less attention with only a few algorithms proposed recently [1, 7, 8, 9, 10, 11, 12, 13]. In practice some particular signals such as spectral data have the limitations of nonnegativity. Experiments show that the previously mentioned algorithms, developed for general signals, cannot be efficiently applied to such nonnegative signals. But there are still few investigations in the context of nonnegative analysis dictionary learning. So we mainly focus on analysis sparse representation for nonnegative dictionary learning.

The remainder of this paper is organized as follows. Section II, we first introduce the analysis model with the sparsity constraints. In Section III, we describe the problem formulation of the analysis sparse model. Then we use ℓ_1 norm as the sparse constraint. The detailed algorithm is described in Section IV. In Section V we present experimental results for the proposed algorithm. Finally, conclusions are drawn in SectionVI .

2 The Analysis Sparse Model

Generally, the problem of finding the analysis dictionary Ω and the corresponding sparse representation $\mathbf{H}$ can be modeled by using the minimization of ℓ_0-norm. $||\mathbf{H}||_0$ counts the number of nonzero elements in the matrix of $\mathbf{H}$. While the results by ℓ_0-norm can lead $\mathbf{H}$ to be the sparsest, the algorithm by it is combinatorial optimization and is usually NP-hard. Therefore, some convex relaxations of ℓ_0, such as ℓ_1, have been proposed in the literature for the convenience of optimization in practice.

For the nonnegative matrix, ℓ_1-norm is either a way to measure the sparsity of the matrix. In the fields of super resolution and signal enhancement, the effectiveness and versatility of the ℓ_1 method indicate that it has a useful role [15]. The ℓ_1-norm of $\mathbf{H}$ can be written as $\|\mathbf{H}\|_1$, which is defined as $\mathbf{H}$. Namely, $\|\mathbf{H}\|_1 = \sum_{ij} |\mathbf{H}_{ij}|$. If $\mathbf{H} \in \mathbb{R}_+$, the ℓ_1 norm of $\mathbf{H}$ can simply written as $\|\mathbf{H}\|_1 \equiv \sum_{ij} \mathbf{H}_{ij}$. Compared with ℓ_0-norm, ℓ_1-norm is easier to solve. For nonnegative matrix, its ℓ_1-norm is derivative and smooth. We can use gradient descent method to solve the nonnegative ℓ_1-norm problem. It is clear that the smaller $\|\mathbf{H}\|_1$ is the sparser $\mathbf{H}$ is. In addition, some authors also impose sparsity constrains by using ℓ_2-norm, because of the particularity of the sparse nonnegative matrix factorization. In this paper, we adopt ℓ_1-norm as the measures of sparsity with analysis model for nonnegative sparse representation.

3 Problem Formulation

In analysis sparse representation dictionary learning, $\mathbf{X}$, Ω, and $\mathbf{H}$ denote the data samples matrix, the analysis dictionary matrix, and the corresponding coefficient matrix, respectively. m is the size of atoms in the analysis dictionary, n is the number of atoms, and N denotes the number of signal data samples.

Generally, the analysis sparse problem is formulated as following: given a signal data matrix $\mathbf{X}$, find two matrices Ω and $\mathbf{H}$ satisfying

$$\Omega\mathbf{X} = \mathbf{H}, \tag{1}$$

where $\Omega \in \mathbb{R}_+^{n\times m}$ $(n \geq m)$ and $\mathbf{H} \in \mathbb{R}_+^{n\times N}$ are all the nonnegative matrices. $\mathbf{X}$ can be generated by

$$\mathbf{X} = \Omega^{\dagger}\mathbf{H}, \tag{2}$$

here $\mathbf{X} \in \mathbb{R}^{m\times N}$, and $\Omega^{\dagger}$ is the pseudo inverse of the analysis dictionary Ω. In this way, we can get observed signal $\mathbf{X}$ from the analysis dictionary Ω and the sparse matrix $\mathbf{H}$. Although elements in Ω and $\mathbf{H}$ are all nonnegative, the pseudo inverse of Ω may contain negative elements. In this case, the observed signal may have negative elements.

