# Learning Overcomplete Dictionaries from Markovian Data

S. Akhavan[1,2] S. Esmaeili[1] M. Babaie-Zadeh[3] H. Soltanian-Zadeh[1,4]

[1]School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France
[3]Department of Electrical and Computer Engineering, Sharif University of Technology, Tehran, Iran
[4]Medical Image Analysis Lab., Henry Ford Health System, Detroit, MI, USA

***Abstract*—We explore the dictionary learning problem for sparse representation when the signals are dependent. In this paper, a first-order Markovian model is considered for dependency of the signals, that has many applications especially in medical signals. It is shown that the considered dependency among the signals can degrade the performance of the existing dictionary learning algorithms. Hence, we propose a method using the Maximum Log-likelihood Estimator (MLE) and the Expectation Minimization (EM) algorithm to learn the dictionary from the signals generated under the first-order Markovian model. Simulation results show the efficiency of the proposed method in comparison with the state-of-the-art algorithms.**

***Index Terms*—Dictionary, sparse representation, Markovian model, state, Markov matrix**

## I. Introduction

Dictionary learning for sparse representation has many signal processing applications such as classification [1], compression [2], denoising [3], and so on. The goal of dictionary learning is factorizing the matrix of signals $\mathbf{X} = [\mathbf{x}_1\,\mathbf{x}_2\,...\,\mathbf{x}_T] \in \mathbb{R}^{M\times T}$ into the dictionary $\mathbf{D} = [\mathbf{d}_1\,\mathbf{d}_2\,...\,\mathbf{d}_N] \in \mathbb{R}^{M\times N}$ with unit norm columns (atoms), and the matrix $\mathbf{S} = [\mathbf{s}_1\,\mathbf{s}_2\,...\,\mathbf{s}_T] \in \mathbb{R}^{N\times T}$ with sparse columns. Each column of the dictionary is usually called an atom. The matrices $\mathbf{D}$ and $\mathbf{S}$ can be estimated by solving the constrained optimization problem

$$\{\mathbf{D},\mathbf{S}\} = \underset{\mathbf{D}\in\mathcal{D},\,\mathbf{S}\in\mathcal{S}}{\operatorname{argmin}}\ \|\mathbf{X}-\mathbf{D}\mathbf{S}\|_F^2, \tag{1}$$

where $\mathcal{D}$ and $\mathcal{S}$ respectively show the set of matrices with unit norm atoms, and the set of matrices with sparse columns, and $\|.\|_F$ denotes the Frobenius norm. This problem is usually solved using alternating minimization. This means that (1) is alternately minimized with respect to $\mathbf{D}$ (*D-Update*) and $\mathbf{S}$ (*S-Update*) until convergence of the parameters.

Most dictionary learning algorithms differ in performing the *D-Update*. Some of the algorithms estimate all atoms of the dictionary simultaneously [4], [5], while some others find the atoms consecutively [6]–[8]. Method of Optimal Direction (MOD) [5] is one of the well-established methods from the first category that considers the following closed form solution for the *D-Update*:

$$\mathbf{D} = \mathbf{X}\,\mathbf{S}^{\dagger} \tag{2}$$

followed by normalizing the atoms, where "†" denotes the pseudo inverse. K-Singular Value Decomposition (K-SVD) [8] is one of the well-known methods from the second category that updates each atom and the corresponding non-zero entries of $\mathbf{S}$ using SVD by keeping all other atoms and entries of $\mathbf{S}$ intact. In fact, the following constrained optimization problem is solved for updating each atom and the associated non-zero entries of $\mathbf{S}$:

$$\{\mathbf{d}_n,\mathbf{s}^r_{[n]}\} = \underset{\mathbf{d}_n,\mathbf{s}^r_{[n]}}{\operatorname{argmin}}\ \|\mathbf{E}^r_n - \mathbf{d}_n(\mathbf{s}^r_{[n]})^T\|_F^2$$
$$s.t.\quad \|\mathbf{d}_n\|_2 = 1,\quad \mathbf{E}_n = \mathbf{X} - \sum_{i\neq n}\mathbf{d}_i(\mathbf{s}_{[i]})^T \tag{3}$$

where $\mathbf{s}_{[i]}$ shows the $i^{th}$ row of $\mathbf{S}$, $\mathbf{s}^r_{[n]}$ represents the non-zero entries of $\mathbf{s}_{[n]}$, and $\mathbf{E}^r_n$ consists of the columns of $\mathbf{E}_n$ which are corresponding to the non-zero entries of $\mathbf{s}_{[n]}$.

