

EFFICIENT RECOVERY OF PRINCIPAL COMPONENTS FROM COMPRESSIVE MEASUREMENTS WITH APPLICATION TO GAUSSIAN MIXTURE MODEL ESTIMATION

Farhad Pourkamali Anaraki, Shannon M. Hughes

Department of Electrical, Computer, and Energy Engineering
University of Colorado at Boulder

ABSTRACT

There has been growing interest in performing signal processing tasks directly on compressive measurements, e.g. low-dimensional linear measurements of signals taken with Gaussian random vectors. In this paper, we present a highly efficient algorithm to recover the covariance matrix of high-dimensional data from compressive measurements. We show that, as the number of data samples increases, the eigenvectors (principal components) of the empirical covariance matrix of a simple matrix-vector multiplication of the compressive measurements converge to the true principal components of the original data. Also, we investigate the perturbation of eigenvalues of the covariance matrix under random projection of the data to find conditions for approximate recovery of them. Furthermore, we introduce an important application of our proposed method for efficient estimation of the parameters of Gaussian Mixture Models from compressive measurements. We present experimental results demonstrating the performance and efficiency of our proposed algorithms.

Index Terms— Random projections, Compressive sensing, Compressive signal processing, Principal component analysis, Gaussian mixture model

1. INTRODUCTION

The high cost of acquiring and processing high-dimensional data has motivated the emerging field of compressive sensing (CS) [1, 2, 3]. CS allows us to reconstruct sparse or compressible signals from a few linear measurements of them, taken with e.g. Gaussian random vectors, known as compressive measurements. However, the high computational complexity of reconstructing the original signals has proven to be a bottleneck of CS for practical applications.

Fortunately, for many applications, e.g. estimation of underlying parameters of data, feature extraction, and signal classification, we may be able to avoid expensive signal reconstruction altogether, using only the partial information embedded in the compressive measurements of data to perform signal processing tasks. Hence, there have been some attempts to perform signal processing tasks directly on compressive measurements. For example, in [4], some initial steps to analyze certain inference problems within the compressed space have been done. In [5, 6, 7], performance limits of compressive sensing-based classification of signals have been studied. Also, in [8, 9, 10, 11], various algorithms for learning sparsifying dictionaries from the compressive measurements have been proposed.

In this paper, we focus on the problem of performing Principal Component Analysis (PCA) using the information embedded in the compressive measurements. PCA is a fundamental tool in data analysis and statistics that finds the linear subspace that best fits the data. It is frequently used for dimensionality reduction, feature extraction,

and as a pre-processing step for classification in many applications such as face recognition [12, 13, 14]. In [15, 16], it has been shown that normal PCA on random projections of data, under certain conditions, returns nearly the same result as PCA on the original data.

However, the major drawback of this method is the high computational complexity of recovering the random projections from the compressive measurements. Indeed, the complexity of this process depends heavily on the dimension of original data, which makes it prohibitively expensive when the data dimension is high.

In this paper, we thus introduce an efficient algorithm that allows us to perform PCA using the compressive measurements of data. We will show that, both theoretically and experimentally, when normal PCA is instead applied to a simple matrix-vector multiplication of each compressive measurement with the matrix consisting of the random vectors, it returns nearly the same result as PCA on the original data under similar conditions to [15].

Furthermore, we explore an immediate and important application of our proposed method to estimate the parameters of Gaussian Mixture Models (GMMs) from compressive measurements. GMMs provide powerful tools in signal processing and machine learning for various applications such as data modeling, classification, segmentation, and a large class of inverse problems [17, 18, 19, 20]. In [20], an algorithm for learning the parameters of GMMs from compressive measurements is proposed. However, this algorithm is quite computationally expensive and typically cannot succeed without an application-specific initialization that is very close to the true solution. Thus, we are motivated to present an efficient algorithm that allows us to estimate the parameters of GMMs from the compressive measurements for a wide variety of signals. Its efficiency makes our proposed framework a contender for important applications such as model-based clustering of gene expression microarray data [21].

In Section 2, we present the notation and a brief review of prior work. Section 3 presents two theorems verifying that our proposed method returns nearly the same result as PCA on the original data. In Section 4, we explain the application of our proposed method for estimation of the parameters of GMMs. In Section 5, we show experimental results on both synthetic and real-world datasets to verify the performance of our proposed method and its application.

