# DICTIONARY LEARNING FOR SPARSE REPRESENTATION USING WEIGHTED $l_1$-NORM


Haoli Zhao, Shuxue Ding, Yujie Li, Zhenni Li, Xiang Li and Benying Tan

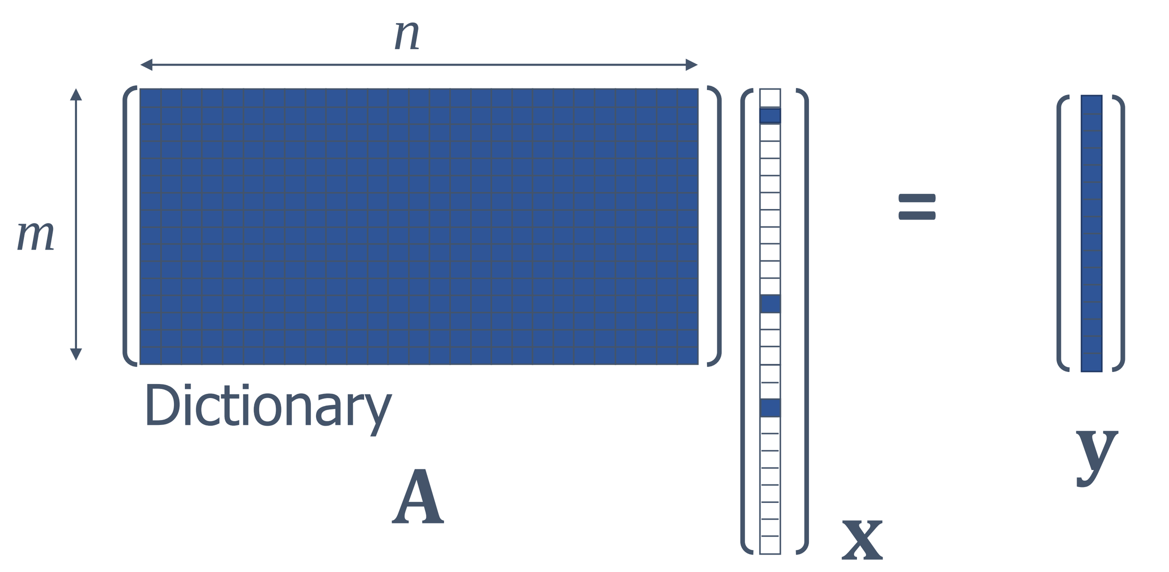
School of Computer Science and Engineering， The University of Aizu


## Introduction

**Goal**: An efficient algorithm for overcomplete dictionary learning with $l_p$-norm$(0 < p < 1)$ as sparsity constraint to achieve sparse representation from a set of known signals in the synthesis model.

### The Synthesis Model

- The signal model: $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$




- To approach sparsest possible $\mathbf{x}$, additional sparse constraint is introduced:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_p \text{ subject to } \|\mathbf{y} - \mathbf{Ax}\|_2 \le \varepsilon,$$

- $0 < p < 1$ was chosen in the proposed algorithm as to enhance sparseness of solutions compared to $l_1$-norm [1].

### Dictionary Learning

- To learn the overcomplete dictionary $\mathbf{A}$ from the signal sample itself,
- Result in better matching to the contents of the signals and representation of signals with fewer atoms of the dictionary [2].

## Formulation

- The problem is formed as:

$$(\mathbf{A}, \mathbf{X}) = \arg\min_{\mathbf{X},\mathbf{A}} \left(\frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|^2 + \lambda\|\mathbf{X}\|_p^p\right),$$

- The concave $l_p$-norm is reformed as a convex weighted $l_1$-norm:

$$d_w\left(\mathbf{x}^{(iter)}\right) = \sum_{k=1}^{n} |\mathbf{x}(k)^{iter-1}|^{p-1} |\mathbf{x}(k)^{iter}|,$$

