Conformity Evaluation of Data Samples by $L_1$-Norm Principal-Component Analysis

Ying Liu$^a$ and Dimitris A. Pados$^b$

$^a$Department of Electrical Engineering, The State University of New York at Buffalo, Buffalo, NY 14260
$^b$Department of Computer and Electrical Engineering and Computer Science & I-SENSE, Florida Atlantic University, Boca Raton, FL 33431

ABSTRACT

We describe an iterative procedure for soft characterization of outlier data in any given data set. In each iteration, data compliance to nominal data behavior is measured according to current $L_1$-norm principal-component subspace representations of the data set. Successively refined $L_1$-norm subspace data set representations lead to successively refined outlier data characterization. The effectiveness of the proposed theoretical scheme is experimentally studied and the results show significantly improved performance compared to $L_2$-PCA schemes, standard $L_1$-PCA, and state-of-the-art robust PCA methods.

Keywords: Conformity evaluation, dimensionality reduction, eigenvector decomposition, feature extraction, $L_1$-norm, outliers, principal-component analysis.

1. INTRODUCTION

In many real-world applications such as video surveillance, processing of remotely sensed images, and network traffic analysis, the dimensionality of data are usually very high, and direct handling of the high-dimensional data is computationally expensive. Fortunately, many high-dimensional data have intrinsic low-rank structures, that can be revealed through dimensionality reduction schemes. Over several decades, principal-component analysis (PCA) has been one of the most widely applied dimensionality reduction methods due to its simplicity and effectiveness. Given a dataset, PCA finds a set of projection vectors (the so-called “principal components”) to maximize the variance of the projected data points, and the structure of the original data could be effectively preserved under the projection. The nominal compliance of each data sample can then be inferred by leveraging the principal components, and be used to perform tasks such as moving object extraction and outlier detection.

Nevertheless, the conventional PCA based on the $L_2$-norm ($L_2$-PCA) is prone to the presence of outliers because the effect of the outliers with a large norm is exaggerated by the use of the $L_2$-norm. To alleviate this problem, there has been a growing interest in robust PCA methods, such as $L_1$-error minimization$^{1-4}$ non-negative matrix factorization via Manhattan distance minimization (MahNMF)$^5$ and the low-rank and sparse decomposition approach$^{6,8}$. Another line of research has focused on robust subspaces calculation by explicit $L_1$ projection maximization$^{9,11}$. The resulting principal components are called $L_1$ principal components. In particular, the exact calculation of $L_1$ principal components was developed$^{11}$ for the first time in the literature. Later, suboptimal algorithms were developed$^{12,13}$ for fast computation of the $L_1$ principal components. The $L_1$-PCA method has been successfully applied to a wide range of research fields such as direction of arrival (DoA) estimation$^{14}$ and robust face recognition$^{15,17}$. Most recently, compressed-sensed-domain $L_1$-PCA methods were developed for low-rank background scene and sparse foreground moving objects extraction from compressed-sensed surveillance video sequences$^{18,20}$.

In spite of their robustness, for a given data set with potential outliers, the existing $L_1$-PCA methods$^{9,20}$ merely calculate a “one-shot” $L_1$ subspace, which can still be away from the true nominal signal subspace of interest when the data sets are severely contaminated. In this paper, we propose an iterative approach that
iteratively generates a sequence of improved $L_1$ subspaces. In each iteration, nominal compliance of each sample is inferred by its position relative to the $L_1$ subspace calculated in the previous iteration and translated to a “weight”. Samples with higher weights tend to be nominal samples and samples with lower weights are more likely to be the outliers. Weighted $L_1$-PCA calculation is then carried out in which the contribution of outlying samples in the data set is suppressed resulting in an improved $L_1$-subspace. The sample weights converge as the iteration number increases and the iterative algorithm terminates when the weights in the current and previous iterations are deemed close enough.

The remainder of this paper is organized as follows. In Section 2, we develop the iterative $L_1$-PCA algorithm, provide convergence analysis, and propose a stopping criterion for practical implementation of the algorithm. In Section 3, the effectiveness of the proposed algorithm is demonstrated through two experiments: (i) moving objects extraction in video surveillance, and (ii) unsupervised outlier detection of the UCI machine learning data sets. Finally, a few conclusions are drawn in Section 4.