2. NOTATION AND RELATION TO PRIOR WORK

Assume that our original data are centered at $\bar{\mathbf{x}} \in \mathbb{R}^p$ and $\{\mathbf{v}_i\}_{i=1}^d \in \mathbb{R}^p$ are the orthonormal principal components (PCs). Then, each data sample is represented as $\mathbf{x}_i = \bar{\mathbf{x}} + \sum_{j=1}^d w_{ij} \sigma_j \mathbf{v}_j + \mathbf{z}_i$, $i=1, \dots, n$, where $\{w_{ij}\}_{i=1}^n$ and $\{\mathbf{z}_i\}_{i=1}^n$ are drawn i.i.d. from $\mathcal{N}(0, \mathbf{I}_{d \times d})$ and $\mathcal{N}(0, \frac{\epsilon^2}{p} \mathbf{I}_{p \times p})$, respectively. Also, $\{\sigma_i\}_{i=1}^d$ are scalar constants reflecting the energy of the data in each principal direction such that $\sigma_1 > \sigma_2 > \dots > \sigma_d > 0$. The additive term \mathbf{z}_i allows for some error in our assumption that the data lie on a d -dimensional subspace of \mathbb{R}^p , and it is easy to see that the signal-to-noise ratio is $\text{SNR} = \sum_j \sigma_j^2 / \epsilon^2$.

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

We then assume that we have access only to compressive measurements of the original data. More precisely, we take measurement matrices $\{\mathbf{R}_i\}_{i=1}^n \in \mathbb{R}^{p \times m}$, $m < p$, with i.i.d. entries drawn from $\mathcal{N}(0, \frac{1}{p})$, which is widely used in the CS framework. Then, each compressive measurement $\mathbf{y}_i \in \mathbb{R}^m$ is obtained by taking the inner product of \mathbf{x}_i with the random vectors comprising the columns of \mathbf{R}_i , i.e. $\mathbf{y}_i = \mathbf{R}_i^T \mathbf{x}_i$. As has been noted previously [9, 15], it is important to consider distinct and multiple measurement matrices $\{\mathbf{R}_i\}_{i=1}^n$. If we use the same measurement matrix, then we can only retrieve information within its column space.

In recent work [15, 16], a simple method for recovering the center and PCs from the compressive measurements has been proposed. The projection matrix onto the subspace spanned by the columns of \mathbf{R}_i is $\mathbf{P}_i = \mathbf{R}_i (\mathbf{R}_i^T \mathbf{R}_i)^{-1} \mathbf{R}_i^T \in \mathbb{R}^{p \times p}$. The compressive projection measurement of each \mathbf{x}_i is then $\mathbf{P}_i \mathbf{x}_i = \mathbf{R}_i (\mathbf{R}_i^T \mathbf{R}_i)^{-1} \mathbf{R}_i^T \mathbf{x}_i = \mathbf{R}_i (\mathbf{R}_i^T \mathbf{R}_i)^{-1} \mathbf{y}_i$. Hence, each $\mathbf{P}_i \mathbf{x}_i$ can be recovered directly from the compressive measurement \mathbf{y}_i and the measurement matrix \mathbf{R}_i without knowing or estimating the original data sample \mathbf{x}_i . It has been shown that, under certain conditions, the center and PCs of $\{\mathbf{P}_i \mathbf{x}_i\}_{i=1}^n$ converge to the true center and PCs of $\{\mathbf{x}_i\}_{i=1}^n$ (up to a known scaling factor) almost surely as the number of data samples increases. Also, the eigenvalues are only changed slightly by the scattering of energy into other directions by the random projections.

The major drawback of this method is the high-computational complexity of recovering the compressive projection measurements $\{\mathbf{P}_i \mathbf{x}_i\}_{i=1}^n$ from $\{\mathbf{y}_i\}_{i=1}^n$. Indeed, finding $\mathbf{P}_i \mathbf{x}_i$ requires calculation of the inverse of the matrix $\mathbf{R}_i^T \mathbf{R}_i \in \mathbb{R}^{m \times m}$, which has computational complexity proportional to m^3 . For example, for a small image of size 32×32 pixels, even if we only measure 10% of the pixels ($m/p = 0.1$), then $m^3 \approx 10^6$.

In this paper, we thus introduce a more efficient method for recovering the center and PCs from the compressive measurements $\{\mathbf{y}_i\}_{i=1}^n$. We will show that when the usual PCA is instead applied to $\{\mathbf{R}_i \mathbf{y}_i\}_{i=1}^n$, it returns nearly the same result as PCA on the original data. This leads to a noticeable speed-up of the method, since we only need to calculate a matrix-vector multiplication as opposed to calculating the matrix inversion.

