NONCONVEX SPARSE DICTIONARY LEARNING

Komal Krishna Mogilipalepu

Indian Institute of Science, Bangalore

1. INTRODUCTION

In the Dictionary learning problem, Given the signal $\mathbf{Y} \in \mathbb{R}^{m \times p}$, represent it into a linear combination of overcomplete set of basis vectors $\mathbf{D} \in \mathbb{R}^{m \times n}$ with sparse coefficients $\mathbf{X} \in \mathbb{R}^{n \times p}$ such as $\|\mathbf{Y} - \mathbf{DX}\|$ is minimized. The optimization problem follows as

$$\min_{\mathbf{D},\mathbf{X}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{DX}\|_F^2 \quad \text{s. t.} \quad \|x_i\|_0 \leq s \;\; \forall i \in p \quad \text{and} \quad \|d_j\|_2^2 = 1 \;\; \forall j \in n \tag{1}$$

the above problem is non-convex because we have to minimize jointly with $\mathbf{D}$ and $\mathbf{X}$. We can solve this using alternating minimization(AltMin), which is the heuristic approach and can provide local convergence with appropriate assumptions. Many works solve Eq (1) using AltMin, where we will keep the dictionary fixed when updating the sparse coefficients and viceversa. Aharon *et al.* [1] proposed KSVD a heuristic approach where the dictionary update is by approximating the residual with the highest singular value vectors. Rubinstein *et al.* [2] extend KSVD by considering the dictionary is sparse in a standard DCT or wavelet domain, which is known as Double sparsity model. Inspired by double sparsity, we model dictionary learning problem as two stage sparse coding approach where we will only solve two sparse recovery optimization problems in each stage. Although our approach is more related to double sparsity model but there are key differences in our approach, which are -

• Double sparsity model is proposed as extension of KSVD which is SparseKSVD but in our approach we are not using any SVD operation. In other words we are solving a two stage sparse recovery optimization problem.

• In SparseKSVD they will update the sparse coefficients including the sparse dictionary in the dictionary learning stage, where we will only update sparse dictionary in our approach.

• KSVD and SparseKSVD uses the OMP, inspired by LASSO formulation we use convex and non-convex iterative methods in our proposed approach.

Dictionary learning using KSVD models are lack of convergence gaurantees. Agarwal *et al.* [3] shown the local convergence of overcomplete dictionary estimation by AltMin approach, where they update the dictionary by least square solution. Inspired by [3] we would like to give local convergence gaurantees to our algorithm.

2. PROBLEM FORMULATION

Given samples $\mathbf{Y} \in \mathbb{R}^{m \times P}$ and a fixed dictionary $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$, estimate sparse matrix $\mathbf{A} \in \mathbb{R}^{n \times K}$ which selects few dictionary atoms in $\mathbf{\Phi}$ for the given samples $\mathbf{Y}$, and sparse coefficients $\mathbf{X} \in \mathbb{R}^{K \times P}$.

$$\min_{\mathbf{X},\mathbf{A}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi AX}\|_F^2 \quad \text{s. t.} \quad \|x_p\|_0 \leq s \;\; \forall p \in P, \quad \|a_k\|_0 \leq s_A \;\; \forall k \in K \quad \text{and} \quad \|\mathbf{\Phi}a_k\|_2^2 = 1 \;\; \forall k \in K. \tag{2}$$

We solve the optimization problem (2) using AltMin, which has two stages dictionary update stage followed by sparse coding stage. Alternatively, we will solve the following two optimization problems until local convergence.

$$\mathbf{X}_{est} = \min_{\mathbf{X}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi A}_{est}\mathbf{X}\|_F^2 \quad \text{s. t.} \quad \|x_p\|_0 \leq s \;\; \forall p \in P,$$

$$\mathbf{A}_{est} = \min_{\mathbf{A}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi A X}_{est}\|_F^2 \quad \text{s. t.} \quad \|a_k\|_0 \leq s_A \;\; \forall k \in K \quad \text{and} \quad \|\mathbf{\Phi}a_k\|_2^2 = 1 \;\; \forall k \in K. \tag{3}$$

