OVERCOMPLETE SPARSIFYING TRANSFORM LEARNING ALGORITHM USING A CONSTRAINED LEAST SQUARES APPROACH

Ender M. EKŞİOĞLU and Özden BAYIR

Electronics and Communications Engineering Department, Istanbul Technical University, Istanbul, Turkey

The Question

For a given data set, how can we learn an overcomplete sparsifying transform?

Introduction

A recent analysis operator learning (AOL) algorithm has been presented in [1]. In [1], the learned analysis operators are constrained to lie in the set of Uniformly Normalized Tight Frames (UNTF). A new framework has been introduced in [2] as a more general paradigm for analysis operator learning. In this new "Sparsifying Transform Learning" framework, the minimization problem for operator learning is formulated in a modified manner when compared to the minimization problems of the analysis operator learning algorithms. The expensive cosparse coding step of the classic analysis operator learning algorithms gets replaced with a thresholding step of much reduced complexity.

A Solution

We develop a new sparsifying transform learning algorithm "Constrained Least Squares Sparsifying Transform Learning (CLS-TL)" by merging the transform learning approach of [2] with the constrained AOL algorithm of [1]. Despite its reduced complexity, the CLS-TL algorithm has performance comparable to the AOL algorithm.

Constrained AOL

Dictionary learning can be formalized as:

$$\min_{\mathbf{D}\in\mathscr{D},\mathbf{X}} \|\mathbf{D}\mathbf{X}-\mathbf{Y}\|_F^2, \text{ s.t. } \|\boldsymbol{x}_n\|_0 \le s \quad (1)$$

A noisy formulation of learning a suitable analysis operator for a given signal set can been given as:

$$\min_{\boldsymbol{\Omega}\in\mathscr{C},\mathbf{X}} \|\mathbf{X}-\mathbf{Y}\|_F^2, \text{ s.t. } \|\boldsymbol{\Omega}\boldsymbol{x}_n\|_0 \le s \quad (2)$$

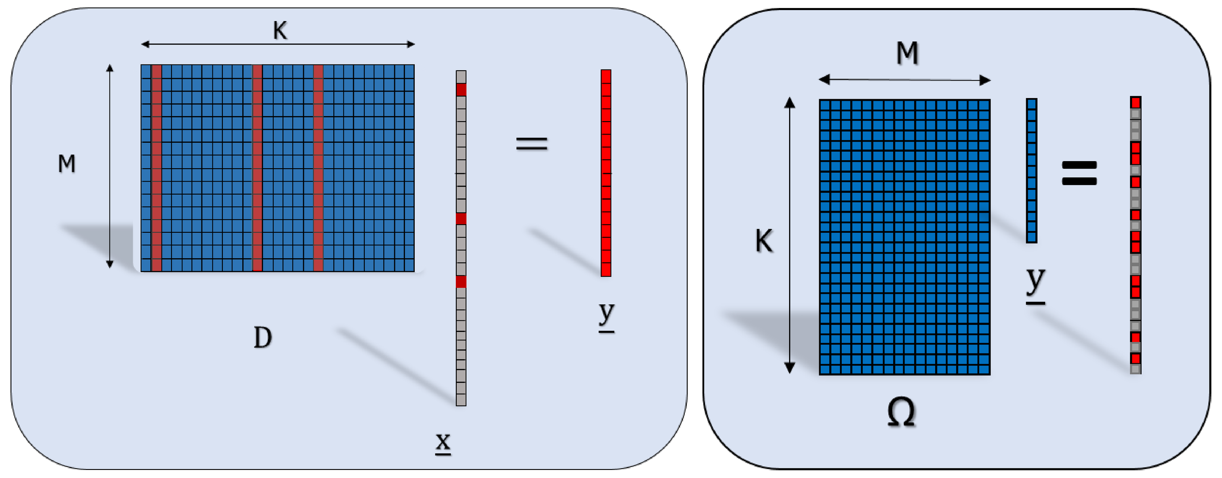

Figure: a) Synthesis model, b) Analysis model.