3. EFFICIENT COMPRESSIVE PCA

In this section, we will show that performing PCA on $\{\mathbf{R}_i \mathbf{y}_i\}_{i=1}^n$ recovers the true center (up to a known scaling factor) and PCs of the original data $\{\mathbf{x}_i\}_{i=1}^n$. We will also explain the effect of parameters of our model on the scattering of energy of each PC into other directions. Hence, we find the conditions for which we can approximately recover the true eigenvalues, or equivalently the true covariance matrix of the data.

Theorem 1. Assume that $\{\mathbf{R}_i\}_{i=1}^n$, $\{\mathbf{x}_i\}_{i=1}^n$, and $\{\mathbf{y}_i\}_{i=1}^n$ are as defined in Section 2. Then as the number of samples $n \rightarrow \infty$, the center of $\{\mathbf{R}_i \mathbf{y}_i\}_{i=1}^n$ (scaled by $\frac{p}{m}$) converges to the true center of the original data $\bar{\mathbf{x}}$ almost surely:

$$\lim_{n \rightarrow \infty} \frac{p}{m} \frac{1}{n} \sum_{i=1}^n \mathbf{R}_i \mathbf{y}_i = \bar{\mathbf{x}} \quad (1)$$

Proof. This was previously shown in [16]. \square

Lemma 2. Let \mathbf{t}_k denote the k^{th} column of the matrix $\mathbf{R} \mathbf{R}^T$ where $\mathbf{R} \in \mathbb{R}^{p \times m}$ is a random matrix with i.i.d. entries drawn from $\mathcal{N}(0, \frac{1}{p})$. Let $\{\mathbf{e}_i\}_{i=1}^p \in \mathbb{R}^p$ denote the standard basis. If we define the matrix $\Lambda_{k,l} \triangleq \mathbb{E}[\mathbf{t}_k \mathbf{t}_l^T]$, $1 \leq k, l \leq p$, then

$$\begin{aligned} \Lambda_{k,k} &= \frac{m}{p^2} \mathbf{I}_{p \times p} + \left(\frac{m^2 + m}{p^2} \right) \mathbf{e}_k \mathbf{e}_k^T \\ \Lambda_{k,l} &= \frac{m^2}{p^2} \mathbf{e}_k \mathbf{e}_l^T + \frac{m}{p^2} \mathbf{e}_l \mathbf{e}_k^T, \quad k \neq l. \end{aligned} \quad (2)$$

Proof. See Appendix. \square

Theorem 3. Assume that $\{\mathbf{R}_i\}_{i=1}^n$, $\{\mathbf{x}_i\}_{i=1}^n$, and $\{\mathbf{y}_i\}_{i=1}^n$ are as defined in Section 2, where $\bar{\mathbf{x}} = \mathbf{0}$. Then as the number of samples $n \rightarrow \infty$, the eigenvectors of the empirical covariance matrix of $\{\mathbf{R}_i \mathbf{y}_i\}_{i=1}^n$ converge to the true eigenvectors of the original data's covariance matrix

$$\lim_{n \rightarrow \infty} \frac{p^2}{(m^2 + m)} \frac{1}{n} \sum_{i=1}^n \mathbf{R}_i \mathbf{y}_i \mathbf{y}_i^T \mathbf{R}_i^T = \mathbf{C}_{\text{true}} + \alpha \mathbf{I}_{p \times p} \quad (3)$$

where $\mathbf{C}_{\text{true}} \triangleq \sum_{j=1}^d \sigma_j^2 \mathbf{v}_j \mathbf{v}_j^T$ is the true underlying covariance matrix of the data and $\alpha \triangleq \frac{\sum_{j=1}^d \sigma_j^2}{(m+1)} + \frac{(m+p+1)}{(m+1)p} \epsilon^2$.

Proof. Consider the model for the original data where we have assumed that $\bar{\mathbf{x}} = \mathbf{0}$, $\mathbf{x} = \sum_{j=1}^d w_j \sigma_j \mathbf{v}_j + \mathbf{z}$. Since w_j , \mathbf{z} , and \mathbf{R} are independent with $\mathbb{E}[w_j] = 0$, $\mathbb{E}[\mathbf{z}] = \mathbf{0}$, and $\mathbb{E}[w_j^2] = 1$, we can verify

$$\begin{aligned} \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{x} \mathbf{x}^T \mathbf{R} \mathbf{R}^T] &= \sum_{j=1}^d \sigma_j^2 \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{v}_j \mathbf{v}_j^T \mathbf{R} \mathbf{R}^T] + \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{z} \mathbf{z}^T \mathbf{R} \mathbf{R}^T] \\ &= \sum_{j=1}^d \sigma_j^2 \mathbf{C}_v^{(j)} + \mathbf{C}_\epsilon \end{aligned} \quad (4)$$

