

Outlier Processing via L1-Principal Subspaces

Shubham Chamadia, Dimitris A. Pados

Department of Electrical Engineering, The State University of New York at Buffalo, NY 14260

E-mail: {shubhamc, pados}@buffalo.edu

Abstract

With the advent of big data, there is a growing demand for smart algorithms that can extract relevant information from high-dimensional large data sets, potentially corrupted by faulty measurements (outliers). In this context, we present a novel line of research that utilizes the robust nature of L_1 -norm subspaces for data dimensionality reduction and outlier processing. Specifically, (i) we use the euclidean-distance between original and L_1 -norm-subspace projected samples as a metric to assign weight to each sample point, (ii) perform ($K=2$)-means clustering over the *one-dimensional* weights discarding samples corresponding to the outlier cluster, and (iii) compute the robust L_1 -norm principal subspaces over the reduced “clean” data set for further applications. Numerical studies included in this paper from the fields of (i) data dimensionality reduction, (ii) direction-of-arrival estimation, (iii) image fusion, and (iv) video foreground extraction demonstrate the efficacy of the proposed outlier processing algorithm in designing robust low-dimensional subspaces from faulty high-dimensional data.

Introduction

In big-data applications, high-dimensional large datasets demand robust algorithms designed to extract meaningful information in the presence of potentially faulty/corrupted data entries. Traditionally, dimensionality reduction and outlier detection are treated separately. In this work, we provide a novel unified approach to effectively deal with outliers and high dimensionality simultaneously.

There is a wide range of applications engaging outlier detection approaches, for example cyber intrusion, video surveillance, and signal processing for wireless networks applications (Kumar 2005; Markou and Singh 2006; Sun, Xiao, and Wang 2007). Numerous techniques exist to address the outlier detection and processing challenges including statistical distance-based learning and clustering approaches (Knox and Ng 1998; Ramaswamy, Rastogi, and Shim 2000; Torr and Murray 1993) to name a few. Here, we focus on principal component analysis (PCA) (Pearson 1901), the celebrated unsupervised linear dimensionality reduction technique. The objective of PCA is to find orthonormal basis vectors/subspaces that capture the maximum pos-

sible variation of the original data. The prominent PCA algorithmic solution comes from the well known L_2 -norm based singular value decomposition (SVD) of the data matrix or equivalently, eigen-value decomposition of the data covariance matrix. However, it is widely known that L_2 -norm guided processing is highly sensitive to the presence of outliers. Recently, preferred solutions focus on pursuing L_1 -norm based principal components design involving absolute-value analysis that imparts significantly less emphasis on the extreme outliers as compared to squared-value L_2 -based solutions. A few algorithms involving L_1 -norm calculations have been proposed for robust subspace design (Ding et al. 2006; Kundu, Markopoulos, and Pados 2014; Kwak 2008; Markopoulos, Karystinos, and Pados 2014; Markopoulos et al. 2016). (Markopoulos, Karystinos, and Pados 2014) presented the first known exact algorithms for the computation of L_1 -norm principal components of data.

In this work, we propose a new outlier processing approach by utilizing the robust L_1 -norm subspace designs. In particular, (i) we compute L_1 -principal subspaces of the given data set, (ii) assign a weight to each sample using L_1 -subspace distance metrics, (iii) implement traditional K -means clustering over the *one-dimensional* weight space to discard samples corresponding to the outlier cluster, and (iv) finally, over the reduced *cleaned* dataset recompute robust L_1 subspaces for applications thereafter. Detailed numerical studies involving applications in dimensionality reduction, direction-of-arrival estimation of an unknown signal in the presence of heavy intermittent jammers, image restoration from corrupted copies (image fusion), and online video foreground tracking presented in this paper demonstrate the effectiveness of the proposed robust L_1 -norm integrated outlier processing algorithm.

