COMPRESSIVE K-SVD

Farhad Pourkamali Anaraki, Shannon M. Hughes
Department of Electrical, Computer, and Energy Engineering
University of Colorado at Boulder

ABSTRACT

Dictionary learning algorithms design a dictionary that is specifically tailored to enable sparse representation of a given set of training signals. In turn, the increased sparsity of the signals with respect to this dictionary enables significantly improved performance in a variety of state-of-the-art signal processing tasks, e.g. compressive sensing. However, while these algorithms typically assume that all training data is fully available, this may not be the case in practice. In fact, the high cost of acquiring each signal or the sheer amount of data to be acquired may motivate us to take a compressive sensing (CS) approach, taking only a few CS measurements of each signal. In this paper, we present a novel algorithm for learning a dictionary on a set of training signals using only compressive sensing measurements of them. Our proposed algorithm is a generalization of the well-known K-SVD algorithm and preserves its convergence properties. Experimental results on synthetically generated data verify that our proposed algorithm can recover the generating dictionary atoms from CS measurements alone (so long as enough measurements of enough training signals are available), even for the case of noisy measurements. Finally, we show that compressive K-SVD (CK-SVD) can also be used to aid in signal reconstruction and compressive classification on the CS measurements.

***Index Terms*—** Dictionary learning, K-SVD, Sparse representation, Compressive sensing, Compressive classification

1. INTRODUCTION

In recent years, there has been increasing interest in developing algorithms for learning a dictionary based on a given set of training signals. The goal is to learn a dictionary that leads to a sparse representation for each signal in the set. While choosing a predetermined dictionary (i.e. wavelets/DCT) is simpler, this signal-adaptive dictionary learning typically leads to a much more compact representation. In fact, the increased sparsity enabled by dictionary learning has been revolutionizing performance in many state-of-the-art signal processing tasks such as compression, feature extraction, classification, image denoising, and compressive sensing. As an illustrative example, wavelet-based compression and denoising strategies have been very successful in the past due to how sparsely images can be represented in the wavelet domain. However, recent results show that denoising using a dictionary learned from patches of a noisy image leads to huge gains in performance over even wavelets [1].

The method of optimal directions (MOD) [2] and K-SVD [3] are two well-known algorithms for learning a dictionary from a set of training signals. Both of these are simple but efficient iterative algorithms that alternate between sparse coding and dictionary update steps to find the dictionary allowing the best possible sparse representation for each signal in the set. Moreover, the K-SVD algorithm has been shown to converge quickly [3] and to outperform other dictionary learning algorithms in practice leading to improved performance in many image processing applications.

One important application of dictionary learning is compressive sensing. Compressive sensing (CS) aims to allow reconstruction of signals that are sparse in some basis from a small number of measurements. Various algorithms for signal recovery from CS measurements have been introduced [4, 5, 6], but typically, the performance of these algorithms, e.g. ℓ^1-minimization, crucially relies on the sparsity of the signals of interest. Hence, a more compact representation for the signals leads to huge gains in performance, improving the reconstruction accuracy or reducing the required number of CS measurements.

Because of this, there have been some initial attempts at dictionary learning directly from CS measurements [7, 8, 9]. For example, the previous work in [7] and [8] tries to estimate principal components of the training data from CS measurements alone. Indeed, in [8] it has been observed that performing normal principal component analysis (PCA) on low-dimensional random projections of data is a simple and effective approach that produces the same result as performing PCA on the original data.

However, sparse representation algorithms like K-SVD typically give a much more compact representation than that obtained through PCA, which is typically not very sparse. This strongly motivates us to develop an algorithm for designing a dictionary that can sparsely represent the data using the CS measurements.

In this paper, our goal is to present such an algorithm for learning a dictionary based on a given set of CS measurements. Our proposed algorithm, compressive K-SVD, is a generalization of the well-known K-SVD algorithm. More precisely, our algorithm is an iterative approach that alternates between sparse coding and dictionary update steps to minimize the error in representation of the CS measurements.

