DICTIONARY LEARNING FOR BLOCK-SPARSE SIGNALS

Ender M. Eksioglu

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

The Question

For a given data set, how can we learn a block-sparsifying dictionary (and block structure)?

Introduction

Sparse representation algorithms approximate the signal vector as a linear combination of a small selection from the atoms constituting the dictionary. A related concern is the design of a dictionary which results in efficient sparse representations for a given set of signal vectors. Dictionary learning algorithms (DLAs) try to deduce from given data a dictionary which is optimally suitable for sparse representation of the data. Recently there are attempts at dictionary learning with a block sparse signal structure assumption. In block-sparse dictionary design, a dictionary and corresponding block structure should be learned. In [1] such a framework where both dictionary and block structure are learned side by side is provided.

The Solution

In a recent work, we developed a method for block structure identification that permitted use of diverse proximity measures between blocks and also the use of clustering techniques from literature [2]. In this paper we further develop a least-squares based framework for block-sparse dictionary learning, where the block structure identification method is used as a sub-step. The online variant of this novel block-sparse dictionary learning framework is also be provided.

Block-Sparse Signal Representation

The distribution of dictionary atoms to blocks can be given by an assignment vector $\mathbf{\Gamma} \in \mathbb{R}^K$. The set of the indices of atoms included in the j^{th} block of $\mathbf{\Gamma}$ will be denoted by $\mathbf{\Omega}_j^{\mathbf{\Gamma}}$, where the size of block j will be denoted as the cardinality $|\mathbf{\Omega}_j^{\mathbf{\Gamma}}|$.

A vector $\boldsymbol{w}$ is said to be block k-sparse over $\mathbf{\Gamma}$, if its non-zero components occur in only k of the total B blocks, hence if $\|\boldsymbol{w}\|_{\mathbf{\Gamma}} = \|\boldsymbol{w}\|_{2,0}^{\mathbf{\Gamma}} = k$. A noise-free formulation for the block-sparse signal representation problem of signal vector $\boldsymbol{x}$ over a dictionary $\mathbf{D}$ and block structure $\mathbf{\Gamma}$ can be given as follows.

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \|\boldsymbol{w}\|_{\mathbf{\Gamma}} \text{ s.t. } \boldsymbol{x} = \mathbf{D}\boldsymbol{w} \qquad (1)$$

Dictionary Learning Problem Formulation

We consider the problem of learning a block-sparsifying dictionary and the corresponding block structure pair for a given data set. We formulate the block-sparsifying dictionary learning problem using the following optimization over a block-sparsity regularized cost function for the given data set.

$$\{\hat{\mathbf{D}}, \hat{\mathbf{\Gamma}}, \hat{\mathbf{W}}\} = \arg\min_{\mathbf{D},\mathbf{\Gamma},\mathbf{W}} \left\{ \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 + \gamma \sum_{n=1}^{N} \|\boldsymbol{w}_n\|_{\mathbf{\Gamma}} \right\} \quad \text{s.t.} \quad |\mathbf{\Omega}_j^{\mathbf{\Gamma}}| \leq s, \forall j \in \mathbf{\Gamma} \qquad (2)$$

Here, $\mathbf{X} = \{\boldsymbol{x}_n\}_{n=1}^N \in \mathbb{R}^{M \times N}$ and $\mathbf{W} = \{\boldsymbol{w}_n\}_{n=1}^N \in \mathbb{R}^{K \times N}$ are the time concatenated data matrix and the corresponding time concatenated representation matrix, respectively. A block-sparsifying dictionary design optimization problem similar to (2) has been introduced in [1].

Learning the Block-Sparsifying Dictionary and Block Structure Pair

We propose an algorithm for the approximate solution of the non-convex optimization problem given in (2). We apply the block coordinate descent approach where the objective function is minimized with respect to a single variable with the other variables being held constant. In (2) we have the block-structure vector $\mathbf{\Gamma}$ as a third variable. Therefore we propose the following three-step algorithm which realizes an optimization procedure, where each step performs a minimization over a single variable with the other two held constant. Different from the two-step algorithm of [1], here are three steps each simplified to optimizations over a single variable.

Block-Sparsifying Dictionary Learning with Block Structure Identification

Input: Data record of length N, $\mathbf{X} = \{\boldsymbol{x}_n\}_{n=1}^N$. Some a priori information on the block structure. We assume maximal block size is given as s.