3. SIMULATION RESULTS

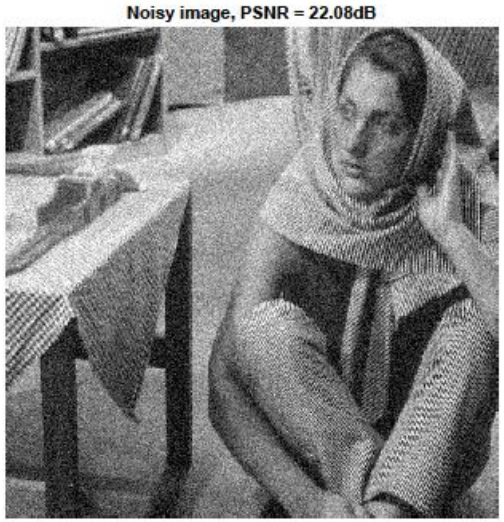
4. CONCLUSION

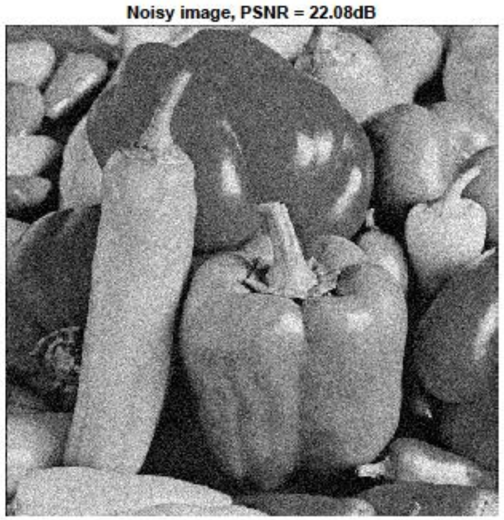
5. REFERENCES

[1] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[2] Ron Rubinstein, Michael Zibulevsky, and Michael Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Transactions on signal processing*, vol. 58, no. 3, pp. 1553–1564, 2009.

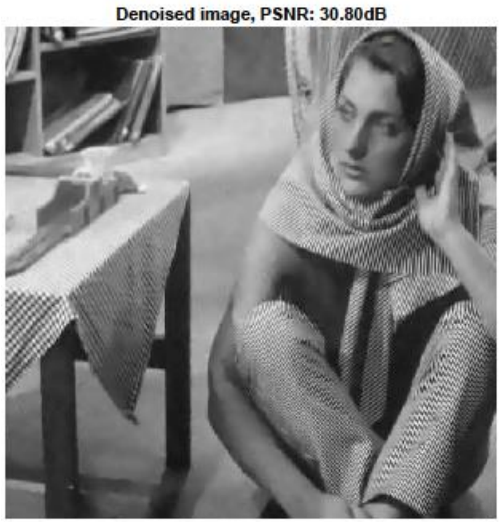
[3] Alekh Agarwal, Animashree Anandkumar, Prateek Jain, Praneeth Netrapalli, and Rashish Tandon, "Learning sparsely used overcomplete dictionaries," in *Conference on Learning Theory*. PMLR, 2014, pp. 123–137.



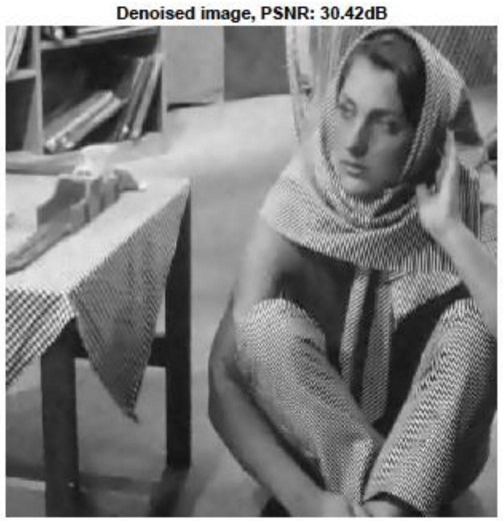


Set 11	KSVD	KSVDS	L1	MCP	LP
Barbara	30.80	30.42	30.73	30.69	30.69
Boat	30.38	30.24	30.35	30.38	30.34
House	33.30	33.14	33.55	33.56	33.44
Lenna	32.43	32.38	32.44	32.46	32.41
Peppers	32.20	32.14	32.32	32.33	32.31