where we have defined $\mathbf{C}_v^{(j)} \triangleq \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{v}_j \mathbf{v}_j^T \mathbf{R} \mathbf{R}^T]$ for $j=1, \dots, d$ and $\mathbf{C}_\epsilon \triangleq \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{z} \mathbf{z}^T \mathbf{R} \mathbf{R}^T]$. First we compute the covariance matrix $\mathbf{C}_v^{(j)}$. For simplicity and without loss of generality, we omit the dependence on j . Consider a PC $\mathbf{v} = [\nu_1, \dots, \nu_p]^T$ with unit norm:

$$\begin{aligned} \mathbf{C}_v &= \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{v} \mathbf{v}^T \mathbf{R} \mathbf{R}^T] \stackrel{(a)}{=} \mathbb{E} \left[\left(\sum_{k=1}^p \nu_k \mathbf{t}_k \right) \left(\sum_{l=1}^p \nu_l \mathbf{t}_l \right)^T \right] \\ &= \sum_{k=1}^p \nu_k^2 \Lambda_{k,k} + \sum_{k \neq l} \nu_k \nu_l \Lambda_{k,l} \stackrel{(b)}{=} \left(\frac{m^2 + m}{p^2} \right) \mathbf{v} \mathbf{v}^T + \frac{m}{p^2} \mathbf{I}_{p \times p} \end{aligned}$$

where in (a) \mathbf{t}_k is the k^{th} column of $\mathbf{R} \mathbf{R}^T$ and in (b) we have used Lemma 2. Hence we have

$$\mathbf{C}_v^{(j)} = \left(\frac{m^2 + m}{p^2} \right) \mathbf{v}_j \mathbf{v}_j^T + \frac{m}{p^2} \mathbf{I}_{p \times p}, \quad j = 1, \dots, d. \quad (5)$$

Next, we need to find the covariance matrix \mathbf{C}_ϵ induced by the noise. Assume that $\mathbf{z} = [z_1, \dots, z_p]^T$, then

$$\begin{aligned} \mathbf{C}_\epsilon &= \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{z} \mathbf{z}^T \mathbf{R} \mathbf{R}^T] = \mathbb{E} \left[\left(\sum_{k=1}^p z_k \mathbf{t}_k \right) \left(\sum_{l=1}^p z_l \mathbf{t}_l \right)^T \right] \\ &= \sum_{k=1}^p \mathbb{E}[z_k^2] \Lambda_{k,k} + \sum_{k \neq l} \mathbb{E}[z_k] \mathbb{E}[z_l] \Lambda_{k,l} \\ &= \frac{\epsilon^2}{p} \sum_{k=1}^p \Lambda_{k,k} = \frac{(m^2 + m(p+1))}{p^3} \epsilon^2 \mathbf{I}_{p \times p}. \end{aligned} \quad (6)$$

Finally, we substitute the covariance matrices calculated in (5) and (6) into the equation (4) to find the covariance matrix of the random projections of data:

$$\begin{aligned} \mathbb{E}[\mathbf{R} \mathbf{R}^T \mathbf{x} \mathbf{x}^T \mathbf{R} \mathbf{R}^T] &= \left(\frac{m^2 + m}{p^2} \right) \sum_{j=1}^d \sigma_j^2 \mathbf{v}_j \mathbf{v}_j^T \\ &+ \left(\frac{m}{p^2} \sum_{j=1}^d \sigma_j^2 + \frac{(m^2 + m(p+1))}{p^3} \epsilon^2 \right) \mathbf{I}_{p \times p} \end{aligned}$$

Hence

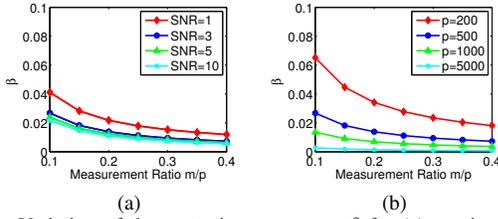


Fig. 1: Variation of the scattering parameter β for (a) varying SNR and measurement ratios m/p for $p=500$, (b) varying p and m/p for fixed SNR=3.

