

COMBINATORIAL SEARCH FOR THE L_p-NORM PRINCIPAL COMPONENT OF A MATRIX

*Dimitris G. Chachlakis and Panos P. Markopoulos**

Dept. of Electrical and Microelectronic Engineering
Rochester Institute of Technology, Rochester NY, USA
E-mail: dimitris@mail.rit.edu and panos@rit.edu

ABSTRACT

We study L_p-norm Principal-Component Analysis (L_p-PCA) of a matrix. For $p = 2$ (standard PCA), the problem can be solved with standard Singular-Value Decomposition (SVD). For $p = 1$ (L1-PCA), the problem was recently solved exactly and approximately with efficient iterative algorithms. For general values of p , the exact solution to L_p-PCA remains to date unknown. In this work, for the first time in the literature, we prove that, for $p \leq 1$, L_p-PCA can be solved exactly through combinatorial optimization and present the first exact solver. Our experimental studies on medical data demonstrate the significant robustness of L_p-PCA, $p \leq 1$, against outliers.

Index Terms— Principal-Component Analysis, PCA, L1-PCA, L_p-norm, L_p-PCA, outliers, robustness.

1. INTRODUCTION

Principal-Component Analysis (PCA) is a fundamental method for data analysis with a plethora of signal processing and machine learning applications [1]. PCA is typically formulated as a L2-norm error minimization, or, equivalently, a L2-norm projection maximization problem. Mathematically, for given matrix $\mathbf{X}$ of size D -by- N , PCA seeks a D -dimensional unit-norm vector $\mathbf{q}$ that minimizes $\|\mathbf{X} - \mathbf{q}\mathbf{q}^T \mathbf{X}\|_2^2$, or equivalently, maximizes $\|\mathbf{X}^T \mathbf{q}\|_2^2$, where the squared L2-norm $\|\cdot\|_2^2$ returns the sum of the squared entries of its argument. The solution to PCA can be obtained by Singular-Value Decomposition (SVD) of $\mathbf{X}$ [2]. Despite its documented success, PCA is known to exhibit severe sensitivity against outliers within the processed data [3, 4]. Outliers are high-magnitude/peripheral data points that lie far away from the nominal data subspace and commonly appear in modern datasets, e.g., due to data storage/transfer errors, faulty sensors, or deliberate data contamination in adversarial environments [5]. The outlier-sensitivity of PCA can be attributed to its L2-norm-based formulation which places squared emphasis to each data point, thus promoting the impact of outliers. Accordingly, applications that rely on

PCA are severely affected when outliers exist in the processed data [6]. To counteract the impact of outliers, researchers have proposed an array of “robust” PCA formulations [6, 7]. One approach considers a L1-norm based residual-error minimization formulation the (approximate) solution to which is computed by means of alternating-optimization [8]. A more straightforward approach considers the projection maximization PCA formulation and replaces the outlier-sensitive L2-norm by the more robust L1-norm. In this way, linear emphasis is placed on each data point, resulting in the popular L1-norm PCA (L1-PCA) formulation. L1-PCA was solved exactly in [9, 10]. Efficient [11, 12], adaptive [13], incremental [14], stochastic [15], and complex-valued data [16, 17] solvers have also been proposed. L1-PCA has been successfully employed in an array of applications [18–22] where it has been documented that it attains similar performance to PCA when the processed data are nominal/clean while it exhibits strong resistance against outliers. L1-norm formulations have recently been proposed for robust tensor analysis (e.g., L1-Tucker [23–29] and L1-Rescal [30]). Similar to L1-PCA, L1-Tucker and L1-Rescal exhibited significant robustness against data corruptions.

Apart from $p = 2$ (standard PCA) and $p = 1$ (L1-PCA), the exact solution to L_p-PCA remains unknown. Motivated by the documented success of L1-PCA, in this work, we study L_p-PCA for $p \leq 1$, expecting that further decreasing p , and thus the emphasis per training sample, can increase the robustness against outliers. Specifically, we show for the first time that the exact solution to L_p-PCA can be found by means of combinatorial optimization and present the first exact algorithm for its solution. Our experimental studies on matrix reconstruction and medical-data classification corroborate the outlier resistance of L_p-PCA.

2. PROBLEM STATEMENT

Consider $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$ with $\text{rank}(\mathbf{X}) = \beta \leq \min\{D, N\}$. For any $p > 0$, L_p-PCA is formulated as

$$\max_{\mathbf{q} \in S_D} \|\mathbf{X}^T \mathbf{q}\|_p^p, \quad (1)$$

*Corresponding author.

