SEQUENTIAL STRUCTURED DICTIONARY LEARNING FOR BLOCK SPARSE REPRESENTATIONS

Abd-Krim Seghouane and Asif Iqbal

Department of Electrical and Electronic Engineering, The University of Melbourne

Karim Abed-Meraim

PRISME Laboratory, The University of Orléans, France

ABSTRACT

Dictionary learning algorithms have been successfully applied to a number of signal and image processing problems. In some applications however, the observed signals may have a multi-subpspace structure that enables block-sparse signal representations. Based on the observation that the observed signals can be approximated as a sum of low rank matrices, a new algorithm for learning a block-structured dictionary for block-sparse signal representations is proposed. It's derived via sequential penalized low rank matrix approximation, where a block coordinate descent approach is used to estimate the matrix pairs that form the different low rank matrix approximations. Experimental results on synthetic and standard gray-scale images illustrating the performance of the proposed algorithm are provided.

Index Terms— Dictionary learning, sequential learning, block sparsity, low rank matrix approximation.

1. INTRODUCTION

Sparse signal modeling has attracted a lot of research interest in recent times because of its wide range of applications. It has been successfully applied to magnetic resonance image reconstruction [1], low-dose X-ray CT reconstruction [2], and functional magnetic resonance imaging (fMRI) data analysis [3, 4, 5, 6]. Sparse signal modeling now generally refers to overcomplete dictionary learning as this leads to improved performance results by adapting a dictionary to fit the signals. Sparse coding and dictionary update constitute the two main building blocks of most overcomplete dictionary learning algorithms. The former seeks sparse solutions $\mathbf{x}_i \in R^K$, $i = 1, ..., N$ to the undetermined systems of linear equations $\mathbf{y}_i = \mathbf{D}\mathbf{x}_i$, where $\mathbf{D} \in R^{n \times K}$ is a given dictionary and $\mathbf{y}_i \in R^n$ are observation signals with $n < K < N$, while the latter aims to learn the overcomplete dictionary$\mathbf{D}$.

In some applications, the observed signals may have a multi-subspace structure that can be further exploited. This is, for example, the case in face recognition or motion segmentation [7] where the entire set of signals $\mathbf{Y} \in \mathbb{R}^{n \times N}$ can be thought as lying in multiple low-dimensional subspaces characterizing the different classes of interest. In these cases, the overcomplete dictionary learning algorithms that uses the ℓ_0 or ℓ_1 norms, for sparse coding, treat the atoms in the dictionary independently, ignoring the underlying subspace structure in the signals. For such signals, instead of looking for the sparsest representation, it may be more appropriate to look for the best representation in a union of a small number of subspaces. Several methods have been proposed to learn block sparse representations for a set of signals given a block-structured dictionary; among them one can cite the block orthogonal matching pursuit (BOMP) [8] or the group lasso [9]. On the other hand very few algorithms have been proposed for overcomplete block-structured dictionary learning. The aim of these algorithms is to learn a dictionary that provides adapted block-sparse representations for a set of signals. In [10], an extension of the K-SVD [11], (named Block-KSVD or BK-SVD for short) was proposed for learning a block-structured dictionary for block-sparsifying a given set of signals. The algorithm doesn't require prior block structure knowledge, but automatically learns it given the maximal number of atoms per group. It is an iterative algorithm that alternates between updating the block structure of the dictionary using the agglomerative clustering algorithm [12] and updating the dictionary atoms using an extension of [11] to better fit the data. The proposed extension, BK-SVD, is itself an iterative alternating minimization procedure that alternates between a block-sparse coding stage solved using BOMP and a dictionary update stage solved using a singular value decomposition (SVD). As for the K-SVD, although this iterating methods generally can guarantee that the dictionary learning objective function value is decreasing, the generated sequence of iterates may diverge [13]. Therefore, there is a need for a new block-structured dictionary algorithm that should be computationally efficient and generates better performance.