2. PROPOSED ITERATIVE $L_1$ PRINCIPAL-COMPONENT ANALYSIS

2.1 Background of $L_1$-PCA

We consider a $D \times N$ $(N < D)$ data matrix that has $N$ data points $x_i \in \mathbb{R}^D$, $i = 1, 2, \ldots, N$, that is

$$X = [x_1 \ x_2 \ \cdots \ x_N]. \quad (1)$$

The conventional $L_2$-PCA seeks $r$ orthonormal projection vectors to describe the rank-$r$ subspace of the data matrix $X$ by

$$P^{L_2} : P_{L_2} = \arg \max_{P \in \mathbb{R}^{D \times r}} \|X^T P\|_2, \quad (2)$$

and the solution is given by the $r$ dominant-singular-value left singular vectors of the original data matrix $X$.

Nevertheless, the $L_2$-norm metric in Problem $P^{L_2}$ is sensitive to outlying samples that are numerically distant from the nominal data, leading to inaccurate low-rank subspace calculation and data conformity evaluation. Motivated by this observed drawback of $L_2$ subspace signal processing, subspace-decomposition approaches that are based on the $L_1$ norm were proposed for robust low-rank subspace computation. Replacing the $L_2$-norm in Problem $P^{L_2}$ by $L_1$-norm, the so-called $L_1$-PCA calculates principal components in the form of

$$P^{L_1} : P_{L_1} = \arg \max_{P \in \mathbb{R}^{D \times r}} \|X^T P\|_1. \quad (3)$$

Since projecting all the data to a subspace deviated by the outliers is less likely to generate a larger projection $L_1$-norm than projecting the data to the correct low-rank subspace, $P_{L_1}$ in (3) is likely to be closer to the true nominal rank-$r$ subspace than $L_2$-PCA. The $r$ columns of $P_{L_1}$ in (3) are the so-called $r$ $L_1$ principal components that describe the rank-$r$ subspace in which $X$ lies. It was shown that the exact calculation of the $L_1$ principal components in Problem $P^{L_1}$ can be recast as a combinatorial problem and solved in exponential time. Besides, a polynomial-time algorithm is developed for any fixed data dimension $D$, and bit-flipping based sub-optimal algorithms were also proposed for the fast calculation of the $L_1$ principal components.

The regular $L_1$-PCA problem in (3) seeks a rank-$r$ subspace from the data matrix $X \in \mathbb{R}^{D \times N}$ by one-shot calculation. Although the adopted $L_1$-norm maximization is less affected by outliers compared to $L_2$-norm maximization in $L_2$-PCA in (2), the produced $L_1$ subspace $P_{L_1}$ can still be away from the true nominal signal low-rank subspace. In the following subsection, we propose an iterative method that generates a sequence of improved $L_1$ subspaces for the same data matrix $X$. 
2.2 Iterative Data Conformity Evaluation

We consider the calculation of $r$ principal components $P_{L_1} \in \mathbb{R}^{D \times r}$, $D > r > 1$. Initially, the direct $L_1$ subspace is computed via (3) and denoted by $P_{L_1}^{(0)}$. Next, the distance of each sample $x_n$ from subspace $P_{L_1}^{(0)}$ is defined as the $L_2$ error between $x_n$ and its rank-$r$ surrogate

$$d_n^{(1)} = \| x_n - P_{L_1}^{(0)} P_{L_1}^{(0)\top} x_n \|_2, \quad n = 1, \ldots, N.$$  (4)

We expect large $d_n^{(1)}$ if $x_n$ is an “outlier” and small $d_n^{(1)}$ if $x_n$ is a nominal sample. Therefore, the conformity of each sample can be measured as the reciprocal of its $L_2$ distance from the subspace, i.e.,

$$w_n^{(1)} = (d_n^{(1)})^{-1}, \quad n = 1, \ldots, N,$$  (5)

followed by normalization

$$\tilde{w}_n^{(1)} = \frac{w_n^{(1)}}{\sum_{n=1}^{N} w_n^{(1)}}, \quad n = 1, \ldots, N.$$  (6)