Problem Formulation & Motivation

Consider N real-valued measurements, each of dimension D that form the data matrix $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_N] \in \mathbb{R}^{D \times N}$. The classical L_2 -norm-based PCA problem can be mathematically formulated as

$$\mathbf{Q}_{L_2} = \arg \max_{\mathbf{Q} \in \mathbb{R}^{D \times P}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_P} \|\mathbf{X}^T \mathbf{Q}\|_2 \quad (1)$$

where \mathbf{Q}_{L_2} represents the P -rank orthonormal subspace ($P \leq \text{rank}(\mathbf{X})$ L_2 -norm principal components) and $\|\mathbf{A}\|_2 =$

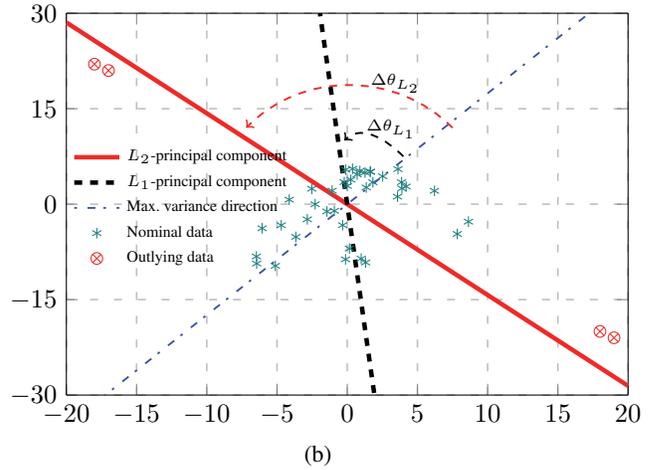
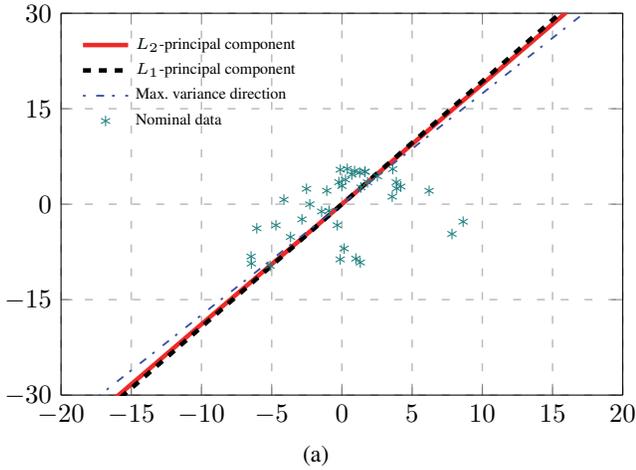


Figure 1: (a) Principal component over the original clean data matrix $\mathbf{X}_{2 \times 40}$ (*). (b) Principal component over data matrix $\mathbf{X}_{2 \times 40}$ corrupted by 4 appended outliers (\otimes) (angular deviation $\Delta\theta_{L_2}=95^\circ$ and $\Delta\theta_{L_1}=63^\circ$). As a benchmark, in both figures we plot the ideal maximum-variance direction of the clean-data distribution (dominant eigenvector of true nominal-data auto-covariance matrix).

$\sqrt{\sum_{i,j} |A_{i,j}|^2}$ is the L_2 -norm (Frobenius norm) of matrix \mathbf{A} with elements $A_{i,j}$. The optimal solution to (1) is well known to be given by the P dominant left-singular vectors of the data matrix \mathbf{X} or equivalently the P dominant eigenvectors of $\mathbf{X}^T \mathbf{X}$.

Regrettably, L_2 -norm based process by (1) is highly sensitive to the presence of faulty measurements (outliers). One natural approach to overcome this drawback is to pursue L_1 -norm based principal-component analysis which modifies the problem in (1) to the new problem

$$\mathbf{Q}_{L_1} = \arg \max_{\mathbf{Q} \in \mathbb{R}^{D \times P}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_P} \|\mathbf{X}^T \mathbf{Q}\|_1 \quad (2)$$

where \mathbf{Q}_{L_1} represents the P ($\leq \text{rank}(\mathbf{X})$) orthonormal L_1 -principal components of \mathbf{X} and $\|\mathbf{A}\|_1 = \sum_{i,j} |A_{i,j}|$ for any matrix \mathbf{A} . Due to the non-convex nature of the problem in (2), it is challenging to calculate the optimal L_1 -principal components. The literature serves us with a few sub-optimal (Ding et al. 2006; Kundu, Markopoulos, and Pados 2014; Kwak 2008; Markopoulos et al. 2016) and recently proposed optimal (Markopoulos, Karystinos, and Pados 2014) algorithms to design the L_1 -principal subspaces.