The main minimization problem for operator learning presented in [3] is of the same form as (2), with $\mathscr{C}$ defined as follows:

$$\mathscr{C} = \{\boldsymbol{\Omega} : \operatorname{rank}(\boldsymbol{\Omega}_{\Lambda_n}) = M - s, \|\boldsymbol{\omega}^k\|_2 = 1 \quad (3)$$

The formulation in [1] convexly relaxes the learning problem by using the ℓ_1 instead of the ℓ_0 norm:

$$\min_{\boldsymbol{\Omega}\in\mathscr{C},\mathbf{X}} \frac{\lambda}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 + \|\boldsymbol{\Omega}\mathbf{X}\|_1. \quad (4)$$

The UNTF constraint is a culmination of row norm and full rank constraints, and it is given as follows:

$$\mathscr{C} = \{\boldsymbol{\Omega} : \boldsymbol{\Omega}^T\boldsymbol{\Omega} = \boldsymbol{I}, \text{ and } \|\boldsymbol{\omega}^k\|_2 = 1, \forall k\}. \quad (5)$$

The AOL algorithm is based on a two-stage alternating minimization solution for (4) which can been given as follows:

$$\boldsymbol{\Omega}^{[i]} = \operatorname*{argmin}_{\boldsymbol{\Omega}\in\mathscr{C}} \|\boldsymbol{\Omega}\mathbf{X}^{[i-1]}\|_1 \quad (6a)$$

$$\mathbf{X}^{[i]} = \operatorname*{argmin}_{\mathbf{X}} \frac{\lambda}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 + \|\boldsymbol{\Omega}^{[i]}\mathbf{X}\|_1 \quad (6b)$$

Constrained Sparsifying TL

Using both the sparsifying transform learning paradigm [2] and the constrained analysis operator learning problem from (4), we now present a new constrained formulation for transform learning.

$$\min_{\boldsymbol{\Omega}\in\mathscr{C},\mathbf{X}} \|\boldsymbol{\Omega}\mathbf{Y}-\mathbf{X}\|_F^2 + \eta\|\mathbf{X}\|_1 \quad (7)$$

We adopt the two-step iterative approach:

$$\boldsymbol{\Omega}^{[i]} = \operatorname*{argmin}_{\boldsymbol{\Omega}\in\mathscr{C}} \|\boldsymbol{\Omega}\mathbf{Y}-\mathbf{X}^{[i-1]}\|_F^2 \quad (8a)$$

$$\mathbf{X}^{[i]} = \operatorname*{argmin}_{\mathbf{X}} \|\boldsymbol{\Omega}^{[i]}\mathbf{Y}-\mathbf{X}\|_F^2 + \eta\|\mathbf{X}\|_1 \quad (8b)$$

(8b) is solved by soft thresholding $\boldsymbol{\Omega}^{[i]}\mathbf{Y}$ as in [2]:

$$(\mathbf{X}^{[i]})_{k,n} = \begin{cases} (\boldsymbol{\Omega}^{[i]}\mathbf{Y})_{k,n} - \frac{\eta}{2}, & (\boldsymbol{\Omega}^{[i]}\mathbf{Y})_{k,n} \ge \frac{\eta}{2} \\ (\boldsymbol{\Omega}^{[i]}\mathbf{Y})_{k,n} + \frac{\eta}{2}, & (\boldsymbol{\Omega}^{[i]}\mathbf{Y})_{k,n} < -\frac{\eta}{2} \\ 0, & \text{else} \end{cases}. \quad (9)$$

This exact solution in (9) is much simpler to obtain than solving (6b). For the problem (8a), we propose the approximate solution of finding the least squares solution followed by a projection onto the UNTF set as given below:

$$\boldsymbol{\Omega}_{\text{ls}}^{[i]} = \mathbf{X}^{[i-1]}\mathbf{Y}^{\dagger} = \mathbf{X}^{[i-1]}\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}. \quad (10)$$

