Low Mutual and Average Coherence Dictionary Learning Using Convex Approximation

Javad Parsa[1], Mostafa Sadeghi[1], Massoud Babaie-Zadeh[1]

and Christian Jutten[2]

1) Electrical Engineering Department, Sharif University of Technology, Tehran, IRAN.
2) GIPSA-Lab, University of Grenoble Alpes, Grenoble, France.

Dictionary learning

- Dictionary learning problem as follows[1]:

1) Sparse representation:

$$\boldsymbol{X}^{(k+1)} = \underset{\boldsymbol{X}\in\chi}{\operatorname{argmin}} \left\| \boldsymbol{Y} - \boldsymbol{D}^{(k)}\boldsymbol{X} \right\|_F^2 \qquad (2)$$

2) Dictionary update:

$$\boldsymbol{D}^{(k+1)} = \underset{\boldsymbol{D}\in\xi}{\operatorname{argmin}} \left\| \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}^{(k+1)} \right\|_F^2 \qquad (3)$$

Stage 1 is an ordinary sparse coding problem [2]-[5], which can be done, for example, by Orthogonal Matching Pursuit (OMP) [6].

Mutual and Average Coherence

- 1) Mutual coherence:

$$\mu = \max_{i \neq j} \frac{|d_i^T d_j|}{\|d_i\|_2 \|d_j\|_2} \text{ [7]}, \; D \in \mathbb{R}^{m \times n} \rightarrow \mu_{welch} = \sqrt{\frac{n-m}{m(n-1)}}$$

$$\mu_{welch} \leq \mu \leq 1 \text{ [8]}$$

- 2) Average coherence:

$$\mu_{avg} = \sqrt{\frac{\|D^T D - I\|_F^2}{n(n-1)}}$$

- Mutual coherence plays an important role in sparse approximation. A signal with a sparse representation x with sparsity level (number of non-zero elements) s, can be recovered from $y = Dx$ when [9]:

$$s \leq \frac{1}{2}\left(1 + \frac{1}{\mu}\right)$$

- According to above equation, dictionaries with low mutual coherence are better for high s.

Decreasing mutual and average coherence

- Some of algorithms reducing mutual coherence of dictionary by solving the following problem [10]:

$$min_{D\in\zeta, X\in\chi, H\in\upsilon}\|Y - DX\|_F^2 + \frac{\lambda}{2}\|D^T D - H\|_F^2 \quad (H)$$

$$\upsilon = \{H \in \mathbb{R}^{n\times n}: H = H^T, h_{ii} = 1, \forall i \max_{i\neq j}|h_{ij}| \leq \mu_0\}$$

This problem does not yield closed form solution for updating the dictionary because the gradient of above cost function over *D* is not linear.

Our proposed method

- Convex approximation[11]:

$$\boldsymbol{D} = \boldsymbol{D}_0 + \boldsymbol{D}\text{-}\, \boldsymbol{D}_0 \quad , \quad \boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{X}\text{-}\, \boldsymbol{X}_0$$

$$\boldsymbol{D}\boldsymbol{X} = (\boldsymbol{D}_0 + \boldsymbol{D} - \boldsymbol{D}_0)(\boldsymbol{X}_0 + \boldsymbol{X} - \boldsymbol{X}_0) \approx \boldsymbol{D_0}\boldsymbol{X} + \boldsymbol{D}\boldsymbol{X_0} - \boldsymbol{D_0}\boldsymbol{X_0}$$

$$\boldsymbol{D}^T\boldsymbol{D} = (\boldsymbol{D_0} + \boldsymbol{D} - \boldsymbol{D_0})^T(\boldsymbol{D_0} + \boldsymbol{D} - \boldsymbol{D_0}) \approx \boldsymbol{D_0^T}\boldsymbol{D} + \boldsymbol{D}^T\boldsymbol{D_0} - \boldsymbol{D_0^T}\boldsymbol{D_0}$$

in which, it is assumed that $\|(\boldsymbol{D} - \boldsymbol{D}_0)(\boldsymbol{X} - \boldsymbol{X}_0)\|_F$ and $\|(\boldsymbol{D} -$