The problem of analysis sparse representation can be formulated as the minimization of an objective function below,

$$\min D(\Omega\mathbf{X}|\mathbf{H}). \tag{3}$$

The popular choice is the Euclidean (EUC) distance which can be defined as,

$$\min D_{EUC}(\Omega\mathbf{X}|\mathbf{H}) = \frac{1}{2}\|\Omega\mathbf{X} - \mathbf{H}\|_F^2, \tag{4}$$

where the operator $\|\cdot\|_F$ denotes the Frobenius norm. The coefficient $\frac{1}{2}$ is in order to offset the derivative coefficient. Besides, there are other cost functions IS divergence and KL divergence, whose expressions are given as follows,

$$\min D_{IS}(\Omega\mathbf{X}|\mathbf{H}) = \frac{\Omega\mathbf{X}}{\mathbf{H}} - \ln\frac{\Omega\mathbf{X}}{\mathbf{H}} - 1, \tag{5}$$

$$\min D_{KL}(\Omega\mathbf{X}|\mathbf{H}) = \Omega\mathbf{X}\ln\frac{\Omega\mathbf{X}}{\mathbf{H}} - \Omega\mathbf{X} + \mathbf{H}. \tag{6}$$

In this paper, we take the popular (EUC) distance as cost function; while in the future work we could compare these different choices.

4 The Analysis Model with the ℓ_1-norm as the Sparsity Measure

We can introduce additional constraints to circumscribe the solutions set. For instance, the smoothness constraint adds for better spectral signatures, sparseness constraint for information processing demands, and volume constraint for meeting convex-geometry model. To solve the problem above, we can introduce additional constraint to circumscribe the solutions set. In this paper, the emphasis is the sparseness of coefficient matrix **H**, and we introduce ℓ_1-norm to enforce the sparseness of coefficient matrix **H**. We will express respectively in detail in the next section.

At first, we discuss the cost function of the analysis sparse representation with constraint of l_1-norm. The cost function imposed constraint can be written as,

$$\begin{aligned} &\min_{\Omega,\mathbf{H}} f(\Omega,\mathbf{H}) = \frac{1}{2}\|\Omega\mathbf{X}-\mathbf{H}\|_F^2 + \lambda\|\mathbf{H}\|_1, \\ &\text{Subject to} \quad \Omega \succeq 0, \mathbf{H} \succeq 0, \end{aligned} \tag{7}$$

where $\lambda \geq 0$ is regularization parameters which can be adjusted for controlling the tradeoff between the approximation error and the sparsity of the coefficient matrix **H**. $\succeq$ means all the elements in the matrices Ω and **H** are nonnegative.

This cost function is a convex, so that we can employ the alternating descent strategy to solve the above two convex optimal problems, finding the optimal factor Ω corresponding to a fixed factor **H** reduces to a convex optimization problem, and vice versa. Given the initial matrices $\Omega^{(0)}$ and $\mathbf{H}^{(0)}$, the sequences $\Omega^{(t)}$ and $\mathbf{H}^{(t)}$ are computed by the following formulas,

$$\Omega_{ij}^{(t)} = \Omega_{ij}^{(t-1)} - \eta\frac{\partial f(\Omega,\mathbf{H})}{\partial\Omega_{ij}}, \quad 1 \leq i \leq m, 1 \leq j \leq n, \tag{8}$$

$$\mathbf{H}_{ij}^{t} = \mathbf{H}_{ij}^{t-1} - \xi\frac{\partial f(\Omega,\mathbf{H})}{\partial\mathbf{H}_{ij}}, \quad 1 \leq i \leq n, 1 \leq j \leq N. \tag{9}$$

Scalar quantities η and ξ are the step lengths to take along the negative gradients. While we should set η and ξ carefully when updating Ω_{ij} and $\mathbf{H}_{ij}$. Setting them as some small positive constants are totally operable [16].

After several algebraic manipulations, the partial derivatives of the objective function $f(\Omega,\mathbf{H})$ with respect to Ω_{ij} and $\mathbf{H}_{ij}$ can be expressed in matrix vector form:

$$\frac{\partial f(\Omega,\mathbf{H})}{\partial\Omega} = \Omega\mathbf{X}\mathbf{X}^T - \mathbf{H}\mathbf{X}^T, \tag{10}$$

$$\frac{\partial f(\Omega,\mathbf{H})}{\partial\mathbf{H}} = -\Omega\mathbf{X} + \mathbf{H} + \lambda. \tag{11}$$

Since $\mathbf{H} \in \mathbb{R}_+$, the partial derivatives of $\lambda\|\mathbf{H}\|_1$ is λ, namely,

$$\frac{\partial(\lambda\|\mathbf{H}\|_1)}{\partial\mathbf{H}} = \lambda. \tag{12}$$