2. BACKGROUND

Given a collection of signals $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\} \in \mathbb{R}^{n \times N}$ and a sparsity constraint s, dictionary learning algorithms

This work was supported by the Australian Research Council; grant FT 130101394.

seek to find a dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$ and a sparse representation matrix $\mathbf{X} \in \mathbb{R}^{K \times N}$ such that $\mathbf{Y} \simeq \mathbf{DX}$ or alternatively $||\mathbf{Y} - \mathbf{DX}||_F \leq \epsilon$. They achieve this objective by attempting to optimize the following cost function

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

where $\mathbf{d}_m$ and $\mathbf{x}_i$ denote the m^{th} column of the dictionary $\mathbf{D}$ and i^{th} column of the sparse code matrix $\mathbf{X}$. Most algorithms for dictionary learning attempt to approximate (1) using an alternating optimization scheme which scheme starts with a sparse coding stage to optimize for $\mathbf{X}$, given a fixed $\mathbf{D}$

$$\mathbf{X} = \min_{\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 \ \text{s.t.} \ \| \mathbf{x}_i \|_0 \leq s, \ \forall\, 1 \leq i \leq N. \tag{2}$$

Approximate solutions of (2) can be found in polynomial time using either greedy strategies, such as orthogonal matching pursuit (OMP), or by replacing the ℓ_0-norm by its convex relaxation; the ℓ_1-norm. This follows with a dictionary update stage to optimize for $\mathbf{D}$ given a fixed $\mathbf{X}$

$$\mathbf{D} = \arg\min_{D} \| \mathbf{Y} - \mathbf{DX} \|_F^2 \quad \text{s.t.} \ \| \mathbf{d}_m \|_2 = 1. \tag{3}$$

Various methods have been proposed for the dictionary update stage; among them, least squares [14], maximum likelihood [15], using a coherence constraint [16], the use of a single SVD [17] or multiple SVDs' [11] on a reduced error matrix to generate the atom updates as follows

$$\{\mathbf{d}_k, \widetilde{\mathbf{x}}_k\} \ = \ \arg\min_{\mathbf{d}_k, \widetilde{\mathbf{x}}_k^{row}} ||\mathbf{E}_k^R - \mathbf{d}_k \widetilde{\mathbf{x}}_k^{row}||_F^2, \tag{4}$$

where $\mathbf{E}_k^R = \mathbf{E}_k \mathbf{I}_{w_k}$ with $\mathbf{E}_k = \mathbf{Y} - \sum_{i=1, i\neq k}^{K} \mathbf{d}_i \mathbf{x}_i^{row}$ and $\mathbf{I}_{w_k}$ the $N \times |w_k|$ submatrix of the $N \times N$ identity matrix obtained by retaining only those columns whose index numbers are in w_k defined by $w_k = \{i | 1 \leq i \leq N; \mathbf{x}_k^{row}(i) \neq 0\}$. Along with the atom $\mathbf{d}_k$ in (4), the corresponding nonzero coefficients $\widetilde{\mathbf{x}}_k^{row} = \mathbf{x}_k^{row} \mathbf{I}_{w_k}$; where $\mathbf{x}_k^{row}$ represents the k^{th} row of $\mathbf{X}$, are also updated. $\mathbf{E}_k^R$ instead of $\mathbf{E}_k$ is used to preserve the sparsity pattern in $\mathbf{X}$ obtained in the sparse coding stage.

Assuming that the atoms of the dictionary $\mathbf{D}$ can be partitioned into blocks; in [10] an extension of the K-SVD that enables block-sparse representations of a set of signals $\mathbf{Y}$ is proposed. Given a dictionary $\mathbf{D} = [\mathbf{D}_1, ..., \mathbf{D}_J]$ with a known J blocks structure[1], each of which consists of p_j atoms; i.e.; $\mathbf{D}_j \in \mathbb{R}^{n \times p_j}$, $j = 1, ..., J$, $p_j < n$, BK-SVD [10] solves the block formulation of (1)

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 \quad \text{s.t.} \ \| \mathbf{x}_i \|_{0,p} \leq s, \ \text{and} \ \| \mathbf{d}_m \|_2 = 1, \tag{5}$$

where $\| \, . \, \|_{0,p}$, in this case, is the group ℓ_0-norm that counts the number of nonzero blocks as defined by $B \in \mathbb{R}^K$; the vector of block assignments for the atoms of $\mathbf{D}$. Similar to [11], BK-SVD uses an iterative alternating scheme to solve (5) where each iteration uses a block coordinate descent approach composed of two stages. A sparse coding stage which uses BOMP to solve (5) for fixed $\mathbf{D}$ as

$$\min_{\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 \quad \text{s.t.} \quad \| \mathbf{x}_i \|_{0,p} \leq s. \tag{6}$$

