BI-SPARSITY PURSUIT FOR ROBUST SUBSPACE RECOVERY

Xiao Bian, Hamid Krim

North Carolina State University
Department of Electrical and Computer Engineering
Raleigh, North Carolina, USA, 27695

ABSTRACT

The success of sparse models in computer vision and machine learning in many real-world applications, may be attributed in large part, to the fact that many high dimensional data are distributed in a union of low dimensional subspaces. The underlying structure may, however, be adversely affected by sparse errors, thus inducing additional complexity in recovering it. In this paper, we propose a bi-sparse model as a framework to investigate and analyze this problem, and provide as a result, a novel algorithm to recover the union of subspaces in presence of sparse corruptions. We additionally demonstrate the effectiveness of our method by experiments on real-world vision data.

Index Terms— Sparse representation, signal recovery, video segmentation, face clustering, non-convex optimization

1. INTRODUCTION

Exploring low dimensional structures embedded in high dimensional spaces has long been an effective and popular methodology to analyze high dimensional data such as images and videos. In particular, low dimensional linear and nonlinear models of varying complexity, have been proposed to address a diverse body of data types. An example of a linear model is that proposed in [1], where high dimensional face images are represented in a low dimensional space, and analyzed by principal component analysis. When accounting for nonlinear structure in the data, a powerful assumption is that a low-dimensional data manifold capturing the essential information, embeds in a high dimensional space. Effective manifold learning methods have been developed to learn data manifolds by exploring relations among neighboring samples [2] [3] [4] [5].

One often arizing issue with the state-of-the-art manifold learning algorithms, is that their performance is highly dependent on the sample density. In other words, dense sample points are needed in order to recover the high curvature area of a data manifold [2] [6]. To address sparsely distributed high dimensional data, sparse models have been proposed on the assumption that data reside in a union of subspaces (UoS) [7] [8] [9] [10]. Specifically, each data class is from one of the subspaces, and can hence be represented as a linear combination of a basis/frames of that subspace. The dimension of each subspace is typically quite small in comparison to the overall dimensions of the UoS, thus yielding a sparse representation of each data class.

Building on work in [11], sparse modeling has been quite successfully applied to high dimensional data analysis, such as images and videos [6] [10]. These prior formulations have, however, assumed noise/perturbation-free data model, which unfortunately limits the scope of UoS model in accounting for corruptions/occlusions, as widely found in practice, and as argued here, well modeled as sparse errors. For example, in video segmentation, the foreground comprised of cars, human, etc, may be interpreted as “perturbations” or outliers relative to the background. Additionally, in face image clustering problem, the change of lighting conditions can be modeled as corruptions/occlusions of the “perfect” face images, and spatially sparse in relation to the entirety of the images.

In this paper, we focus on the problem that data samples can be approximately distributed in a union of subspaces with the presence of sparse corruptions, and propose a bi-sparse model/paradigm to address the problem. Furthermore, we develop a corresponding algorithm, referred to as Robust Subspace Recovery via a bi-sparsity pursuit (RoSuRe), to effectively recover data from corruptions and while simultaneously revealing the relations among the data samples. This algorithm is shown to have a superior performance on the video segmentation and face clustering problems.

2. PROBLEM FORMULATION

Consider a set of data points \( l \in R^d \) uniformly sampled from a union of subspaces \( S = \bigcup_{i=1}^{J} S^i \). Assuming a sufficient sampling density, each sample can then be represented by a linear combination of other samples from the same subspace. Mathematically, we represent the data matrix by \( L = [I_1|I_2| \ldots |I_n] \), yielding

\[
L = LW,
\]

(1)

where \( W \) is (up to a permutation) a \( n \times n \) block-diagonal matrix.
More specifically, let \( n_i \) be the number of samples from \( S^i \), and \( b_i \) the dimension of block \( W_i \) of \( W \), then \( n_i \geq b_i \). It follows that \( b_i \leq \max\{n_i\} \). This condition constrains \( W \) to be a sparse matrix, since \( \rho(W) = \|W\|_0/n^2 \leq \max\{b_i\}/n \leq \max\{n_i\}/n \).