Furthermore, experimental results verify that our proposed algorithm can recover the generating dictionary atoms of both synthetic and real-world datasets from CS measurements alone. Moreover, the dictionary obtained from our proposed algorithm can then be used to greatly improve the reconstruction accuracy of each data point from its CS measurements. It can also be used for classification of signals based on CS measurements of them [10].

In Section 2, we review the original K-SVD algorithm briefly. Section 3 presents our proposed algorithm, compressive K-SVD, in detail, with discussion on convergence of the algorithm. In Section 4, we show experimental results that verify the performance of our proposed algorithm for learning dictionaries on both synthetic and real-world datasets. We also explore two important applications of our proposed algorithm: (1) signal recovery and (2) signal classification based on the CS measurements.

2. REVIEW OF THE K-SVD ALGORITHM

Given a set of n training signals $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ in $\mathbb{R}^p$, a dictionary $\mathbf{D} \in \mathbb{R}^{p \times d}$ that leads to the best representation under a strict

This material is based upon work supported by the National Science Foundation under Grant CCF-1117775.

sparsity constraint for each member in the set is obtained by minimizing the following representation error

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 \quad \text{subject to} \quad \forall i, \quad \left\|\mathbf{x}^{(i)}\right\|_0 \le T \tag{1}$$

where $\mathbf{X} = \left[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right]$ is the coefficient matrix corresponding to the training signals $\mathbf{Y}$ and $\left\|\mathbf{x}^{(i)}\right\|_0$ counts the number of nonzero entries of the coefficient vector $\mathbf{x}^{(i)} \in \mathbb{R}^d$. Solving this optimization problem may lead to an exact representation for the ith signal $\mathbf{y}_i$ such that $\mathbf{y}_i = \mathbf{D}\mathbf{x}^{(i)}$ or each $\mathbf{y}_i$ may be well-approximated by the learned dictionary, satisfying $\left\|\mathbf{y}_i - \mathbf{D}\mathbf{x}^{(i)}\right\|_2 \le \epsilon$ for some small ϵ.

In [3], an iterative algorithm, K-SVD, was presented that alternates between sparse coding of the signals with respect to a fixed dictionary and the process of updating dictionary atoms. The main advantages of this algorithm are its simplicity, efficiency, and fast convergence with respect to other previously proposed dictionary learning methods. Indeed, the sparse coding step is common between different dictionary learning methods, so the process of updating dictionary atoms is the main topic of interest. In the sparse coding step, the optimization problem in (1) is solved with respect to a fixed $\mathbf{D}$ to find the best coefficient matrix $\mathbf{X}$. The penalty term in (1) can be written as

$$\|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 = \sum_{i=1}^{n} \left\|\mathbf{y}_i - \mathbf{D}\mathbf{x}^{(i)}\right\|_2^2 \tag{2}$$

Minimizing (2) with respect to $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ can be solved as n distinct optimization problems for each signal $\mathbf{y}_i$ in the set as follows

$$\min_{\mathbf{x}^{(i)}} \left\|\mathbf{y}_i - \mathbf{D}\mathbf{x}^{(i)}\right\|_2^2 \quad \text{subject to} \quad \left\|\mathbf{x}^{(i)}\right\|_0 \le T \tag{3}$$

Finding the sparsest solution $\mathbf{x}^{(i)}$ is an NP-hard problem, but the approximate solution can be obtained by pursuit algorithms such as the simple greedy algorithm orthogonal matching pursuit (OMP) [11] used in K-SVD.