Similar to the K-SVD, the blocks in $\mathbf{D}$ are updated sequentially, along with the corresponding nonzero coefficients in $\mathbf{X}$. For every block $j = 1, ..., J$, the update uses the error matrix $\mathbf{R}_j = \mathbf{Y} - \sum_{i=1, i\neq j}^{J} \mathbf{D}_i \mathbf{X}_i$ of the signals excluding the contribution of j^{th} block $\mathbf{D}_j$. The corresponding sparse code $\mathbf{X}_j \in R^{p_j \times N}$ is also updated along $\mathbf{D}_j$ by only using the column of $\mathbf{R}_j^R = \mathbf{R}_j \mathbf{I}_{w_j}$ associated with the nonzero columns of $\mathbf{X}_j$, i.e. $\tilde{\mathbf{X}}_j = \mathbf{X}_i \mathbf{I}_{w_j}$. Both $\mathbf{D}_j$ and $\mathbf{X}_j$ are updated by finding the matrix $\mathbf{D}_j \mathbf{X}_j$ of rank p_j that best approximates $\mathbf{R}_j^R$ by minimizing $||\mathbf{R}_j^R - \mathbf{D}_j \tilde{\mathbf{X}}_j||_F$ using the SVD. Thus, $\mathbf{D}_j$ is updated using the left p_j singular vectors, and $\mathbf{X}_j$ as the right p_j singular vectors times the associated singular values.

3. PROPOSED APPROACH

The design of the proposed block-structured dictionary learning algorithm for block-sparse signal representation is derived by using a variant of (5) where the ℓ_2-norm $\| \, . \, \|_2$ [9][18] group sparse penalty[2] is used instead of the $\| \, . \, \|_{0,p}$ to give

$$\min_{\mathbf{D},\mathbf{X}} \sum_{i=1}^{N} \left(\left\| \mathbf{y}_i - \sum_{j=1}^{J} \mathbf{D}_j \mathbf{x}_{ij} \right\|_2^2 + \lambda \sum_{j=1}^{J} \sqrt{p_j} \ \| \mathbf{x}_{ij} \|_2 \right) \tag{7}$$

$$\text{and} \ \| \mathbf{d}_m \|_2 = 1 \ \forall\, 1 \leq m \leq K,$$

where $\sqrt{p_j}$ is used to rescale the penalty with respect to the dimensionality of the vector $\mathbf{x}_{ij}$ [9]. The ℓ_2-norm achieves shrinkage as the ℓ_1-norm does, but works with blocks of coefficients. The cost function (7) provides a model where some blocks of coefficients are exactly zero making (7) a good candidate convex relaxation or variant of (5), suitable for the design of a block-structured dictionary learning algorithm. The cost (7) is an extension of the objective used in [19][20][21][22] but adapted for block sparse representation. From (7), we may be tempted to adopt an iterative alternating minimization scheme to solve (7), by minimizing over one variable while keeping the other one fixed in a similar way as it was done for the minimization of (5) in [10] where the BOMP was used instead of the group lasso [9].

However, we adopt a different approach which leads to a more efficient algorithm because it avoids using the SVD on a big matrix in the algorithms iterations. It is based on rewriting the first term of (7) as $\|\mathbf{Y} - \mathbf{DX}\|_F^2$ and observing that the matrix $\mathbf{DX}$ can be expressed as a sum of block sparse

[1] Block structure is defined as the atom-block correspondence.