To illustrate the importance of solving (2) and motivate the need for further research on the robustness of L_1 -subspaces, we consider the following simple example. We generate a 2D-data matrix $\mathbf{X}_{2 \times 40}$ whose 40 data points (*) are drawn from the Gaussian distribution $\mathcal{N}\left(\mathbf{0}_2, \begin{bmatrix} 15 & 12 \\ 12 & 29 \end{bmatrix}\right)$. We calculate and plot the optimal L_2 (SVD) and optimal L_1 (Markopoulos, Karystinos, and Pados 2014) principal component of \mathbf{X} in Figure 1(a). For reference, we also plot the ideal maximum-variance direction, i.e., the direction of the dominant eigenvector of the auto-covariance matrix $\begin{bmatrix} 15 & 12 \\ 12 & 29 \end{bmatrix}$. Next, we add 4 outlier points (\otimes) to \mathbf{X} and recalculate the L_2 and L_1 principal component

as shown in Figure 1(b). It can be observed that under the nominal data scenario in Figure 1(a), both the L_1 and L_2 principal components follow a close line along the ideal direction. However, in the presence of the outlying data points in Figure 1(b), the L_2 principal direction steers almost orthogonally to the true direction, i.e. $\Delta\theta_{L_2}=95^\circ$, and the L_1 -principal direction shifts by $\Delta\theta_{L_1}=63^\circ$. This visual analysis clearly exhibits the limitation of L_2 -PCA (and to some lesser extent of direct L_1 -PCA) in combating outlying data. In the following section, we explain how L_1 -PCA can be used to excise outlying data and lead to a most robust subspace design.

Proposed Algorithm

In this section, we discuss step by step the proposed integrated outlier processing and robust subspace design approach via L_1 -principal component analysis. No prior information about the data statistics or the number of outliers is assumed. For the benefit of the readers, we also outline the pseudo-code of the proposed scheme in Figure 2.

Step 1: Obtain L_1 -principal components $(\mathbf{Q}_{L_1})_{D \times P}$ of data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$

Given a real-valued data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ (possibly corrupted), we calculate P ($\leq \min(D, N)$) L_1 -principal components by

$$\mathbf{Q}_{L_1} = \arg \max_{\mathbf{Q} \in \mathbb{R}^{D \times P}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_P} \|\mathbf{X}^T \mathbf{Q}\|_1.$$

A computationally light-weight approach to approximate \mathbf{Q}_{L_1} would be, for instance, the fast iterative suboptimal algorithm of (Kundu, Markopoulos, and Pados 2014) to evaluate the first L_1 -principal component of the data and design the rest $P-1$ principal components via sequentially replacing \mathbf{X} by its projection onto the nullspace of the

Input: data matrix $\mathbf{X} \in \mathbb{R}^{(D \times N)}$, # of PCs. = P

- 1: $\mathbf{Q}_{L_1} = \text{L1-PCs}(\mathbf{X}, P)$
- 2: $\bar{w}_i = \|\mathbf{x}_i - \mathbf{Q}_{L_1} \mathbf{Q}_{L_1}^T \mathbf{x}_i\|_2^2, \forall i \in \{1, 2, \dots, N\}$
 $w_i = \frac{\bar{w}_i}{\sum_{i=1}^N \bar{w}_i}, \mathbf{w} = [w_1, w_2, \dots, w_N]$
- 3: K -means cluster(\mathbf{w}) = $[\mu_h, \mu_l]$
 $\mathcal{I} = \{i\} : i \in \text{extract_indices}(\mathbf{w}, \mu_h)$
 $\mathbf{X}(:, \mathcal{I}) = \emptyset$ and $\mathbf{X}_{clean} \leftarrow \mathbf{X}$
- 4: $\mathbf{Q}_{L_1}^{clean} = \text{L1-PCs}(\mathbf{X}_{clean}, P)$

Output: P robust L_1 -subspace $\mathbf{Q}_{L_1}^{clean}$

Function: $\text{L1-PCs}(\mathbf{X}, P)$

Input: $\mathbf{X}_{D \times N}, P$

for $p = 1 : P$

$\mathbf{q}_{L_1} = \arg \max_{\mathbf{q} \in \mathbb{R}^D, \|\mathbf{q}\|_2=1} \|\mathbf{X}^T \mathbf{q}\|_1$, using (Kundu, et al. 2014)

$\mathbf{X} = \mathbf{X} - \mathbf{q}_{L_1} \mathbf{q}_{L_1}^T \mathbf{X}$

$\mathbf{Q}_{L_1}(:, p) = \mathbf{q}_{L_1}$

end

Output: \mathbf{Q}_{L_1}

Figure 2: Algorithm to detect/remove outliers and compute $P \leq \min(D, N)$ robust L_1 -principal components for a real data matrix $\mathbf{X}_{D \times N}$.

previously calculated components. Figure 2, function: L1-PCs presents the details.