$$(\boldsymbol{D}^*, \boldsymbol{X}^*) = \underset{\boldsymbol{D}\in\xi, \boldsymbol{X}\in\chi}{\operatorname{argmin}} \|\boldsymbol{Y} + \boldsymbol{D_0}\boldsymbol{X_0} - \boldsymbol{D_0}\boldsymbol{X} - \boldsymbol{D}\boldsymbol{X_0}\|_F^2 + \frac{\lambda}{2}\left\|\boldsymbol{D}^T\boldsymbol{D_0} + \boldsymbol{D_0^T}\boldsymbol{D} - \boldsymbol{D_0^T}\boldsymbol{D_0} - \boldsymbol{H}\right\|_F^2$$

Our proposed method *(continued)*

1) Sparse representation:

$$\mathbf{D} = \mathbf{D}_k, \mathbf{D_0} = \mathbf{D}_{k-1}, \mathbf{X_0} = \mathbf{X}_k, \mathbf{Z}_k = \mathbf{Y} - (\mathbf{D}_k - \mathbf{D}_{k-1})\mathbf{X}_k$$

$$\mathbf{X}_{k+1} = \underset{\mathbf{X} \in \chi}{\operatorname{argmin}} \|\mathbf{Z}_k - \mathbf{D}_{k-1}\mathbf{X}\|_{\mathrm{F}}^2$$

2) Dictionary updating:

$$\mathbf{X} = \mathbf{X_0} = \mathbf{X}_{k+1}, \mathbf{D_0} = \mathbf{D}_k$$

$$\boldsymbol{D}_{k+1} = \underset{\boldsymbol{D} \in \xi}{argmin} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}_{k+1}\|_F^2 + \frac{\lambda}{2} \left\|\boldsymbol{D}^T\boldsymbol{D}_k + \boldsymbol{D}_k^T\boldsymbol{D} - \boldsymbol{D}_k^T\boldsymbol{D}_k - \boldsymbol{H}_k\right\|_F^2$$

Our proposed method *(continued)*

- Dictionary updating:

$$M(\boldsymbol{D}) = \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}_{k+1}\|_F^2 + \frac{\lambda}{2}\left\|\boldsymbol{D}^T\boldsymbol{D}_k + \boldsymbol{D}_k^T\boldsymbol{D} - \boldsymbol{D}_k^T\boldsymbol{D}_k - \boldsymbol{H}_k\right\|_F^2$$

$$\nabla_{\boldsymbol{D}} M(\boldsymbol{D}) = (\boldsymbol{D}\boldsymbol{X}_{k+1} - \boldsymbol{Y})\boldsymbol{X}_{k+1}^T + \boldsymbol{\lambda}\boldsymbol{D}_k(\boldsymbol{D}^T\boldsymbol{D}_k + \boldsymbol{D}_k^T\boldsymbol{D} - \boldsymbol{D}_k^T\boldsymbol{D}_k - \boldsymbol{H}_k)$$

$$\boldsymbol{D}\boldsymbol{W}_k + \lambda\boldsymbol{D}_k\boldsymbol{D}^T\boldsymbol{D}_k + \lambda\boldsymbol{D}_k\boldsymbol{D}_k^T\boldsymbol{D} = \boldsymbol{C}_k$$

After some calculation (the details are in the paper), we can write:

$$vec(\boldsymbol{D}_{k+1}) = \left(\boldsymbol{W}_{\boldsymbol{k}} \otimes I_m + I_n \otimes \lambda\boldsymbol{D}_k^T\boldsymbol{D}_k + \lambda\boldsymbol{B}_k\right)^{-1} vec(\boldsymbol{C}_k)$$

Summary of our proposed method(Convex-GSD)

1) Sparse coding:

$$\mathbf{Z}_k = \mathbf{Y} - (\mathbf{D}_k - \mathbf{D}_{k-1})\mathbf{X}_k \Longrightarrow \boldsymbol{X}_{k+1} = \mathrm{OMP}(\boldsymbol{Z}_k, \boldsymbol{D}_{k-1}, s)$$