- A smoothed approximation for the absolute value is introduced: $|x| \approx \frac{1}{c}\log\cosh(cx)$,
- The hierarchically alternating update strategy [3] is employed:

$$(\{\mathbf{x}_{i:}\}, \{\mathbf{a}_{:i}\}) = \arg\min_{\mathbf{x}_{i:},\mathbf{a}_{:i}} \left(\frac{1}{2}\left\|\mathbf{a}_{:i}\mathbf{x}_{i:} - \left(\mathbf{y} - \sum_{k=1,k\neq i}^{n} \mathbf{a}_{:k}\mathbf{x}_{k:}\right)\right\|^2 + \lambda d_w(\mathbf{x}_{i:})\right).$$

## Algorithm

**Algorithm 1**

**Task**: To find proper $\mathbf{A}$ and $\mathbf{X}$ to minimize$(\mathbf{A}, \mathbf{X}) = \arg\min_{\mathbf{X},\mathbf{A}} \left(\frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|^2 + \lambda\|\mathbf{X}\|_p^p\right)$

**Initialization**: input data set **Y** and proper λ and c

- Initialize matrices $\mathbf{A}$ and $\mathbf{X}$

**Main iteration**: increment iter by 1

- Sparse representation: calculate value of X entry by entry,

$$x_{i:}(\mathrm{l})^{(\mathrm{iter})} = \frac{\mathbf{a}_{:i}{}^T\mathbf{y}_i - \lambda\left|x_{i:}(l)^{(iter-1)}\right|^{p-1}\tanh\left(cx_{i:}(l)^{(iter)}\right)}{\mathbf{a}_{:i}{}^T\mathbf{a}_{:i}}; l = 1, \dots, N; i = 1, \dots, n,$$

set $p = 1$ at the first iteration;
- Dictionary learning: update A column by column, $\mathbf{a}_{:i}{}^{(iter)} \leftarrow \tilde{\mathbf{y}}_i\mathbf{x}_{i:}{}^T(\mathbf{x}_{i:}\mathbf{x}_{i:}{}^T)^{-1}; i = 1, \dots, n,$
- Stopping rule: stop if both $\mathbf{A}$ and $\mathbf{X}$ have converged or iter reach preset max iteration.

**Output**: $\mathbf{A}$ and $\mathbf{X}$

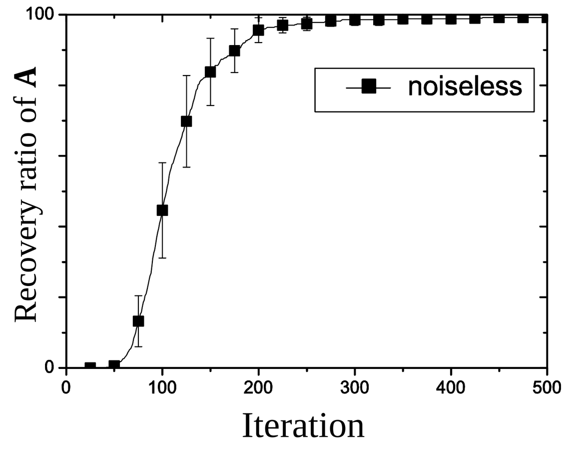
## Experiments

### Experiment Setup

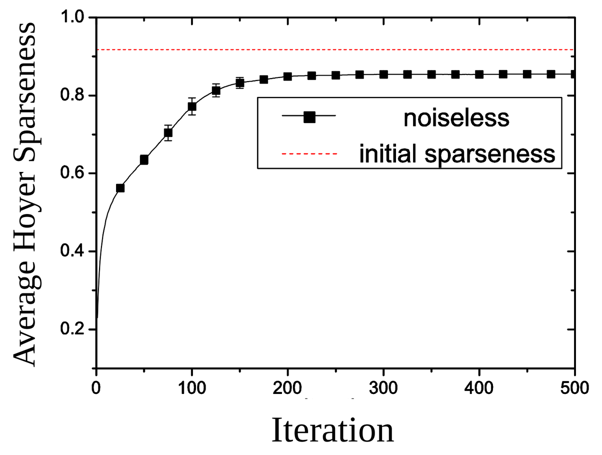
- Dimension of dictionary: $m = 20, n = 50$,
- Sample size of signals: $N = 1500$.