The *S-Update* is divided into $T$ different sparse recovery problems from the signals, i.e., the columns of $\mathbf{X}$, which can be solved by any of the sparse recovery algorithms such as Basis Pursuit (BP) [9], Orthogonal Matching Pursuit (OMP) [10], Smoothed $l_0$ (SL0) [11], and so on. Since there is no information about the dependency of the columns of $\mathbf{X}$, they are assumed to be independent, and the *S-Update* decouples to $T$ different sparse recovery problems. However, in some applications in medical signals and images, e.g., electroencephalography recordings in some specific diseases or diffusion weighted images [12]–[14], the signals are dependent and hence, the current strategy in the *S-Update* degrades the dictionary learning performance. In this paper, we consider a well-known model, especially in medical signals [14], [15], for the dependency of signals and explain how to perform the dictionary learning when the signals are generated under this model.

## II. Proposed Model

We assume that there are $Q$ states that are activated under the first-order Markovian model with a fixed Markov matrix $\mathbf{P} \in \mathbb{R}^{Q\times Q}$. Each sample of the signals is associated with one of these states. Hence, we can assign a sequence of states $\{q_1, q_2, ..., q_T\}$ to the signals, where $q_t \in \{1, 2, ..., Q\}$ for $t = 1, 2, ..., T$. Each state has its own dictionary $\mathbf{D}^{(q)} = [\mathbf{d}^{(q)}_1\,\mathbf{d}^{(q)}_2\,...\,\mathbf{d}^{(q)}_L] \in \mathbb{R}^{M\times L}$ with $L$ atoms. In fact, the concatenation of the dictionaries in different states makes the main dictionary, i.e., $\mathbf{D} = [\mathbf{D}^{(1)}\,\mathbf{D}^{(2)}\,...\,\mathbf{D}^{(Q)}] \in \mathbb{R}^{M\times N}$

978-1-5386-4752-3/18/$31.00 ©2018 IEEE

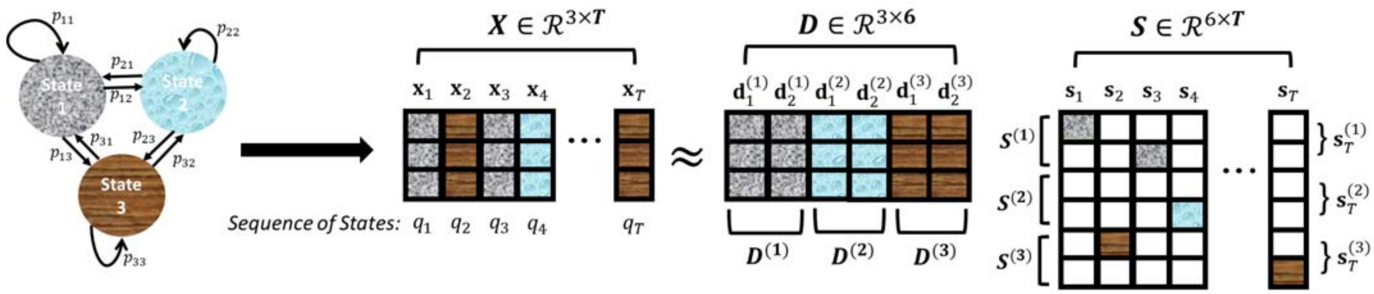


Fig. 1. Considered model for $M = 3$ dimensional signals when there are $Q = 3$ states with $L = 2$ atoms in each one.

($N = LQ$). It should be noted that the dictionaries $\mathbf{D}^{(q)}$ for $q = 1, 2, ..., Q$ have no common atoms. When a state is activated, a sparse linear combination from the atoms of the corresponding dictionary generates the signal. The considered model is schematically shown in Fig. 1 when $M = Q = 3$ and $L = 2$.