When computing the $L_1$ subspace, data samples with larger conformity should contribute more and samples with smaller conformity should be suppressed such that the resulting calculated $L_1$ subspace is more accurate. In this direction, we propose that each data sample $x_n$ is scaled by its conformity $\tilde{w}_n^{(1)}$. We then form a weight matrix with the conformity values,

$$\tilde{W}^{(1)} = \begin{bmatrix} \tilde{w}_1^{(1)} & 0 & 0 & \cdots \\ 0 & \tilde{w}_2^{(1)} & 0 & \cdots \\ \vdots \\ 0 & 0 & \cdots & \tilde{w}_N^{(1)} \end{bmatrix},$$  (7)

and update the $L_1$ subspace as

$$P_{L_1}^{(1)} = \arg \max_{P \in \mathbb{R}^{D \times r}} \| (X \tilde{W}^{(1)})^\top P \|_1.$$  (8)

The above weighted $L_1$-subspace calculation can be performed iteratively. In the $k$th iteration, the weights are computed using the $L_1$-subspace $P_{L_1}^{(k-1)}$ computed at the $(k-1)$th iteration, i.e.

$$d_n^{(k)} = \| x_n - P_{L_1}^{(k-1)} P_{L_1}^{(k-1)\top} x_n \|_2, \quad 1 \leq n \leq N,$$  (9)

$$w_n^{(k)} = (d_n^{(k)})^{-1},$$  (10)

$$\tilde{w}_n^{(k)} = \frac{w_n^{(k)}}{\sum_{n=1}^{N} w_n^{(k)}},$$  (11)

$$\tilde{W}^{(k)} = \begin{bmatrix} \tilde{w}_1^{(k)} & 0 & 0 & \cdots \\ 0 & \tilde{w}_2^{(k)} & 0 & \cdots \\ \vdots \\ 0 & 0 & \cdots & \tilde{w}_N^{(k)} \end{bmatrix},$$  (12)

Subsequently, the $L_1$-subspace at the $k$th iteration is updated to

$$P_{L_1}^{(k)} = \arg \max_{P \in \mathbb{R}^{D \times r}} \| (X \tilde{W}^{(k)})^\top P \|_1.$$  (13)
In this section, we assess the effectiveness of the proposed iterative re-weighted $L_1$-norm Principal-Components Calculation (rank-$r$)

**Algorithm 1** Iterative Re-weighted $L_1$-norm Principal-Components Calculation (rank-$r$)

**Input:** $X = [x_1 \cdots x_N] \in \mathbb{R}^{D \times N}$, $r$, $\beta$, $\epsilon$.

**Initialization:** Find $P_{L_1}^{(0)} \in \mathbb{R}^{D \times r}$ by (3).

**Iterative $L_1$ Subspace Calculation:**

for $k = 1, 2, \cdots,$ do

1. Compute the $L_2$ error between $x_n$ and its rank-$r$ surrogate $d_n^{(k)} = \|x_n - P_{L_1}^{(k-1)}P_{L_1}^{(k-1)T}x_n\|_2$.
2. Define $w_n^{(k)} = (d_n^{(k)})^{-1}$.
3. If $k = 1$, $w_n^{(k)} \leftarrow w_n^{(k)}$; if $k > 1$, update $w_n^{(k)}$ by (15).
4. Check stopping criterion: if $\|w^{(k)} - w^{(k-1)}\|_2 < \epsilon$, exit.
5. Update $\tilde{w}_n^{(k)}$ and $\tilde{W}^{(k)}$ by (11) and (12).
6. Update the $L_1$ subspace $P_{L_1}^{(k)}$ by (13).