[2] also known as the ℓ_2^1-norm in signal processing

low rank matrices or $\sum_{i=1}^{J} \mathbf{D}_j \mathbf{X}_j$, where $\mathbf{X}_j$ is a matrix with some full sparse columns associated with the block dictionary $\mathbf{D}_j$. This is a natural block-structured representation of the dataset $\mathbf{Y}$ because it separates out contributions of the various dictionary blocks to represent $\mathbf{Y}$. Furthermore, to favor the within block incoherence and unit ℓ_2-norm requirement on the atoms, we introduce the block ortho-normality constraint into the model. A similar constraint is used in [10] as well. Thus we can rewrite the optimization problem (7) as follows

$$\sum_{j=1}^{J} \min_{\mathbf{D}_j, \mathbf{X}_j} \|\mathbf{E}_j - \mathbf{D}_j \mathbf{X}_j\|_F^2 + \lambda \sum_{i=1}^{N} \sqrt{p_j} \parallel \mathbf{x}_{ij} \parallel_2 \tag{8}$$

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

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. In (8), each column of $\mathbf{X}_j$ is treated as a group and $\parallel \mathbf{x}_{ij} \parallel_2 = 0$ is equivalent to setting the i^{th} column of $\mathbf{X}_j$ as zeros. Thus the $\parallel . \parallel_2$ penalty encourages columnwise sparsity on the $\mathbf{X}_j$ matrices.

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 $\parallel \boldsymbol{D}_m - \boldsymbol{D}_{m-1} \parallel_F / \parallel \boldsymbol{D}_{m-1} \parallel_F \geq \epsilon_1$ **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 $\parallel \boldsymbol{D}_j^t - \boldsymbol{D}_j^{t-1} \parallel_F \geq \epsilon_2$ **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 $\mathbf{E}_j {\mathbf{X}^t}_j^\top = \mathbf{U}\Lambda\mathbf{V}^\top$
      Update $\mathbf{D}_j^t = \mathbf{U}\mathbf{V}\top$
      Output: $\mathbf{D}_j, \mathbf{X}_j$
Result: $\mathbf{D}$ and $\mathbf{X}$

3.1. Algorithm Derivation

The variable $\mathbf{D}_j$ and $\mathbf{X}_j$ in (8) are obtained by a block coordinate descent method. For each $j = 1, ..., J$, the proposed algorithm has two stages, a sparse coding stage, where (8) is optimized with respect to $\mathbf{X}_j$ while keeping $\mathbf{D}_j$ fixed, and a dictionary update stage, where (8) is optimized with respect to $\mathbf{D}_j$ with $\mathbf{X}_j$ fixed. The algorithm updates the factors of the various low rank matrices one by one.

The sparse coding stage is defined by

$$\min_{\mathbf{X}_j} \|\mathbf{E}_j - \mathbf{D}_j \mathbf{X}_j\|_F^2 + \lambda \sum_{i=1}^{N} \sqrt{p_j} \parallel \mathbf{x}_{ij} \parallel_2 \tag{9}$$

which is equivalent to

$$\min_{\mathbf{x}_{ij}} \sum_{i=1}^{N} \|(\mathbf{e}_{ij}) - \mathbf{D}_j \mathbf{x}_{ij}\|_2^2 + \lambda \sum_{i=1}^{N} \sqrt{p_j} \parallel \mathbf{x}_{ij} \parallel_2 . \tag{10}$$

where $\mathbf{e}_{ij}$ is the i^{th} column of $\mathbf{E}_j$. This step can be solved with an extension of the soft thresholding approach [9], which after some algebraic manipulations, gives

$$\mathbf{x}_{ij} = \left(1 - \frac{\lambda\sqrt{p_j}}{2 \parallel \mathbf{D}_j^\top \mathbf{e}_{ij} \parallel_2}\right)_+ \mathbf{D}_j^\top \mathbf{e}_{ij} \tag{11}$$

and corresponds to a vector soft-thresholding rule.
The dictionary update stage is found by solving the following minimization problem

$$\min_{\mathbf{D}_j} \|\mathbf{E}_j - \mathbf{D}_j \mathbf{X}_j\|_F^2 \text{ s.t. } \mathbf{D}_j^\top \mathbf{D}_j = \mathbf{I}_{p_j} \tag{12}$$

which is an orthogonal Procrustes problem [23]. The solution is $\mathbf{D}_j = \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are obtained from the SVD of $\mathbf{E}_j \mathbf{X}_j^\top = \mathbf{U}\Lambda\mathbf{V}^\top$, where $\mathbf{U}$ is $n \times p_j$ and $\mathbf{V}$ is $p_j \times p_j$.