Substitute it to above partial derivatives functions, then we can obtain the update rules as follows,

$$\Omega_{ij}^{(t)} = \Omega_{ij}^{(t-1)} - \eta\Omega_{ij}^{(t-1)}\mathbf{X}\mathbf{X}^T + \eta\mathbf{H}\mathbf{X}^T, \tag{13}$$

$$\mathbf{H}_{ij}^{t} = \mathbf{H}_{ij}^{(t-1)} + \xi\Omega\mathbf{X} - \xi\mathbf{H}_{ij}^{t-1} - \xi\lambda. \tag{14}$$

According to the analysis above, the proposed analysis sparse dictionary learning algorithm is summarized as follows, Analysis L_1 norm algorithm $(A - L_1)$

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

1) Initialize Ω and $\mathbf{H}$ as random nonnegative matrices, $t = 1, \eta, \xi$;

2) Scale rows of $\mathbf{H}$ to a unit ℓ_2-norm:$\sum_{j=1}^{N} \mathbf{H}_{ij}^2 = 1, \forall i$;

3) Iterate until converge or stop;

$$\Omega_{ij}^{(t)} \leftarrow \Omega_{ij}^{(t-1)} - \eta \Omega_{ij}^{(t-1)} \mathbf{X}\mathbf{X}^T + \eta \mathbf{H}\mathbf{X}^T, \qquad 1 \le i \le m, 1 \le j \le n$$

Any negative valued components in Ω^t are set to zeros;
Rescale each rows of $\Omega^{(t)}$ to a unit ℓ_2-norm: $\sum_{i=1}^{n} \Omega_{ij}^2 = 1, \forall j$;

$$\mathbf{H}_{ij}^{(t)} \leftarrow \mathbf{H}_{ij}^{(t-1)} + \xi \Omega \mathbf{X} - \xi \mathbf{H}_{ij}^{(t-1)} - \xi \lambda, \qquad 1 \le i \le n, 1 \le j \le N$$

Any negative valued components in $\mathbf{H}^t$ are set to zeros;
Rescale each rows of $\mathbf{H}^{(t)}$ to a unit ℓ_2-norm: $\sum_{j=1}^{N} \mathbf{H}_{ij}^2 = 1, \forall i$;

5 The Case with Noise

If the observation signal is provided by $\mathbf{Y} = \mathbf{X} + \varepsilon$. ε corresponds to noise. We want recovery $\mathbf{X}$ from its noisy version $\mathbf{Y}$ [17].

Suppose we measure a signal of the form

$$\mathbf{Y}_i = \mathbf{X}_i + \varepsilon_i, \qquad i = 1, 2, ..., K, \tag{15}$$

where ε_i is the noise vector with a bounded ℓ_2-norm, namely, $\|\varepsilon_i\|_2 \le \sigma$, here σ denotes the noise level. In most practical situations, the noise is stationary and bounded. We assume that the noiseless signal $\mathbf{X}$ can be modeled by the sparse-land model,

$$\Omega \mathbf{X}_i = \mathbf{H}_i. \tag{16}$$

Thus we can get the denoised signal $\mathbf{X}$ from Ω and $\mathbf{H}$ using the former equation $\mathbf{X} = \Omega^{\dagger}\mathbf{H}$. Here $\mathbf{H}$ is the sparse coefficient. We can estimate $\mathbf{H}$ and Ω from the following optimization problem,

$$\min_{\Omega, \mathbf{H}} f(\Omega, \mathbf{H}) = \frac{1}{2} \|\Omega \mathbf{Y} - \mathbf{H}\|_F^2 + \lambda \|\mathbf{H}\|_1. \tag{17}$$

This implicate that we can solve this problem with the method in Sec. 4.

6 Numerical Experiments

In this section, we made experiments for evaluating the proposed algorithm. From the experiments we test whether this algorithm can recover the original nonnegative dictionary.