Step 2: Evaluate reliability weight

We calculate the reliability weight corresponding to each sample using the previously designed L_1 -principal components as

$$\bar{w}_n = \|\mathbf{x}_n - \mathbf{Q}_{L_1} \mathbf{Q}_{L_1}^T \mathbf{x}_n\|_2^2 \quad \forall n \in \{1, 2, \dots, N\}. \quad (3)$$

In signal processing terminology, the weight \bar{w}_n is often referred to as the rank- P reconstruction error for the n^{th} sample. We further normalize each weight \bar{w}_n in (3) by

$$w_n = \frac{\bar{w}_n}{\sum_{i=1}^N \bar{w}_i} \quad (4)$$

and collect the normalized weights in a single reliability weight vector \mathbf{w}

$$\mathbf{w} \triangleq [w_1, w_2, \dots, w_N]. \quad (5)$$

The normalized weights will be used as a metric to judge the quality of each available sample. Ideally, a corrupted sample will lie far away from the L_1 -principal directions exhibiting a high relative reconstruction error value w_n .

Step 3: Outlier removal via clustering

Figure 3: Average representation error versus percentage of outlying data (total number of samples $N=300$, $P = 2$ principal components).

There are many ways to detect and remove the outlying samples. For instance, if the algorithm has a priori knowledge of the number of corrupted samples, say t , we can simply look for the t -highest weights in (5) and remove the corresponding samples. However, in real-world applications it is quite rare to have such a priori knowledge about the number of existing outliers, especially in high-dimensional datasets. Therefore, we design an intuitive clustering method for separating the nominal samples from likely outliers. In particular, we implement conventional ($K=2$)-means clustering over the weight space which will furnish two clusters and their corresponding means as

$$K\text{-means cluster}(\mathbf{w}) = [\mu_h, \mu_l]. \quad (6)$$

The cluster having lower mean value (μ_l) would retain the true data samples whereas the cluster with higher mean (μ_h) is expected to contain outlier samples which would be subsequently discarded. In this context, we define the function $\text{extract_indices}(\mathbf{w}, \mu_h)$ that inputs the weight vector \mathbf{w} and the higher cluster mean μ_h and outputs the indices of the weights that are contained in the cluster with mean (μ_h),

$$\mathcal{I} = \{i : i \in \text{extract_indices}(\mathbf{w}, \mu_h)\}.$$

The samples corresponding to indices contained in index set \mathcal{I} are collected and removed from the original corrupted data \mathbf{X} to obtain the ‘‘clean’’ data set \mathbf{X}_{clean} ,

$$\mathbf{X}(:, \mathcal{I}) = \emptyset \text{ and } \mathbf{X}_{clean} \leftarrow \mathbf{X}.$$

It is important to highlight that to separate nominal from outlier data, K -means clustering ($K=2$) is performed over the N scalar weights and not over the original high D -dimensional data which would be computationally very expensive.

Step 4: Recalculate L_1 -principal components ($\mathbf{Q}_{L_1}^{clean}$) over outlier processed/removed data \mathbf{X}_{clean}

The final step involves recalculating the L_1 -principal components (repeating Step 1 only) over the obtained ‘‘cleaned’’ data set \mathbf{X}_{clean} by

$$\mathbf{Q}_{L_1}^{clean} = \arg \max_{\mathbf{Q} \in \mathbb{R}^{D \times P}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_P} \|\mathbf{X}_{clean}^T \mathbf{Q}\|_1.$$

It is expected that the so designed $\mathbf{Q}_{L_1}^{clean}$ principal subspaces reveal a deeper insight into the given data matrix than the original \mathbf{Q}_{L_1} (or \mathbf{Q}_{L_2}) principal subspaces.

