Isfahan University of Technology

UNIVERSITY OF MICHIGAN

COHERENCE REGULARIZED DICTIONARY LEARNING

M. Nejati[1], S. Samavi[1], S.M.R. Soroushmehr[2], K. Najarian[2,3]

[1]Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran
[2]Emergency Medicine Department, University of Michigan, Ann Arbor, USA
[3]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, USA

Introduction

Sparsifying Dictionary:

- Dictionary plays a critical role in a successful sparse representation modeling.
- Learned overcomplete dictionaries have become popular in recent years.

Introduction

Dictionary Learning:

Objective: adapting dictionary to data for their sparse representations.

$$(\widehat{\mathbf{D}}, \widehat{\mathbf{X}}) = \underset{\mathbf{D},\mathbf{X}}{\operatorname{argmin}} \ \underbrace{\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2}_{\text{Data fitting}} + \lambda \underbrace{\sum_{i=1}^{N} \|\mathbf{x}_i\|_0}_{\text{Sparsity Regularizer}}$$

$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K]$ Dictionary
$\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ Training signals
$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ Sparse representation of **Y**

Mutual Coherence of Dictionary

- Important dictionary property which measures the maximal correlation of any two distinct atoms in the dictionary: $\mu(\mathbf{D}) \stackrel{\text{def}}{=} \max_{i \neq j} |\langle \mathbf{d}_i, \mathbf{d}_j \rangle| = \max_{i \neq j} |\mathbf{d}_i^T \mathbf{d}_j|$

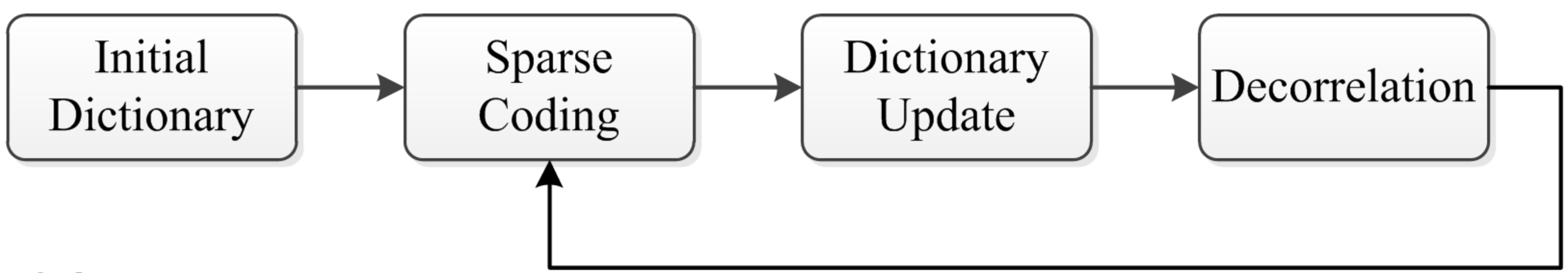
Mutual Coherence of Dictionary

- Importance of mutual coherence:
  - ✓ direct impact on stability and performance of sparse coding algorithms.
  - ✓ lower coherence permits better sparse recovery.
  - ✓ reduction of over-fitting to the training data.

Coherence Reduction Strategies

a) Atom Decorrelation

Adding a decorrelation step to the existing methods.



Disadvantages:

- extra computation cost of decorrelation step.
- approximation error is not considered in decorrelation step.

Coherence Reduction Strategies

b) Coherence Penalty

Augmenting the dictionary learning objective with a coherence penalty (regularization):

$$\left(\widehat{\mathbf{D}}, \widehat{\mathbf{X}}\right) = \underset{\mathbf{D},\,\mathbf{X}}{\operatorname{argmin}} \, \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \sum_{i=1}^{N} \|\mathbf{x}_i\|_0 + \operatorname{Corr}(\mathbf{D})$$

Proposed Learning Model

Our **Co**herence **Re**gularized (CORE) model:

$$\min_{\mathbf{D}\in\mathcal{D},\,\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \sum_{i=1}^{N} \|\mathbf{x}_i\|_0 + \eta \underbrace{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \left(\mathbf{d}_i^T \mathbf{d}_j\right)^2}_{\text{Coherence regularization}}$$

Proposed Learning Model

Alternate minimization scheme is used:

- Sparse coding: Orthogonal matching pursuit (OMP)
- Dictionary update: The focus of this paper
  - ✓ It is performed in a block coordinate fashion.
  - ✓ Simultaneous updating of an arbitrary subset of atoms is allowed.