2) Dictionary updating:

$$vec(\boldsymbol{D}_{k+1}) = \left(\boldsymbol{W}_k \otimes I_m + I_n \otimes \lambda \boldsymbol{D}_k^T \boldsymbol{D}_k + \lambda \boldsymbol{B}_k\right)^{-1} vec(\boldsymbol{C}_k)$$

3) Updating matrix **H** [5]:

$$\mu_0 \geq \mu_{welch} \ and \ \eta = \mathbf{D}_{k+1}^T \mathbf{D}_{k+1}$$

$$\mathbf{H}_{k+1} = \begin{cases} 1 & i = j \\ \eta & i \neq j, |\eta| \leq \mu_0 \\ sgn(\eta)\mu_0 & i \neq j, |\eta| \geq \mu_0 \end{cases}$$

Simulation Results:

- Apply our innovation on the GSD[10] and RAMC[12] and results in Convex-GSD and Convex-RAMC, respectively.
- The criteria to evaluate:

1) Root Mean Square Error (RMSE) :

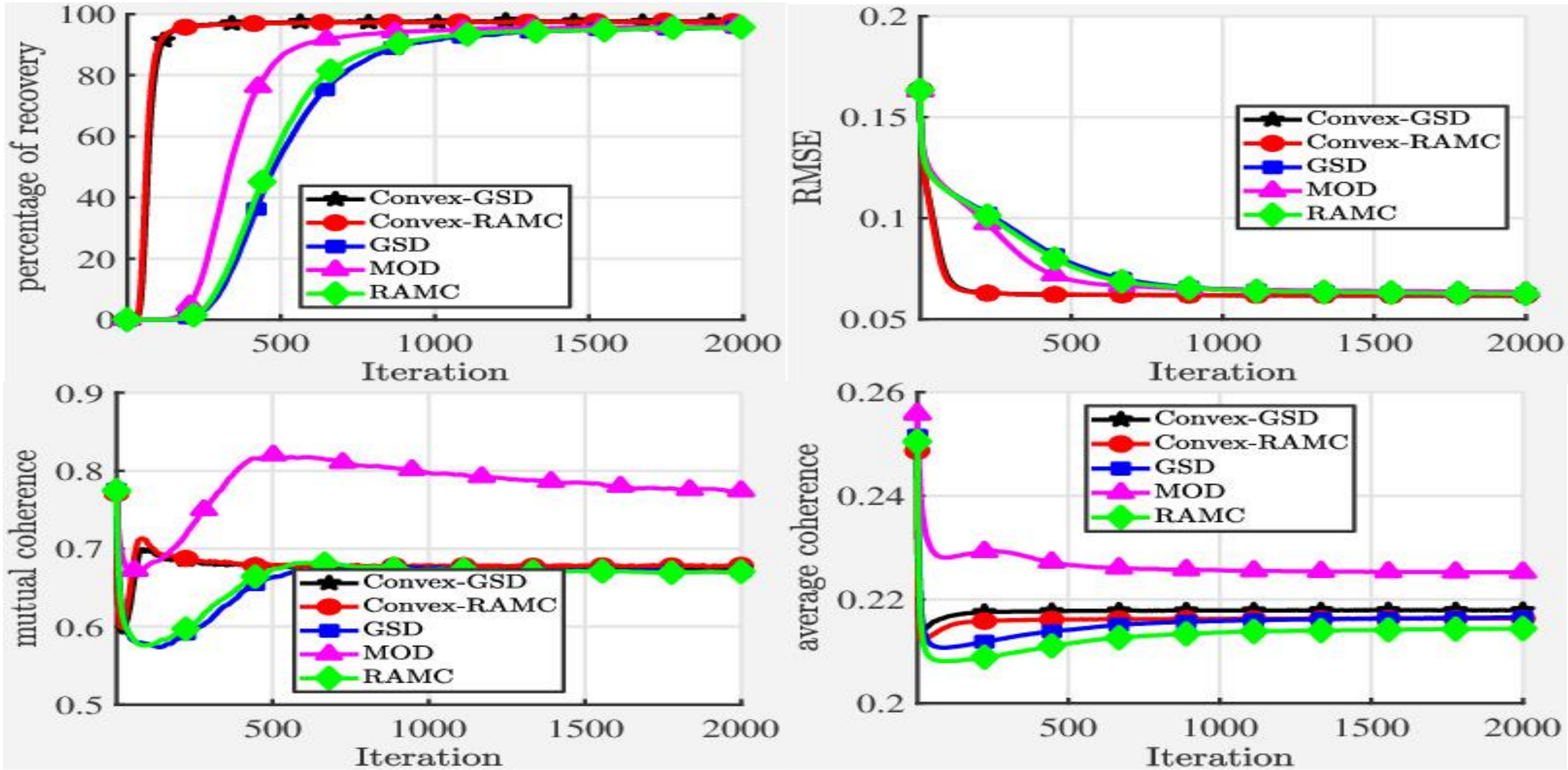
$$\varepsilon_K = \frac{\|\boldsymbol{Y} - \boldsymbol{D}^k \boldsymbol{X}^k\|_F}{\sqrt{mL}}$$