end for

**Output:** Rank-$r$ $L_1$ subspace sequence $P_{L_1}^{(k)} \in \mathbb{R}^{D \times r}$, $k = 1, 2, \cdots$.

### 2.3 A Convergent Weight Sequence

Inspired by, we modify the weight update formula as follows to guarantee a convergent weight sequence for practical algorithmic implementation. In the $k$th iteration, we first compute the $\ell_2$ error (distance) for each sample as in (9). Then we define

$$w_n^{(k)} = (d_n^{(k)})^{-1}$$

and update the weight $w_n^{(k)}$ based on $w_n^{(k)}$. If $k = 1$, let $w_n^{(k)} = u_n^{(k)}$. For $k > 1$, $w_n^{(k)}$ is updated by

$$w_n^{(k)} = \begin{cases} 
  w_n^{(k-1)}(1 - \beta^{k-1}), & \text{if } u_n^{(k)} < w_n^{(k-1)}(1 - \beta^{k-1}), \\
  u_n^{(k)} & \text{if } w_n^{(k-1)}(1 - \beta^{k-1}) \leq u_n^{(k)} \leq w_n^{(k-1)}(1 + \beta^{k-1}), \\
  w_n^{(k-1)}(1 + \beta^{k-1}), & \text{if } u_n^{(k)} > w_n^{(k-1)}(1 + \beta^{k-1})
\end{cases}$$

(15)

where $0 < \beta < 1$ is a pre-defined parameter. Intuitively, we avoid updating the weights too aggressively by restricting the new weight $w_n^{(k)}$ to be within a small neighborhood of the weight in the previous iteration $w_n^{(k-1)}$. The size of the neighborhood depends on $\beta$. Subsequently, $w_n^{(k)}$ is normalized as in (11), followed by weight matrix construction in (12). The convergence of the weight sequence can be verified by

$$\lim_{k \to \infty} \beta^{k-1} = 0,$$

(16)

$$\lim_{k \to \infty} (w_n^{(k)} - w_n^{(k-1)}) = 0,$$

(17)

$$\lim_{k \to \infty} (\tilde{w}_n^{(k)} - \tilde{w}_n^{(k-1)}) = 0.$$  

(18)

### 2.4 Stopping Criterion

In implementing the proposed iterative algorithm, we exit the algorithm when the difference between the weight vectors at the $k$th and $(k-1)$th iteration is smaller than a predefined threshold $\epsilon$, that is,

$$\|w^{(k)} - w^{(k-1)}\|_2 < \epsilon,$$

(19)

where $w^{(k)} = [w_1^{(k)}, w_2^{(k)}, \cdots, w_N^{(k)}]^T$ and $w^{(k-1)} = [w_1^{(k-1)}, w_2^{(k-1)}, \cdots, w_N^{(k-1)}]^T$.

The complete algorithm is summarized in Algorithm 1.

### 3. EXPERIMENTAL STUDIES

In this section, we assess the effectiveness of the proposed iterative re-weighted $L_1$-PCA (IRW $L_1$-PCA) algorithm through two experimental studies: (i) moving object extraction in video surveillance, and (ii) unsupervised outlier detection of the UCI data sets.
3.1 Moving Object Extraction

Consider a sequence of surveillance video frames $X_t \in \mathbb{R}^{m \times n}$ with frame resolution of $m \times n$ pixels and time index $t = 1, ..., N$. A typical surveillance video sequence is consisted of a static background scene that can be modeled as a low-rank component, and sparse foreground moving objects superimposed on the static background scene that are regarded as the outliers. For security monitoring, the objective is to extract the moving objects. In our experiment, we perform block-by-block IRW $L_1$-PCA for low-rank background modeling and foreground extraction. We divide each frame $X_t$, into $J$ blocks $X^j_t \in \mathbb{R}^{m_b \times n_b}$, $j = 1, ..., J$. We let $x^j_t \in \mathbb{R}^D$, $D = m_b n_b$, represent vectorization of $X^j_t$ via column concatenation. For each sequence of co-located blocks, $x^j_t$, $t = 1, ..., N$, we model the static background scene as a low-rank component $z^j_t$ and the foreground moving objects as an outlying component $s^j_t$. That is,