Experimental Studies

In this section, we carry out experimental studies drawn from the research fields of data dimensionality reduction, direction-of-arrival estimation, robust image fusion, and video foreground modeling. Here, we compare (i) conventional L_2 subspaces (SVD), (ii) L_1 subspaces (Kundu, Markopoulos, and Pados 2014), (iii) the proposed outlier-processed robust subspace denoted by $L_1 + L_1$ (the term is motivated by the computation of L_1 subspaces twice), and (iv) the analogous L_2 version of (iii) denoted by $L_2 + L_2$.

Experiment 1: Data dimensionality reduction

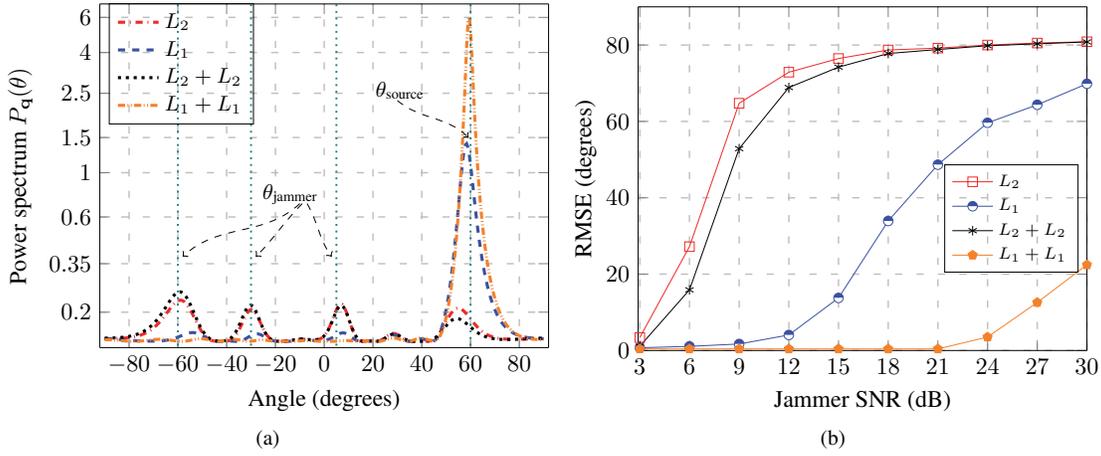


Figure 4: MUSIC analysis: (a) Instantaneous spectrum with jammers located at $\theta_{\text{jammer}} = \{-60^\circ, -30^\circ, 5^\circ\}$ and signal of interest at $\theta_{\text{source}} = 60^\circ$ (b) Root-mean-square-error (RMSE) versus jammer SNR.

We generate a real-valued data matrix $\mathbf{X} \in \mathbb{R}^{25 \times 300}$ whose 300 observation points have independent, identically distributed coordinates drawn from the Gaussian distribution $\mathcal{N}(0, 4)$. We then arbitrarily select and contaminate a certain percentage of samples by adding an independent additive white Gaussian noise outlier vector $\mathbf{v} \in \mathbb{R}^{300 \times 1}$ whose entries are drawn from the high variance Gaussian distribution $\mathcal{N}(0, 35)$. We denote the corrupted data matrix by \mathbf{X}_{crpt} . Our performance metric is the true data representation error

$$RE = \|\mathbf{X} - \mathbf{Q}\mathbf{Q}^T \mathbf{X}\|_2^2$$

averaged over $M = 2^{14}$ independent experiments. In each experiment, we seek $P=2$ principal components of \mathbf{X}_{crpt} by the competing algorithms.

In Figure 3, we plot RE as a function of the percentage of outlying data. As expected, the proposed $L_1 + L_1$ outlier processing scheme exhibits lowest reconstruction error over the whole range of corruption percentage.

