Sequential structured dictionary learning for block sparse representations

Abd-Krim Seghouane⋆, Asif Iqbal⋆, & Karim Abed-Meraim†

⋆ Department of Electrical and Electronic Engineering
Melbourne School of Engineering, The University of Melbourne, Australia

† PRISME Laboratory,
The University of Orléans, France

THE UNIVERSITY OF MELBOURNE

Introduction

- Dictionary learning methods have been applied to a number of signal and image processing applications.
- In some applications, the observed signal may have a multi-subspace structure that can be well exploited under the block-sparse signal representation framework.
- Using the observation that observed signals can be approximated as a sum of low rank matrices, we propose a noew algorithm for learning a block-structured dictionary for block-sparse signal representations.
- Proposed algorithm not only performs better, but is also more efficient.
- This work is an extension of [2] and [3].

Background

Consider a data matrix $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\} \in \mathbb{R}^{n\times N}$ and a sparsity constraint s, DL algorithms seek to find a dictionary $\mathbf{D} \in \mathbb{R}^{n\times K}$ and a sparse representation matrix $\mathbf{X} \in \mathbb{R}^{K\times N}$ by optimizing the following cost function

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 \text{ s.t. } \| \mathbf{x}_i \|_0 \leq s,\ \forall\, 1 \leq i \leq N, \quad \text{and } \| \mathbf{d}_m \|_2 = 1,\ \forall\, 1 \leq m \leq K. \tag{1}$$

Optimization of the above cost function is carried out via alternating optimization of $\mathbf{D}$ and $\mathbf{X}$. Given a block-structured $\mathbf{D} = [\mathbf{D}_1, ..., \mathbf{D}_J]$ with a known J blocks structure, i.e.; $\mathbf{D}_j \in \mathbb{R}^{n\times p_j}$, $j = 1, ..., J$, $p_j < n$, BK-SVD [4] solves the block formulation of (1)

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 \quad \text{s.t. } \| \mathbf{x}_i \|_{0,p} \leq s, \tag{2}$$

using an iterative alternating scheme. For every block j, they use SVD to solve

$$||\mathbf{R}_j^R - \mathbf{D}_j\tilde{\mathbf{X}}_j||_F \tag{3}$$

where $\mathbf{R}_j^R = \mathbf{R}_j\mathbf{I}_{w_j}$, $\mathbf{R}_j = \mathbf{Y} - \sum_{i=1,i\neq j}^J \mathbf{D}_i\mathbf{X}_i$, and $\tilde{\mathbf{X}}_j = \mathbf{X}_i\mathbf{I}_{w_j}$.

The Proposed Approach

The proposed algorithm is based on a variant of (2) where the ℓ_2-norm $\| \,.\, \|_2$ [1] group sparse penalty is used instead of the $\| \,.\, \|_{0,p}$ to give

$$\min_{\mathbf{D},\mathbf{X}} \sum_{i=1}^N \left(\left\| \mathbf{y}_i - \sum_{j=1}^J \mathbf{D}_j\mathbf{x}_{ij} \right\|_2^2 + \lambda \sum_{j=1}^J \sqrt{p_j} \| \mathbf{x}_{ij} \|_2 \right) \quad \text{and } \| \mathbf{d}_m \|_2 = 1\ \forall\, 1 \leq m \leq K. \tag{4}$$

Our method is based on rewriting the first term of (4) as $\|\mathbf{Y} - \mathbf{DX}\|_F^2$ and observing that the matrix $\mathbf{DX}$ can be expressed as a sum of block sparse low rank matrices or $\sum_{i=1}^J \mathbf{D}_j\mathbf{X}_j$. With an additional block incoherence constraint, the optimization problem (4) becomes

$$\sum_{j=1}^J \min_{\mathbf{D}_j,\mathbf{X}_j} \|\mathbf{E}_j - \mathbf{D}_j\mathbf{X}_j\|_F^2 + \lambda \sum_{i=1}^N \sqrt{p_j} \| \mathbf{x}_{ij} \|_2 \tag{5}$$

$$\text{s.t. } \mathbf{D}_j^\top\mathbf{D}_j = \mathbf{I}_{p_j}$$

where $\mathbf{x}_{ij}$ is the i^{th} column of $\mathbf{X}_j$, $\mathbf{E}_j = \mathbf{Y} - \sum_{i=1,i\neq j}^J \mathbf{D}_i\mathbf{X}_i$, and $\mathbf{I}_{p_j}$ is a p_j dimensional identity matrix.