This research is supported in part by the U. S. National Science Foundation, under grant OAC-1808582.

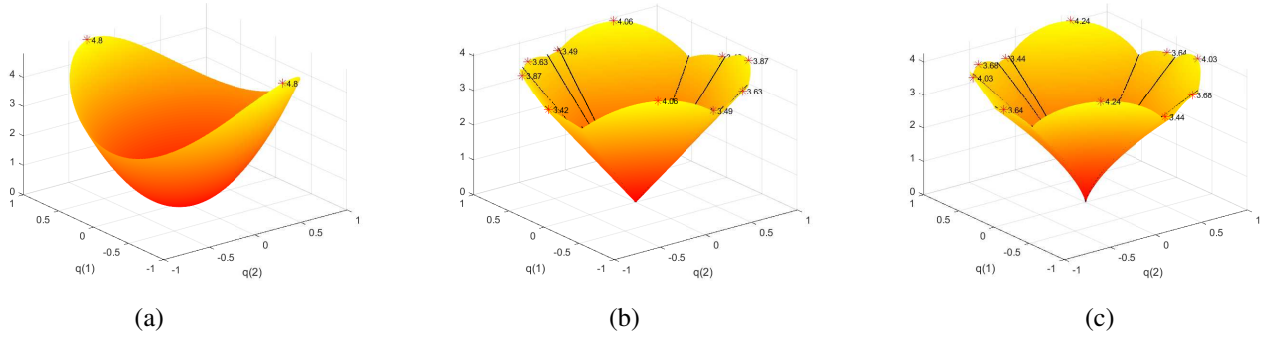


Fig. 1. Visual illustration of $f_p(\mathbf{q}) = \|\mathbf{X}^\top \mathbf{q}\|_p^p$, $\mathbf{q} \in \mathcal{B}_D$, for (a) $p = 2$, (b) $p = 1$, and (c) $p = 0.75$.

where, for any $\mathbf{x} \in \mathbb{R}^D$, $\|\mathbf{x}\|_p := \left(\sum_{d=1}^D |\mathbf{x}_d|^p\right)^{\frac{1}{p}}$ and $\mathcal{S}_D := \{\mathbf{q} \in \mathbb{R}^D : \|\mathbf{q}\|_2 = 1\}$. To date, Lp-PCA has been solved exactly only for $p = 2$ and $p = 1$. For general p , the exact solution to (1) remains unknown. For $p > 1$, only approximate solvers exist that provably converge to local maxima [31]. For $p < 1$, there exist neither exact solvers nor converging iterative approximators. In the sequel, we briefly review the special case of $p = 1$ (L1-PCA). Then, we present new theory on Lp-PCA ($p \leq 1$) that enables us to compute its exact solution.

2.1. L1-PCA BACKGROUND

For the special case of $p = 1$, (1) simplifies to the popular L1-PCA formulation [10]:

$$\max_{\mathbf{q} \in \mathcal{S}_D} \|\mathbf{X}^\top \mathbf{q}\|_1. \quad (2)$$

The exact solution to L1-PCA was presented in [10], where it was shown that if $\mathbf{b}_{\text{opt}}$ is a solution to $\max_{\mathbf{b} \in \{\pm 1\}^N} \|\mathbf{X}\mathbf{b}\|_2$, then, the exact solution to (2) is given by $\mathbf{q}_{\text{L1}} = \omega(\mathbf{X}\mathbf{b}_{\text{opt}})$, where $\omega(\mathbf{x}) := \mathbf{x} \|\mathbf{x}\|_2^{-1}$ for any $\mathbf{x} \in \mathbb{R}^D$. Accordingly, the computational cost for solving L1-PCA exactly was found to be $\mathcal{O}(N^\beta)$ [10]. For practical applications, an efficient algorithm that approximates $\mathbf{q}_{\text{L1}}$ with cost $\mathcal{O}(DN \min\{D, N\} + N^2\beta)$, comparable to that of SVD, was offered in [11]. Motivated by the outlier-robustness of L1-PCA, in the sequel, we study Lp-norm PCA for $p \leq 1$, which places *sub-linear* ($p < 1$) emphasis to each data point of $\mathbf{X}$ compared to the *linear* emphasis of L1-PCA and *squared* emphasis of standard PCA.