Goal: $\{\hat{\mathbf{D}}, \hat{\mathbf{\Gamma}}, \hat{\mathbf{W}}\} = \arg\min_{\mathbf{D},\mathbf{\Gamma},\mathbf{W}} \Big\{ \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 + \gamma \sum_{n=1}^N \|\boldsymbol{w}_n\|_{\mathbf{\Gamma}} \Big\}$ s.t. $|\mathbf{\Omega}_j^{\mathbf{\Gamma}}| \leq s, \forall j \in \mathbf{\Gamma}$.

Initialize the dictionary, possibly as $\mathbf{D}^{(0)} = \mathbf{X}_K = \{\boldsymbol{x}_n\}_{n=1}^K$.
for $i := 1, 2, \ldots$ **do** ▷ iteration
  for $n := 1, 2, \ldots, N$ **do** ▷ batch sparse representation step
    $\tilde{\boldsymbol{w}}_n = \underset{\boldsymbol{w}}{\operatorname{argmin}} \|\boldsymbol{x}_n - \mathbf{D}^{(i-1)}\boldsymbol{w}\|_2^2 + \gamma\|\boldsymbol{w}\|_0$
  end for
  $\mathbf{\Gamma}^{(i)} = \underset{\mathbf{\Gamma}}{\operatorname{argmin}} \Big\{ \sum_{n=1}^N \|\tilde{\boldsymbol{w}}_n\|_{\mathbf{\Gamma}} \Big\}$ s.t. $|\mathbf{\Omega}_j^{\mathbf{\Gamma}}| \leq s, j \in \mathbf{\Gamma}$. ▷ block structure identification
  for $n := 1, 2, \ldots, N$ **do** ▷ batch block sparse representation step
    $\boldsymbol{w}_n^{(i)} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \Big(\|\boldsymbol{x}_n - \mathbf{D}^{(i-1)}\boldsymbol{w}\|_2^2 + \gamma\|\boldsymbol{w}\|_{\mathbf{\Gamma}^{(i)}} \Big)$
  end for
  $\mathbf{D}^{(i)} = \mathbf{X}\mathbf{W}^{(i)^{\dagger}}$ ▷ dictionary update
end for ▷ end of iteration

The block-sparsifying dictionary learning algorithm as introduced in Alg. 1 works in the batch mode, using the entire data set in each iteration. Online dictionary learning algorithms have also been proposed [3]. These online algorithms draw one single data vector (or a mini-batch) at each time iteration from the data set and update the dictionary using the innovation gathered from this single vector (or mini-batch). In this section we propose an online RLS version for Alg. 1. We replace the batch sparse representation, batch block-sparse representation and the batch dictionary update steps of Alg. 1 with their online counterparts. Alg. 2 outlines the complete new online block-sparse dictionary learning algorithm.

Online RLS based Block-Sparsifying Dictionary Learning Algorithm (Block-RLS-DLA)

Initialize the dictionary, possibly as $\mathbf{D}_N^{(0)} = \mathbf{X}_K = \{\boldsymbol{x}_n\}_{n=1}^K$.
for $i := 1, 2, \ldots$ **do** ▷ iteration
  $\mathbf{D}_0^{(i)} = \mathbf{D}_N^{(i-1)}$, $\mathbf{C}_0 = \mathbf{I}_K$ ▷ initialization
  for $t := 1, 2, \ldots, N$ **do** ▷ time iteration
    $\tilde{\boldsymbol{w}}_t = \underset{\boldsymbol{w}}{\operatorname{argmin}} \|\boldsymbol{x}_t - \mathbf{D}_{t-1}^{(i)}\boldsymbol{w}\|_2^2 + \gamma\|\boldsymbol{w}\|_0$ ▷ sparse representation
    $\mathbf{\Gamma}_t = \underset{\mathbf{\Gamma}}{\operatorname{argmin}} \Big\{ \sum_{n=1}^t \|\tilde{\boldsymbol{w}}_n\|_{\mathbf{\Gamma}} \Big\}$ s.t. $|\mathbf{\Omega}_j^{\mathbf{\Gamma}}| \leq s, j \in \mathbf{\Gamma}$ ▷ block structure ident.
    $\boldsymbol{w}_t = \underset{\boldsymbol{w}}{\operatorname{argmin}} \Big(\|\boldsymbol{x}_t - \mathbf{D}_{t-1}^{(i)}\boldsymbol{w}\|_2^2 + \gamma\|\boldsymbol{w}\|_{\mathbf{\Gamma}_t} \Big)$ ▷ block-sparse representation
    $\boldsymbol{r} = \boldsymbol{x}_t - \mathbf{D}_{t-1}^{(i)}\boldsymbol{w}_t$, $\mathbf{C}^*_{t-1} = \lambda^{-1}\mathbf{C}_{t-1}$
    $\boldsymbol{u} = \mathbf{C}^*_{t-1}\boldsymbol{w}_t$, $\alpha = \dfrac{1}{1 + \boldsymbol{w}_t^T\boldsymbol{u}}$
    $\mathbf{C}_t = \mathbf{C}^*_{t-1} - \alpha\boldsymbol{u}\boldsymbol{u}^T$
    $\mathbf{D}_t^{(i)} = \mathbf{D}_{t-1}^{(i)} + \alpha\boldsymbol{r}\boldsymbol{u}^T$ ▷ RLS update
  end for ▷ end of time iteration
end for ▷ end of iteration