2) Percentage of atom recovery:

$$min_j(1 - |\boldsymbol{D}(:,i)^T \boldsymbol{D_t}(:,j)|) < 0.01.$$

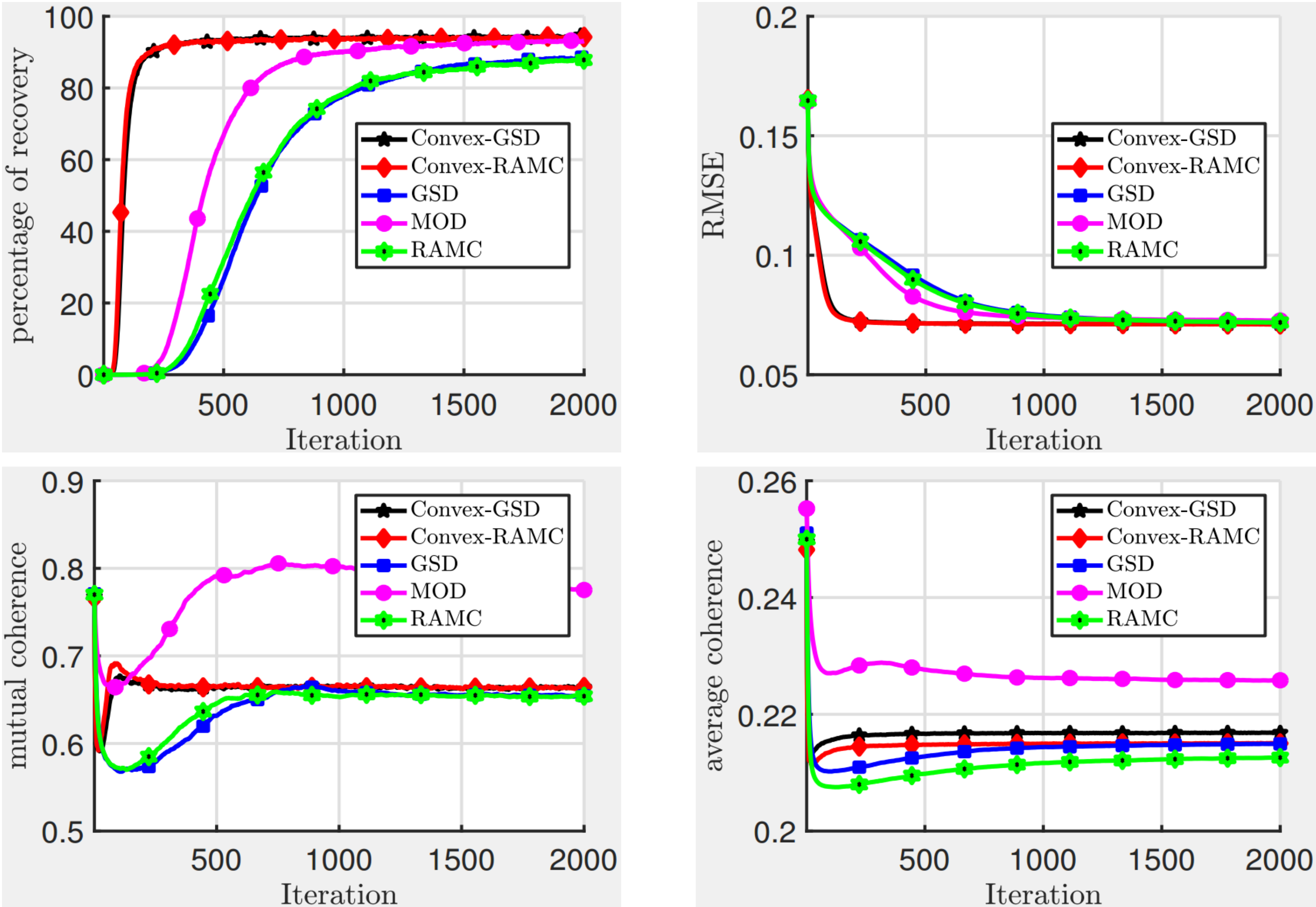
3) Mutual and average coherence.