The main contribution of the K-SVD algorithm is the dictionary update process which minimizes the penalty term in (1) using a simple iterative approach. This penalty term can be written as

$$\begin{aligned} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 &= \left\| \left(\mathbf{Y} - \sum_{j\neq k} \mathbf{d}^{(j)}\mathbf{x}_T^j \right) - \mathbf{d}^{(k)}\mathbf{x}_T^k \right\|_F^2 \\ &= \left\|\mathbf{R}_k - \mathbf{d}^{(k)}\mathbf{x}_T^k\right\|_F^2 \end{aligned} \tag{4}$$

where $\mathbf{d}^{(j)}$ is the jth dictionary atom, $\mathbf{x}_T^j$ is the corresponding coefficients for it for each signal (the jth row of $\mathbf{X}$), and $\mathbf{R}_k = \mathbf{Y} - \sum_{j\neq k} \mathbf{d}^{(j)}\mathbf{x}_T^j$ is the representation error for the training signals when the kth dictionary atom is removed. In the dictionary update process, it is assumed at each step k that $\mathbf{d}^{(j)}$ and $\mathbf{x}_T^j$, $j \neq k$, are fixed. We then minimize the criterion over $\mathbf{d}^{(k)}$ and $\mathbf{x}_T^{(k)}$ which is equivalent to finding the best rank-1 approximation of $\mathbf{R}_k$. The minimizer might typically be obtained by applying SVD to the matrix $\mathbf{R}_k$, but the strict sparsity constraint also must be considered. Therefore, in the K-SVD algorithm, we shrink the matrix $\mathbf{R}_k$ by eliminating columns corresponding to those training signals for which $\mathbf{x}_k^{(i)} = 0$ and then find the best rank-1 approximation of the shrunken $\mathbf{R}_k$ to update $\mathbf{d}^{(k)}$ and $\mathbf{x}_T^{(k)}$. This strategy preserves the support of the coefficient matrix $\mathbf{X}$. Indeed, joint optimization of the dictionary atoms and the corresponding coefficients in the dictionary update step leads to a much more efficient minimization of (4) compared to other dictionary learning methods.

3. COMPRESSIVE K-SVD

In this section, we present our proposed compressive K-SVD algorithm (CK-SVD). Given a set of CS measurements for n training signals $\{\mathbf{m}_i\}_{i=1}^n$ in $\mathbb{R}^m$, the goal is to find a dictionary $\mathbf{D} \in \mathbb{R}^{p\times d}$ that leads to the best possible representation for the original training signals $\mathbf{Y}$ under the strict sparsity constraint.

The universal CS measurement matrix is the random Gaussian matrix with each entry drawn i.i.d from $\mathcal{N}(0,1)$. Let $\mathbf{E}_i \in \mathbb{R}^{p\times m}$ denote the measurement matrix used for the ith training signal. Then, each measurement vector can be written as follows

$$\mathbf{m}_i = \mathbf{E}_i^T \mathbf{y}_i \in \mathbb{R}^m \quad \text{for } i = 1, 2, \ldots, n \tag{5}$$

It is very important to consider different measurement matrices for different training signals here. Otherwise, by projecting all the training signals onto one low-dimensional random subspace, the original signal space is lost, and we will not be able to find the sparse representation model for the original training signals.

We assume that the dictionary model is $\mathbf{D} = \mathbf{B}\mathbf{A}$ where $\mathbf{B}$ is a fixed matrix and the matrix $\mathbf{A}$ is the atom representation dictionary [12]. The matrix $\mathbf{B}$ can contain some prior knowledge about the training signals, e.g. principal components of them learned through another method, or it can be the identity matrix. In this model, the ith column of $\mathbf{A}$ denoted by $\mathbf{a}^{(i)}$ corresponds to the ith dictionary atom $\mathbf{d}^{(i)} = \mathbf{B}\mathbf{a}^{(i)}$, for $i = 1, 2, \ldots, d$. Therefore, each measurement vector $\mathbf{m}_i$ is

$$\mathbf{m}_i = \mathbf{E}_i^T \mathbf{y}_i = \mathbf{E}_i^T \mathbf{B}\mathbf{A}\mathbf{x}^{(i)} \quad \text{for } i = 1, \ldots, n \tag{6}$$