Experiment 2: Direction-of-Arrival (DoA) estimation

We consider a uniform linear antenna array of $D = 7$ elements. The array collects $N = 30$ observations of a binary phase-shift-keying (BPSK) signal arriving at an angle $\theta_1 = 60^\circ$ in the presence of additive white complex Gaussian noise $\mathbf{v} \in \mathbb{C}^{7 \times 1}$,

$$\mathbf{x}_n = A_1 b_n \mathbf{s}_{\theta_1} + \mathbf{v}_n, \quad n = 1, 2, \dots, 30, \quad (7)$$

where \mathbf{s}_{θ_1} is the array response vector, $b_n \in \{\pm 1\}$ is the Bernoulli equiprobable information bit, $A_1 > 0$, and the signal-to-noise ratio (SNR) is set at $\text{SNR}_1 = 3$ dB. We further assume that any 3 observations (out of 30) are corrupted by 3 jammers having $\text{SNR}_j = 9$ dB and angle of arrival $-60^\circ, -30^\circ$ and 5° . We call the resulting corrupted observation data matrix $\mathbf{X}_{\text{crpt}} \in \mathbb{C}^{7 \times 30}$ and transform \mathbf{X}_{crpt} by concatenation to its real-domain version $\mathbf{X}'_{\text{crpt}} = \begin{bmatrix} \text{real}(\mathbf{X}_{\text{crpt}}) \\ \text{imag}(\mathbf{X}_{\text{crpt}}) \end{bmatrix} \in \mathbb{R}^{14 \times 30}$.

We then calculate and plot in Figure ??, the MUSIC-type DoA estimation spectrum function (Schmidt 1986)

$$P_{\mathbf{q}}(\theta) = \frac{1}{\mathbf{s}_{\theta}^T (\mathbf{I} - \mathbf{q}\mathbf{q}^T) \mathbf{s}_{\theta}}, \quad \theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right),$$

where $\mathbf{q} \in \mathbb{R}^{14 \times 1}$ represents the single principal component of $\mathbf{X}'_{\text{crpt}}$ produced by the competing algorithms and $\mathbf{s}_{\theta} \in \mathbb{R}^{14 \times 1}$ is the concatenated real-valued version of the array response scanning vector. As observed, the L_2 -based schemes completely steer toward the jammers, whereas the proposed $L_1 + L_1$ scheme is virtually unaffected indicating the true active signal direction of arrival.

In Figure ??, for the same experiment, we plot the root-mean-squared-error $\text{RMSE} = \sqrt{\frac{1}{M} \sum_{m=1}^M (\theta_1 - \hat{\theta}^{(m)})^2}$ as a function of the jammer SNR. Here, $\hat{\theta}^{(m)}$ is the estimated angle (highest peak of $P_{\mathbf{q}}(\theta)$) in the m^{th} experiment and $\theta_1 = 60^\circ$ is the true fixed source angle. The jammers' location θ_j is chosen independently and uniformly in $\theta_j \in (-\frac{\pi}{2}, \frac{\pi}{2})$. $M = 2^{14}$ independent experiments are conducted. It is again interesting to observe the superiority of the proposed $L_1 + L_1$ scheme in estimating the true angle of arrival.

Experiment 3: Robust image fusion

We consider $N = 10$ copies of the grayscale *Lenna* image (256×256) in Figure 5(a), each of which is corrupted per pixel by i.i.d. zero mean AWGN noise of variance $\sigma^2 = 50$ and further 8-bit quantized (Figure 5(b) example). Next, we arbitrarily choose 8 (out of the 10) noisy images and overwrite 40% of pixels by salt-and-pepper corruption as in Figure 5(c). Finally, to make image recovery even more challenging, we append to the data set the 256×256 grayscale *baboon* image in Figure 5(d) as an extreme outlier. It is important to note that none of the tested algorithms has knowledge of the type of corruption or the presence of a mislabeled (baboon) image.

