LEARNING SPARSE DICTIONARIES WITH A POPULARITY-BASED MODEL

Jianzhou Feng*, Li Song*, Xiaoming Huo†, Xiaokang Yang* and Wenjun Zhang*
*Shanghai Jiaotong University, Institute of Image Comm. & Information Proc.Shanghai, China † Georgia Institute of Technology, School of Industrial and Systems Engineering, Atlanta, USA

Contribution

We propose a novel dictionary learning algorithm—the PK-SVD algorithm—which assumes prior probabilities on the dictionary atoms and learns a sparse dictionary under a popularity-based model. The prior distribution brings the flexibility that is desirable in applications. We examine our algorithm in both synthetic tests and image denoising experiments.

Model

The general sparse representation model assumes x to be the solution of

$$\hat{x} = \arg\min_{x} \|y - Dx\|_2^2 + \lambda\|x\|_0, \tag{1}$$

We proposed a popularity based model which assumes different popularity for different atoms. We define

$$p_{ij} \triangleq P(|x^{(j)}| \text{ is the } i^{th} \text{ largest among } \{x^{(1)}, \dots, x^{(K)}\}) \tag{2}$$

The MAP estimate becomes

$$\hat{x} = \arg\min \|y - Dx\|_2^2 + \lambda \sum_{i=1}^{\|x\|_0} \log(\frac{1}{p_{i,sel_i}}), \text{subject to } \|x\|_0 \le T_0, \tag{3}$$

where sel_i is the entry of x that has the ith largest absolute value.
Suppose $Y = \{y_i\}_{i=1}^N$, $y_i \in \mathbb{R}^n$ is the example set, the best dictionary D, the popularity matrix $P = \{p_{ij}\}$ and the decomposition $X = \{x_i\}_{i=1}^N$, $x_i \in \mathbb{R}^K$ are the solution of

$$\arg\min_{D,X,P} \|Y - DX\|_F^2 + \lambda \sum_{j=1}^{N} \sum_{i=1}^{\|x_j\|_0} \log(\frac{1}{p_{i,sel_i^{(j)}}}), \text{subject to } \forall j, \|x_j\|_0 \le T_0. \tag{4}$$

The POMP Algorithm

The POMP algorithm for solving (3) is

- Initialization: $x^0 = 0$, support set $\Gamma^0 = \varnothing$.
- For $J = 1, 2, \dots, T_0$
  1. *Update Residual*: $r^J = y - Dx^{J-1}$.
  2. *Sweep*: For $k = 1, 2, \dots, K$,
     $$m(k) = \langle r^J, d_k\rangle^2 + \lambda \ln(p_{J,k}).$$
  3. *Update Support Set*: Find a maximizer
     $$k^* = \arg\max m(k).$$
     If $m(k^*) > 0$, set $\Gamma^J = \Gamma^{J-1} \cup \{k^*\}$. Else, break out of the iteration.
  4. *Update weight vector*:
     $$x^J = \arg\min_{\text{Support}(x)=\Gamma^J} \|y - Dx\|_2^2.$$

The PK-SVD algorithm

The PK-SVD dictionary learning algorithm for solving (4) is

- Initialization: Set the random normalized dictionary matrix $D^0 \in \mathbb{R}^{n\times K}$. Set the popularity matrix $P^0 = (\frac{1}{K})_{T_0\times K}$. Set $\lambda_0 = \max(\frac{\|Y\|_F^2}{3nN}, \lambda)$. Set $J = 1$.
- For $J = 1, 2, \dots, M$
  1. *Sparse Coding Stage*: Compute the initial $X^J = \{x_j^J\}_{j=1}^N$ using POMP algorithm with D^{J-1}, P^{J-1} and λ_{J-1} as parameter.
  2. *Dictionary Update Stage*: Update the initial X^J to the updated X^J and D^{J-1} to D^J using the same step as in K-SVD[1].
  3. *Popularity Matrix Update Stage*:
     - Define $N_{ij} = \#\{k | sel_i^{(k)} = j, k \in \{1, 2, \dots, N\}\}$.
     - For all the element of $P^{(J)}$, set $p_{i,j}^{(J)} = \frac{N_{ij}}{\sum_{k=1}^K N_{ik}}$.
- Set $\lambda_J = \lambda_{J-1} + \frac{\lambda - \lambda_0}{M-1}$.
- Set $J = J + 1$.