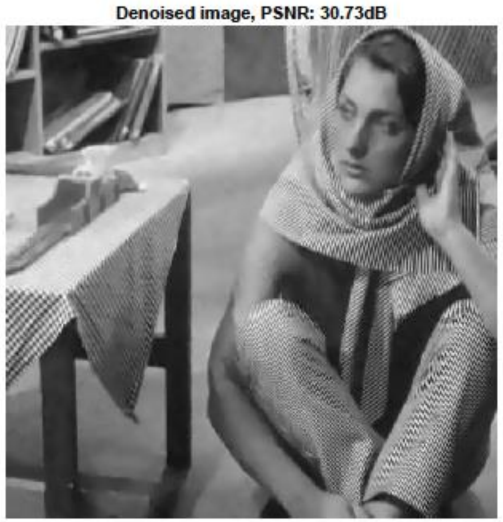
PSNR Denoising Performance Comparison



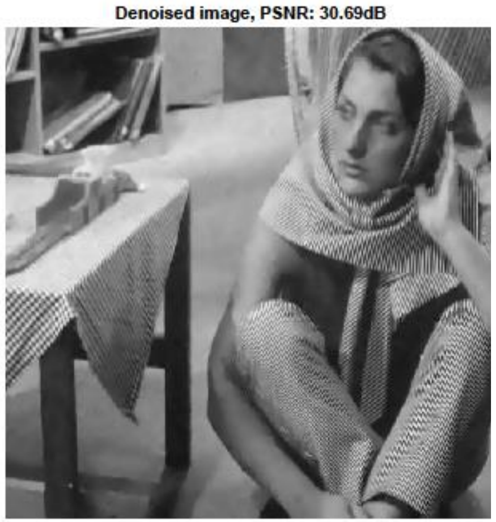
KSVD



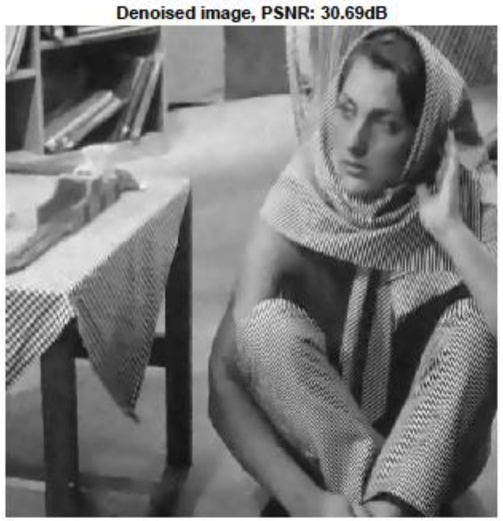
KSVDS



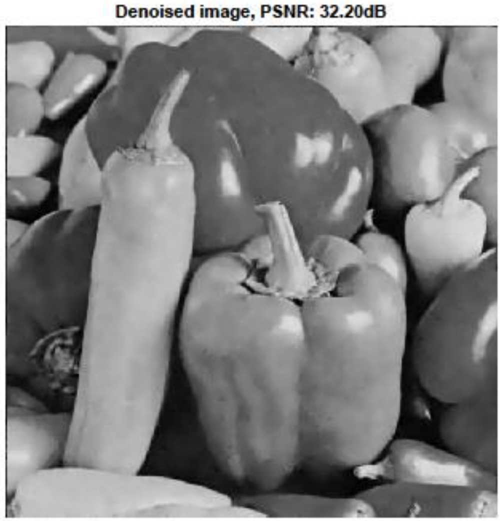
