Imperial College London

CSP @ EEE Imperial

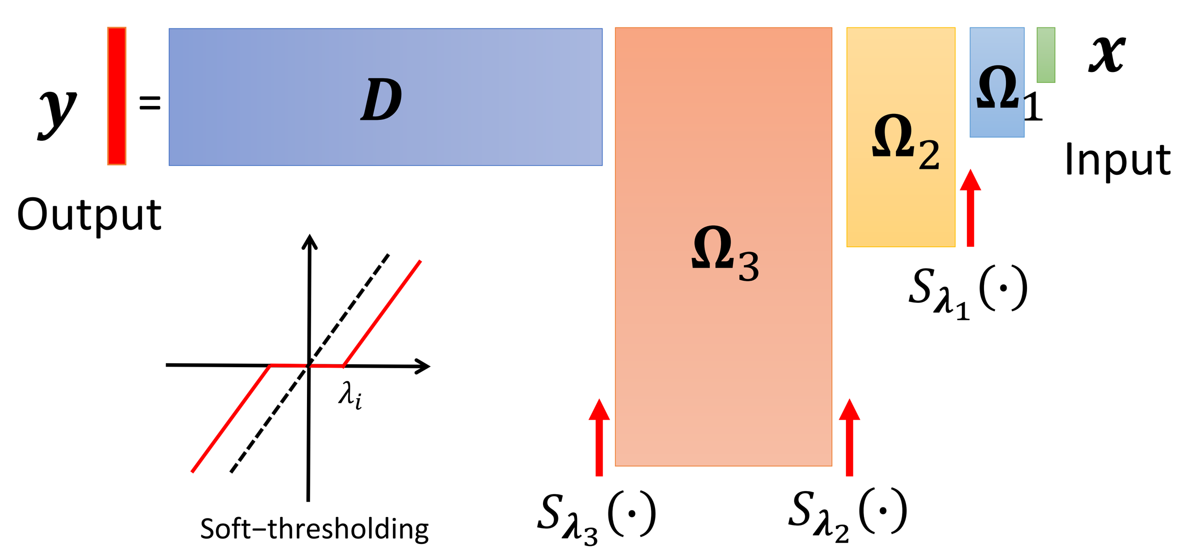
A Deep Dictionary Model to Preserve and Disentangle Key Features in a Signal

Jun-Jie Huang and Pier Luigi Dragotti
Imperial College London, UK

ICASSP 2019

Introduction

A deep dictionary model [1] consists of multiple layers of analysis dictionaries $\{\boldsymbol{\Omega}_i\}_{i=1}^{L-1}$ interlaced with corresponding soft-thresholding operations $\{S_{\lambda_i}(\cdot)\}_{i=1}^{L-1}$ and a single synthesis dictionary $\boldsymbol{D}$.



Objective:

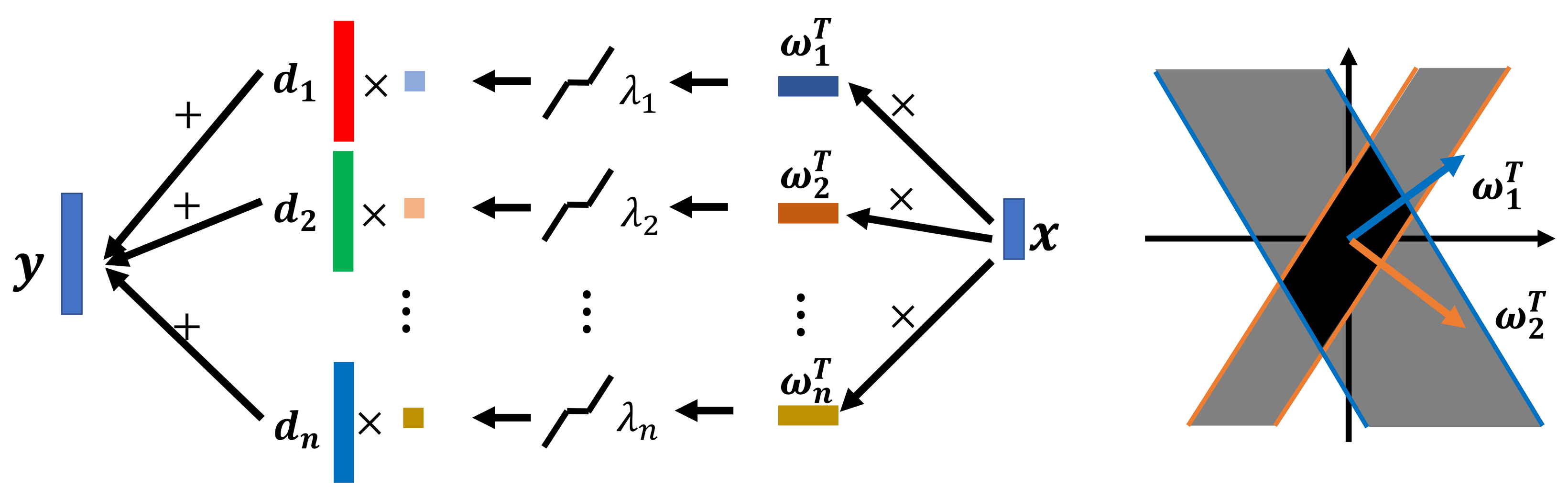
- Learning interpretable linear transforms and non-linear operations.