$$\left(\frac{p^2}{m^2 + m}\right) \mathbb{E} \left[\mathbf{R} \mathbf{R}^T \mathbf{x} \mathbf{x}^T \mathbf{R} \mathbf{R}^T \right] = \mathbf{C}_{true} + \alpha \mathbf{I}_{p \times p}$$

where $\mathbf{C}_{true} \triangleq \sum_{j=1}^d \sigma_j^2 \mathbf{v}_j \mathbf{v}_j^T$ and $\alpha \triangleq \frac{\sum_{j=1}^d \sigma_j^2}{(m+1)} + \frac{(m+p+1)}{(m+1)p} \epsilon^2$. Theorem 3 then follows from the law of large numbers. \square

From Theorem 3, we observe that the eigenvectors of our proposed estimator (i.e. PCs) converge to the true eigenvectors. However, the eigenvalues are changed somewhat to reflect the scattering of energy into other directions by the random projections as observed in [15]. We can write $\alpha = \beta (\sum_{j=1}^d \sigma_j^2)$, where we have defined the scattering parameter $\beta \triangleq \frac{1}{p \times m/p + 1} (1 + \frac{1}{\text{SNR}}) + \frac{1}{p} \times \frac{1}{\text{SNR}}$. Let us assume that the total signal power, i.e. $\sum_{j=1}^d \sigma_j^2$, is fixed. First, we can see that when SNR increases, then the scattering parameter β decreases to a limit. Also, if we assume that p is fixed, then as the measurement ratio m/p increases, β decreases. This also holds when the measurement ratio m/p is fixed, where in this case as p increases, β decreases. This is illustrated in Fig. 1. Therefore, in the case that the original data is high-dimensional, the amount of perturbation of eigenvalues is negligible even for small measurement ratios.

As a final note, the estimation accuracy of the center and PCs from the compressive measurements can be improved slightly using an iterative procedure mentioned in [15]. For example, at each iteration, one PC can be estimated using Theorem 3 and removed from the compressive measurements by a least-squares estimation.

4. COMPRESSIVE GAUSSIAN MIXTURE MODEL

In this section, we will show that our proposed method can be used for efficient estimation of the parameters of GMMs from the compressive measurements as well.

We consider the framework for GMMs discussed in [20, 3]. Assume that there exist K Gaussian distributions $\{\mathcal{N}(\mu_k, \Sigma_k)\}_{k=1}^K$ where each data sample \mathbf{x}_i belongs to only one cluster (the hard assignment case) with equal probability of each. Given the compressive measurements of data $\{\mathbf{y}_i = \mathbf{R}_i^T \mathbf{x}_i\}_{i=1}^n$, our goal is both to estimate the Gaussian parameters $\{(\mu_k, \Sigma_k)\}_{k=1}^K$ and to identify the Gaussian distribution that each data sample \mathbf{x}_i belongs to.

We introduce an efficient algorithm, Fast Compressive Expectation Maximization (FC-EM), which is analogous to the classical Expectation Maximization (EM) algorithm [22, 23, 24]. FC-EM is an iterative algorithm that, following an initialization, alternates between two steps (E-step and M-step) to increase the maximum-likelihood (ML) probability of the compressive measurements.

In [20], Sapiro *et al.* have proposed an algorithm for this setting. Their approach recovers the high-dimensional signals from compressive measurements by maximizing the MAP probability given the compressive measurements. This allows them to then calculate the empirical mean and covariance matrix for updating the parameters of each Gaussian cluster as usual. However, we will instead aim to

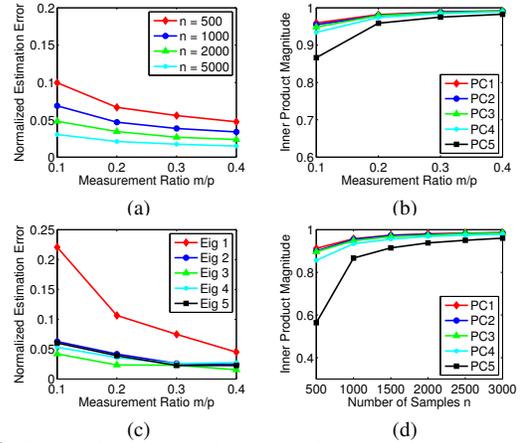


Fig. 2: Results for synthetic data. Plot of (a) normalized estimation error for the center for varying n and m/p , (b) magnitude of the normalized inner product between the estimated and true PCs for varying m/p and fixed $n=1000$, (c) normalized estimation error for the eigenvalues for varying m/p and fixed $n=1000$, and (d) magnitude of the normalized inner product between the estimated and true PCs for varying n and fixed $m/p=0.1$.

update the parameters of GMMs directly from the compressive measurements without estimating the original data samples. This allows us to avoid the expensive signal reconstruction step in each iteration of our proposed FC-EM algorithm.