3. PROPOSED EXACT SOLUTION

We consider $p \leq 1$ and commence our developments with Lemma 1. The proofs of Lemma 1 and all subsequently presented Lemmas are omitted due to lack of space.

Lemma 1. For any $\mathbf{X} \in \mathbb{R}^{D \times N}$, it holds that

$$\max_{\mathbf{q} \in \mathcal{S}_D} \|\mathbf{X}^\top \mathbf{q}\|_p^p = \max_{\mathbf{q} \in \mathcal{B}_D} \|\mathbf{X}^\top \mathbf{q}\|_p^p, \quad (3)$$

where $\mathcal{B}_D := \{\mathbf{q} \in \mathbb{R}^D : \|\mathbf{q}\|_2 \leq 1\}$.

In view of Lemma 1, the solution to Lp-PCA in (1) can equivalently be pursued over the closed unit-radius hyperball. A visual illustration of $f_p(\mathbf{q}) = \|\mathbf{X}^\top \mathbf{q}\|_p^p$, $\mathbf{q} \in \mathcal{B}_D$, is offered in Fig. 1 for arbitrary matrix¹ $\mathbf{X}$ and $p \in \{2, 1, 0.75\}$. Next, for every $\mathbf{b} \in \mathcal{A} := \{\pm 1\}^N$, we define

$$\mathcal{P}(\mathbf{b}) := \{\mathbf{q} \in \mathcal{B}_D : \text{sgn}(\mathbf{X}^\top \mathbf{q}) = \mathbf{b}\} \quad (4)$$

and notice that there exist instances of $\mathbf{b} \in \mathcal{A}$ for which $\mathcal{P}(\mathbf{b}) = \emptyset$. Moreover, the following Lemma 2 holds true.

Lemma 2. For every $\mathbf{b}, \mathbf{b}' \in \mathcal{A}$ such that $\mathbf{b} \neq \mathbf{b}'$, it holds that $\mathcal{P}(\mathbf{b}) \cap \mathcal{P}(\mathbf{b}') = \emptyset$.

We define $\mathcal{N}(\mathbf{X}) := \{\mathbf{q} \in \mathcal{B}_D : \exists n \in [N] \text{ such that } \mathbf{q}^\top \mathbf{x}_n = 0\}$, where $[N] := \{1, 2, \dots, N\}$. By Lemma 2 and the definition of $\mathcal{N}(\mathbf{X})$, the following Lemma 3 straightforwardly follows.

Lemma 3. It holds that $\mathcal{B}_D = \bigcup_{\mathbf{b} \in \mathcal{A}} \mathcal{P}(\mathbf{b}) \cup \mathcal{N}(\mathbf{X})$.

Lemma 3 states that, for any $\mathbf{X}$, the closed unit-radius hyperball can be partitioned in a finite number of non-overlapping sets. Moreover, the following Lemma 4 holds.

Lemma 4. Let $\mathbf{q}_{\text{Lp}}$ denote the exact solution to (3) for $p \leq 1$. Then, $\mathbf{q}_{\text{Lp}} \notin \mathcal{N}(\mathbf{X})$.

In view of Lemmas 1-4, we find

$$\max_{\mathbf{q} \in \mathcal{S}_D} \|\mathbf{X}^\top \mathbf{q}\|_p^p = \max_{\mathbf{q} \in \mathcal{B}_D} \|\mathbf{X}^\top \mathbf{q}\|_p^p \quad (5)$$

$$= \max_{\mathbf{q} \in \bigcup_{\mathbf{b} \in \mathcal{A}} \mathcal{P}(\mathbf{b}) \cup \mathcal{N}(\mathbf{X})} \|\mathbf{X}^\top \mathbf{q}\|_p^p \quad (6)$$

$$= \max_{\mathbf{q} \in \bigcup_{\mathbf{b} \in \mathcal{A}} \mathcal{P}(\mathbf{b})} \|\mathbf{X}^\top \mathbf{q}\|_p^p. \quad (7)$$

By Lemma 2, the solution to (7) can be pursued by combinatorial search over $\mathcal{A}$: i.e., by separately solving

$$\max_{\mathbf{q} \in \mathcal{P}(\mathbf{b})} \|\mathbf{X}^\top \mathbf{q}\|_p^p \quad (8)$$

for each $\mathbf{b} \in \mathcal{A}$ for which $\mathcal{P}(\mathbf{b}) \neq \emptyset$. The question of interest now is how to solve (8). Below we show that the solution to

$$\mathbf{X} = \begin{bmatrix} 1.7876 & -0.1967 & 0.9107 & 0.0728 & 0.6752 \\ -0.8204 & -0.8901 & -0.0123 & 0.9394 & 0.7860 \end{bmatrix} \in \mathbb{R}^{2 \times 5}.$$