As shown here, similar to the main dictionary, the matrix $\mathbf{S} \in \mathbb{R}^{N\times T}$ is divided into $L$ submatrices, $\mathbf{S}^{(q)} \in \mathbb{R}^{L\times T}$ for $q = 1, 2, ..., Q$, corresponding to the states. Also, the contribution of the $q^{th}$ state for the generation of $\mathbf{x}_t$, or in other words, the $t^{th}$ column of $\mathbf{S}^{(q)}$ is shown by $\mathbf{s}_t^{(q)}$.

The set of unknown parameters is

$$\Omega = \underbrace{\{\mathbf{D}, \mathbf{S}, \mathbf{P}\}}_{\Theta} \cup \{q_1, q_2, ..., q_T\}, \tag{4}$$

which must be estimated from the signals. According to the dictionary learning problem and the considered model, the following constraints must be considered for the elements of $\Theta$:

$$\mathbf{D} \in \mathcal{D},\ \mathbf{S} \in \mathcal{S},\ \mathbf{P} \in \mathcal{P} \rightarrow \Theta \in \Lambda \tag{5}$$

where $\mathcal{P}$ shows the set of matrices with each row summing to one and positive entries, and $\Lambda$ represents the feasible space for $\Theta$. Here, the model definition is complete.

Now, using a simple example, we show why the *S-Update* cannot be decoupled into $T$ independent sparse recovery problems in the considered model. Assume that we are in the *S-Update*, and we want to find the activated state for $\mathbf{x}_t$, and then, select the proper atoms from the dictionary corresponding to the activated state. Due to the presence of the first-order Markovian model, the following optimization problem must be solved for finding $q_t$:

$$q_t = \underset{q}{\operatorname{argmax}}\ \ p(q_t = q|\mathbf{x}_t, q_{t-1}) \tag{6}$$

where the objective function shows the probability of being in state $q$, given the sample $\mathbf{x}_t$ and the previous activated state $q_{t-1}$. According to the first-order Markovian model, the objective function can be written and simplified as

$$\begin{aligned} p(q_t = q|\mathbf{x}_t, q_{t-1}) &= \frac{p(\mathbf{x}_t, q_t = q, q_{t-1})}{p(\mathbf{x}_t, q_{t-1})} \\ &= \frac{p(\mathbf{x}_t|q_t = q)\, p(q_t = q|q_{t-1})}{p(\mathbf{x}_t)} \end{aligned} \tag{7}$$

where $p(q_t = q|q_{t-1})$ is the $(q_{t-1}, q_t)^{th}$ entry of the Markov matrix. Therefore, estimating the activated state for the $t^{th}$ sample depends directly on the activated state for the $(t-1)^{th}$ sample, and we cannot apply the sparse recovery algorithms on each of the samples independently form other samples in the considered model.

It is worth noting that the decision making for $q_t$ also depends on the probability of observing $\mathbf{x}_t$ given $q_t$, according to (7). Based on this observation, it can be easily shown that when the signal to noise ratio increases, the effect of the considered model in the dictionary learning procedure decreases.

## III. Proposed Method

We first estimate $\Theta$ using the Maximum Log-likelihood Estimator (MLE) and the Expectation Minimization (EM) algorithm. Then, we find the sequence of states using Viterbi algorithm [15].

Each sample of the signals is expressed as

$$\textit{If}\ q_t = q \rightarrow \mathbf{x}_t = \mathbf{D}^{(q)}\mathbf{s}_t^{(q)} + \mathbf{n}_t, \tag{8}$$

where $\mathbf{n}_t \in \mathbb{R}^M$ shows the additive noise with $\mathcal{N}(0, \sigma_0^2\mathbf{I})$ distribution. The noise vectors for $t = 1, 2, ..., T$ are independent and identically distributed (i.i.d). Hence, the following objective function must be maximized for finding the MLE solution:

$$g(\Theta) = \log(f(\mathbf{X}|\Theta)) = \sum_{t=1}^{T} \log(f(\mathbf{x}_t|\Theta)). \tag{9}$$