The condition that data samples are generally corrupted by sparse errors, yields the following problem:

**Problem 1.** Given a set of data samples \( X = \{x_1, x_2, \ldots, x_n\} \), find a partition of \( X \), such that each part \( X_i \) can be decomposed into a low dimensional subspace (represented as a low rank matrix \( L_i \)) spanned by an associated set of samples \( \{x_i\} \), and a sparse error (represented as a sparse matrix \( E_i \)), such that

\[
X_i = L_i + E_i, \quad i = 1, \ldots, J
\]

In this case, each sample \( l_i \) is corrupted by some sparse error \( e_i \). Intuitively, we want to separate the sparse errors from the data matrix \( X \), and view the remainder as Eqn(1).

To properly formulate this problem, we leverage the parsimonious property of \( l_i \)-norm to approximate \( \|\cdot\|_0 \), and define a functional on \( X \) as follows,

**Definition 1.** (\( \mathcal{W}_1 \)-function on a matrix space). For any \( d \times n \) matrix \( X \) that rank(\( X \)) \( \leq n \),

\[
\mathcal{W}_1(X) = \min_W \|W\|_1, \quad \text{s.t.} \quad X = XW, \quad W_{ii} = 0.
\]

We then have the following formulation for Problem 1,

\[
\min \mathcal{W}_1(L) + \lambda \|E\|_1 \tag{2}
\]

\[
\text{s.t.} \quad X = L + E.
\]

It is worth noting that formulation Eqn(2) bears a similar form to the problem of robust PCA in [12]. Conceptually, both problems attempt to decompose the data matrix into two parts: one with a parsimonious support, and the other also with a sparse support, albeit in a different domain. For robust PCA, the parsimonious support of the low rank matrix lies in the singular values. In our case, the sparse support of \( L \) lies in the matrix \( W \) via the \( \mathcal{W}_1 \) function, meaning that the columns of \( L \) can be sparsely self-represented.

Furthermore, substituting \( \mathcal{W}_1(L) \) per Definition 1 in Eqn(2), allows us to work on the following optimization problem,

\[
\min_{W, E} \|W\|_1 + \lambda \|E\|_1, \tag{3}
\]

\[
\text{s.t.} \quad X = L + E, \quad L = LW, \quad W_{ii} = 0.
\]

### 3. Algorithm: Robust Subspace Recovery via Bi-Sparsity Pursuit

Obtaining an algorithmic solution to Eqn(3) is complicated by the bilinear term in the constraints which makes the problem non-convex. In this section, we leverage the successes of the alternating direction method (ADMM) [13], and the linearized ADMM (LADM) [14] in large scale sparse representation problem, and focus on designing an adapted algorithm to solve Eqn(3).

Our method, also referred to as robust subspace recovery via bi-sparsity pursuit (RoSuRe) [15], is based on linearized ADMM [14]. Concretely, we pursue the sparsity of \( E \) and \( W \) alternately until convergence. Besides the effectiveness of ADMM on \( l_1 \) minimization problems, a more profound rationale for this approach is that the augmented Lagrange multiplier (ALM) method can better address the non-convexity of Eqn(3) [16] [17]. Although there is no guarantee on the convergence of general non-convex problems, [17] states that under the ALM setting, and mild conditions on the objective function and the constraints, the duality gap can be reduced even to zero with a sufficiently large augmented Lagrange multiplier \( \mu \).