Simulations

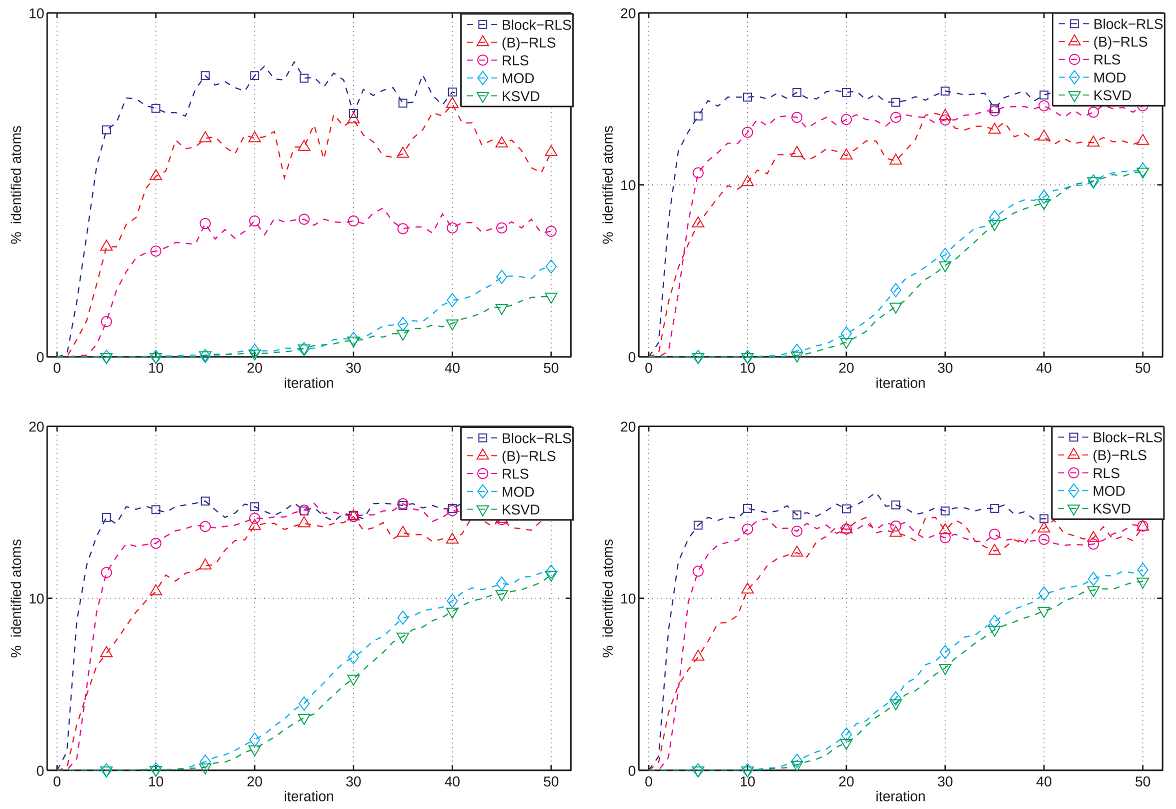

Figure 1: Dictionary learning performance of the various algorithms under different SNR scenarios.

References

[1] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Dictionary optimization for block-sparse representations," *IEEE Trans. Signal Process.*, vol. 60, no. 5, May 2012.

[2] E. M. Eksioglu, "A clustering based framework for dictionary block structure identification," in *ICASSP 2011*, pp. 4044–4047.

[3] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 4, Apr. 2010.