4. EXPERIMENTAL RESULTS

4.1. Synthetic Experiment

In this section, we compare the signal representation performance of the aforementioned algorithms when the data, under consideration, is block-sparse. Consider a generating dictionary $\mathbf{D}_0 \in \mathbb{R}^{50\times 120}$ with normally distributed entries and a constant block-size $p = 3$ (each block with 3 atoms). The block structure (atom-block correspondence) is assumed to be known and set as $B = [1, 1, 1, 2, 2, 2, \ldots, J, J, J]$ with $J = 40$ total blocks. Each entry of $B \in \mathbb{R}^{120}$ represents the atom-block correspondence. We ortho-normalize every sub-block to yield J (p-dimensional) subspaces. Next we generate $N = 2000$ block-sparse signals $\mathbf{Y}$ via linear combination of s $\mathbf{D}_0$ blocks with coefficients from $\mathcal{U}(-0.5, 0.5)$. The resulting signal matrix $\mathbf{Y} \in \mathbb{R}^{50\times N}$ was then corrupted by additive white Gaussian noise (AWGN), leading to the noisy signal matrix $\mathbf{Y}_n$ with a specific signal to noise ratio (SNR). This noisy dataset was passed to K-SVD, BK-SVD, and the proposed dictionary learning method for decomposition. All algorithms were iterated 30 times to learn the dictionaries[3].
Starting with a random initial dictionary, the entire learning process was repeated 100 times, with multiple block-sparsity parameter $s = [2, 3, 4, 5]$ and SNR levels of $[-10, -5, 0]$ dB.

[3]The learning process would stop once the relative dictionary change fell below $\epsilon_1 = 0.01$, i.e.,$\parallel \mathbf{D}_m - \mathbf{D}_{m-1} \parallel_F / \parallel \mathbf{D}_{m-1} \parallel_F > \epsilon_1$.

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

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**

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**
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

For every trial, we calculated the normalized reconstruction error as $\|\mathbf{Y} - \mathbf{D}\,\mathbf{X}\|_F / \|\mathbf{Y}\|_F$. The mean results are outlined in Table 1. For a fair comparison, the block-sparsity controlling parameter λ was chosen such that the final block-sparsity $\hat{s}$ in $\mathbf{X}$ matches the signal block-sparsity level s, where the final block-sparsity is calculated using $\| \mathbf{X} \|_0 / (N \times p)$. The selected λ and the respective final block-sparsity level $\hat{s}$ are also provided in Table 1. The results show that the proposed algorithm was able to outperform both K-SVD and BK-SVD algorithms over every SNR and block-sparsity level. Furthermore, the run times over a single trial for all algorithms are also shown in Table 2.

4.2. Experiment on Real Images

Here, we evaluate the algorithms performance using gray-scale images. Starting with a standard gray-scale image, we extracted all non-overlapping 8×8 patches, vectorized them, and placed them as column vectors to generate a signal matrix $\mathbf{Y} \in \mathbb{R}^{64 \times N}$. All except *'boat'* image were of size 512×512 pixels. The pixel values were normalized to stay between $[0, 1]$. Starting with a randomly initialized dictionary $\mathbf{D} \in \mathbb{R}^{64 \times 96}$ with ℓ_2-normalized columns, we iterated K-SVD, BK-SVD and the proposed algorithms 50 times to learn the dictionaries. For BK-SVD and the proposed algorithms, we assume that the dictionary block structure is known, i.e., $B = [1, 1, 1, 2, 2, 2, \ldots, J, J, J]$ with $J = 32$ total blocks and $p = 3$ atoms per block. Furthermore, block-sparsity parameter for BK-SVD was set to s and the sparsity parameter for K-SVD was set to $s \times p$. Whereas, for the proposed algorithm, we tried multiple values for λ and selected the ones giving best performance in terms of normalized reconstruction error as defined next. Once the learning process was complete, we generated a sensing matrix $\mathbf{A} \in \mathbb{R}^{16 \times 64}$ as outlined in [24], and used it to compress the image patch data and the learned dictionary atoms from 64 to 16 dimensions. Using the compressed data and dictionaries, we found s block-sparse solution for BK-SVD and proposed algorithm, whereas, for K-SVD, we found a $s \times p$ sparse solution. Using the resulting sparse approximations, and the uncompressed image data and dictionaries, we calculated the normalized reconstruction error (NRE) for $s = [2, 3]$. This entire process was repeated 10 times and the mean NRE results for all algorithms are presented in Table 3. Here it is evident that the proposed algorithm was able to achieve the lowest NRE for all the images.