Since the activated state for the $t^{th}$ sample is unknown, we express the probability density function (pdf) of $\mathbf{x}_t$ given $\Theta$ as

$$f(\mathbf{x}_t|\Theta) = \sum_{q=1}^{Q} p(q_t = q) f(\mathbf{x}_t|q_t = q, \Theta), \tag{10}$$

where the pdf of $\mathbf{x}_t$, given $q_t = q$ and $\Theta$, can be written as follows according to (8):

$$f(\mathbf{x}_t|q_t = q, \Theta) = (\frac{1}{\sqrt{2\pi\sigma_0^2}})^M \exp(-\frac{\|\mathbf{x}_t - \mathbf{D}^{(q)}\mathbf{s}_t^{(q)}\|_2^2}{2\sigma_0^2}). \tag{11}$$

By substituting (11) in (10), and the obtained result in (9), we get

$$g(\Theta) = \sum_{t=1}^{T} \log\big(\sum_{q=1}^{Q} p(q_t = q) f(\mathbf{x}_t | q_t = q, \Theta)\big). \tag{12}$$

Optimizing $g(\Theta)$ is difficult due to the presence of the term "$\log \sum_{q=1}^{Q}$". Since the summation over $q$ is like an expectation operator, we consider the following lower bound for the objective function according to the Jensen's inequality:

$$g(\Theta) \geq \sum_{t=1}^{T} \sum_{q=1}^{Q} p(q_t = q) \, \log\big(f(\mathbf{x}_t | q_t = q, \Theta)\big). \tag{13}$$

By substituting (11) in the obtained lower bound, it can be shown that the following constrained optimization problem must be solved for estimating $\Theta$:

$$\Theta = \underset{\Theta \in \Lambda}{\operatorname{argmin}} \; h(\Theta)$$
$$h(\Theta) = \sum_{t=1}^{T} \sum_{q=1}^{Q} p(q_t = q) \, \|\mathbf{x}_t - \mathbf{D}^{(q)} \mathbf{s}_t^{(q)}\|_2^2. \tag{14}$$

Since both $\Theta$ and $p(q_t = q)$ are unknown, we use the EM method to estimate the parameters, i.e., the following two steps are alternately performed until the convergence of the parameters.

*1) Expectation Step:* In this step, we assume that $\Theta$ is fixed, and estimate $p(q_t = q)$ for $t = 1, 2, ..., T$ and $q = 1, 2, ..., Q$.

Since $\Theta$ is known, the Markov matrix and the conditional pdf of the signals $f(\mathbf{x}_t | q_t = q, \Theta)$ for $t = 1, 2, ..., T$ and $q = 1, 2, ..., Q$ are known according to (11). Hence, we can use the forward-backward procedure for estimating $p(q_t = q)$ as explained with details in [15] for the first-order Markovain model. It should be mentioned that $p(q_t = i, q_{t+1} = j)$ that shows the probability of being in state $i$ for the $t^{th}$ sample and bing in state $j$ for the $(t + 1)^{th}$ sample is also calculated during the forward-backward procedure, which will be used in the next step to estimate the Markov matrix.

*2) Minimization Step:* In this step, we assume that $p(q_t = q)$ is fixed, and estimate $\Theta$. We first show how to estimate $\mathbf{P}$, and then $\mathbf{D}$ and $\mathbf{S}$.

The $(i, j)^{th}$ entry of the Markov matrix can be estimated as

$$[\mathbf{P}]_{i,j} = p_{ij} \rightarrow p_{ij} = \frac{\sum_{t=1}^{T-1} p(q_t = i, q_{t+1} = j)}{\sum_{t=1}^{T-1} p(q_t = i)}. \tag{15}$$