### Performance of Algorithm

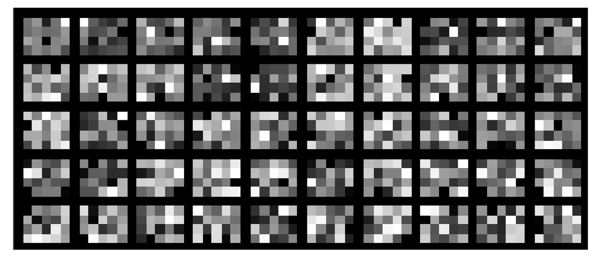
- $p = 0.9$, no noise:



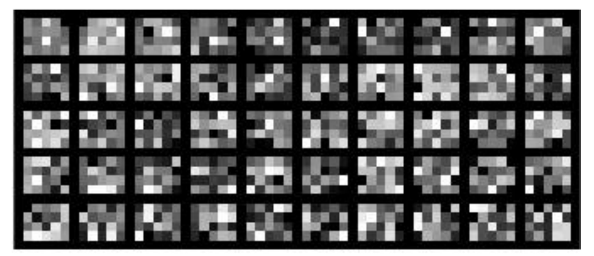

(a)




(b)



(c)

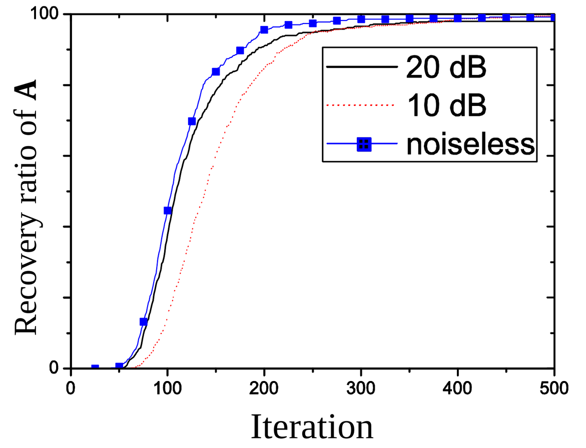


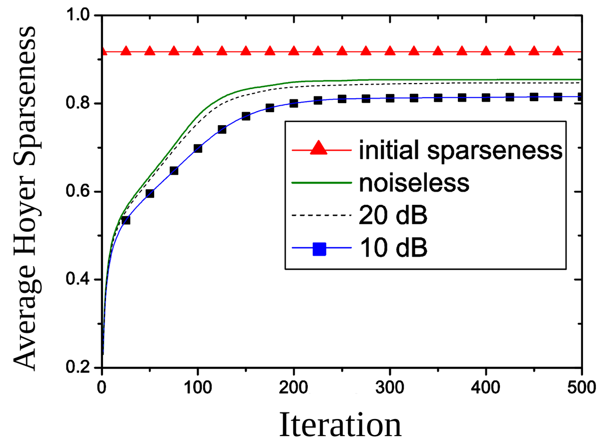
(d)

Experimental result of recovered dictionary after 500 iterations (c) and the ground true dictionary (d) are present in $4 \times 5$-dimensional subspaces.