L1



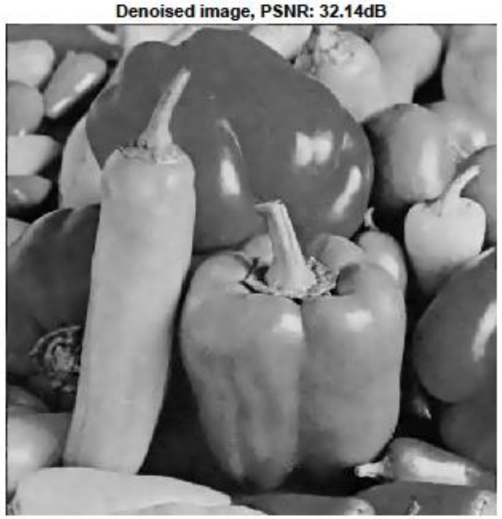
MCP



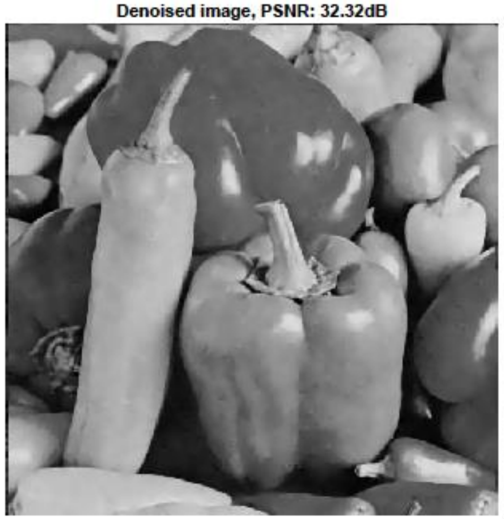
LP



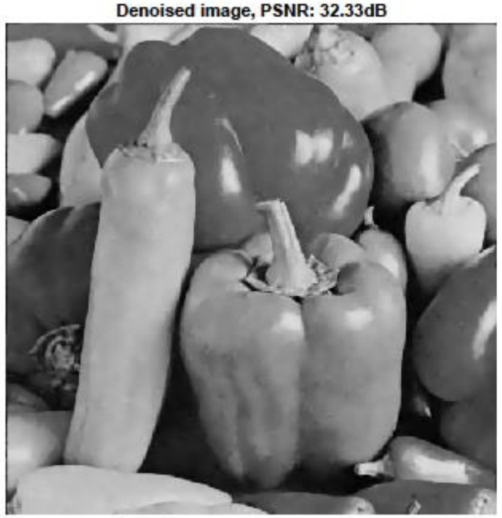
KSVD



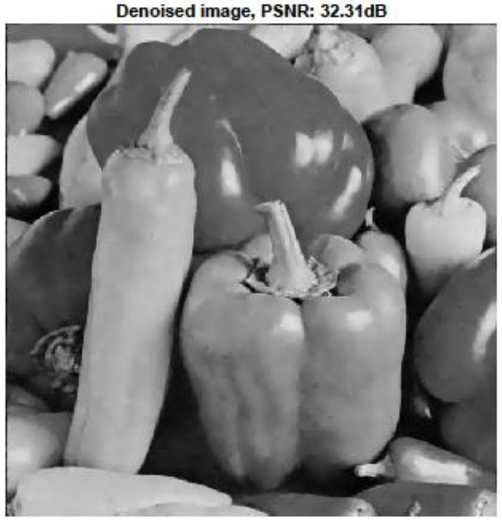