A generalization of the objective function in (1) for the case that we have access only to the CS measurements is to seek the dictionary $\mathbf{D}$ that leads to the best representation for the CS measurements, rather than the original signals, under the strict sparsity constraint. Our proposed objective function thus becomes a well-defined penalty term that measures the quality of our sparse representation in terms of the only information available from the original data. This leads to the criterion:

$$\min_{\mathbf{A},\mathbf{X}} \sum_{i=1}^{n} \left\|\mathbf{m}_i - \mathbf{E}_i^T \mathbf{B}\mathbf{A}\mathbf{x}^{(i)}\right\|_2^2 \quad \text{subject to} \quad \forall i, \quad \left\|\mathbf{x}^{(i)}\right\|_0 \le T \tag{7}$$

Clearly, for $\mathbf{E}_i = \mathbf{I}_{p\times p}$, $i = 1, \ldots, n$, this objective function reduces to the usual K-SVD criterion. Solving the optimization problem in (7) will give us the atom representation dictionary $\mathbf{A}$ and the corresponding dictionary $\mathbf{D} = \mathbf{B}\mathbf{A}$ such that $\mathbf{y}_i = \mathbf{D}\mathbf{x}^{(i)}$ or $\left\|\mathbf{y}_i - \mathbf{D}\mathbf{x}^{(i)}\right\|_2 \le \epsilon$ for some small ϵ.

In our proposed CK-SVD algorithm, the penalty term in (7) is minimized in a simple iterative approach, similar to K-SVD's, that alternates between sparse coding and dictionary update steps.

3.1. Sparse Coding

In the sparse coding step, the penalty term in (7) is minimized with respect to a fixed $\mathbf{A}$ to find the best coefficient matrix $\mathbf{X}$ under the strict sparsity constraint. This can be written as

$$\min_{\mathbf{X}} \sum_{i=1}^{n} \left\|\mathbf{m}_i - \Psi_i \mathbf{x}^{(i)}\right\|_2^2 \quad \text{subject to} \quad \forall i, \quad \left\|\mathbf{x}^{(i)}\right\|_0 \le T \tag{8}$$

where $\Psi_i = \mathbf{E}_i^T \mathbf{B}\mathbf{A}$ is a fixed equivalent dictionary for representation of the ith measurement vector $\mathbf{m}_i$. This optimization problem can be considered as n distinct optimization problems for each measurement vector. We can then use OMP to find the approximate solution $\mathbf{x}^{(i)}$ similar to the K-SVD algorithm.

3.2. Dictionary Update

Assume that the $\mathbf{a}^{(j)}$, $j \neq k$ and coefficients $\mathbf{x}_T^j$, $j \neq k$ are fixed. The goal is to update the kth dictionary atom, or equivalently $\mathbf{a}^{(k)}$, and its corresponding coefficients sequentially for $k = 1, \ldots, d$. The penalty term in (7) can be written as

$$\begin{aligned}
&\sum_{i=1}^{n} \left\| \mathbf{m}_i - \mathbf{E}_i^T \mathbf{B} \mathbf{A} \mathbf{x}^{(i)} \right\|_2^2 \\
&= \sum_{i=1}^{n} \left\| \mathbf{m}_i - \mathbf{E}_i^T \mathbf{B} \sum_{j=1}^{d} \mathbf{a}^{(j)} \mathbf{x}_j^{(i)} \right\|_2^2 \\
&= \sum_{i=1}^{n} \left\| \left(\mathbf{m}_i - \mathbf{E}_i^T \mathbf{B} \sum_{j \neq k} \mathbf{a}^{(j)} \mathbf{x}_j^{(i)} \right) - \mathbf{E}_i^T \mathbf{B} \mathbf{a}^{(k)} \mathbf{x}_k^{(i)} \right\|_2^2 \\
&= \sum_{i \in \mathcal{I}_k} \left\| M_k^{(i)} - \mathbf{E}_i^T \mathbf{B} \mathbf{a}^{(k)} \mathbf{x}_k^{(i)} \right\|_2^2 + \sum_{i \notin \mathcal{I}_k} \left\| M_k^{(i)} \right\|_2^2 \qquad (9)
\end{aligned}$$

where $\mathbf{x}_k^{(i)}$ is a scalar corresponding to the coefficient of the kth dictionary atom in the representation of $\mathbf{m}_i$ with respect to Ψ_i, $\mathcal{I}_k$ is a set of indices of measurement vectors using the kth dictionary atom defined as follows