**Algorithm 1** Subspace Recovery via Bi-Sparsity Pursuit (RoSuRe)

1. **Initialize:** Data matrix \( X \in \mathbb{R}^{m \times n} \), \( \lambda \), \( \rho \), \( \eta_1 \), \( \eta_2 \)
2. **while** not converged **do**
   1. **Update** \( W \) by linearized soft-thresholding
      \[
      W_{k+1} = X - E_k,
      \]
      \[
      W_{ii}^{k+1} = 0.
      \]
   2. **Update** \( E \) by linearized soft-thresholding
      \[
      E_{k+1} = I - W_k,
      \]
      \[
      E_{kk}^{k+1} = \rho \mu_k^k.
      \]
   3. **Update** the lagrange multiplier \( Y \) and the augmented Lagrange multiplier \( \mu \)
      \[
      Y_{k+1} = Y_k + \mu_k (L_{k+1} - L_k)
      \]
      \[
      \mu_{k+1} = \rho \mu_k
      \]
3. **end while**

Specifically, substituting \( L \) by \( X - E \), and using \( L = LW \), we can reduce Eqn(3) to a two-variable problem, and hence write the augmented Lagrange function of Eqn(3) as follows,

\[
L(E, W, Y, \mu) = \lambda \|E\|_1 + \|W\|_1 + \langle LW - L, Y \rangle + \frac{\mu}{2} \|X - E\| W - (X - E)\|_F^2, \tag{4}
\]

where \( Y \) is the Lagrange multiplier. Letting \( \hat{W} = I - W \), we alternatively update \( W \) and \( E \),

\[
W_{k+1} = \arg \min_W \|W\|_1 + \|L_{k+1} W - L_{k+1}, Y_k\|_F^2 + \frac{\mu}{2} \|L_{k+1} W - L_{k+1}\|_F^2, \tag{5}
\]

\[
E_{k+1} = \arg \min_E \|E\|_1 + \frac{1}{\rho} \|E - X\| \hat{W}_{k+1}, Y_k\|_F^2 + \frac{\mu}{2} \|E - X\| \hat{W}_{k+1}\|_F^2. \tag{6}
\]
The solution of Eqn(5) and Eqn(6) can be well approximated in each iteration by linearizing the augmented Lagrange term [14],

\[
W_{k+1} = \mathcal{T}_{\frac{1}{\eta_1}} \left( W_k + \frac{L_{k+1}^T (L_{k+1} W_k - Y_k / \mu_k)}{\eta_1} \right),
\]

(7)

\[
E_{k+1} = \mathcal{T}_{\frac{1}{\eta_2}} \left( E_k + \frac{(L_{k+1} W_{k+1} - Y_k / \mu_k) W_{k+1}^T}{\eta_2} \right),
\]

(8)

where \( \eta_1 \geq \|L\|_2^2 \), \( \eta_2 \geq \|\hat{W}\|_2^2 \), and \( \mathcal{T}_\alpha(\cdot) \) is a soft-thresholding operator.

In addition, the Lagrange multipliers are updated as follows,

\[
Y_{k+1} = Y_k + \mu_k (L_{k+1} W_{k+1} - L_{k+1})
\]

(9)

\[
\mu_{k+1} = \rho \mu_k
\]

(10)

4. EXPERIMENTS AND VALIDATION

Since the UoS model has been intensively researched and successfully applied to many computer vision and machine learning problems [18] [19] [6], we expect that our model may also address a class of problems. Specifically, we investigate the video background subtraction and face clustering problem, as exemplars of this promising novel method.

4.1. Video background subtraction

Surveillance videos can be naturally modeled as UoS model due to their relatively static background and sparse foreground. The power of our proposed UoS model lies in coping with both a static camera and a panning one with periodic motion. Here we test our method in both scenarios using surveillance videos from MIT traffic dataset [20]. In Fig.1, we show the segmentation results with a static background. For the scenario of a "panning camera", we generate a sequence by cropping the previous video. The cropped region is swept from bottom right to top left and then backward periodically, at the speed of 5 pixels per frame. The results are shown in Fig.2. We can see that the results in the moving camera scenario are only slightly worse than the static case.