KSVDS



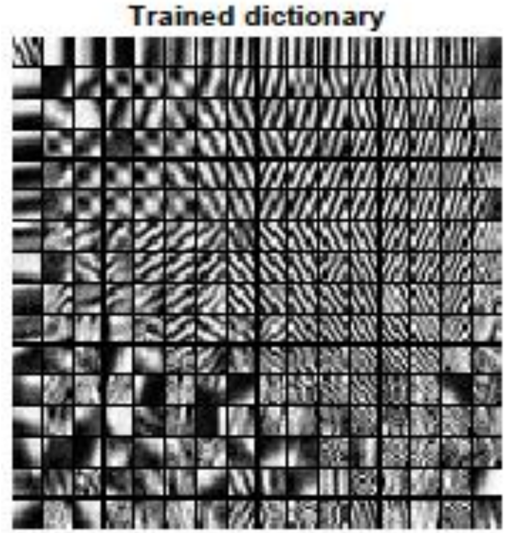
L1



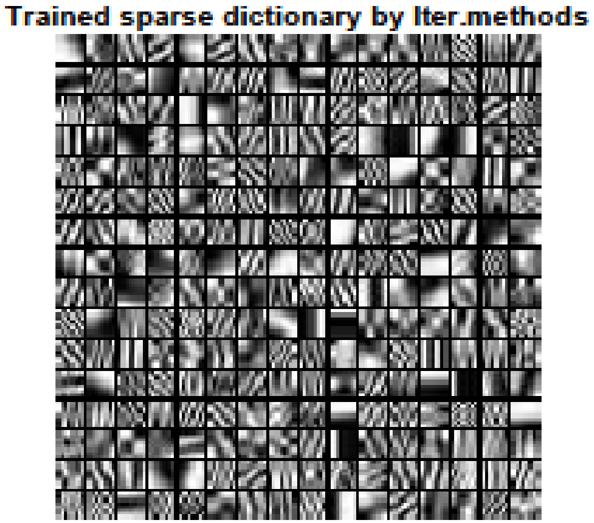
MCP



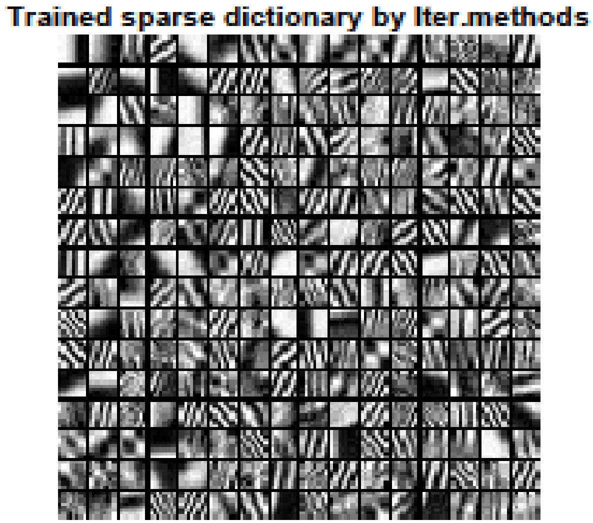
LP



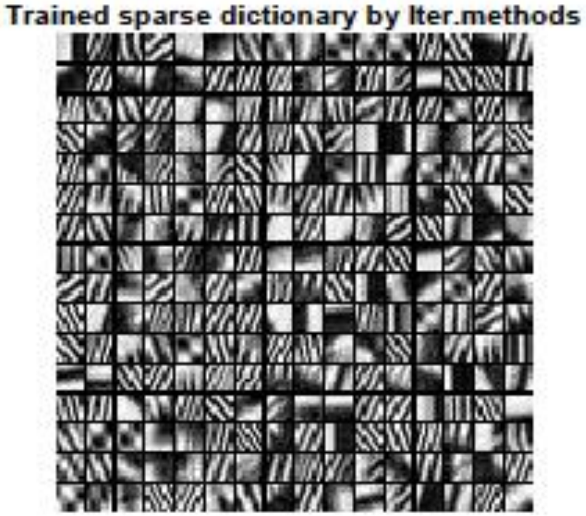
KSVD