$$\mathcal{I}_k = \left\{ i \mid 1 \leq i \leq n,\ \mathbf{x}_k^{(i)} \neq 0 \right\} \qquad (10)$$

and $M_k^{(i)}$ is the representation error for the ith measurement vector when the kth dictionary atom is removed. Because the $\mathbf{E}_i$s are distinct this problem cannot be solved by SVD as before. However, the penalty term in (9) is a quadratic function of $\mathbf{a}^{(k)}$ and the minimizer is obtained by setting the derivative of the penalty term with respect to $\mathbf{a}^{(k)}$ equal to zero to obtain

$$\mathbf{B}^T \left(\sum_{i \in \mathcal{I}_k} \left(\mathbf{x}_k^{(i)} \right)^2 \mathbf{E}_i \mathbf{E}_i^T \right) \mathbf{B} \mathbf{a}^{(k)} = \mathbf{B}^T \left(\sum_{i \in \mathcal{I}_k} \mathbf{x}_k^{(i)} \mathbf{E}_i M_k^{(i)} \right) \qquad (11)$$

Defining the matrix $\mathbf{G}_k$ and the vector $\mathbf{b}_k$ as follows

$$\begin{aligned}
\mathbf{G}_k &\triangleq \mathbf{B}^T \left(\sum_{i \in \mathcal{I}_k} \left(\mathbf{x}_k^{(i)} \right)^2 \mathbf{E}_i \mathbf{E}_i^T \right) \mathbf{B} \\
\mathbf{b}_k &\triangleq \mathbf{B}^T \left(\sum_{i \in \mathcal{I}_k} \mathbf{x}_k^{(i)} \mathbf{E}_i M_k^{(i)} \right) \qquad (12)
\end{aligned}$$

we thus find that the update of the kth column of matrix $\mathbf{A}$, $\mathbf{a}^{(k)}$, is obtained by

$$\mathbf{a}^{(k)} = \mathbf{G}_k^{+} \mathbf{b}_k \qquad (13)$$

Given the new $\mathbf{a}^{(k)}$, the optimal $\mathbf{x}_k^{(i)}$ for each $i \in \mathcal{I}_k$ is given by least squares as follows. It is clear that the support of the coefficient matrix $\mathbf{X}$ is preserved as in the K-SVD algorithm.

$$\mathbf{x}_k^{(i)} = \frac{\langle M_k^{(i)}, \mathbf{E}_i^T \mathbf{B} \mathbf{a}^{(k)} \rangle}{\| \mathbf{E}_i^T \mathbf{B} \mathbf{a}^{(k)} \|_2^2} \qquad (14)$$

3.3. Convergence of the CK-SVD Algorithm

In the sparse coding step, optimization of the penalty term in (8) leads to a reduction in MSE conditioned on the success of the OMP algorithm. Furthermore, in the dictionary update process, each dictionary atom and its corresponding coefficients are updated by minimizing quadratic functions, and then these updates are used for updating the rest of the dictionary atoms. Therefore, a monotonic reduction in MSE is guaranteed. Since MSE is bounded from below by zero, we can conclude that, conditional on OMP's success, convergence to a local minimum is guaranteed. We typically run a few times initializing the matrix $\mathbf{A}$ to a few different initial random matrices to avoid getting stuck in a local minima.

Input: CS measurement vectors $\{\mathbf{m}_i\}_{i=1}^n$, CS measurement matrices $\{\mathbf{E}_i\}_{i=1}^n$, matrix $\mathbf{B}$, sparsity level T, number of dictionary atoms d.
Initialization: matrix $\mathbf{A}$, set $J = 1$.
Procedure: Repeat untill convergence (or maximum number of iterations)