In the E-step of our proposed algorithm, we assume that the parameters $\{(\mu_k, \Sigma_k)\}_{k=1}^K$ are known. For each $\mathbf{y}_i = \mathbf{R}_i^T \mathbf{x}_i$, the likelihood function conditioned on the k^{th} Gaussian cluster $p(\mathbf{y}_i | c_i = k)$ has the distribution $\mathcal{N}(\mathbf{R}_i^T \mu_k, \mathbf{R}_i^T \Sigma_k \mathbf{R}_i)$. The ML estimate, i.e. maximizing over k $p(\mathbf{y}_i | c_i = k)$, is then equivalent to minimizing the Mahalanobis distance [23],

$$\hat{c}_i = \arg \min_k (\mathbf{y}_i - \mathbf{R}_i^T \mu_k)^T (\mathbf{R}_i^T \Sigma_k \mathbf{R}_i)^{-1} (\mathbf{y}_i - \mathbf{R}_i^T \mu_k) \quad (7)$$

where \hat{c}_i is the ML estimate of the generating Gaussian cluster for the i^{th} data sample. Then, in the M-step, given that we have estimated the identity of the generating Gaussian cluster c_i for each compressive measurement, we update the mean and covariance matrix of each Gaussian cluster. It is known that the ML estimate of these parameters is the empirical estimate [23]. Thus, we can use our method for updating the parameters of each Gaussian cluster. Let us denote the set of compressive measurements that belong to the k^{th} cluster by \mathcal{I}_k , then $\hat{\mu}_k = \frac{p}{m} \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \mathbf{R}_i^T \mathbf{y}_i$ and $\hat{\Sigma}_k = \mathbf{U} \mathbf{S} \mathbf{U}^T$, where \mathbf{U} and \mathbf{S} contain the PCs and eigenvalues obtained by Theorem 3 on $\{\mathbf{y}_i | i \in \mathcal{I}_k\}$.

It is well-known that having a reasonable set of initialization parameters for fitting the GMM is critical [24]. One typical approach is to use the K-means algorithm [24, 23] as a pre-processing step to get a set of initialization parameters. This algorithm iteratively partitions data into K clusters assigning each data sample to the cluster with the nearest center in Euclidean distance. We can easily perform the same algorithm on the compressive measurements for initialization of our FC-EM.

5. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of our proposed method and its application for both synthetic and real-world datasets. For our first experiment, we synthetically generate data samples $\{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{200}$ with 5 principal components such that $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5) = (30, 25, 20, 15, 10)$. We also choose $\epsilon = 27.38$

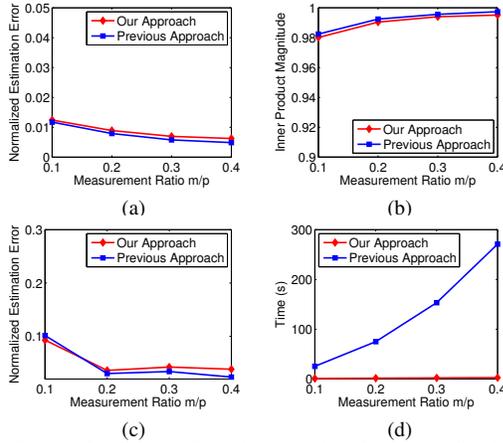


Fig. 3: Results for the Frey Face dataset. Plot of (a) normalized estimation error for the center, (b) magnitude of the normalized inner product between the first estimated and true PC, (c) normalized estimation error for the first eigenvalue, and (d) run-time of each method, for varying measurement ratios m/p . We see that our proposed method has two orders of magnitude less computation time in comparison with [15] with nearly the same performance.

such that the SNR=3. Each entry of the center is drawn from the uniform distribution $[0, 10)$. We then apply our Compressive PCA to estimate the center, PCs, and eigenvalues. To get a sense of the average performance, we report the results averaged over 20 independent trials. Fig. 2(a) shows the accuracy of the estimated center, where the error is the distance between the estimated and the true center normalized by the true center's norm. We can see that this error becomes very small for n sufficiently large, even for small measurement ratios m/p .