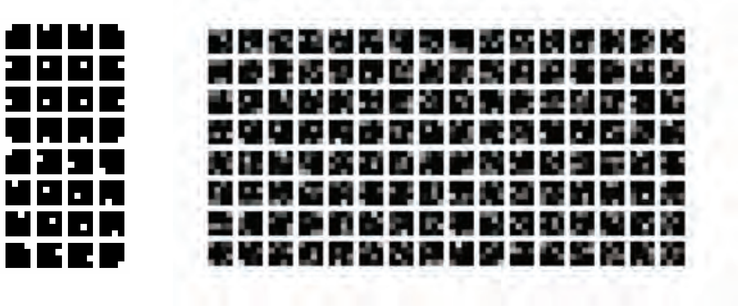

Figure 1: Left: The dictionary of size 32 $\times$ 16 , corresponding to image patched of size 4 $\times$ 4 pixels. Right: The corresponding observed signals.

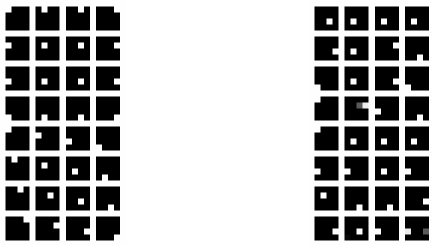

Figure 2: Left: The ground truth dictionary of size 32 $\times$ 16 , corresponding to image patched of size 4 $\times$ 4 pixels. Right: The learnt analysis dictionary.

We first get the results with the analysis sparse representation with the proposed algorithm. What is more, we add noise to the observed signals and get results in the noisy situation. In the experiments, all problems were coded in Matlab, and were run in the environment of Matlab 7.8 (R2009a).

6.1 Generating Analysis Dictionary and the Signals

In our experiments, we began with generating an analysis dictionary Ω as a ground truth. We chose a dictionary of $2d$ rows and d columns; every row contained only one nonnegative element and $d-1$ zeros. The dictionary was of size $2d \times d$, thus it was twice redundant. Naturally, elements of the dictionary became nonnegative. Then scaled rows of the dictionary to a unit ℓ_2-norm.

Then we generated the observational signals from the nonnegative analysis dictionary. This was done by $\mathbf{X} = \Omega^{\dagger}\mathbf{H}$, where the coefficients $\mathbf{H}$ were randomly chosen for each component. Fig. 1 shows the nonnegative analysis dictionary Ω for $d = 16$ and the related observational signals.

6.2 Results of Experiments using $A - L_1$ Algorithm

We applied the $A - L_1$ algorithm to the above generated signals. The initialized dictionary and corresponding coefficients were composed with random entries that were i.i.d. uni-

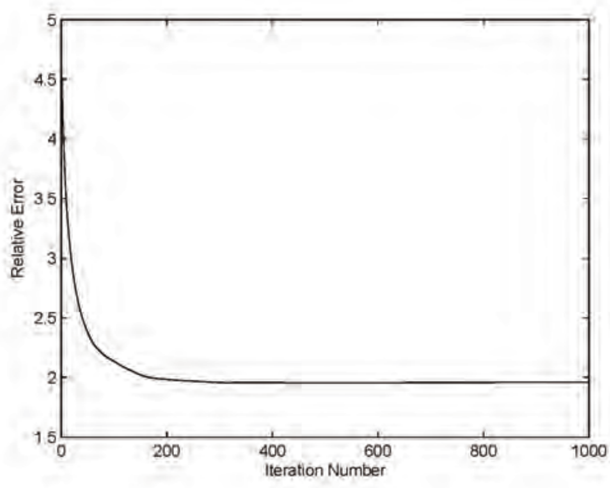

Figure 3: The relative error curve of mean $||\mathbf{H}-\Omega\mathbf{X}||_F^2$.

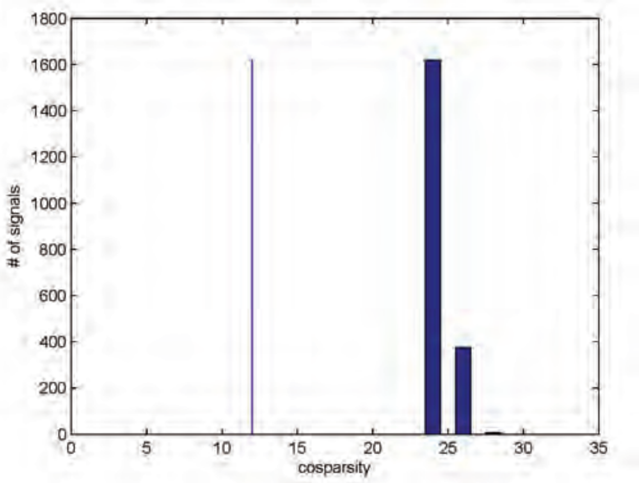

Figure 4: A histogram of the cosparsities of the 2000 sample signals.

formly distributed and nonnegative. Scalar quantities η and ξ were all set to 0.0001. The positive regularization parameter λ was set to 8. The iteration repeated until satisfying the stopping criterion. In our experiment the criterion was a fixed number (1000) of iterations.