1. **Sparse Coding Step:** For each $i = 1, \ldots, n$
   - $\Psi_i \leftarrow E_i^T \mathbf{B} \mathbf{A}$
   - $\mathbf{x}^{(i)} \leftarrow \arg\min_{\mathbf{x}^{(i)}} \left\| \mathbf{m}_i - \Psi_i \mathbf{x}^{(i)} \right\|^2$ s.t. $\left\| \mathbf{x}^{(i)} \right\|_0 \leq T$
2. **Dictionary Update Step:** For each $k = 1, \ldots, d$
   - $\mathcal{I}_k \leftarrow \left\{ i \mid 1 \leq i \leq n,\ \mathbf{x}_k^{(i)} \neq 0 \right\}$
   - For each $i \in \mathcal{I}_k$:
     $M_k^{(i)} \leftarrow \left(\mathbf{m}_i - \mathbf{E}_i^T \mathbf{B} \sum_{j \neq k} \mathbf{a}^{(j)} \mathbf{x}_j^{(i)} \right)$
   - $\mathbf{G}_k \leftarrow \mathbf{B}^T \left(\sum_{i \in \mathcal{I}_k} \left(\mathbf{x}_k^{(i)} \right)^2 \mathbf{E}_i \mathbf{E}_i^T \right) \mathbf{B}$
     $\mathbf{b}_k \leftarrow \mathbf{B}^T \left(\sum_{i \in \mathcal{I}_k} \mathbf{x}_k^{(i)} \mathbf{E}_i M_k^{(i)} \right)$
   - $\mathbf{a}^{(k)} \leftarrow \mathbf{G}_k^{+} \mathbf{b}_k$
   - $\mathbf{x}_k^{(i)} \leftarrow \frac{\langle M_k^{(i)}, \mathbf{E}_i^T \mathbf{B} \mathbf{a}^{(k)} \rangle}{\| \mathbf{E}_i^T \mathbf{B} \mathbf{a}^{(k)} \|^2}$
   - $J \leftarrow J + 1$.

Output: Atom representation dictionary $\mathbf{A}$.

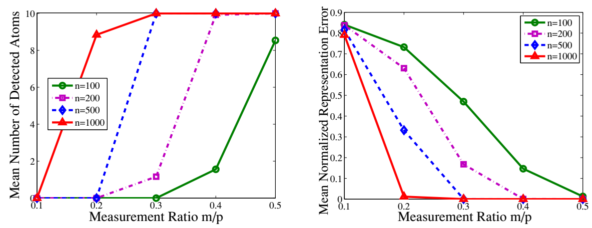
Fig. 1: The Compressive K-SVD Algorithm.

4. EXPERIMENTAL RESULTS

In this section, we demonstrate the effectiveness of our proposed algorithm for both synthetic and real-world datasets. First, we repeat the synthetic data experiment of [3]. We generate a random matrix $\mathbf{D}$ of size 50×10 with i.i.d entries drawn from the $\mathcal{N}(0, 1)$. This matrix, which is also known as the generating dictionary, contains 10 random vectors in $\mathbb{R}^{50}$, each normalized to have unit ℓ^2-norm. A set of training signals $\{\mathbf{y}_i\}_{i=1}^n$ is generated where each element in this set is a linear combination of three distinct generating dictionary atoms chosen uniformly i.i.d from the generating dictionary, and the corresponding coefficients are chosen i.i.d from the normal distribution $\mathcal{N}(0, 80)$. Then, each training signal $\mathbf{y}_i$ is projected onto an m-dimensional random subspace using the measurement matrix $\mathbf{E}_i$ leading to CS measurements.

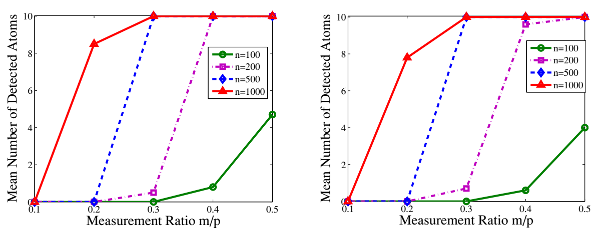
CK-SVD is then applied on the set of CS measurements to learn the generating dictionary atoms of the original training signals. The maximum number of iterations is set to 100 except for the case where we have very few training signals, $n = 100$, where it is set to 200. In fact, increasing the number of training signals increases convergence speed. The matrix $\mathbf{B}$ is set to the identity matrix and the matrix $\mathbf{A}$ is initialized with a random matrix of size 50×10 with i.i.d entries drawn from the $\mathcal{N}(0, 1)$ distribution. To reduce the chance of getting stuck in local minima, we use 5 different initializations for $\mathbf{A}$ to run the algorithm and choose the one with the minimum value of the objective function in (7).