CVX code snippet for the solution to (8)

Input: $\mathbf{X} \in \mathbb{R}^{D \times N}$, $\mathbf{b} \in \mathcal{A}$, $0 < p \leq 1$

```

1:  $\mathbf{Y} = \mathbf{X} \cdot \text{repmat}(\mathbf{b}', D, 1)$ ;
2: cvx_begin
3: variable q(D)
4: maximize( sum( pow_p( Y'*q, p ) ) )
5: subject to
6:  $\mathbf{q}' * \mathbf{q} \leq 1$ 
7:  $\mathbf{Y}' * \mathbf{q} \geq \mathbf{0}$ 
8: cvx_end

```

Return: $\mathbf{q} \in \mathbb{R}^D$

Fig. 2. CVX code snippet for the solution to (8).

(8) exists and can be found exactly, in accordance with the following Lemmas 5 and 6.

Lemma 5. For any $\mathbf{b} \in \mathcal{A}$, $\mathcal{P}(\mathbf{b})$ is a convex set.

Lemma 6. For $p \leq 1$ and any $\mathbf{b} \in \mathcal{A}$, $\|\mathbf{X}^\top \mathbf{q}\|_p^p$ is concave with respect to $\mathbf{q} \in \mathcal{P}(\mathbf{b})$. Moreover, $\forall \mathbf{q} \in \mathcal{P}(\mathbf{b})$ it holds that $\|\mathbf{X}^\top \mathbf{q}\|_p^p = \sum_{n \in [N]} |\mathbf{x}_n^\top \mathbf{q}|^p = \sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{q}^\top \mathbf{x}_n)^p$ and $[\mathbf{b}]_n \mathbf{q}^\top \mathbf{x}_n > 0 \forall n \in [N]$.

By Lemmas 5 and 6, Proposition 1 follows.

Proposition 1. The problem in (8) is a convex problem and can be equivalently rewritten as

$$\min_{\mathbf{q} \in \mathcal{P}(\mathbf{b})} - \sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{q}^\top \mathbf{x}_n)^p. \quad (9)$$

Proposition 1 states that, for every $\mathbf{b} \in \mathcal{A}$ such that $\mathcal{P}(\mathbf{b}) \neq \emptyset$, (8) is a convex problem and, thus, it can be solved exactly, e.g., by interior point methods. For instance, considering a standard primal-dual interior-point solver based on Newton's method [32, 33], (8) can be solved with about cubic cost in D, N . More sophisticated solvers with lower computational cost can also be derived, but this is beyond the scope of this paper. For simplicity in presentation, in this work we solve (8) using CVX [34, 35]. In Fig. 2, we offer a CVX code snippet for Matlab for the solution to (8). The proposed algorithm for exact Lp-PCA is summarized in Fig. 3.

4. EXPERIMENTAL STUDIES

4.1. Matrix Reconstruction

We consider rank-1 matrix $\mathbf{X} = \mathbf{q}_{\text{nom}} \mathbf{u}^\top \in \mathbb{R}^{(D=5) \times (N=7)}$, where $\mathbb{E}\{\|\mathbf{u}\|_2^2\} = DN$ and $\mathbf{q}_{\text{nom}}$ denotes the nominal signal subspace. We disrupt the structure of $\mathbf{X}$ by Additive White Gaussian Noise, forming $\mathbf{X}_n = \mathbf{X} + \mathbf{N}$, where $\mathbf{N}$ draws entries from $\mathcal{N}(0, D^{-1}N^{-1})$. Moreover, we further corrupt one column of $\mathbf{X}$ by adding to it a high-magnitude outlying point. That is, we form $\mathbf{X}_{n+o} = \mathbf{X}_n + \mathbf{O} = \mathbf{X} + \mathbf{N} + \mathbf{O}$, where $[\mathbf{O}]_{:,i} = \mathbf{o}$, $i \in [N]$ is arbitrarily chosen and for any $j \in [N] \setminus i$, $[\mathbf{O}]_{:,j} = \mathbf{0}_D$. Vector $\mathbf{o}$ draws entries from $(0, \sigma^2)$. We vary $\sigma^2 \in \{0, 0.4, 0.8, 1.2\}$, $p \in \{2, 0.8, 0.5\}$, and for every pair (σ^2, p) we compute the exact Lp-norm principal

Algorithm 1. The proposed exact Lp-PCA solution.