The final result is obtained by an approximate projection of $\boldsymbol{\Omega}_{\text{ls}}^{[i]}$ onto the UNTF:

$$\boldsymbol{\Omega}^{[i]} = \boldsymbol{\mathcal{P}}_{\text{UN}}\{\boldsymbol{\mathcal{P}}_{\text{TF}}\{\boldsymbol{\Omega}_{\text{ls}}^{[i]}\}\}. \quad (11)$$

CLS-TL Algorithm

Constrained Least Squares Sparsifying Transform Learning (CLS-TL)

Input: Data record of length N, $\mathbf{Y} = \{\boldsymbol{y}_n\}_{n=1}^N$. Regularization constant η.

Goal: $\min_{\boldsymbol{\Omega}\in\mathscr{C},\mathbf{X}} \|\boldsymbol{\Omega}\mathbf{Y}-\mathbf{X}\|_F^2 + \eta\|\mathbf{X}\|_1$

Initialize $\boldsymbol{\Omega}^{[0]}$ and calculate $\mathbf{X}^{[0]} = \lfloor\boldsymbol{\Omega}^{[0]}\mathbf{Y}\rfloor_\eta$.

Calculate $\mathbf{Y}^{\dagger} = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}$.

for $i := 1, 2, \ldots$ **do** ▷ main iteration

  $\boldsymbol{\Omega}^{[i]} = \boldsymbol{\mathcal{P}}_{\text{UN}}\{\boldsymbol{\mathcal{P}}_{\text{TF}}\{\mathbf{X}^{[i-1]}\mathbf{Y}^{\dagger}\}\}$ ▷ Transform update step, complete with LS solution and UNTF projection.

  $\mathbf{X}^{[i]} = \lfloor\boldsymbol{\Omega}^{[i]}\mathbf{Y}\rfloor_\eta$ ▷ transform sparse coding step realized by soft thresholding.

end for ▷ end of main iteration

Related Work : Transform K-SVD

In a related work we proposed an algorithm called as 'Transform K-SVD'. This algorithm brings the transform learning and the K-SVD based analysis dictionary learning approaches together.Transform K-SVD has much reduced complexity.

E.M. Eksioglu and O. Bayir, K-SVD meets Transform Learning: Transform K-SVD, *IEEE Signal Process. Letters*, vol.21, no.3, pp.347-351, March 2014.