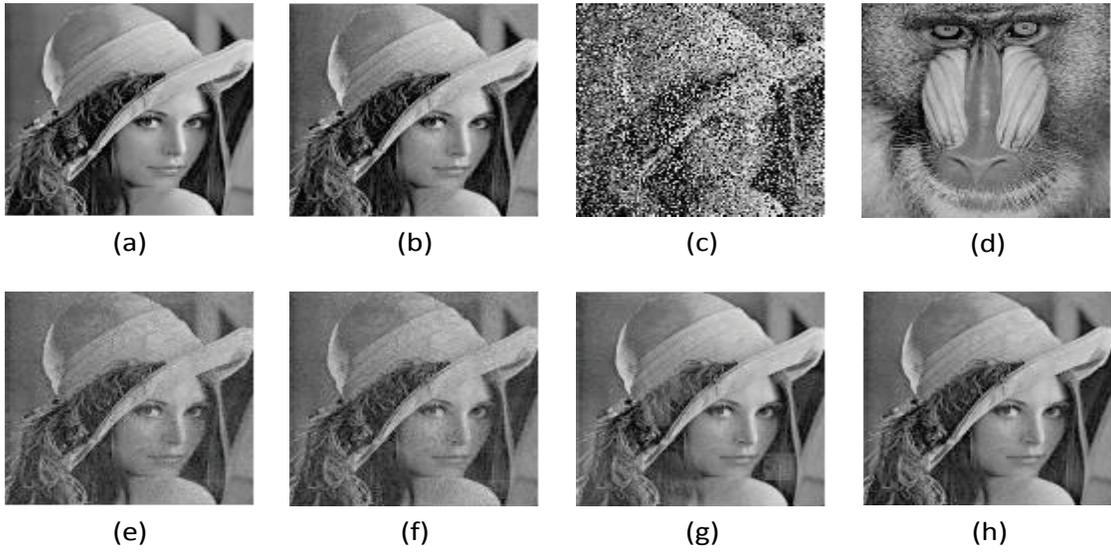


Figure 5: The image restoration via single principal component over reconstruction block size (32×32). (a) original *Lenna* image (256×256); (b) noisy instance; (c) instance corrupted (40% pixels) by “salt and pepper” corruption; (d) original *baboon* image (outlier); (e) optimal L_1 fused (Markopoulos, Kundu, and Pados 2015); (f) optimal L_2 fused; (g) $L_2 + L_2$ fused; (h) proposed $L_1 + L_1$ fused.

We analyze the resulting set of 11 images patch-wise, i.e. divide each image into squared patches of dimension $d \times d$ ($d = 32$, $\frac{256 \times 256}{32 \times 32} = 64$ patches per image) and form the data matrix \mathbf{X}^p as

$$\mathbf{X}^p = [\mathbf{i}_1^p, \mathbf{i}_2^p, \dots, \mathbf{i}_{11}^p]_{(1024 \times 11)} \quad (8)$$

where $\mathbf{i}_n^p = \text{vec}(\mathbf{I}_n^p)$ with \mathbf{I}_n^p being the p^{th} -patch of n^{th} -image. We extract the single principal component \mathbf{q}^p of \mathbf{X}^p and use it to evaluate the normalized reliability weight r_n^p corresponding to each vectorized patch sample

$$\bar{r}_n^p = \|\mathbf{i}_n^p - \mathbf{q}^p (\mathbf{q}^p)^T \mathbf{i}_n^p\|_2^{-2}$$

and then normalize $r_n^p = \frac{\bar{r}_n^p}{\sum_{n=1}^{11} \bar{r}_n^p}$. (9)

Finally, via weighting, we fuse the 11 patches to restore $\hat{\mathbf{I}}^p$ as (Markopoulos, Kundu, and Pados 2015)

$$\hat{\mathbf{I}}^p = \sum_{n=1}^{11} r_n^p \mathbf{I}_n^p. \quad (10)$$

Figure 5(e) shows an instance of the restored image. Figure 5(f) shows the outcome when the L_2 principal component is used instead. Figures 5(g) and (h) show the outcome upon outlier excision by the $L_2 + L_2$ and the proposed $L_1 + L_1$ procedure. The visual comparative success of $L_1 + L_1$ is striking. In Figure 6, we also plot the peak signal-to-noise ratio (PSNR) of the restored image as a function of the percentage of corrupted pixels.

Experiment 4 - Online foreground tracing over surveillance video

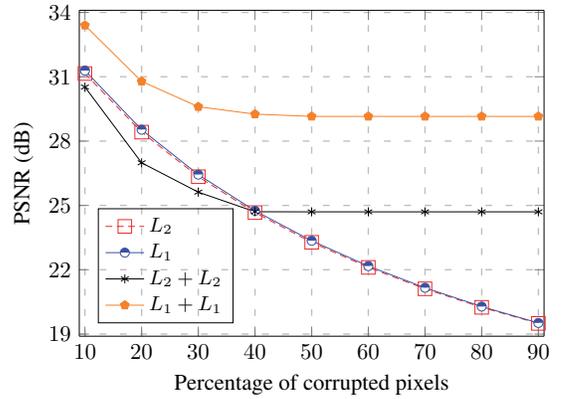


Figure 6: IPSNR of restored *Lenna* image with varying percentage of corrupted pixels.