The results of these experiments are shown in the following figures. Fig. 2 is the learnt dictionary, from which, we can see the recovery status visually. Fig. 3 describes the relative error $||\mathbf{H}-\Omega\mathbf{X}||_F^2$. The experiment was repeated 10 times and we draw the curve with the averaged value of these experiment results. Fig. 4 shows the histogram of the sparse coefficient of observational signals. Fig. 5 presents a recovery histogram of the sparse coefficient of recovered signals. All these show the effectiveness of the proposed algorithm.

Next, the learned dictionary was compared with the ground truth dictionary. By sweep-ing through the rows, which represent atoms, of the ground truth and the learned dictionary,

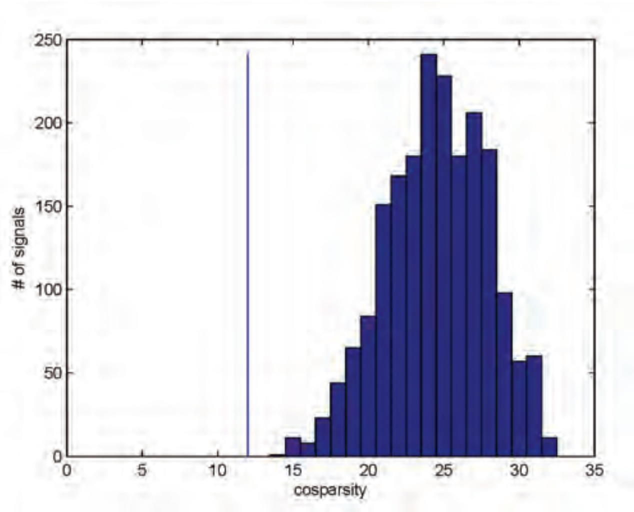

Figure 5: A histogram of the cosparsities of the 2000 recovered signals.

find the closest row between the two dictionaries. A distance less than 0.01 was considered as a success of the recovery. Namely, a row w_j in the ground truth dictionary Ω is regarded as successful recovered if

$$\min_i (1 - |\hat{w}_i^T w_j|) < 0.01, \tag{18}$$

where $\hat{w}_i$ are the atoms of the learnt dictionary. Fig. 6 shows the recovery curve of analysis dictionary in noiseless and noisy situations.

6.3 Results of Experiments in noisy

Besides the noiseless situation, we also made experiments in which the uniformly distributed positive noise of signal-to-noise ratio (SNR) was added to the observed signals in order to evaluate the performance and robustness of anti-noise. Signal-to-noise ratio (SNR) is a measure used in science and engineering that compares the level of a desired signal to the level of background noise. In our experiments, we set the SNR as 80 dB.

All experiments were repeated 10 times, and we use the average value. Fig. 6 shows the results of the experiment for noise levels of 80 dB and for the noiseless case using $A - L_1$ Algorithm and compare with ADT Algorithm [6], the recovery of the dictionary can nearly reach 80% in noisy situation and 85% in the noiseless situation. We can find that in the situation of noiseless the recovery is a little higher than that in a noise of 80 dB. As we see from the results, the algorithms can work well compared with the ADT algorithm.

7 Conclusions

In this paper, we proposed an algorithm for learning the nonnegative dictionary in the analysis model, which is parallel to the synthesis model in its rationale and structure. The proposed algorithm utilizes Euclidean distance as the error measure with the ℓ_1-norm as sparse constraint. Results of numerical experiments demonstrated the ability of our method to correctly and effectively learn the nonnegative analysis dictionary. The proposed algorithm is quite simple but very efficient. By the results of comparing with the situation in

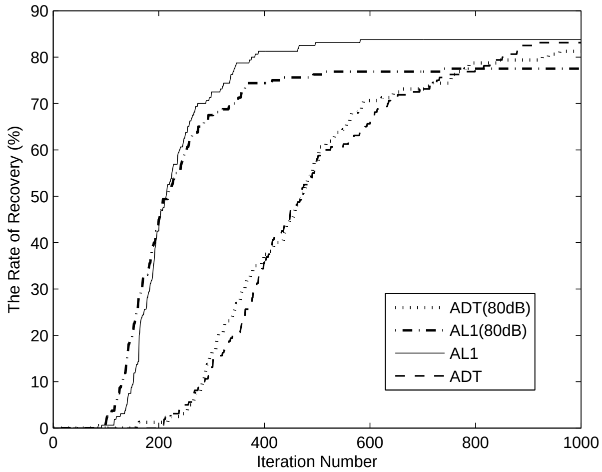

Figure 6: The recovery curves of the analysis dictionary.