Now, we back to (14) for estimating $\mathbf{D}$ and $\mathbf{S}$. Since the parameters of the states appear in $h(\Theta)$ independently from each other, the following constrained optimization problem can individually be solved for all of the states (i.e., $q = 1, 2, ..., Q$) to find $\mathbf{D}$ and $\mathbf{S}$:

$$\{\mathbf{D}^{(q)}, \mathbf{S}^{(q)}\} = \underset{\mathbf{D}^{(q)} \in \mathcal{D}, \, \mathbf{S}^{(q)} \in \mathcal{S}}{\operatorname{argmin}} \sum_{t=1}^{T} p(q_t = q) \|\mathbf{x}_t - \mathbf{D}^{(q)} \mathbf{s}_t^{(q)}\|_2^2. \tag{16}$$

The objective function has a weighted least square form which can be solved using alternating minimization. In fact, due to the similarity of (16) and (1), any dictionary learning algorithms, e.g. MOD or K-SVD, can be employed for solving (16) following the *S-Update* and the *D-Update*.

The noticeable point is the presence of $p(q_t = q)$ in (16) that acts as a weight and causes a minor modification in the *D-Update* and does not change the *S-Update*.

If we use MOD for solving (16), the following closed form solution is obtained in the *D-Update*:

$$\mathbf{D}^{(s)} = (\sum_{t=1}^{T} p(q_t = q) \, \mathbf{s}_t^{(q)} {\mathbf{s}_t^{(q)}}^T)^{-1} (\sum_{t=1}^{T} p(q_t = q) \, \mathbf{x}_t {\mathbf{s}_t^{(q)}}^T), \tag{17}$$

followed by normalizing the atoms. We call the proposed method as "New MOD" when (16) is solved using MOD.

We can also use K-SVD for solving (16). We define $\dot{\mathbf{x}}_t = \sqrt{p(q_t = q)} \, \mathbf{x}_t$ and $\dot{\mathbf{s}}_t^{(q)} = \sqrt{p(q_t = q)} \, \mathbf{s}_t^{(q)}$, and consider them as the columns of $\dot{\mathbf{X}}$ and $\dot{\mathbf{S}}^{(q)}$, respectively. Hence, the following constrained optimization problem must be solved for updating each atom and its associated non-zero entries:

$$\{\mathbf{d}_n, \dot{\mathbf{s}}_{[n]}^r\} = \underset{\mathbf{d}_n, \dot{\mathbf{s}}_{[n]}^r}{\operatorname{argmin}} \; \|\dot{\mathbf{E}}_n^r - \mathbf{d}_n (\dot{\mathbf{s}}_{[n]}^r)^T\|_F^2$$
$$s.t. \quad \|\mathbf{d}_n\|_2 = 1, \quad \dot{\mathbf{E}}_n = \dot{\mathbf{X}} - \sum_{i \neq n} \mathbf{d}_i (\dot{\mathbf{s}}_{[i]})^T, \tag{18}$$

where $\dot{\mathbf{s}}_{[i]}$ represents the $i^{th}$ row of $\dot{\mathbf{S}}$ (i.e., the vertical concatenation of $\dot{\mathbf{S}}^{(q)}$ for $q = 1, 2, ..., Q$), $\dot{\mathbf{s}}_{[i]}^r$ shows the non-zero entries of $\dot{\mathbf{s}}_{[i]}$, and $\dot{\mathbf{E}}_n^r$ consists of the columns of $\dot{\mathbf{E}}_n$ which are corresponding to the non-zero entries of $\mathbf{s}_{[n]}$. It is worth noting that when (18) is solved and $\dot{\mathbf{s}}_{[i]}^r$ is obtained, the corresponding $\mathbf{s}_{[i]}^r$ is also calculated according to the definition of $\dot{\mathbf{s}}_t^{(q)}$. We call the proposed method as "New K-SVD" when (16) is solved using K-SVD.

By performing a few iterations between the expectation and the minimization steps, $\Theta$ is estimated. By determination of $\Theta$, we can use Viterbi algorithm [15] to find the sequence of states $\{q_1, q_2, ..., q_T\}$. Once the sequence of states is determined, we apply a minor modification on each column of $\mathbf{S}$. We keep the entries corresponding to the activated state, and make the other entries equal to zero. Hence, the final values of the parameters are determined.

## IV. Experimental Results

We compare the performance of MOD and K-SVD with "New MOD" and "New K-SVD" in recovering a known dictionary in this Section.

