Learning Overcomplete Dictionaries from Markovian Data

Saeed Akhavan[1], Samaneh Esmaeili[1], Masoud Babaie-Zadeh[2], Hamid Soltanian-Zadeh[1]

[1] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

[2] School of Electrical and Computer Engineering, University of Sharif, Tehran, Iran

1. Dictionary Learning Problem

✓ **Target:** factorizing the matrix of training signals into the dictionary with unit norm columns (atoms), and the coefficient matrix with sparse columns, i.e.,

$$\{\mathbf{D},\mathbf{X}\} = \underset{\mathbf{D},\mathbf{X}}{\text{argmin}} \ \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2$$
$$s.t. \quad \|\mathbf{x}_k\|_0 \le N_0, \quad 1\le k\le K$$
$$\|\mathbf{d}_n\|_2 = 1, \quad 1\le n \le N$$

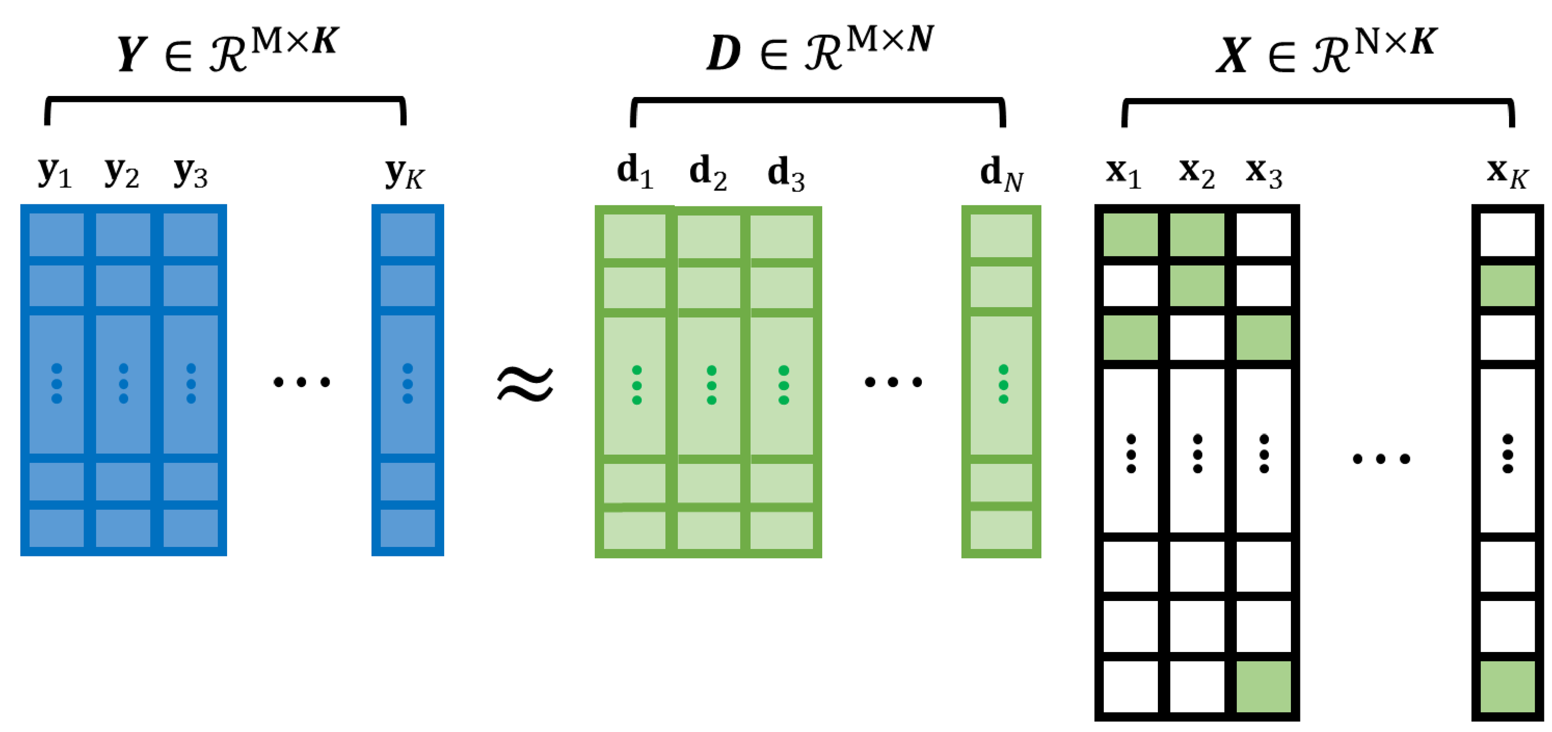


Fig 1. Schematic diagram of dictionary learning problem.

➢ **Typical Solution.** -----> **Alternation Minimization**

I. Sparsification:

$$\mathbf{X} = \underset{\mathbf{X}}{\text{argmin}} \ \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2$$
$$s.t. \ \|\mathbf{x}_k\|_0 \le N_0, \ \ 1\le k\le K$$

The training signals are considered statistically independent.

$$\mathbf{x}_k = \underset{\mathbf{x}_k}{\text{argmin}} \ \|\mathbf{y}_k-\mathbf{D}\mathbf{x}_k\|_F^2 \quad s.t. \ \|\mathbf{x}_k\|_0 \le N_0$$

I. Dictionary Update:

$$\mathbf{D} = \underset{\mathbf{D}}{\text{argmin}} \ \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2$$
$$s.t. \ \|\mathbf{d}_n\|_2 = 1, \ \ 1\le n\le N$$

2. Considered Model

✓ **Target:** Performing dictionary learning when the training signals are not statistically independent, and have the first-order Markovian dependency.

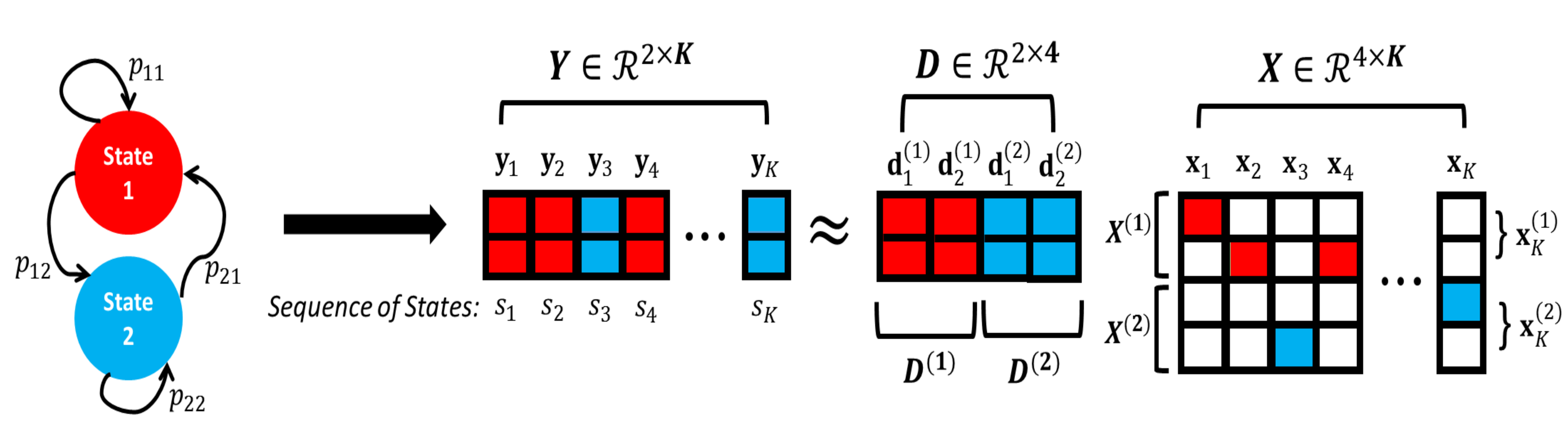
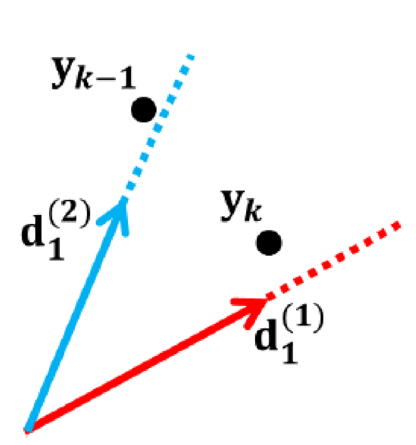


Fig 2. Schematic diagram of the considered model for dictionary learning problem.