### Comparison in varied noise levels

- $p = 0.9$

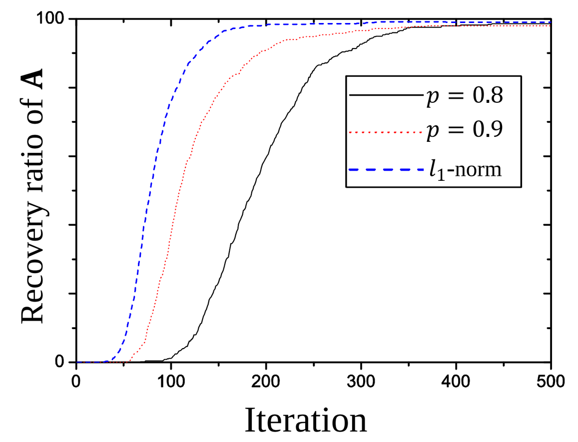


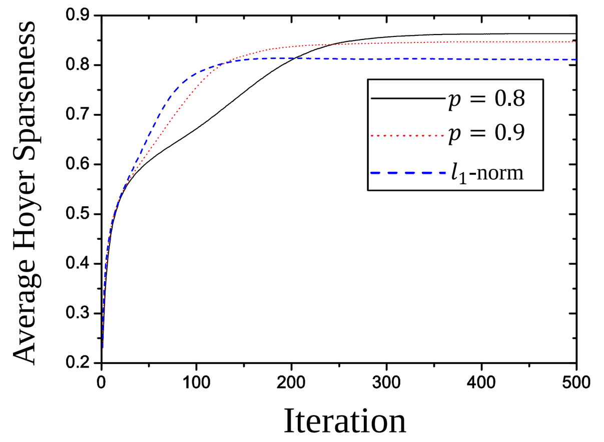




## Experiments

### Comparison with varied $p$ values

- SNR $= 20$dB




### Comparison with different algorithms

| Algorithms | Experiments results | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Recovery ratio of **A** (%) | | | Sparseness ($10^{-2}$) | | | Time spent until convergence (sec) | | |
| | 10 dB | 20 dB | noiseless | 10 dB | 20 dB | noiseless | 10 dB | 20 dB | noiseless |
| Proposed Algorithm | 99.5 ± 1.2 | 98.0 ± 1.8 | 99.2 ± 1.0 | 81.5 ± 0.2 | 84.7 ± 0.2 | 85.5 ± 0.4 | 19.5 ± 1.9 | 15.6 ± 2.2 | 15.4 ± 2.3 |
| K-SVD [4] | 84.8 ± 9.0 | 95.6 ± 3.6 | 94.5 ± 3.0 | 91.4 ± 0.1 | 91.7 ± 0.1 | 91.8 ± 0.1 | 14.5 ± 5.7 | 14.6 ± 7.0 | 11.9 ± 4.3 |
| MOD [5] | 85.4 ± 6.2 | 91.0 ± 3.0 | 91.7 ± 3.3 | 91.4 ± 0.1 | 91.6 ± 0.1 | 91.7 ± 0.1 | 9.8 ± 2.0 | 9.5 ± 3.8 | 10.0 ± 4.9 |
| FOCUSS-CNDL [6] | 84.6 ± 3.8 | 90.0 ± 2.4 | 90.2 ± 3.0 | 88.0 ± 0.5 | 90.5 ± 0.3 | 90.9 ± 0.3 | 29.9 ± 3.8 | 24.8 ± 7.3 | 27.8 ± 4.7 |

### Application in image denoising

- $p = 0.9$, initial SNR $= 22.10$dB, after 20 iterations, SNR $= 27.59$dB

## Conclusion

- The combination of hierarchically alternating update strategy and weighted $l_1$-norm method is effective in the synthesis model;
- The weighted $l_1$-norm is valid to reduce sparseness with lower $p$ value;
- The proposed algorithm has good robustness against noise.
- The proposed algorithm can be capable with image denoising
- Future work: mathematically rigorous convergence analysis

## Key Reference


[1] Li, Z., Ding, S., Li, Y., & Chen, W. (2015). Dictionary learning with ℓ1/2 regularizer for sparsity based on proximal operator. *2015 IEEE 7th International Conference on Awareness Science and Technology*.

[2] Elad, M. & Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Transactions on Image Processing*, 15(12), 3736–3745.

[3] Gillis, N., & Glineur, F. (2012). Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization. *Neural Computation*, 24(4), 1085-1105.

[4] Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*. 54(11), 4311–4322.

[5] Engan, K., Aase, S. O., & Husoy, J. H. (1999). Method of optimal directions for frame design. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*. 5, 2443-2446

[6] Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T., & Sejnowski, T. J. (2003). Dictionary Learning Algorithms for Sparse Representation. *Neural Computation*, 15(2), 349-396


## Contact


Haoli Zhao, Mail: D8172101@u-aizu.ac.jp

Professor Shuxue Ding, Mail: Sding@u-aizu.ac.jp