### *A. Data Generation*

We assume that there are $Q = 3$ states which are activated under the first-order Markovian model with a fixed Markov matrix $\mathbf{P} \in \mathbb{R}^{3\times 3}$. We consider two different Markov matrices to present the results:

$$\mathbf{P}_1 = \begin{bmatrix} 0.34 & 0.33 & 0.33 \\ 0.33 & 0.34 & 0.33 \\ 0.33 & 0.33 & 0.34 \end{bmatrix}$$

$$\mathbf{P}_2 = \begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.10 & 0.80 & 0.10 \\ 0.05 & 0.05 & 0.90 \end{bmatrix} \tag{19}$$

The states are approximately activated independently from each other in $\mathbf{P}_1$, while they are highly dependent in $\mathbf{P}_2$. Then, a sequence of Markovian states $\{q_1, q_2, ..., q_T\}$ with length of $T = 3000$ samples is generated according to the considered Markov matrix. The MATLAB function "hmmgenerate" can be employed for generating the sequence.

Then, the sparse matrix $\mathbf{S} \in \mathbb{R}^{60\times 3000}$ is produced. For generating each column of $\mathbf{S}$, $k$ non-zero entries are randomly chosen from $\mathbf{s}_t^{(q_t)}$, then, their values are selected from zero-mean and unit variance Gaussian variables.

Then, the dictionary is generated by a random matrix $\mathbf{D}^{15\times 60}$ with zero-mean and unit variance Gaussian entries which are i.i.d, followed by normalizing the columns. The dictionary is partitioned into three matrices of size $15 \times 20$ which are corresponding to the dictionaries of the states. Finally, the signals are generated as follows:

$$\begin{aligned} &\mathbf{x}_t = \mathbf{D}\mathbf{s}_t + \mathbf{n}_t \\ &t = 1, 2, ..., 3000, \end{aligned} \tag{20}$$

where the noise vectors for $t = 1, 2, ..., T$, are i.i.d with $\mathcal{N}(0, \sigma_0^2 \mathbf{I})$ distribution. We can adjust $\sigma_0^2$ to achieve the desired signal to noise ratio (SNR), which is defined as follows:

$$\text{SNR} = \frac{\|\mathbf{DS}\|_F^2}{\|\mathbf{X} - \mathbf{DS}\|_F^2}. \tag{21}$$

### B. Results

We use the correct recovery rate, which is the number of correctly estimated atoms divided by the number of atoms, for presenting the results [8]. The average of the correct recovery rate over 100 trails, in different SNRs and for different sparsity levels, are reported in Table I and Table II when we respectively consider $\mathbf{P}_1$ and $\mathbf{P}_2$ as the Markov matrices.

As shown in Table I, the results of MOD and K-SVD are respectively similar to ones in New MOD and New K-SVD. The reason is that the states are activated independently from each other. Hence, decoupling the *S-Update* to $T$ different sparse recovery problems is a correct strategy.

On the other hand, when the states are highly dependent, the new versions of MOD and K-SVD perform better than the initial versions, especially in low SNR$s$, as shown in Table II. As mentioned before, it can also be seen that whatever the SNR increases, the effect of the Markovian model in the dictionary learning procedure decreases.

TABLE I
PERCENTAGE OF CORRECT RECOVERY RATE WHERE THE STATES ARE ACTIVATED ALMOST INDEPENDENTLY FROM EACH OTHER.

| SNR$_{dB}$ | Algorithm | $k=3$ | $k=4$ | $k=5$ |
|---|---|---|---|---|
| 7 | MOD | 78.7 | 70.5 | 2.5 |
| | New MOD | 79.4 | 70.6 | 2.6 |
| | K-SVD | 81.3 | 79.7 | 10.6 |
| | New K-SVD | 81.3 | 80.4 | 12.2 |
| 15 | MOD | 84.6 | 83.5 | 75.6 |
| | New MOD | 84.9 | 83.9 | 75.8 |
| | K-SVD | 86.1 | 85.6 | 81.6 |
| | New K-SVD | 86.9 | 86.4 | 81.9 |
| 30 | MOD | 87.3 | 84.2 | 81.3 |
| | New MOD | 87.4 | 86.6 | 83.2 |
| | K-SVD | 88.4 | 87.8 | 84.3 |
| | New K-SVD | 89.5 | 88.6 | 84.8 |
| 60 | MOD | 90.3 | 88.6 | 85.9 |
| | New MOD | 90.8 | 88.9 | 86.7 |
| | K-SVD | 91.5 | 90.1 | 89.4 |
| | New K-SVD | 92.7 | 91.2 | 89.6 |