Inter- and Intra-coherence Penalties:

Suppose we want to update a subset $\mathbf{D}_\Omega = [\mathbf{d}_i]_{i\in\Omega}$ and the rest $\mathbf{D}_{\bar{\Omega}} = [\mathbf{d}_i]_{i\in\bar{\Omega}}$ is fixed. Then we have to solve:

$$\min_{\mathbf{D}_\Omega} \left\|\mathbf{E}_\Omega - \mathbf{D}_\Omega \mathbf{X}_{[\Omega]}\right\|_F^2 + \underbrace{\eta\left\|\mathbf{D}_{\bar{\Omega}}^T \mathbf{D}_\Omega\right\|_F^2}_{\text{Inter-Coherence}} + \underbrace{\frac{\eta}{2}\left\|\mathbf{D}_\Omega^T \mathbf{D}_\Omega - \mathbf{I}\right\|_F^2}_{\text{Intra-Coherence}}$$

where $\mathbf{X}_{[\Omega]} = \mathbf{X}(\Omega, :)$ and $\mathbf{E}_\Omega = \mathbf{Y} - \mathbf{D}_{\bar{\Omega}} \mathbf{X}_{[\bar{\Omega}]}$.

Proposed CORE-I Update

Consider the inter-coherence penalty.

By differentiation w.r.t $\mathbf{D}_{\Omega}$ we have:

$$\eta \underbrace{\mathbf{D}_{\overline{\Omega}}\mathbf{D}_{\overline{\Omega}}^{T}}_{\mathbf{A}}\mathbf{D}_{\Omega} + \mathbf{D}_{\Omega}\underbrace{\mathbf{X}_{[\Omega]}\mathbf{X}_{[\Omega]}^{T}}_{\mathbf{B}} = \underbrace{\mathbf{E}_{\Omega}\mathbf{X}_{[\Omega]}^{T}}_{\mathbf{C}}$$

$$\mathbf{A}\mathbf{D}_{\Omega} + \mathbf{D}_{\Omega}\mathbf{B} = \mathbf{C}$$

Sylvester Equation

This matrix equation can be solved by standard methods.

Proposed CORE-II Update

Consider the both inter- and intra-coherence terms.

Proposed CORE-II Update

By differentiation of objective w.r.t $\mathbf{D}_\Omega$ we have:

$$\underbrace{\eta \mathbf{D}_{\overline{\Omega}} \mathbf{D}_{\overline{\Omega}}^T}_{\mathbf{A}} \mathbf{D}_\Omega + \mathbf{D}_\Omega \underbrace{\left(\mathbf{X}_{[\Omega]} \mathbf{X}_{[\Omega]}^T + \eta \mathbf{D}_\Omega^T \mathbf{D}_\Omega - \eta \mathbf{I}_m \right)}_{\mathbf{B}} = \underbrace{\mathbf{E}_\Omega \mathbf{X}_{[\Omega]}^T}_{\mathbf{C}}$$

We use an iterative scheme to update:

$$\mathbf{D}_\Omega^{(t+1)} \leftarrow \mathrm{Sylv}\left(\mathbf{D}_\Omega^{(t)}; \mathbf{A}, \mathbf{B}^{(t)}, \mathbf{C}^{(t)} \right)$$