Detection of anomalies and moving objects in a frame is the essence of surveillance video processing. Here, we consider the popular dataset “Hall of a business building” (grayscale 144×176 frames) often used for foreground modeling. We begin from frame number 1000 and collect and vectorize the next 150 frames for training purpose. Pursuing *online* background processing (Pierantozzi et al. 2016), we add 1 incoming frame to our previous frame set and discard frames¹ which are more likely to contain moving objects by implementing our proposed scheme with 2 principal components. To render foreground moving objects, we subtract frames from their low dimensional subspace representation

¹At any realization, frame excision is performed only if the number of frames is greater than 10.



Figure 7: Hallway sequence : (a) original frames; (b) L_2+L_2 reconstructed foreground; (c) proposed L_1+L_1 reconstructed foreground.

(background) using the extracted principal components.

In Figure. 7, we show the extracted foreground by L_2+L_2 and L_1+L_1 ; L_1+L_1 is significantly superior.

Conclusion

In this paper, we developed a simple, yet effective unsupervised outlier processing algorithm for robust feature extraction from faulty high-dimensional data. Motivated by the robust nature of the L_1 -subspace, we integrated L_1 -subspace design with sample weighting and ($K=2$)-means cluster sample excision over the scalar weights domain to discard “corrupted” samples and design a robust low-dimensional subspace over the “cleaned” dataset. Extensive numerical studies were carried out to compare the proposed robust subspace design against conventional L_2 and L_1 -based schemes. The proposed algorithm showed excellent performance in designing robust subspaces over heavily corrupted data sets with end applications in data dimensionality reduction, direction-of-arrival estimation, image fusion, and video foreground extraction.

Acknowledgment

The authors would like to thank Dr. Sandipan Kundu and Ms. Colleen Bailey for their valuable comments and suggestions on this manuscript.

References

- Ding, C.; Zhou, D.; He, X.; and Zha, H. 2006. R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. In *Proceedings ICML*, 281–288.
- Knox, E. M., and Ng, R. T. 1998. Algorithms for mining distancebased outliers in large datasets. In *Proceedings VLDB*, 392–403.
- Kumar, V. 2005. Parallel and distributed computing for cybersecurity. *Distrib. Syst. Online, IEEE* 6(10).

Kundu, S.; Markopoulos, P. P.; and Pados, D. A. 2014. Fast computation of the L1-principal component of real-valued data. In *Proceedings ICASSP*, 8028–8032.

Kwak, N. 2008. Principal component analysis based on L1-norm maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(9):1672–1680.

Markopoulos, P. P.; Kundu, S.; Chamadia, S.; and Pados, D. A. 2016. Efficient L1-norm principal-component analysis via bit flipping. *arXiv preprint arXiv:1610.01959*.

Markopoulos, P. P.; Karystinos, G. N.; and Pados, D. A. 2014. Optimal algorithms for L1-subspace signal processing. *IEEE Trans. Signal Process.* 62(19):5046–5058.

Markopoulos, P. P.; Kundu, S.; and Pados, D. A. 2015. L1-fusion: Robust linear-time image recovery from few severely corrupted copies. In *Proceedings ICIP*, 1225–1229.

Markou, M., and Singh, S. 2006. A neural network-based novelty detector for image sequence analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(10):1664–1677.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Lond. Phil. Mag. and J. Sci.* 2(11):559–572.

Pierantozzi, M.; Liu, Y.; Pados, D. A.; and Colonnese, S. 2016. Video background tracking and foreground extraction via L1-subspace updates. In *Proceedings SPIE*.

Ramaswamy, S.; Rastogi, R.; and Shim, K. 2000. Efficient algorithms for mining outliers from large data sets. In *ACM Sigmod Record*, 427–438.

Schmidt, R. 1986. Multiple emitter location and signal parameter estimation. *IEEE Antennas Propag. Mag.* 34(3):276–280.

Sun, B.; Xiao, Y.; and Wang, R. 2007. Detection of fraudulent usage in wireless networks. *IEEE Trans. Veh. Tech.* 56(6):3912–3923.

Torr, P. H., and Murray, D. W. 1993. Outlier detection and motion segmentation. In *Proceedings SPIE*, 432–443.