80 SNR, we can see the algorithms $A - L_1$ that we proposed can work well in the noisy situation as well as noiseless situation. However, we require revealing more properties of the analysis dictionary; and designing more efficient algorithms suitable for general cases. Further applications remain as our future work.

References

[1] R. Rubinstein, T. Peleg, and M. Elad, Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model, IEEE Trans. on Signal Processing, vol. 61, no. 3, pp. 661-677, 2013.

[2] Z. Tang, S. Ding, Z. Yang, Dictionary learning for sparse representation by nonnegative matrix factorization with constraint of determinant-type of maximization, ICIC Express Letter. vol. 4, no. 5, October 2010.

[3] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, The cosparse analysis model and algorithms, Appl. Comput. Harmon. Anal., vol. 34, no. 1, pp. 30-56, 2013.

[4] M. Elad, Sparse and Redundant Representations From Theory to Applications in Signal and Image Processing. NewYork: Springer, 2010.

[5] J. Starck, F. Murtagh, and M. Fadili, Sparse Image and Signal Processing Wavelets, Curvelets, Morphological Diversity. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[6] Y. Li, S. Ding, and Z. Li, A dictionary-learning algorithm for the analysis sparse model with a determinant-type of sparsity measure, in IEEE International Conference on Digital Signal Processing (DSP), 2014, pp. 152-156.

[7] S. Ravishankar, and Y. Bresler, Learning Sparsifying Transforms, IEEE Trans. on Signal Processing, vol. 61, no. 5, pp. 1072-1086, 2013.

[8] R. Giryesa, S. Namb, M. Elada, R.Gribonvalb, and M. E. Daviesc, Greedy-Like Algorithms for the Cosparse Analysis Model, Linear Alg. Appl., 2013, [online] Available: http://www.sciencedirect.com/science/article/pii/S0024379513001870.

[9] T. Peleg and M. Elad, Performance guarantees of the thresholding algorithm for the cosparse analysis model, IEEE Trans. Inf. Theory., Vol. 59, no. 3, pp. 1832-1845, 2013.

[10] B. Ophir, M. Elad, N. Bertin, and M. D. Plumbley, Sequential minimal eigenvalues: an approach to analysis dictionary learning, in Proceedings of the 9th European Signal Processing Conference, 2011, pp. 1465-1469.

[11] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, Constrained overcomplete analysis operator learning for cosparse signal modelling, IEEE Trans. on Signal Processing, vol. 61, no. 9, pp. 2341-2355, 2013.

[12] Y. Zhang, H. Wang, T. Yu, and W. Wang, Subset pursuit for analysis dictionary learning, in Signal Processing Conference (EUSIPCO), 2013, pp. 1-5.

[13] Y. Zhang, W. Wang, H. Wang, and S. Sanei, Kplane clustering algorithmfor analysis dictionary learning, in Proc. of the IEEE International Workshop on Machine Learning for Signal Processing, 2013, pp. 1-4.

[14] J. M. Duarte-Carvajalino and G. Sapiro, Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization, IEEE Trans. on Image Processing, vol. 18, no. 7, pp. 1395-1408, 2009.

[15] Cedric A. Zala, Ian Barrodale, and J. S. Kennedy, High-Resolution Signal and Noise Field Estimation Using the L1 (Least Absolute Values) Norm, IEEE Journal of Oceanic Engineering, Vol. OE-12, no. 1, January 1987.

[16] P. O. Hoyer, Non-negative matrix factorization with sparseness constraints, J. Mach. Learn. Res.,vol. 5, no. 37, pp. 1457-1469, 2004.

[17] Y. Zhang, H. Wang, T. Yu, and W. Wang, Subset pursuit for analysis dictionary learning, in Signal Processing Conference (EUSIPCO), 2013, pp. 1-5.