Fig. 2(b) shows the magnitude of the normalized inner product between each estimated and true PC for varying measurement ratios for fixed $n=1000$. Here, a value of one represents the best estimate. We can see that our method is able to recover the true PCs. Fig. 2(c) represents the accuracy of the estimated eigenvalues for varying measurement ratios for fixed $n=1000$, where the error is $\frac{|\sigma_{true}^2 - \sigma_{estimated}^2|}{\sigma_{true}^2}$. As we discussed, when the measurement ratio increases, the estimation error of eigenvalues decreases. Fig. 2(d) shows the effect of number of data samples n on the performance of our method. The magnitude of the normalized inner product between each PC and the corresponding true PC for varying n for fixed measurement ratio $m/p=0.1$ is plotted. As we expected, when n increases, the performance of our method goes to 1.

Next, we consider the Frey Face dataset. This is a nice example of a potential real-world application of our work. This dataset includes 1965 images of the face of a single person with different emotions. It is widely used for evaluating the performance of dimensionality reduction techniques [25]. Each image is 20×28 , hence $p=560$. We compare the performance of our proposed method with the previous method [15]. Fig. 3 shows (a) the normalized estimation error for the center, (b) the magnitude of the normalized inner product between the first estimated PC and the true empirical estimate using the original dataset, (c) the normalized estimation error for the first eigenvalue, and (d) the run-time of each method respectively, for varying measurement ratios m/p . We can see that our proposed method is much more efficient than the previous one in [15]. In fact, for this example, our method has two orders of magnitude less computation time with nearly the same performance.

Finally, we demonstrate the performance and efficiency of our

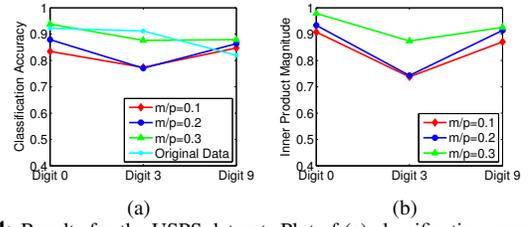


Fig. 4: Results for the USPS dataset. Plot of (a) classification accuracy for each cluster, (b) magnitude of the normalized inner product between the first estimated and true PC for each cluster, for varying measurement ratios.

proposed Compressive Gaussian Mixture Model setting for clustering of signals from the compressive measurements on an example using the USPS dataset. The USPS dataset contains ten classes of handwritten digits with size 16×16 ($p=256$). We consider three clusters including 3 digits of zero, three, and nine each containing 1100 data samples. We then apply the FC-EM algorithm both to assign each compressive measurement to one of the clusters and estimate the parameters of each Gaussian cluster. For initialization, we use K-means on the compressive measurements in m -dimensional space. We set the number of iterations for our FC-EM to 5. In Fig. 4(a), we compare classification accuracy for each cluster for varying measurement ratios m/p and for the EM algorithm on the original data. Also, Fig. 4(b) shows the normalized inner product between the first PC of each cluster after performing FC-EM and the true PC, i.e. the PC obtained by performing PCA on each class of digits. We see that our proposed approach leads to nearly the same performance, even for small measurement ratios such as $m/p=0.3$. The run-time of our method is (107, 138, 174) seconds for $m/p=(0.1, 0.2, 0.3)$, where it takes 101 seconds to run the EM algorithm on the original data. Therefore, we see that the additional complexity of working with the compressive measurements is negligible in our method, unlike if we had recovered all 3300 signals first and then performed GMM.

6. CONCLUSIONS

We have presented an efficient algorithm for performing PCA directly on compressive measurements. In fact, our proposed method is able to recover the center, and PCs much more efficiently than previous approaches. Also, we have studied the perturbation of eigenvalues for reliable recovery of them. Finally, we introduced an important application of our proposed method for estimation of the parameters of GMMs. We present results on the USPS handwritten digit dataset which verify that our approach is able to classify this dataset accurately, even for small measurement ratios.

APPENDIX

Proof of Lemma 2: The k^{th} column of the matrix $\mathbf{R}\mathbf{R}^T$ has the form $\mathbf{t}_k = \left[\sum_{j=1}^m r_{1j}r_{kj}, \dots, \sum_{j=1}^m r_{pj}r_{kj} \right]^T$, where $r_{ij} \triangleq (\mathbf{R})_{ij}$. We compute the entries of $\Lambda_{k,k}$ in three cases using $\mathbb{E}[r]=0$, $\mathbb{E}[r^2]=\frac{1}{p}$, and $\mathbb{E}[r^4]=\frac{3}{p^2}$ [26]: (i) The l^{th} diagonal entry such that $l \neq k$: $\mathbb{E} \left[\left(\sum_{j=1}^m r_{lj}r_{kj} \right)^2 \right] = \sum_{j=1}^m \mathbb{E}[r_{lj}^2] \mathbb{E}[r_{kj}^2] = \frac{m}{p^2}$. (ii) The k^{th} diagonal entry: $\mathbb{E} \left[\left(\sum_{j=1}^m r_{kj}^2 \right)^2 \right] = \sum_{j=1}^m \mathbb{E}[r_{kj}^4] + \sum_{i \neq j} \mathbb{E}[r_{kj}^2] \mathbb{E}[r_{ki}^2] = \frac{m^2+2m}{p^2}$. (iii) The off-diagonal entries: $\mathbb{E} \left[\left(\sum_{j=1}^m r_{lj}r_{kj} \right) \left(\sum_{j=1}^m r_{l'j}r_{kj} \right) \right] = 0$.