To evaluate the performance of CK-SVD, after learning dictionary atoms and normalizing them to unit ℓ^2-norm, we measure the correlation between each generating dictionary atom and all the learned dictionary atoms. If the absolute value of the correlation is greater than 0.98 for one or more learned atoms, we say that the corresponding generating dictionary atom is detected. Fig. 2 shows the results obtained by applying CK-SVD on 50 independent trials for varying measurement ratios m/p, and varying number of signals n. Also, the average normalized representation error for all the train-



(a) (b)

Fig. 2: Results for synthetic data. (a) Plot of the mean number of detected atoms in 50 trials for varying n and ${}^{m}/_{p}$. (b) Plot of mean normalized representation error of the training signals in 50 trials for varying n and ${}^{m}/_{p}$.



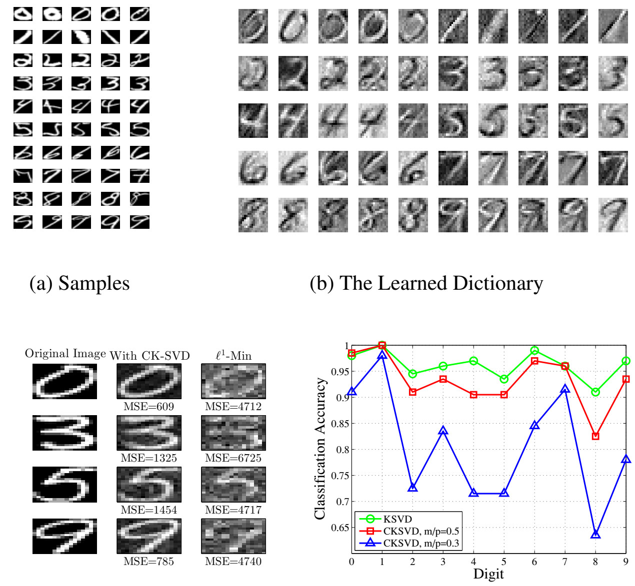
(a) (b)

Fig. 3: Results for noisy CS measurements. Plot of mean number of detected atoms in 10 trials for the case that (a) SNR $= 50$ dB, and (b) SNR $= 30$ dB.

ing signals after decomposing each signal with respect to the learned dictionary $\widetilde{\mathbf{D}}$, ${}^{1}/_{n} \sum_i {\|\mathbf{y}_i - \widetilde{\mathbf{D}}\mathbf{x}^{(i)}\|^2}/{\|\mathbf{y}_i\|^2}$, is plotted in Fig. 2b. Note that CK-SVD is able to recover almost all the generating dictionary atoms for sufficiently many training signals or CS measurements.

For effective real-world data acquisition, our recovery must also be robust to noisy CS measurements. Therefore, we consider the case where our CS measurements are corrupted by white Gaussian noise as, $\mathbf{m}_i = \mathbf{E}_i^T \mathbf{y}_i + \mathbf{n}_i$, for $i = 1, \ldots, n$. In Fig. 3, results for applying CK-SVD on noisy CS measurements for varying signal-to-noise ratios (SNR), varying measurement ratios, and varying number of signals are shown. All the parameters are set as in the previous case and the maximum number of iterations is set to 100. We see that CK-SVD is robust in the case of noisy CS measurements, recovering the generating dictionary atoms very well from the CS measurements.