Overview

1. Analysis Thresholding

Analysis thresholding partitions the input space.

$$\boldsymbol{y} = \boldsymbol{D}S_{\lambda}(\boldsymbol{\Omega x})$$



The soft-thresholding thresholds can not all be too large. There should be at least k analysis-thresholding pairs for information preserving if the input data spans a k dimensional subspace of the output data.

2. Information Preserving and Clustering

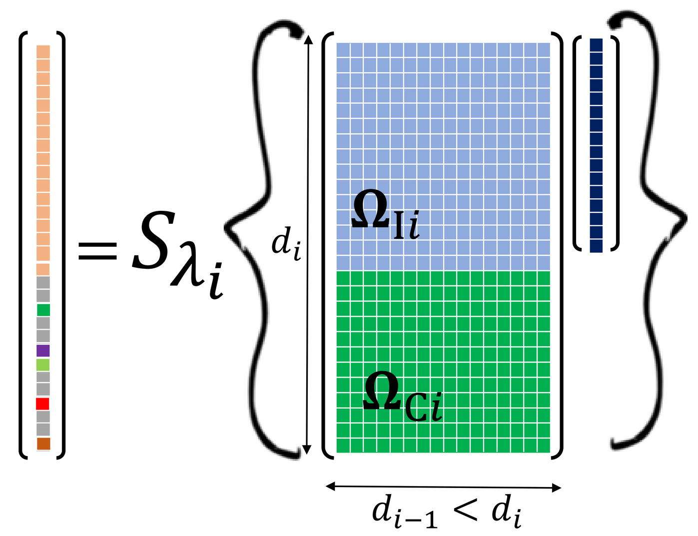
In order to not lose essential information, each analysis dictionary $\boldsymbol{\Omega}_i$ is designed to consist of two sub-dictionaries $\boldsymbol{\Omega}_i = [\boldsymbol{\Omega}_{Ii}, \boldsymbol{\Omega}_{Ci}]$.

- The information preserving analysis dictionary (IPAD) and threshold pair $(\boldsymbol{\Omega}_{Ii}, \boldsymbol{\lambda}_{Ii})$ aims at passing key information from its previous layer.
- The clustering analysis dictionary (CAD) and threshold pair $(\boldsymbol{\Omega}_{Ci}, \boldsymbol{\lambda}_{Ci})$ is to facilitate the separation of key feature in the signal.

Proposed Method

1. IPAD Learning

$\boldsymbol{\Omega}_{Ii}$ is a sparsifying analysis dictionary which is learned using an extension of the geometric analysis operator learning method [1] [2] with the input training data of this layer.