We now compute entries of the matrix $\Lambda_{k,l}$, $k \neq l$. Using the same argument, all entries are zero except two: (i) The entry in the k^{th} row and the l^{th} column: $\mathbb{E} \left[\left(\sum_{j=1}^m r_{kj}^2 \right) \left(\sum_{j=1}^m r_{lj}^2 \right) \right] = \left(\sum_{j=1}^m \mathbb{E}[r_{kj}^2] \right) \left(\sum_{j=1}^m \mathbb{E}[r_{lj}^2] \right) = \frac{m^2}{p^2}$. (ii) The entry in the l^{th} row and the k^{th} column: $\mathbb{E} \left[\left(\sum_{j=1}^m r_{lj}r_{kj} \right)^2 \right] = \frac{m}{p^2}$. \square

7. REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [2] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [3] Y. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, Cambridge University Press, 2012.
- [4] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 445–460, 2010.
- [5] T. Wimalajeewa, H. Chen, and P. Varshney, "Performance limits of compressive sensing-based signal classification," *IEEE Transactions on Signal Processing*, vol. 60, pp. 2758–2770, 2012.
- [6] H. Reboredo, F. Renna, R. Calderbank, and M. R. D. Rodrigues, "Compressive classification," in *IEEE Int. Symp. Information Theory (ISIT)*, July 2013.
- [7] B. Coppà, R. Héliot, D. David, and O. Michel, "Classification from compressive representations of data," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1454–1458.
- [8] S. Gleichman and Y. Eldar, "Blind compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, pp. 6958–6975, 2011.
- [9] J. Silva, M. Chen, Y. Eldar, G. Sapiro, and L. Carin, "Blind compressed sensing over a structured union of subspaces," *arXiv preprint arXiv:1103.2469*, 2011.
- [10] C. Studer and R. Baraniuk, "Dictionary learning from sparsely corrupted or compressed signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3341–3344.
- [11] F. Pourkamali-Anaraki and S. Hughes, "Compressive K-SVD," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 5469–5473.
- [12] C. Burges, "Dimension reduction: A guided tour," *Machine Learning*, vol. 2, pp. 275–365, 2009.
- [13] M. Hubert and S. Engelen, "Robust PCA and classification in biosciences," *Bioinformatics*, vol. 20, pp. 1728–1736, 2004.
- [14] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
- [15] H. Qi and S. Hughes, "Invariance of principal components under low-dimensional random projection of the data," in *19th IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 937–940.
- [16] H. Qi and S. Hughes, "Invariance of principal components under low-dimensional random projection of data," *Submitted to IEEE Transactions on Image Processing*, 2012.
- [17] Z. Ghahramani and M. Jordan, "Supervised learning from incomplete data via an em approach," in *Advances in Neural Information Processing Systems*, 1994.
- [18] D. Reynolds, T. Quatieri, and R. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, pp. 19–41, 2000.
- [19] H. Permuter, J. Francos, and H. Jermyn, "Gaussian mixture models of texture and colour for image database retrieval," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003, vol. 3, pp. III–569–72.
- [20] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, pp. 2481–2499, 2012.
- [21] W. Pan, J. Lin, and C. T. Le, "Model-based cluster analysis of microarray gene-expression data," *Genome Biol.*, vol. 3, no. 2, pp. 1–9, 2002.
- [22] Y. Chen and M. Gupta, "EM demystified: An expectation-maximization tutorial," *Electrical Engineering*, vol. 206, 2010.
- [23] K. Murphy, *Machine learning: a probabilistic perspective*, The MIT Press, 2012.
- [24] C. Bishop, *Pattern recognition and machine learning (information science and statistics)*, Springer, 2007.
- [25] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [26] A. Winkelbauer, "Moments and absolute moments of the normal distribution," *arXiv preprint arXiv:1209.4340*, 2012.