Input: $\mathbf{X} \in \mathbb{R}^{D \times N}$, $0 < p \leq 1$

```

1:  $v \leftarrow 0$ 
2: For every  $\mathbf{b} \in \mathcal{A}$ 
3:    $\mathbf{q} \leftarrow \arg\min_{\mathbf{q} \in \mathcal{P}(\mathbf{b})} - \sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{q}^\top \mathbf{x}_n)^p$ 
4:   If  $\|\mathbf{X}^\top \mathbf{q}\|_p^p > v$ ,  $v \leftarrow \|\mathbf{X}^\top \mathbf{q}\|_p^p$ ,  $\mathbf{q}_{\text{Lp}} \leftarrow \mathbf{q}$ 

```

Return: $\mathbf{q}_{\text{Lp}} \in \mathbb{R}^D$

Fig. 3. Proposed algorithm for exact Lp-PCA ($p \leq 1$).

component of $\mathbf{X}_{n+o}$ as $\hat{\mathbf{q}} = \arg \max_{\mathbf{q} \in \mathcal{B}_D} \|\mathbf{X}_{n+o}^\top \mathbf{q}\|_p^p$ using the proposed exact algorithm in Fig. 3. Then, we compute the Normalized Squared Error (NSE) $\|\hat{\mathbf{q}} \hat{\mathbf{q}}^\top \mathbf{X}_{n+o} - \mathbf{X}\|_2^2 \|\mathbf{X}\|_2^{-2}$. In Fig. 4, we plot the Mean NSE (MNSE) versus σ^2 , computed over 40 statistically independent realizations of $\mathbf{X}_{n+o}$. We observe that, in the case of nominal/clean data ($\sigma^2 = 0$), Lp-PCA exhibits high performance for every $p = 2, 0.8, 0.5$. As σ^2 increases, the MNSE increases. Lp-PCA for $p = 0.5$ and $p = 0.8$ exhibits strong resistance against the outlier compared to standard PCA, the performance of which significantly deteriorates as σ^2 increases.

4.2. Classification of Biomedical Data

We consider the Breast Cancer Wisconsin (Diagnostic) dataset [36, 37] which comprises 569 samples of 30 features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. Each sample is labeled and corresponds to a malignant or benign tissue. Overall, there are 212 malignant and 357 benign tissue-samples available. We arrange all the samples in data matrix $\mathbf{X} \in \mathbb{R}^{(D=30) \times (N=569)}$ and define label-vector $\mathbf{z} \in \{0, 1\}^{569}$, where $[\mathbf{z}]_i = 1$ if $\mathbf{x}_i$ corresponds to a malignant tissue and $[\mathbf{z}]_i = 0$ if $\mathbf{x}_i$ corresponds to a benign tissue. We consider availability of $N_{\text{train}} = 14$ points from each class for training and $N_{\text{test}} = 60$ for testing. We let $\mathbf{X}_b = [\mathbf{X}]_{:, \mathcal{I}_b} \in \mathbb{R}^{D \times N_{\text{train}}}$ and $\mathbf{X}_m = [\mathbf{X}]_{:, \mathcal{I}_m} \in \mathbb{R}^{D \times N_{\text{train}}}$ denote the benign and malignant training data samples, respectively, where $\mathcal{I}_b \subset [569]$, $|\mathcal{I}_b| = N_{\text{train}}$, $\mathcal{I}_m \subset [569]$, and $|\mathcal{I}_m| = N_{\text{train}}$. Similarly, $\mathbf{Y}_b = [\mathbf{X}]_{:, \mathcal{J}_b} \in \mathbb{R}^{D \times N_{\text{test}}}$ and $\mathbf{Y}_m = [\mathbf{X}]_{:, \mathcal{J}_m} \in \mathbb{R}^{D \times N_{\text{test}}}$ denote the benign and malignant data samples, respectively, for testing, where $\mathcal{J}_b \subset [569]$, $|\mathcal{J}_b| = N_{\text{test}}$, $\mathcal{J}_m \subset [569]$, and $|\mathcal{J}_m| = N_{\text{test}}$. There is no overlap between the training and testing data of each class: i.e., $\mathcal{I}_m \cap \mathcal{J}_m = \emptyset$ and $\mathcal{I}_b \cap \mathcal{J}_b = \emptyset$. During training, we compute $\mathbf{q}_b = \arg \max_{\mathbf{q} \in \mathcal{B}_D} \|\mathbf{X}_b^\top \mathbf{q}\|_p^p$ and $\mathbf{q}_m = \arg \max_{\mathbf{q} \in \mathcal{B}_D} \|\mathbf{X}_m^\top \mathbf{q}\|_p^p$ by means of the proposed algorithm. Given a testing sample $\mathbf{y}$ from $\mathbf{Y}_m$ or $\mathbf{Y}_b$, we classify it according to