With the learned analysis dictionary, the distribution of inner product between a row atom and the input data can be well characterized by an i.i.d. zero-mean Laplacian distribution. The soft-threshold for $\boldsymbol{\Omega}_{Ii}$ is then defined as:

$$\boldsymbol{\lambda}_{Ii} = \rho[{}^{1}/_{\sigma_1}, \ldots, {}^{1}/_{\sigma_m}]^T.$$

The optimization for ρ is formulated as:

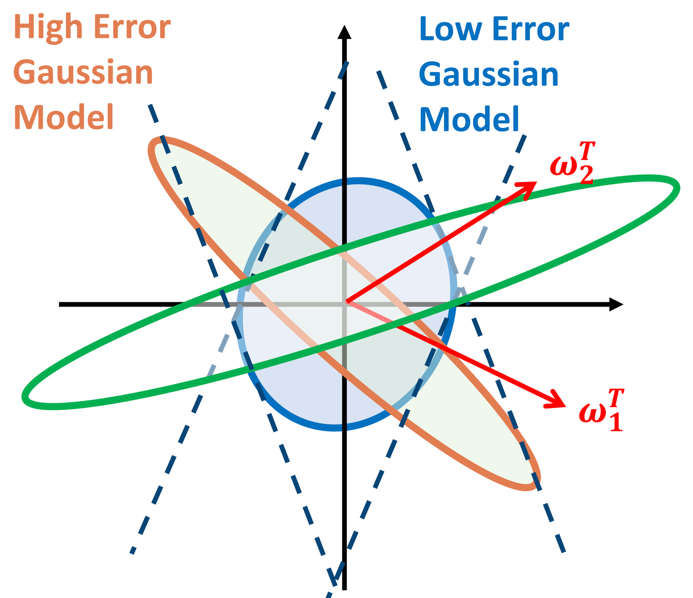
$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} ||\boldsymbol{Y} - \boldsymbol{G}S_{\rho\boldsymbol{\lambda}}(\boldsymbol{\Omega}_{Ii}\boldsymbol{Z}_{i-1})||_F^2,$$

where $\boldsymbol{\lambda} = [{}^{1}/_{\sigma_1}, \ldots, {}^{1}/_{\sigma_m}]^T$, $\boldsymbol{G} = \boldsymbol{YZ}^T(\boldsymbol{ZZ}^T)^{-1}$, $\boldsymbol{Z} = S_{\rho\boldsymbol{\lambda}}(\boldsymbol{\Omega}_{Ii}\boldsymbol{Z}_{i-1})$ and ρ belongs to a discrete set of values.

2. CAD Learning

Idea:

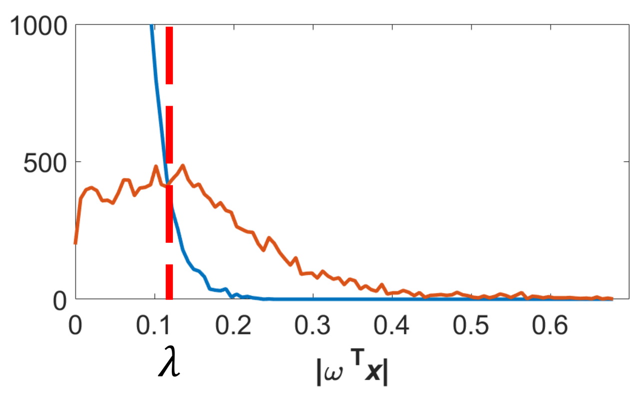
- Model the input training data as a Mixture of Gaussians
- CAD atoms and their thresholds identify data that belong to different Gaussians



Problem formulation:

$$\boldsymbol{\omega}^T = \underset{\boldsymbol{\omega}^T}{\operatorname{argmax}} \sum_{\boldsymbol{x}_i \in X(HE)} \frac{|r_{iHE}\boldsymbol{\omega}^T\boldsymbol{x}_i|}{R_{HE}} - \sum_{\boldsymbol{x}_j \in X(LE)} \frac{|r_{jLE}\boldsymbol{\omega}^T\boldsymbol{x}_i|}{R_{LE}},$$

$$s.t.\ \boldsymbol{\omega}^T\boldsymbol{\omega} = 1.$$



The histogram of $|\boldsymbol{\omega}^T\boldsymbol{x}|$ for the data from LEGMs (blue) and from a HEGM (orange).