$$x^j_t = z^j_t + s^j_t, \quad t = 1, ..., N,$$

and $s^j_t$ appears only a few times in the block sequence. In matrix-form representation of the $j$th block across $N$ frames, $X^j \triangleq [x^j_1, ..., x^j_N] \in \mathbb{R}^{D \times N}$ and

$$X^j = Z^j + S^j.$$

To extract the low-rank background information, we carry out IRW $L_1$-PCA on $X^j$ and obtain the rank-$2$ $L_1$ subspace $P^j_{L_1}$ at convergence. Afterwards, the background blocks can be approximated by $\hat{Z}^j = P^j_{L_1} P^j_{L_1} X^j$ and the foreground blocks can be extracted as $S^j = X^j - \hat{Z}^j$, $j = 1, ..., J$.

We test the method on the Daniel_light video sequence with 80 frames, each of $120 \times 160$ pixels. We process $N = 8$ successive frames at a time. To mitigate the “blockiness” artifact, we divide each frame into $J = 1296$ overlapping blocks of size $15 \times 20$, and apply the proposed IRW $L_1$-PCA method independently to each group of co-located blocks across 8 frames. The final background and foreground scenes are obtained by averaging the extracted background pixels (as well as the foreground pixels) for which multiple results are available.

Fig. 1 displays the background and foreground extracted at multiple distinct time slots $t = 5, 15, 35, 47$ by the proposed IRW $L_1$-PCA, DECOLOR,\textsuperscript{7} MahNMF,\textsuperscript{5} and the regular $L_1$-PCA\textsuperscript{11} methods. In addition to low-rank and sparse decomposition, the DECOLOR\textsuperscript{7} approach also uses Markov random-field (MRF) modeling to improve the accuracy of detecting contiguous outliers. The results show that the foreground scenes extracted by DECOLOR are somewhat affected by the desks and computers that belong to the background scene, and the regular $L_1$-PCA extracted foreground scenes suffer from severe “ghosts” that are formed when an object initially in the background begins to move. Although MahNMF performs as well as the proposed algorithm for $t = 15$ and 35, it incorrectly recovers the background scene for $t = 5$ and 47 and results in problematic foreground extraction. In contrast, the proposed IRW $L_1$-PCA offers much clearer foreground scenes than the other three methods consistently across multiple frames.

3.2 Unsupervised Outlier Detection of UCI Data Sets

The problem of unsupervised outlier detection is to determine data samples that significantly deviate from the majority of data samples. It has been widely studied due to its numerous applications in intrusion detection, land resource exploration, criminological investigation, and medical diagnostics. Different outlier detection paradigms were proposed in the literature, such as density-based methods\textsuperscript{23,24}, feature bagging,\textsuperscript{25} and subspace methods\textsuperscript{26,27}. Among these approaches, the classic local outlier factor (LOF) algorithm\textsuperscript{25} compares the density of each data sample with that of its local neighbors, considering that an outlier has low density compared to its local neighborhood. The state-of-the-art LODES algorithm\textsuperscript{27} combines spectral embedding with LOF. It embeds data samples in a low-dimensional subspace where similar samples are pulled close together. Subsequently, LOF algorithm is applied to the embedded data samples such that outliers can be more easily distinguished from regular objects.