Recently, there has been growing interest in developing algorithms to perform other signal processing tasks such as classification in addition to recovery on CS measurements. To verify the practicality of our proposed approach, CK-SVD is applied on the real-world USPS dataset. The USPS dataset contains ten classes of handwritten digits with size 16×16 ($p = 256$). We assume known training labels but undersampled training data as in [10]. We thus take CS measurements of a randomly chosen 500 samples from each class and apply CK-SVD on them to learn a dictionary for that class. The maximum number of iterations is set to 100, number of dictionary atoms is set to $d = 50$, and the sparsity level is set to $T = 20$. In Fig. 4b, 5 samples from each of the ten classes of the learned dictionary atoms are shown. It is clear that CK-SVD can successfully recover dictionary atoms from the CS measurements. We have also used the learned dictionary for each class to recover signals from the CS measurements using the OMP recovery method [6]. The recovery results in Fig. 4c indicate that the dictionary learned using CK-SVD can lead to huge gains for signal recovery, compared to using a



(a) Samples (b) The Learned Dictionary

(c) CS Recovery (d) Classification

Fig. 4: Results for the USPS dataset. (a) Samples from the USPS dataset. (b) Samples from the learned dictionary atoms using CK-SVD on ten classes of the USPS dataset for the measurement ratio ${}^{m}/_{p} = 0.5$. (c) Reconstruction accuracy comparison of OMP using the CK-SVD learned dictionary and ℓ^1-minimization using Daubechies 8 wavelets. (d) Plot of classification accuracy for K-SVD on the data versus CK-SVD on the CS measurements.

non-signal adaptive dictionary such as wavelets . Finally, we explore an application of CK-SVD to classification of signals based on CS measurements. After learning the dictionaries for all ten classes, we decompose the measurement vector of each testing sample with respect to $\Psi_i = \mathbf{E}_i^T \mathbf{D}$ for all the learned dictionaries. Then, we label each testing sample as that class that leads to the minimum representation error for the measurement vector. Fig. 4d compares the classification accuracy for K-SVD on the original data versus CK-SVD on the CS measurements. As we can see, classification accuracy for the K-SVD strategy is extremely high (~ 0.95), and for measurement ratio ${}^{m}/_{p} = 0.5$, the classification accuracy gets very close to that obtained with access to the full training data. This shows that this classification strategy can be implemented on CS measurements instead of the full data without much loss of performance.

5. CONCLUSIONS

We have presented a method for learning a dictionary that enables a sparse signal representation using only compressive sensing measurements of the signals. Our method successfully recovered the generating dictionary for a synthetically generated dataset, even for the cases of measurement-to-original-dimension ratios as low as 0.3, low number of training samples, and noisy measurements. Additionally, we were able to learn a convincing set of dictionary atoms for the USPS handwritten digit dataset, and to use these dictionaries to facilitate improved signal reconstruction and compressive classification. We foresee the application of our dictionary learning method to improving compressive sensing results in a variety of signal processing tasks, as well as to learning signal dictionaries in other inverse problem settings, which we hope to explore in future work.

6. REFERENCES

[1] M. Elad, M.A.T. Figueiredo, and Yi Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972 –982, June 2010.

[2] K. Engan, S.O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, vol. 5, pp. 2443 –2446.

[3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311 –4322, Nov. 2006.

[4] Yonina C. Eldar and Gitta Kutyniok, Eds., *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[5] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21 –30, March 2008.

[6] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655 –4666, Dec. 2007.

[7] J.E. Fowler, "Compressive-projection principal component analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2230 –2242, Oct. 2009.

[8] H. Qi and S. Hughes, "Invariance of principal components under low-dimensional random projection of the data," *IEEE International Conference on Image Processing (ICIP)*, October 2012.

[9] 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, July 2009.

[10] T. Wimalajeewa, Hao Chen, and P.K. Varshney, "Performance limits of compressive sensing-based signal classification," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2758 –2770, June 2012.

[11] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231 – 2242, Oct. 2004.

[12] H.V. Nguyen, V.M. Patel, N.M. Nasrabadi, and R. Chellappa, "Kernel dictionary learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 2021 –2024.