Synthetic experiments

We randomly choose a normalized dictionary $D \in \mathbb{R}^{20\times 50}$, set the sparsity $T_0 = 3$ and assign the hidden popularity matrix P to be $\{p_{1,j}\}_{j=1}^{10} = \frac{1}{10}$, $\{p_{2,j}\}_{j=11}^{30} = \frac{1}{20}$ and $\{p_{3,j}\}_{j=31}^{50} = \frac{1}{20}$, and generate $Y = \{y_i\}_{i=1}^{4000}$. We then add white noises with various strengths to those signals. For each noise level, we execute 5 experiments. Both the K-SVD and PK-SVD was executed for a maximum number of 80 iterations to learn a new dictionary $\hat{D}$, which is compared with D, using the same method as in [1].

SNR	10db	20db	30db	No Noise
K-SVD	76.8%	82.8%	80.8%	85.6%
PK-SVD	88.8%	94.4%	92.4%	94.4%

As shown above, the difference between PK-SVD and K-SVD is obvious, because PK-SVD tend to adapt to the special popularity matrix P.

Image denoising experiments

For a certain image, we first randomly extract thousands of image patches for dictionary learning. The parameter λ is set as $2.5\sigma^2$, where σ is the noise variance. Assume Y is the observed image, X is the underlying "true" noise-free image, $\hat{D}$ is the learnt dictionary, $R_{ij}X$ is the image patch at position (i, j) and $\hat{\alpha}_{ij}$ is the corresponding linear representation using POMP. We need to solve

$$\hat{X} = \arg\min_{x} \mu\|X - Y\|_F^2 + \sum_{ij} w_{ij}\|R_{ij}X - \hat{D}\hat{\alpha}_{ij}\|_2^2, \tag{5}$$

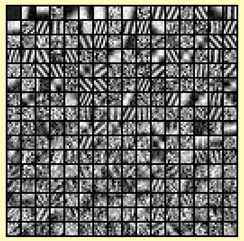
where w_{ij} is calculated by

$$w_{ij} = \frac{\sigma^2}{\|R_{ij}Y - \hat{D}\hat{\alpha}_{ij}\|_2^2 - \lambda\sum_{k=1}^{\|\hat{\alpha}_{ij}\|_0} \ln(p_{k,sel_k})}.$$

The weight represents the confidence of $R_{ij}X$ to be $\hat{D}\hat{\alpha}_{ij}$. Better decomposition yields larger w_{ij}.

σ	Lena		Barbara		Boats		House	
5	**38.61**	38.48	**38.08**	37.98	**37.24**	36.88	39.39	**39.48**
10	**35.50**	35.49	34.41	**34.50**	33.64	**33.67**	35.91	**36.18**
15	33.70	**33.84**	32.40	**32.52**	31.70	**31.89**	34.29	**34.42**
25	31.37	**31.58**	29.60	**29.88**	29.30	**29.64**	32.09	**32.10**
50	27.80	**27.94**	25.55	**26.07**	26.00	**26.38**	**28.01**	27.97

As shown above, in each cell, two denoising results are reported and the higher one is bolded. Left: Results of Elad etc. [3]. Right: Results of the method($\mu = 2.5/\sigma^2$). Our method outperforms the K-SVD in most cases.



In the image "Barbara", which contains high frequency texture features, PK-SVD can better learn the specific characteristics of the image. The learnt dictionary is shown above for $\sigma = 50$. In the figure, we sort the atoms according to its popularity, the upper ones are used more often.

Funding

This work is supported by National Natural Science Foundation of China (60702044, 60932006, 60625103) and 973 Program (2010CB731401).

References

[1] M. Aharon, M. Elad and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation In *IEEE Trans. on Signal Processing 2006*
[2] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation In *IEEE Trans. on Inform. Theory 2004*
[3] M. Elad and M. Aharon. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries In *IEEE Trans. on Image Processing 2006*