▪ **The set of unknown parameters:** $\Omega = \underbrace{\{\mathbf{P},\mathbf{D},\mathbf{X}\}}_{\Theta} \cup \{s_1, s_2, \dots, s_K\}$



✓ Important Factors:

❖ Signal to noise ratio (SNR)

❖ Probability of transition between states.

Fig 3. In the considered model, assigning y_k to one of the atoms is not independent form the activated state (or atom) for y_{k-1}.

3. Model Parameters Estimation

$$\Theta = \underset{\Theta}{\text{argmax}} \sum_{k=1}^{K} log\{\sum_{s=1}^{S} p(s_k = s|\mathbf{Y},\Theta) f(\mathbf{y}_k|s_k = s,\Theta)\}$$
$$s.t. \quad \|\mathbf{d}_l^{(s)}\|_2 = 1, \quad \|\mathbf{x}_k^{(s)}\|_0 \le N_0$$
$$s = 1,2,\dots,S, \quad l = 1,2,\dots,L, \quad k = 1,2,\dots,K$$

Expectation - Minimization

⚠ By determination of θ, the sequence of states, i.e.,$\{s_1, s_2, \dots, s_K\}$ is determined using Viterbi algorithm.

4. Results

➢ **First Scenario:** Independent training signals.

$$\mathbf{P}_1 = \begin{bmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{bmatrix}$$

SNR$_{dB}$	Method	$N_0=3$	$N_0=4$	$N_0=5$
10	MOD	80.6	72.4	4.3
	New MOD	81.2	72.8	4.8
	K-SVD	83.4	81.8	12.9
	New K-SVD	83.7	82.2	14.1
20	MOD	86.5	85.3	77.9
	New MOD	86.9	85.8	77.1
	K-SVD	88.3	87.5	83.5
	New K-SVD	88.4	88.2	82.9
30	MOD	89.1	86.6	83.4
	New MOD	89.6	88.4	85.7
	K-SVD	90.5	89.5	86.2
	New K-SVD	91.7	90.4	86.8
100	MOD	90.1	88.3	85.8
	New MOD	90.3	88.6	86.8
	K-SVD	92.3	90.7	89.5
	New K-SVD	92.4	91.1	89.6

Table 1. Percentage of successful recovery rate in the first scenario where the states are activated almost independently from each other.

➢ **Second Scenario:** Dependent training signals.

$$\mathbf{P}_2 = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix}$$

SNR$_{dB}$	Method	$N_0=3$	$N_0=4$	$N_0=5$
10	MOD	70.3	63.6	$\simeq 0$
	New MOD	81.4	71.9	4.6
	K-SVD	75.6	69.3	3.9
	New K-SVD	82.8	83.1	14.4
20	MOD	75.3	68.4	51.8
	New MOD	87.1	84.9	77.2
	K-SVD	79.6	76.4	72.4
	New K-SVD	88.8	86.5	81.6
30	MOD	85.6	84.2	80.6
	New MOD	88.6	87.7	86.4
	K-SVD	89.8	86.7	83.1
	New K-SVD	92.1	91.3	85.9
100	MOD	88.2	87.1	84.9
	New MOD	89.8	87.5	85.1
	K-SVD	91.3	90.1	88.2
	New K-SVD	92.5	92.3	88.7

Table 2. Percentage of successful recovery rate in the first scenario where the states are dependent.

7. Conclusion

✓ Dependency among the training signals degrade the performance of current dictionary learning algorithm.

✓ We investigated the dictionary learning problem when there is the first-order Markovian model in the generation of signals.