$$(\mathbf{q}_m^\top \mathbf{y})^2 \|\mathbf{y}\|^{-2} \underset{\text{malignant}}{\overset{\text{benign}}{\leq}} (\mathbf{q}_b^\top \mathbf{y})^2 \|\mathbf{y}\|^{-2}. \quad (10)$$

In Fig. 5, we plot the average classification accuracy (computed over 300 distinct realizations of $\mathcal{I}_m, \mathcal{I}_b, \mathcal{J}_m, \mathcal{J}_b$) when p varies in $\{0.1, 0.3, 0.5, 0.7, 0.9, 2\}$. Moreover, we include as a benchmark the classification accuracy of the k -nearest

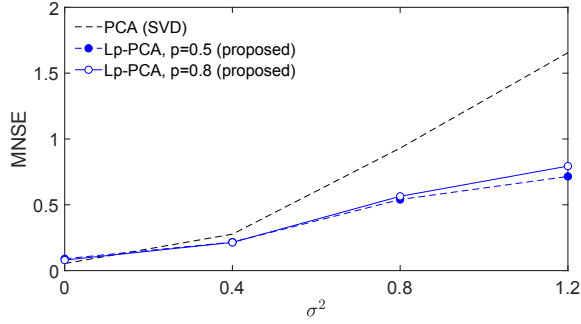


Fig. 4. MNSE v.s. corruption variance σ^2 .

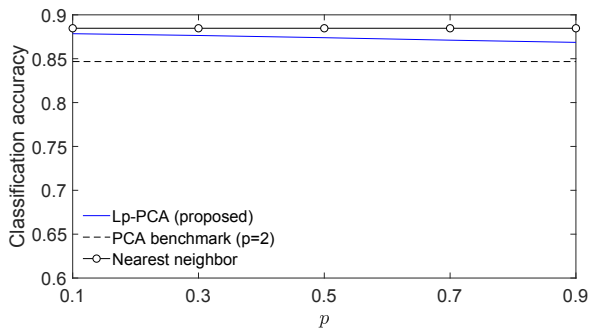


Fig. 5. Classification accuracy v.s. p with no mislabelings.

neighbor classifier for $k = 1$ (NN). We observe that all methods exhibit high performance. NN exhibits the best performance, slightly higher than that of Lp-PCA for $p = 0.1$. Standard PCA, implemented by means of SVD, exhibits the lowest performance. In Fig. 6, we repeat the above experiment but this time we consider that $m = 2$ malignant samples have mistakenly been labeled as benign samples and $m = 2$ benign samples have been mislabeled as malignant. We notice that the performances of NN and SVD are significantly compromised by the mislabelings: i.e., the performance of NN dropped to 0.77 from 0.89 and the performance of PCA dropped to 0.73 from 0.85. On the other hand, Lp-PCA for $p < 1$ remained almost unaffected. Also, we notice that all values of $p < 1$ are similarly robust, with $p = 0.1$ exhibiting the best performance in this study. Finally, in Fig. 7, we fix $p = 0.3$ and let the number of mislabelings, m , vary in $\{0, 1, 2, 3, 4\}$.² Expectedly, when $m = 0$ the performances of all methods are similar to their performances in Fig. 5. However, as m increases, the performances of NN and PCA are compromised. Interestingly, the performance of Lp-PCA ($p = 0.3$) appears to be affected very little by the mislabelings for all values of m .

²That is, m benign samples have wrongly been labeled as malignant and m malignant samples have been labeled as benign samples.

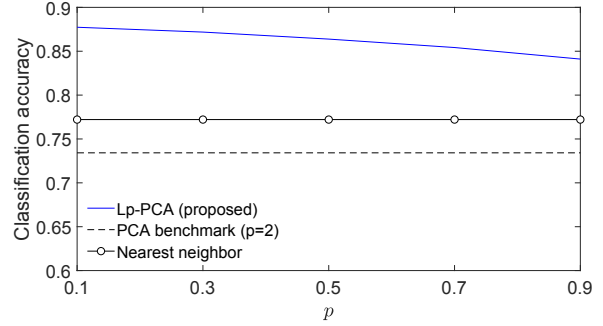


Fig. 6. Classification accuracy v.s. p , for $m = 2$.