In this section, we apply the proposed IRW $L_1$-PCA algorithm to detect the outliers in two real-world data sets taken from the UCI machine learning repository,\textsuperscript{21} Satimage and Glass. The original Statlog (Landsat Satellite) data set is a multi-class classification data set generated from multi-spectral scanner image data. Each sample has $D = 36$ attributes and each class is one land type. In our study, class 2 (cotton crop) is down-sampled to 71 outliers, while all the other classes (red soil, grey soil, damp grey soil, soil with vegetation stubble, and
Figure 1: Daniel light sequence: Original frame [row (i)] of time slot $t = 5, 15, 35$, and 47; proposed IRW $L_1$-PCA reconstructed background and moving objects [rows (ii) and (iii)]; DECOLOR [iv] reconstructed background and moving objects [rows (iv) and (vi)]; MahNMF [v] reconstructed background and moving objects [rows (v) and (vii)]; regular $L_1$-PCA [vi] reconstructed background and moving objects [rows (viii) and (ix)].
very damp grey soil) are combined to form an inlier class. The modified data set is referred to as Satimage. Each time we took \( N_o = 5 \) random samples from the outlier class to construct the outlier data set \( X_o \in \mathbb{R}^{D \times N_o} \), and \( N_i = 10 \) random samples from the inlier class to construct the inlier data set \( X_i \in \mathbb{R}^{D \times N_i} \). The overall data set is formed as \( X = [X_i \ X_o] \in \mathbb{R}^{(D=36) \times (N=15)} \). The proposed IRW \( L_1 \)-PCA algorithm with \( r = 2 \) principal components is applied to \( X \) to obtain the \( L_1 \) subspace \( P_{L_1} \in \mathbb{R}^{(D=36) \times (r=2)} \) at convergence. Subsequently, the \( \ell_2 \)-norm distance from data sample \( x_n \) to the proposed \( L_1 \) subspace \( P_{L_1} \), that is, \( \|x_n - P_{L_1} P_{L_1}^T x_n \|_2 \) is calculated for each data sample as the outlier score, and the samples with highest scores are detected as outliers, i.e. samples that belong to the cotton crop land type. We carried out 20 independent experiments, and plot the receiver operating characteristic curves (ROC) of the proposed algorithm and four existing algorithms, LODES,\(^{27}\) LOF,\(^{23}\) \( L_1 \)-PCA\(^{11}\) and \( L_2 \)-PCA. The ROC plots the true positive rate versus the false positive rate, which is a common measure to evaluate the performance of outlier detection methods. As shown in Fig. 2a, the proposed IRW \( L_1 \)-PCA method demonstrates the ability to capture the outliers significantly earlier than all other algorithms. A summary measure of the detection accuracy is the area under the ROC curve (AUC) shown in table 1. Again we observe that the proposed method achieves the highest AUC value for the Satimage data set among all algorithms in comparison.

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>LODES(^{27})</th>
<th>LOF(^{23})</th>
<th>( L_1 )-PCA(^{11})</th>
<th>( L_2 )-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satimage</td>
<td>0.998</td>
<td>0.951</td>
<td>0.861</td>
<td>0.968</td>
<td>0.545</td>
</tr>
<tr>
<td>Glass</td>
<td>0.971</td>
<td>0.963</td>
<td>0.946</td>
<td>0.920</td>
<td>0.730</td>
</tr>
</tbody>
</table>

Figure 2: The receiver operating characteristics (ROC) curves of (a): the Satimage data set, and (b) the Glass data set.

We carried out a similar study for the Glass data set. For this data set, we are interested in identifying tableware glasses (outlier class) from building window glasses (inlier class). The study of glass-type identification was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence, if correctly identified. In our study, each time we took \( N_o = 5 \) random samples from the outlier class and \( N_i = 20 \) random samples from the inlier class to form a data set of size \( (D = 9) \times (N = 25) \). We applied the proposed IRW \( L_1 \)-PCA method with \( r = 2 \) \( L_1 \) principal components to the data set, and calculate the outlier scores for outlier detection with the same method used for the Satimage data set. The ROC curves are shown in Fig. 2b for the algorithms in comparison, and the corresponding AUC values are shown in table 1. Again we observe
that the proposed algorithm achieved the maximum true positive rate 1 at the lowest false positive rate 0.2, while the LODES and LOF schemes achieve the maximum true positive rate 1 at a higher false positive rate 0.4.

4. CONCLUSION

In the presented work, an iterative re-weighted $L_1$ principal-component analysis algorithm is developed to measure the compliance of data samples to the nominal low-rank structure. In each iteration, the $L_1$ subspace computed in the previous iteration is utilized to infer the conformity of each data sample, and the resultant conformity value is used to weigh the data sample, followed by $L_1$-subspace update. The procedure generates successively refined $L_1$-norm subspaces, which present significantly enhanced performance in tasks such as moving object extraction and outlier detection, compared to the regular $L_1$-PCA and state-of-the-art robust PCA methods.

REFERENCES