TABLE II
PERCENTAGE OF CORRECT RECOVERY RATE WHERE THE STATES ARE DEPENDENT.

| SNR$_{dB}$ | Algorithm | $k=3$ | $k=4$ | $k=5$ |
|---|---|---|---|---|
| 7 | MOD | 68.7 | 61.4 | $\simeq 0$ |
| | New MOD | 78.3 | 69.5 | 2.2 |
| | K-SVD | 73.2 | 67.6 | 1.7 |
| | New K-SVD | 81.9 | 80.3 | 12.3 |
| 15 | MOD | 73.1 | 66.6 | 49.3 |
| | New MOD | 85.5 | 82.8 | 75.1 |
| | K-SVD | 77.4 | 74.3 | 70.6 |
| | New K-SVD | 86.9 | 84.3 | 79.3 |
| 30 | MOD | 83.4 | 82.1 | 78.4 |
| | New MOD | 86.3 | 85.1 | 84.2 |
| | K-SVD | 87.6 | 84.3 | 81.5 |
| | New K-SVD | 90.4 | 89.7 | 83.5 |
| 60 | MOD | 87.3 | 86.4 | 83.5 |
| | New MOD | 88.8 | 86.5 | 84.1 |
| | K-SVD | 91.3 | 89.2 | 87.2 |
| | New K-SVD | 92.6 | 92.4 | 88.3 |

## V. CONCLUSION

We explored the dictionary learning problem for sparse representation when the signals are activated under the first-order Markovian model. We showed that the considered dependency among the signals can degrade the performance of the existing dictionary learning algorithms especially when 1) the Markovian states are highly dependent and 2) the SNR is low. Then, we proposed a method using MLE and EM to learn the dictionary from signals generated under the first-order Markovian model. Simulation results verified the efficiency of the proposed method in comparison with state-of-the-art dictionary learning algorithms in recovering a known dictionary.

## REFERENCES

[1] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 791–804, April 2012.

[2] O. Bryt and M. Elad, "Compression of facial images using the k-svd algorithm," *J. Vis. Comun. Image Represent.*, vol. 19, no. 4, pp. 270–282, May 2008.

[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, Dec 2006.

[4] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, “Dictionary learning for sparse representation: A novel approach,” *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1195–1198, Dec 2013.

[5] K. Engan, S. O. Aase, and J. H. Husoy, “Method of optimal directions for frame design,” in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 5, 1999, pp. 2443–2446 vol.5.

[6] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, “Learning overcomplete dictionaries based on atom-by-atom updating,” *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 883–891, Feb 2014.

[7] L. N. Smith and M. Elad, “Improving dictionary learning: Multiple dictionary updates and coefficient reuse,” *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 79–82, Jan 2013.

[8] 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, Nov 2006.

[9] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[10] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[11] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, “A fast approach for overcomplete sparse decomposition based on smoothed l0 norm,” *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, Jan 2009.

[12] A. R. Mohammadi-Nejad, G. A. Hossein-Zadeh, and H. Soltanian-Zadeh, “Structured and sparse canonical correlation analysis as a brain-wide multi-modal data fusion approach,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1438–1448, July 2017.

[13] M. Afzali, E. Fatemizadeh, and H. Soltanian-Zadeh, “Sparse registration of diffusion weighted images,” *Computer Methods and Programs in Biomedicine, in press*, 2017.

[14] S. Akhavan, R. Phlypo, H. Soltanian-Zadeh, F. Studer, A. Depaulis, and C. Jutten, “Characterizing absence epileptic seizures from depth cortical measurements,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 444–448.

[15] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.


978-1-5386-4752-3/18/$31.00 ©2018 IEEE 