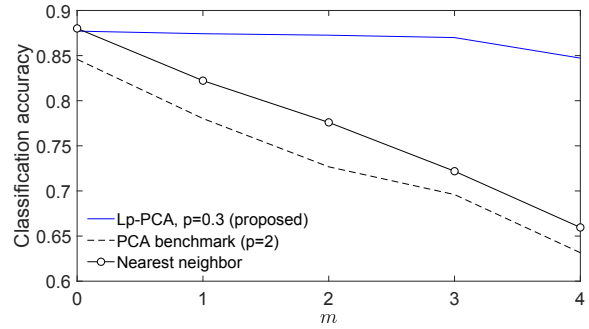


Fig. 7. Classification accuracy v.s. m , for $p = 0.3$.

5. CONCLUSIONS

In this work, we studied Lp-PCA for $p \leq 1$ and showed that it can be solved exactly through combinatorial search. Our experimental studies on data reconstruction and classification illustrate the remarkable outlier-resistance of the proposed method.

References

- [1] I. Jolliffe, "Principal component analysis," in *International encyclopedia of statistical science*. Berlin, Heidelberg: Springer, 2011, pp. 1094-1096.
- [2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.
- [3] V. Barnett and T. Lewis, *Outliers in Statistical Data*, New York, NY, USA: Wiley, 1994.
- [4] J. W. Tukey, "The future of data analysis," *Ann. Math. Statist.*, vol. 33, pp. 1-67, 1962.
- [5] H. Xiao, B. Biggio, G. Brown, G. Fumera, C. Eckert, and F. Roli, "Is feature selection secure against training data poisoning?" in *Proc. Int. Conf. Mach. Learn.*, Lille, France, June 2015, pp. 1689-1698.
- [6] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, May 2011.
- [7] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," in *Proc. Adv. Neural Info. Process. Syst.*, Vancouver, Canada, Dec. 2010, pp. 2496-2504.