$\boldsymbol{D} \in \mathbb{R}^{20\times 50}, s = 7, \lambda = 5, SNR = 30\text{dB}, \boldsymbol{Y} \in \mathbb{R}^{20\times 2500}$



According to this slide and slide 12 too, our two algorithms (include Convex) have higher convergence rate and lower RMSE in comparison to the other algorithms while mutual and average coherence are same.

$\boldsymbol{D} \in \mathbb{R}^{20\times 50}, s = 7, \lambda = 5, SNR = 20\text{dB}, \boldsymbol{Y} \in \mathbb{R}^{20\times 2500}$



Number of iterations and average running time (in seconds) for achieving percentage of recovery= 80. Average running times are reported in parentheses. In this table, $s = 7$ and SNR= 30dB are supposed.

Algorithm	$\lambda = 5$	$\lambda = 10$
Convex-GSD	91 (8.3s)	123 (12.1s)
Convex-RAMC	83 (7.5s)	102 (10.1s)
GSD	697 (18.6s)	1913 (59.3s)
RAMC	643 (17.2s)	626 (19.4s)
MOD	452 (9.4s)	449 (10.5s)

According to this table, our algorithms (include Convex) have higher convergence rate in comparison to the other algorithms.

Conclusions

- Convex method for dictionary learning with low mutual coherence.
- Our approach increases the convergence rate and decreases RMSE.
- Mutual and average coherence of our algorithms are reduced well.

References

1) R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," Proceedings of the IEEE, vol. 98, no. 6, pp. 1045–1057, 2010.

2) J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," Proceedings of the IEEE, vol. 98, no. 6, pp. 948–958, 2010.

3) H. Mohimani, M. Babaie-Zadeh, and Ch. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed `0 norm," IEEE Trans. on Signal Processing, vol. 57, pp. 289–301, 2009.

4) T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," Applied and Computational Harmonic Analysis, vol. 27, no. 3, pp. 265–274, 2009.

5) I. Daubechies, R. DeVore, M. Fornasier, and C. S. G¨unt¨urk, "Iteratively re-weighted least squares minimization for sparse recovery," Communications on Pure and Applied Mathematics, vol. 63, no. 1, pp. 1–38, Jan. 2010..

6) J. A. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Trans. Info. Theory, vol. 53, no. 12, pp. 4655–4666, 2007

7) D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

8) T. Strohmer and Heath R. W., "Grassmannian frames with applications to coding and communication," *Applied and Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257–275, 2003.

9) D. L. Donoho and M. Elad, "Optimally sparse representation in general (non orthogonal) dictionaries via $\ell 1$ minimization," *Proc. Nat. Aca. Sci*, vol. 100, no. 5, pp.2197–2202, 2003.

10) G. li, Z. Zhu, H. Bai, and A. Yu, "A new framework for designing incoherent sparsifying dictionaries," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

11) M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Dictionary learning for sparse representation: A novel approach," IEEE Signal Proc. Letters, vol. 20, no. 12, pp. 1195–1198, 2013.

12) J. Parsa, M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Joint low mutual and average coherence dictionary learning," in Proceeding of 26th European Signal Processing Conference(EUSIPCO 2018), 2018.

13) K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in Proceedings of IEEE ICASSP, 1999, vol. 5.

Thank you for your attention