3. Synthesis Dictionary Learning

With the learned analysis dictionaries, the synthesis dictionary $\boldsymbol{D}$ which is to map $\boldsymbol{Z}_{L-1}$ to the HR patches $\boldsymbol{Y}$ can be obtained using least squares:

$$\boldsymbol{D} = \boldsymbol{YZ}_{L-1}^T(\boldsymbol{Z}_{L-1}\boldsymbol{Z}_{L-1}^T)^{-1}.$$

Numerical Results

The standard 91 training images [3] are applied for training and Set 14 [4] is used for evaluation. For image super-resolution, the up-scaling factor is set to 2. The LR and HR patch size is 3×3 and 6×6, respectively. The input LR feature is the raw pixel values with removed mean.

The deep dictionary model is set to have $L = 4$ layers. The dictionary size for $\boldsymbol{\Omega}_1$, ..., $\boldsymbol{\Omega}_3$ and $\boldsymbol{D}$ is set to 16×9, 36×16, 144×36, and 36×144, respectively.

For comparison, DNNs with the same structure are learned using the same training data. Let us denote DNN-R and DNN-S as the DNN with ReLU non-linearity and soft-thresholding non-linearity, respectively.

Learned Dictionaries

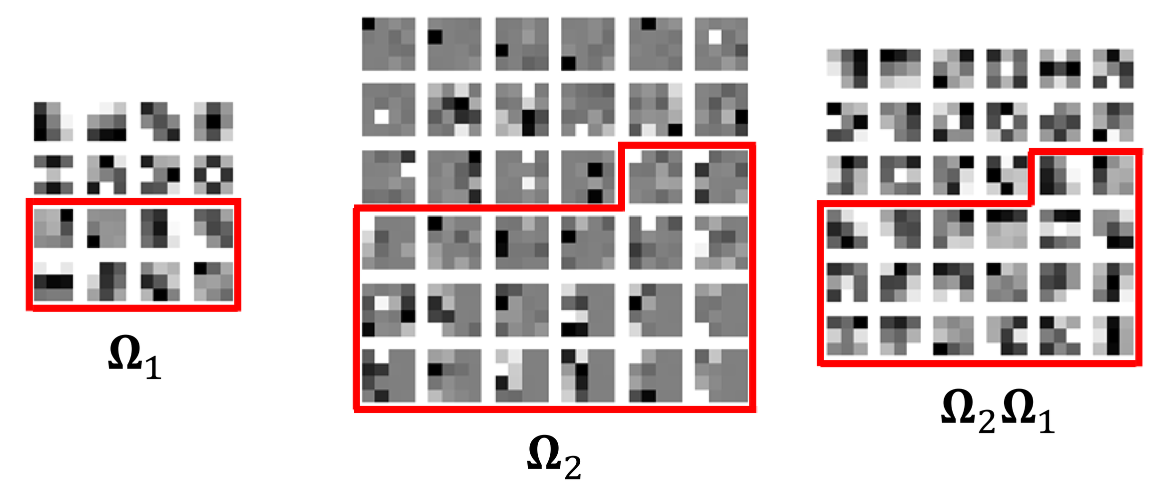


Figure 1. Learned analysis dictionaries. Each atom is displayed as a 2D patch. The atoms within red box are clustering atoms.

Evaluation Results

Image	Bicubic	DNN-R	DNN-S	DDM
baboon	24.86	25.46	25.48	25.42
barbara	27.88	28.41	28.41	28.43
bridge	26.62	27.37	27.45	27.40
costguard	29.26	30.17	30.21	30.17
comic	24.63	27.28	28.45	27.19
face	34.73	35.33	35.42	35.37
flowers	30.20	31.72	31.97	31.73
foreman	35.21	37.36	38.11	37.56
lenna	34.57	35.87	36.04	35.86
man	29.16	30.16	30.29	30.15
monarch	32.77	35.12	35.67	35.25
pepper	34.98	36.23	36.50	36.28
ppt3	24.66	28.31	28.47	28.12
zebra	28.03	32.61	32.84	32.59
Average	**29.83**	**31.53**	**31.74**	**31.54**

Table I. PSNR (dB) by different methods evaluated on Set 14.

References

[1] J.-J. Huang and P. L. Dragotti, "A deep dictionary model for image super-resolution," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'18).

[2] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," IEEE Transactions on Image Processing, vol. 22, no. 6, pp. 2138–2150, 2013.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE transactions on image processing, vol. 19, no. 11, pp. 2861–2873, 2010.

[4] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in International conference on curves and surfaces. Springer, 2010, pp. 711–730.