5. CONCLUSION

A new algorithm for learning block-structured dictionary for block-sparse signal representations is proposed. It is obtained by exploiting the observation that the observed data matrix can be approximated by a sum of low rank matrix approximations. The proposed algorithm is obtained via sequential penalized low rank matrix approximation with a block sparsity promoting penalty. Within this algorithm a block coordinate descent approach is adopted for updating both the block sparse codes and the associated block dictionaries with simple closed form solutions in both stages. The experimental section highlights the superior performance of the proposed algorithm compared to other state of the art algorithms.

6. REFERENCES

[1] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1028–1041, 2011.

[2] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 1682–1696, 2012.

[3] A. K. Seghouane and A. Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Transactions on Medical Imaging*, pp. 1796–1807, 2017.

[4] A. K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Transactions on Image Processing*, pp. 3002–3015, 2017.

[5] A. Iqbal, A. K. Seghouane, and T. Adali, "Shared and subject-specific dictionary learning algorithm for multi-subject fMRI data analysis (ShSSDL)," *IEEE Transactions on Biomedical Engineering*, vol. PP, no. 99, pp. 1–10, 2018.

[6] A. Iqbal and A. K. Seghouane, "A dictionary learning algorithm for multi-subject fMRI analysis based on a hybrid concatenation scheme," *Digital Signal Processing*, pp. 249–260, 2018.

[7] V. Bolon-Canedo, N. Sanchez-Marono, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and Information Systems*, vol. 34, pp. 483–519, 2013.

[8] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3042–3054, 2010.

[9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, pp. 49–67, 2006.

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

[11] M. Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311–4322, 2006.

[12] S. C. Johnson, "Hierarchical clsutering schemes," *Psychometrika*, vol. 32, pp. 241–254, 1967.

[13] C. Bao, H. Ji, Y. Quan, and Z. Shen, "Dictionary learning for sparse coding: Algorithms and convergence analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1356–1369, 2016.

[14] K. Engan, S. O. Aase, and J. Hakon-Husoy, "Method of optimal directions for frame design," *IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, pp. 2443–2446, 1999.

[15] M. Hanif and A. K. Seghouane, "Maximum likelihood orthogonal dictionary learning," *IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 141–144, 2014.

[16] S. Ubaru, A. K. Seghouane, and Y. Saad, "Improving the incoherence of learned dictionary via rank shrinkage," *Neural Computation*, vol. 29, pp. 263–285, 2017.

[17] M. U. Khaled and A. K. Seghouane, "A single SVD sparse dictionary learning algorithm for fMRI data analysis," *In Proceedings of the IEEE Workshop on Statistical Signal Processing, SSP*, pp. 65–68, 2014.

[18] 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*, pp. 1–12, 2018.

[19] A. K. Seghouane and M. Hanif, "A sequential dictionary learning algorithm with enforced sparsity," *IEEE International Conference on Acoustic Speech and signal Processing, ICASSP*, pp. 3876–3880, 2015.

[20] A. K. Seghouane and A. Iqbal, "A regularized sequential dictionary learning algorithm for fMRI data analysis," *In Proceedings of the IEEE Workshop on Machine Learning for Signal Processing, MLSP*, pp. 1–6, 2017.

[21] A. Iqbal and A. K. Seghouane, "An approach for sequential dictionary learning in nonuniform noise," *In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, DICTA*, pp. 1–6, 2017.

[22] A. K. Seghouane and A. Iqbal, "Consistent adaptive sequential dictionary learning," *Signal Processing*, vol. 153, pp. 300–310, 2018.

[23] G. H. Golub and C. f. Van Loan, *Matrix Computations*, Johns Hopkins, 1996.

[24] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, 2009.