Algorithm Overview

Algorithm 1: The proposed Sequential Structured Dictionary Learning Algorithm.

Input: $\mathbf{Y}$, $\mathbf{D}_{ini} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_J]$, J, ϵ_1, λ, B

Ortho-normalize blocks of $\mathbf{D}_{ini}$, set $\epsilon_2 = 0.01$
while $\| \boldsymbol{D}_m - \boldsymbol{D}_{m-1} \|_F \,/\, \| \boldsymbol{D}_{m-1} \|_F \geq \epsilon_1$ **do**
  for j = *1:J* **do**
    $\mathbf{E}_j = \mathbf{Y} - \sum_{i=1,i\neq j}^J \mathbf{D}_i\mathbf{X}_i$,
    while $\| \boldsymbol{D}_j^t - \boldsymbol{D}_j^{t-1} \|_F \geq \epsilon_2$ **do**
      Sparse Coding:
      for i = *1:N* **do**
        $\mathbf{x}_{ij}^t = \left(1 - \frac{\lambda\sqrt{p_j}}{2\|\mathbf{D}_j^{\top t-1}\mathbf{e}_{ij}\|_2}\right)_+ \mathbf{D}_j^{\top t-1}\mathbf{e}_{ij}$
      Dictionary Update:
      Compute the SVD of $\mathbf{E}_j\mathbf{X}_j^{t\top} = \mathbf{U}\Lambda\mathbf{V}^\top$
      Update $\mathbf{D}_j^t = \mathbf{UV}\top$
      Output: $\mathbf{D}_j, \mathbf{X}_j$

Result: $\mathbf{D}$ and $\mathbf{X}$

Synthetic Experiment Results

Table 1: Mean normalized reconstruction error over 100 trials, for multiple signal to noise ratios and signal block-sparsity levels s. λ is the block-sparsity controlling parameter and $\hat{s}$ is the final block-sparsity level of the representation.

SNR in dB	-10				-5				0			
s	2	3	4	5	2	3	4	5	2	3	4	5
K-SVD	2.691	2.879	2.991	3.058	1.562	1.647	1.698	1.729	0.940	0.964	0.978	0.987
BK-SVD	2.316	2.543	2.702	2.819	1.406	1.498	1.567	1.619	0.913	0.938	0.949	0.960
Proposed	**1.350**	**1.275**	**1.328**	**1.394**	**0.970**	**0.975**	**0.971**	**0.980**	**0.801**	**0.807**	**0.798**	**0.789**
$\hat{s}$	2.25	3.22	4.17	5.08	2.17	3.19	4.20	4.99	2.11	3.14	4.14	4.94
λ	0.96	1.07	1.12	1.15	0.60	0.66	0.68	0.71	0.42	0.45	0.47	0.49

Computational Efficiency on Synthetic Exp

Table 2: Run-time for a single trial in seconds.

SNR	0 dB			
s	2	3	4	5
K-SVD	9.9	10.8	12.5	13.7
BK-SVD	4.8	6.4	8.3	10.2
Proposed	**3.6**	**4.3**	**3.5**	**3.5**

Experiment on Real Images

Table 3: Mean NRE results for all algorithms over 10 trials. s the block-sparsity parameter, the best results are in **BOLD**.

	K-SVD		BK-SVD		Proposed	
s	2	3	2	3	2	3
Baboon	0.199	0.279	0.181	0.202	**0.177**	0.187
Barbara	0.196	0.254	0.138	0.216	**0.131**	0.133
Boat	0.143	0.255	0.113	0.158	0.107	**0.104**
Flinstones	0.176	0.226	0.184	0.190	0.165	**0.154**
House	0.169	0.273	0.096	0.191	0.074	**0.070**
Lena	0.141	0.223	0.113	0.105	0.075	**0.072**
Cameraman	0.399	0.460	0.440	0.448	**0.391**	0.414

References

[1] S. Bakin.
Adaptive regression and model selection in data mining problems.
PhD Thesis, The Australian National University, Canberra, 1999.

[2] A. K. Seghouane and A. Iqbal.
Consistent adaptive sequential dictionary learning.
Signal Processing, 153:300–310, 2018.

[3] A. K. Seghouane, A. Iqbal, and K. Abed Meraim.
A sequential block-structured dictionary learning algorithm for block sparse representations.
IEEE Transactions on Computational Imaging, pages 228 – 239, 2018.

[4] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar.
Dictionary optimization for block-sparse representations.
IEEE Transactions on Signal Processing, 60:2386–2395, 2012.