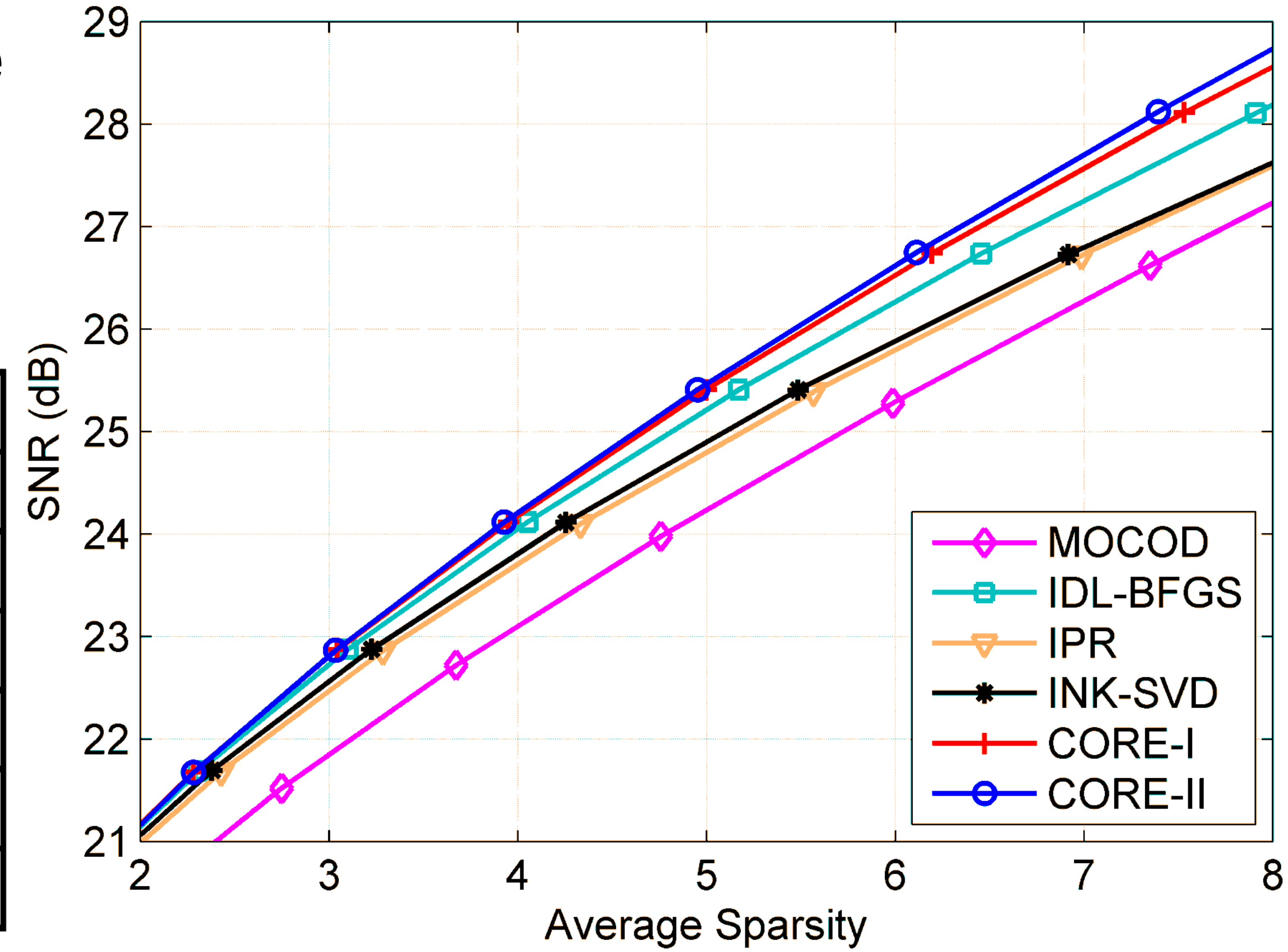
Experimental Results

- Comparison to several incoherent dictionary learning algorithms.
  - ✓ INK-SVD [1], IPR [2], MOCOD [3], IDL-BFGS [4].
- Training on 8x8 image patches and evaluating of sparse approximation's SNR (dB) on test set.

Experimental Results

Table 1. Comparison results in terms of average mutual coherence of trained dictionary, sparse reconstruction performance on test set, and learning run time

Algorithm	μ_{avg}	SNR (dB)	Run Time (s)
CORE-I	0.1080	28.56	149
CORE-II	0.0919	28.73	201
INK-SVD [1]	0.1915	27.62	402
IPR [2]	0.2169	27.59	731
MOCOD [3]	0.1388	27.23	120
IDL-BFGS [4]	0.1258	28.18	608



References

[1] B. Mailhé, D. Barchiesi, and M. D. Plumbley, "INK-SVD: Learning incoherent dictionaries for sparse representations," in *Proc. ICASSP*, 2012, pp. 3573–3576.

[2] D. Barchiesi and M. D. Plumbley, "Learning incoherent dictionaries for sparse approximation using iterative projections and rotations," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2055–2065, 2013.

[3] I. Ramirez, F. Lecumberry, and G. Sapiro, Sparse modeling with universal priors and learned incoherent dictionaries, Tech. Rep., IMA, University of Minnesota, 2009.

[4] C. D. Sigg, T. Dikk, and J. M. Buhmann, "Learning dictionaries with bounded self-coherence," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 861–864, Dec. 2012.