- [8] Q. Ke and T. Kanade, "Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Comp. Vision Patt. Recogn.*, San Diego, CA, Jun. 2005, pp. 739-746.
- [9] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Some options for L1-subspace signal processing," in *Proc. Intern. Sym. Wireless Commun. Sys.*, Ilmenau, Germany, Aug. 2013, pp. 1-5.
- [10] P. P. Markopoulos, G. N. Karystinos and D. A. Pados, "Optimal algorithms for L1-subspace signal processing," *IEEE Trans. Signal Process.*, vol. 62, pp. 5046-5058, Oct. 2014.
- [11] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, "Efficient L1-norm Principal-Component Analysis via bit flipping," *IEEE Trans. Signal Process.*, vol. 65, pp. 4252-4264, Aug. 2017.
- [12] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy l1-norm maximization," in *Proc. Int. Joint Conf. Artif. Intell.*, Barcelona, Spain, Jul. 2011, pp. 1433-1438.
- [13] P. P. Markopoulos, M. Dhanaraj, and A. Savakis, "Adaptive L1-norm principal-component analysis with online outlier rejection," *IEEE J. Selected Topics Signal Process.*, vol. 12, no. 6, pp. 1-13, Dec. 2018.
- [14] M. Dhanaraj and P. P. Markopoulos, "Novel algorithm for incremental L1-norm principal-component analysis," in *Proc. IEEE/EURASIP European Signal Process. Conf.*, Rome, Italy, Sep. 2018, pp. 2020-2024.
- [15] M. Dhanaraj and P. P. Markopoulos, "Stochastic Principal Component Analysis via Mean Absolute Projection Maximization," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Ottawa, Canada, Nov. 2019.
- [16] M. Dhanaraj, D. G. Chachlakis, and P. P. Markopoulos, "Incremental complex L1-PCA for direction-of-arrival estimation," in *Proc. IEEE Western New York Signal Image Process. Workshop*, Rochester, NY, Oct. 2018, pp. 1-5.
- [17] N. Tsagkarakis, P. P. Markopoulos, G. Sklivanitis, and D. A. Pados, "L1-norm principal-component analysis of complex data," *IEEE Trans. Signal Process.*, vol. 66, pp. 3256-3267, Jun. 2018.
- [18] N. Tsagkarakis, P. P. Markopoulos, and D. A. Pados, "Direction finding by complex L1-principal component analysis," in *Proc. IEEE Int. Workshop on Signal Proc. Adv. Wireless Comm.*, Stockholm, Sweden, Jun. 2015, pp. 475-479.
- [19] P. P. Markopoulos, S. Kundu, and D. A. Pados, "L1-fusion: Robust linear-time image recovery from few severely corrupted copies," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, Canada, Sep. 2015, pp. 1225-1229.
- [20] Y. Liu and D. A. Pados, "Compressed-sensed-domain l1-pca video surveillance," *IEEE Trans. Multimedia*, vol. 18, pp. 351-363, Mar. 2016.
- [21] D. G. Chachlakis, P. P. Markopoulos, R. J. Muchhala, and A. Savakis, "Visual tracking with L1-Grassmann manifold modeling," in *Proc. SPIE Defense and Commercial Sens.*, Anaheim, CA, Apr. 2017, pp. 1021102:1-1021102:12.
- [22] P. P. Markopoulos and F. Ahmad, "Indoor human motion classification by L1-norm subspaces of micro-Doppler signatures," in *Proc. IEEE Radar Conf.*, Seattle, WA, May 2017, pp. 1807-1810.
- [23] P. P. Markopoulos, D. G. Chachlakis, and E. E. Papalexakis, "The exact solution to rank-1 L1-norm TUCKER2 decomposition," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 511-515, Jan. 2018.
- [24] D. G. Chachlakis and P. P. Markopoulos, "Robust decomposition of 3-way tensors based on L1-norm," in *Proc. SPIE Def. Comm. Sens.*, Orlando, FL, Apr. 2018, pp. 1065807:1-1065807:15.
- [25] D. G. Chachlakis and P. P. Markopoulos, "Novel algorithms for exact and efficient L1-norm-based TUCKER2 decomposition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Calgary, Canada, Apr. 2018, pp. 6294-6298.
- [26] P. P. Markopoulos, D. G. Chachlakis, and A. Prater-Bennette, "L1-norm Higher-Order Singular-value Decomposition," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Anaheim, CA, Nov. 2018, pp. 1353-1357.
- [27] D. G. Chachlakis, M. Dhanaraj, A. Prater-Bennette, and P. P. Markopoulos, "Options for multimodal classification based on L1-Tucker decomposition," in *Proc. SPIE Defense and Commercial Sens.*, Baltimore, MD, Apr. 2019, pp. 109 8900:1-109 8900:13.
- [28] D. G. Chachlakis, A. Prater-Bennette, and P. P. Markopoulos, "L1-norm Tucker Tensor Decomposition," *IEEE Access*, DOI: 10.1109/ACCESS.2019.2955134.
- [29] K. Tountas, D. G. Chachlakis, P. P. Markopoulos, and D. A. Pados, "Iterative Re-weighted L1-norm PCA of Tensor Data," in *Proc. IEEE Asilomar Conf. Signals, Syst., and Comput.*, Pacific Grove, CA, Nov. 2019.
- [30] D. G. Chachlakis, Y. Tsitsikas, E. E. Papalexakis, and P. P. Markopoulos, "Robust multi-relational learning with absolute projection RESCAL," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Ottawa, Canada, Nov. 2019.
- [31] N. Kwak, "Principal component analysis by L_p -norm maximization," *IEEE Trans. Cybernetics*, vol. 44, no.5, pp. 594-609, 2013.
- [32] S. J. Wright, *Primal-dual interior-point methods*, SIAM, 1997.
- [33] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*, Reading, MA: Addison-wesley, 1984.
- [34] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. <http://cvxr.com/cvx>, Sep. 2013.
- [35] M. Grant and S. Boyd, "Graph implementations for non smooth convex programs," Recent Advances in Learning and Control (a tribute to M. Vidyasagar), Eds. V. Blonel, S. Boyd, and H. Kimura, pp 95-110, Lecture Notes in Control and Inf. Sci., Springer, 2008.
- [36] W. N. Street, W. H. Wolberg, and O. L. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis", in *Proc. IS&T/SPIE Int. Symp. Electronic Imaging: Sci. Technol.*, San Jose, CA, 1993, pp. 861-870.
- [37] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," *Operations Res.*, vol. 43, no. 4, pp. 570-577, Jul.-Aug. 1995.