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. Recently there are attempts at dictionary learning with a block sparse signal structure assumption [1], [2]. 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 [3]. 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 will also be provided, and its effectiveness in block-sparse setting will be demonstrated via simulations.

II. BLOCK-SPARSE DICTIONARY LEARNING

II-A. Block-Sparse Signal Representation

Block-sparse representation considers recovering a sparse decomposition of observed signals over a given over-determined dictionary and corresponding block structure pair. We assume the observed signal at each time point $n$ is a real vector of dimension $M$, where there are a total of $N$ observations, $x_n \in \mathbb{R}^M$ for $n = 1, \ldots, N$. When the atoms $D_k \in \mathbb{R}^M$, $k = 1, \ldots, K$ are ordered as columns, the dictionary is given as a matrix $D \in \mathbb{R}^{M \times K}$ [1]. $D(k) \in \{1, 2, \ldots, B\}$ is the block the atom $d_k$ belongs to. The indices of the atoms included in the $j^{th}$ block is denoted by the set $\Omega_j^F$. The size of block $j$ is the cardinality $|\Omega_j^F|$ with $\sum_{j=1}^B |\Omega_j^F| = K$. The block-sparsity of vector $w$ over block structure $\Gamma$ will be given by $||w||_\Gamma = ||w||_{\Omega_j^F}^F$, where $||\cdot||_F$ the pseudo mixed $(\ell, 0)$ norm. A noise-free formulation for the block-sparse signal representation problem of signal vector $x$ over a dictionary $D$ with block structure $\Gamma$ can be given as follows.

$$\hat{w} = \text{argmin}_{w} \|w\|_{\Gamma} \text{ s.t. } x = Dw \tag{1}$$

II-B. Dictionary Learning Problem Formulation

In this paper 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.

$$\{D, \Gamma, \hat{W}\} = \text{argmin}_{D, \Gamma, W} \left\{ \|X - DW\|_F^2 + \gamma \sum_{n=1}^N \|w_n\|_{\Gamma} \right\} \tag{2}$$

s.t. $|\Omega_j^F| \leq s, \forall j \in \Gamma$

Here, $X = \{x_n\}_{n=1}^N \in \mathbb{R}^{M \times N}$ and $W = \{w_n\}_{n=1}^N \in \mathbb{R}^{K \times N}$ 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].