Simulations

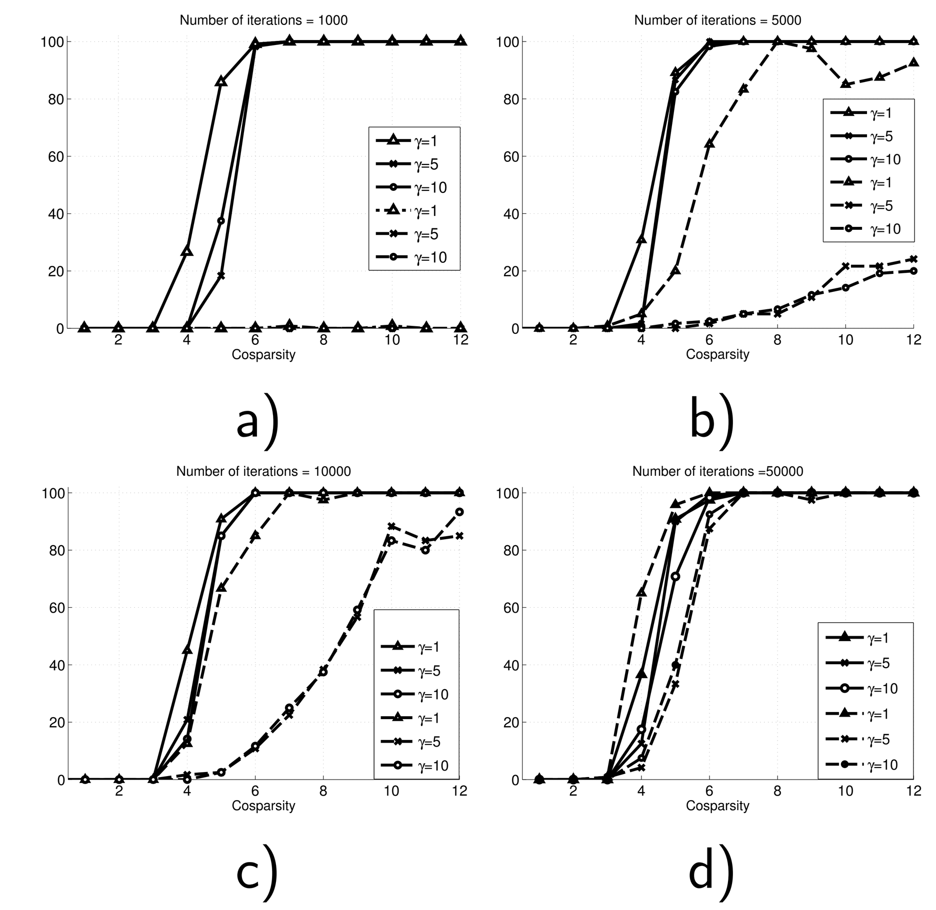

Figure: Average percentage of analysis operator recovery versus cosparsity, CLS-TL vs. AOL of [1]. a) 1000 iterations, b) 5000 iterations, c) 10000 iterations, d) 50000 iterations.

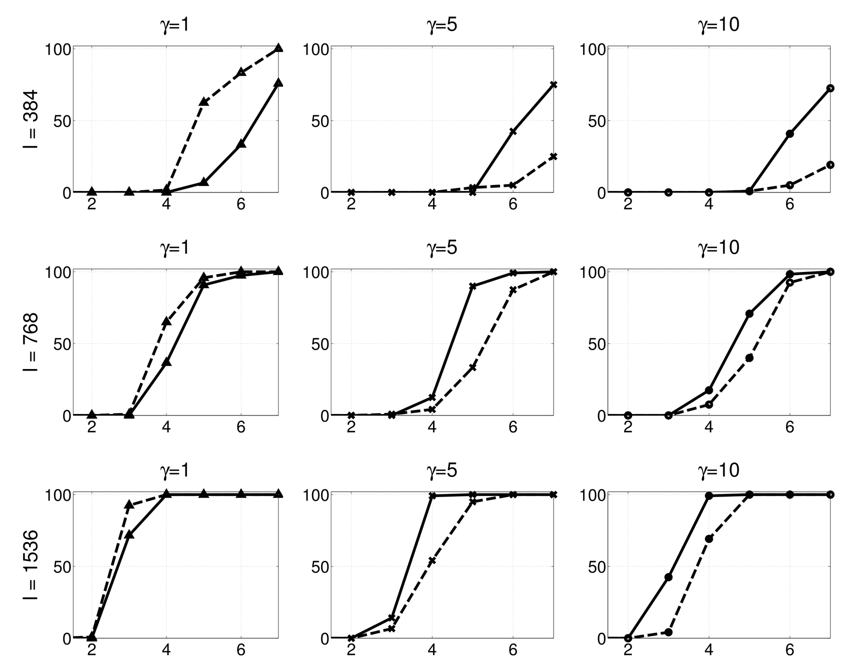

Figure: Average percentage of analysis operator recovery versus cosparsity for different training data set sizes l, CLS-TL vs. AOL of [1]. (Number of iterations: 50000).

References

[1] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Constrained overcomplete analysis operator learning for cosparse signal modelling," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2341–2355, 2013.

[2] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2013.

[3] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, 2013.

[4] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Trans. Signal Process.*, vol. 22, no. 6, pp. 2138–2150, 2013.