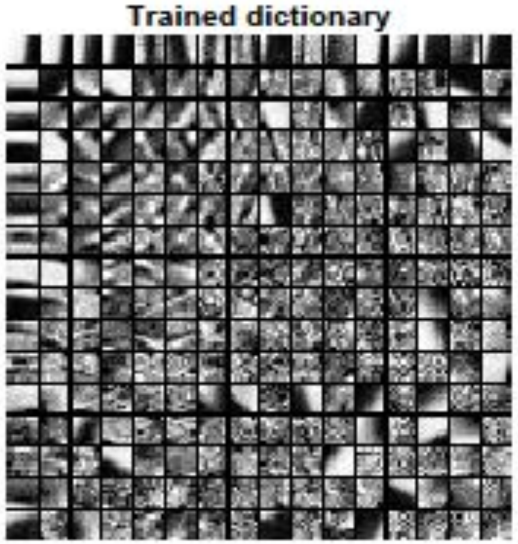
KSVDS



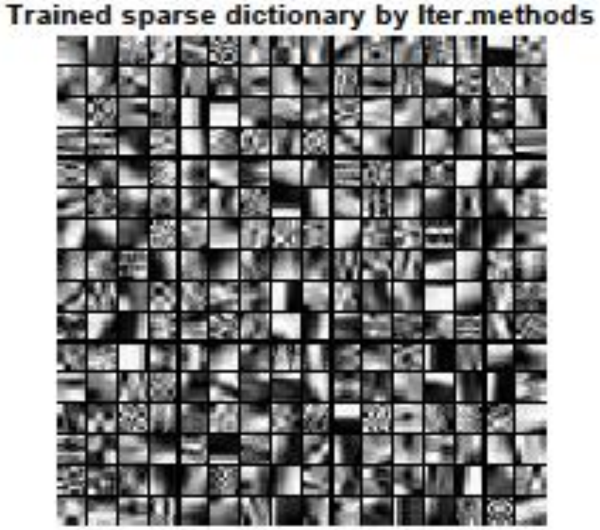
L1



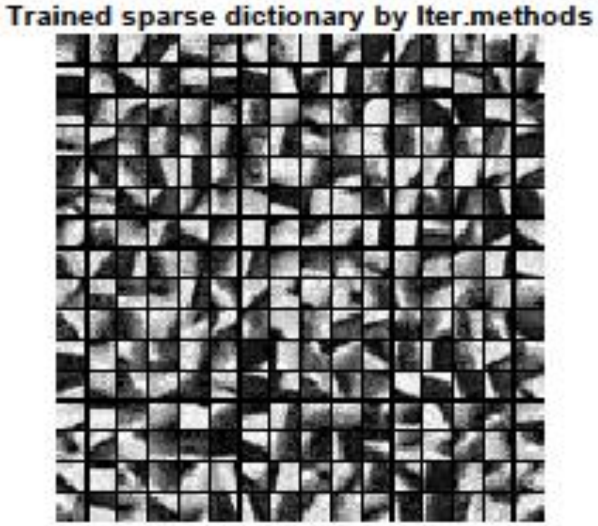
MCP



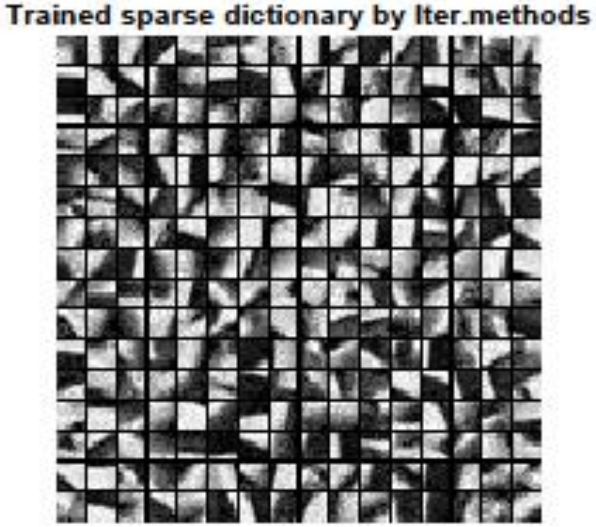
KSVD



KSVDS



L1



MCP