More interestingly, the sparse coefficient matrix \( W \) provides important information about the relations among data points, which potentially may be used to cluster data into individual clusters. In Fig. 3(a), we can see that, for each column of the coefficient matrix \( W \), the nonzero entries appear periodically. Considering the periodic motion of the camera, every frame is mainly represented by the frames when the camera is in a similar position, i.e. a similar background, with the foreground moving objects as “corruptions”. We hence permute the rows and columns of \( W \) according to the position of cameras, as shown in Fig. 3(b). A block-diagonal structure then emerges, where images with similar backgrounds are clustered as one subspace.

4.2. Face clustering under various illumination conditions

Recent research on sparse models implies that a parsimonious representation may be a key factor for classification [6] [21]. Indeed, the sparse coefficients pursued by our method shows clustering features in experiments of both synthetic and real-
Table 1. Clustering error (%) on the Extended Yale Face Database B compared to state-of-the-art methods [19] [18] [23]

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>LSA</th>
<th>LRR</th>
<th>SSC</th>
<th>RoSuRe</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-subjects Mean</td>
<td>38.20</td>
<td>2.54</td>
<td>1.86</td>
<td><strong>0.71</strong></td>
</tr>
<tr>
<td>Median</td>
<td>47.66</td>
<td>0.78</td>
<td><strong>0.00</strong></td>
<td>0.39</td>
</tr>
<tr>
<td>5-subjects Mean</td>
<td>58.02</td>
<td>6.90</td>
<td>4.31</td>
<td><strong>3.24</strong></td>
</tr>
<tr>
<td>Median</td>
<td>56.87</td>
<td>5.63</td>
<td>2.50</td>
<td><strong>1.72</strong></td>
</tr>
<tr>
<td>10-subjects mean</td>
<td>60.42</td>
<td>22.92</td>
<td>10.94</td>
<td><strong>5.62</strong></td>
</tr>
<tr>
<td>Median</td>
<td>57.50</td>
<td>23.59</td>
<td>5.63</td>
<td><strong>5.47</strong></td>
</tr>
</tbody>
</table>

We compare the clustering performance of RoSuRe with the state-of-the-art methods such as local subspace analysis(LSA) [23], sparse subspace clustering (SSC) [19], and low rank representation(LRR) [18]. The best performance of each method is referenced in Table 1 for comparison. As shown in the table, RoSuRe has the lowest mean clustering error rate in all three settings, i.e. 2 subjects, 5 subjects and 10 subjects. In particular, in the most challenging case of 10 subjects, the mean clustering error rate is as low as 5.62% with the median 5.47%. Additionally, we show the robustness of our method with respect to $\lambda$ in a 10-subject scenario. In Fig. 5, the correlation between the value of $\lambda$ and the cluster accuracy maintains above 98% with $\lambda$ varying from 500 to 15000.

In Fig. 6, we present the recovery results of some sample faces from the 10-subject clustering scenario. In most cases, the sparse term $E$ compensates the missing information caused by lighting condition. This is especially true when the shadow area is small, i.e. a sparser support of error term $E$, we can see a visually good recovery of the missing area. This result validates the effectiveness of our method to solve the problem of subspace clustering with corrupted data.

5. CONCLUSION

We have proposed in this paper a novel approach to recover underlying subspaces of data samples from measured data corrupted by general sparse errors. We formulated the problem as a non-convex optimization problem, and proposed an effective algorithm named RoSuRe to well approximate the global solution of the optimization problem. Furthermore, experiments on real-world vision data are presented to show a broad range of applications of our method.

Future work may include several aspects across computer vision and machine learning. It would first be interesting to understand and extend this work from a dictionary learning angle, to learn a feature set for high dimensional data representation and recognition. Additionally, a sufficient condition to guarantee the recovery of corrupted data is not only theoretically interesting, but would also be helpful for better understanding the problem.
REFERENCES


